



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

An international Delphi study to assess the need for multiaxial criteria in diagnosis and management of functional gastrointestinal disorders

Philip Austin

Doctor of Philosophy

University of Edinburgh

2015

Preface

This Ph.D. thesis contains the results of research undertaken using web-based Delphi technique to determine levels of agreement among experts on:

1. the suitability of multiaxial assessment criteria for the evaluation of patients presenting with functional gastrointestinal disorders
2. areas of current information that are important when evaluating patients presenting with functional gastrointestinal disorders
3. areas of information requiring further research which may benefit future evaluation of patients presenting with functional gastrointestinal disorders
4. the importance of the patient-practitioner relationship when evaluating patients presenting with functional gastrointestinal disorders

I developed the initial research question from the following review written by myself prior to the Ph.D. as part of an MSc in pain management at The University of Edinburgh.

1. AUSTIN, P. D. & HENDERSON, S. E. 2011. Biopsychosocial Assessment Criteria for Functional Chronic Visceral Pain: A Pilot Review of Concept and Practice. *Pain Medicine*, 12, 552-564

The following article and poster presentations have been published using data presented in this thesis.

1. AUSTIN, P., HENDERSON, S., POWER, I., JIRWE, M. & ALANDER, T. 2013.
An international Delphi study to assess the need for multiaxial criteria in diagnosis and management of functional gastrointestinal disorders. *J Psychosom Res*, 75, 128-34.
2. AUSTIN PD, HENDERSON, S., POWER, I., JIRWE M & ÅLANDER T 2012. The Need for Multiaxial Assessment and the Importance of Psychosocial Factors in Functional Gastrointestinal Disorders: An Analysis of Worldwide Expert Opinion. *American College of Gastroenterology ASM*. Las Vegas: American College of Gastroenterology. Suppl S1-880.
3. AUSTIN PD, HENDERSON, S., POWER, I., JIRWE M & ÅLANDER T 2012. The Development of Multiaxial Assessment Criteria for Chronic Unexplained Visceral Pain: A Worldwide Delphi Expert Consensus Study (Preliminary Findings). Dundee: Scottish Pain Research Community
4. AUSTIN PD, HENDERSON, S. & POWER, I. 2011. Biopsychosocial Assessment Criteria for Chronic Unexplained Visceral Pain: A Pilot Consensus Study. *Joint British and Canadian Pain Societies ASM*. Edinburgh: British Pain Society

Acknowledgments

I would like to thank Professor Ian Power and Dr. Sarah Henderson of the Department of Anaesthesia and Pain Medicine in Edinburgh for their supervision and support before and during the Ph.D. I would also like to thank Dr. Maria Jirwe and Dr. Ture Ålander at Karolinska Institutet in Stockholm for their expertise and support during their supervision of this thesis. I also wish to acknowledge and thank Dr. Gordon Drummond who not only generously stepped into the role of the primary supervisor over last year of the Ph.D., but

who also gave me the benefit of his knowledge and enthusiasm toward research and meticulous attention to detail and scrutiny of my work. I wish to thank Prof Douglas Drossman, the Chair of the Rome Foundation for his advice on past, current and future practice concerning functional gastrointestinal disorders, throughout this study. I am also grateful to Dr. Tobias Sundberg at Karolinska Institutet, a friend and researcher who gave me valuable advice and support before, during and hopefully after this study. I am also sincerely thankful to Nina Cryne, Maxine Pepper and Duncan Paxton for their administrative help and friendly support. I would like to say a special thanks to my father, Ralph Austin and my wife Fiona for encouraging and convincing me to always believe in myself even when I did not. It was very much appreciated. Finally, I would like to dedicate this thesis to my mother, Marlene Austin, who sadly passed away during the course of the Ph.D.

Abstract

Purpose: While there are diagnostic criteria for functional gastrointestinal disorders (FGIDs), their evaluation is challenging. This is because criteria are based on symptoms, and the underlying pathophysiology is not clear; as such, there are no gold standard tests. Diagnosis is further challenged by considerable clinical overlap between different FGIDs as well as other organic diseases, while many people with FGIDs have more anxiety and depression than healthy individuals. I hypothesised that assessment of separate components of FGIDs that also indicate their effect on the patient could improve diagnosis. My aim was to investigate the evolution of opinions from experts involved in the development of FGID diagnostic criteria on the proposal for the development of multiaxial assessment criteria (MAC) for FGIDs.

Methods: I conducted a web-based Delphi study using a group of purposively sampled experts identified from committees of the Rome Foundation and the International Foundation for Gastrointestinal Disorders. From a systematic search of relevant articles, I generated 132 items that were sent to experts as a first round survey. The items assessed risk and contributing factors, the therapeutic relationship, areas of evaluation and the advantages and disadvantages of multiaxial assessment. Consensus on an item was reached when 75% of experts indicated that they agreed or strongly agreed with the statement.

Key results: 36 of 68 eligible participants (52%) responded to the first round. Consensus was reached on 96 items. Using participant feedback, thematic analysis was used to generate 33 additional items for round two. Thirty-one of 36 participants (86%) replied to rounds two and three. In round two, 19 items gained consensus, and in round three, nine items gained consensus. Participants agreed that multiaxial assessment was needed, using a systematic approach to establish the physiological and psychosocial components of FGIDs. Participants

were unable to agree on the importance of physical risk factors such as previous surgery and genetic association. Overall, 124 of the 167 items achieved consensus.

Conclusion and inferences: The key finding from my study shows that experts agree that multiaxial assessment of FGIDs is needed. I also identified expert agreement on the consideration of psychological risk factors and the importance of the impact of FGID symptoms on daily life. Findings also show that experts disagreed on the impact of physical risk factors, socioeconomic status and spirituality on people with FGIDs. While experts could not agree on genetic and gender-based risk factors, they considered that these areas are important and require further research.

Table of Contents

Preface	2
Acknowledgments.....	3
Abstract.....	5
Chapter 1. Introduction and thesis overview	19
1.1 Introduction.....	19
1.2 Study Objectives	19
1.3 Generation and testing of hypotheses	20
1.4 Hypotheses.....	21
1.5 Overview of thesis chapters	21
Chapter 2. Functional gastrointestinal disorders: origins and perceptions	24
2.1 Overview.....	24
2.2 A brief historical perspective on FGIDs	24
2.3 Abdominal symptoms and psychosocial factors.....	24
2.4 Functional gastrointestinal disorders	26
2.4.1 Symptom-based evaluation of FGIDs	26
2.4.2 The Rome process.....	27
2.4.3 The current FGID definition	28
2.4.4 Irritable bowel syndrome	30
2.4.5 Functional Dyspepsia.....	30
2.4.6 Functional abdominal pain syndrome (FAPS).....	31
2.4.7 Noncardiac chest pain (NCCP)	31
2.4.8 Overlap between FGID conditions	32
2.5 Multidimensional assessment	33
Chapter 3. A Systematic search for literature relevant to the development of MAC for FGIDs 35	
3.1 Overview.....	35
3.2 Introduction.....	35
3.2.1 Existing literature.....	35
3.3 Methods.....	36
3.3.1 Search protocols.....	39
3.3.2 Limitations of systematic search protocols in the present study	41
3.4 Risk and contributing factors of FGIDs.....	44

3.4.1	Background	44
3.4.2	Methods.....	44
3.4.3	Results.....	48
3.5	The patient-practitioner relationship	48
3.5.1	Background	48
3.5.2	Methods.....	49
3.5.3	Results.....	51
3.6	Areas for consideration and possible measurement	52
3.6.1	Background	52
3.6.2	Methods.....	52
3.6.3	Results.....	56
3.7	Multiaxial assessment	56
3.7.1	Background	56
3.7.2	Methods.....	56
3.7.3	Results.....	60
3.8	Summary	60
Chapter 4.	Philosophy and Methodology Underlying the Delphi Technique	62
4.1	Overview	62
4.2	Introduction.....	62
4.3	Inquiry Systems	62
4.3.1	Inquiry systems and the Delphi technique	63
4.3.2	Inquiry systems and the present study	65
4.4	Delphi technique	65
4.4.1	Overview.....	65
4.4.2	Anonymity of Delphi participants.....	66
4.4.3	Iteration and controlled feedback.....	66
4.4.4	Application of The Delphi technique.....	67
4.4.5	Theory Building	67
4.4.6	Comparisons with other consensus building techniques.....	69
4.5	Methodological considerations	70
4.6	Establishing rigour in qualitative research.....	71
4.7	Methodological rigour in Delphi research	73
4.8	Reliability and validity.....	74

4.8.1	Reliability.....	74
4.8.2	Validity	76
4.9	Trustworthiness.....	79
4.9.1	Credibility (Internal validity)	80
4.9.2	Transferability (external validity)	83
4.9.3	Confirmability and Dependability (objectivity and reliability).....	84
4.10	Delphi panels	85
4.10.1	Does size matter?	85
4.10.2	Group expertise	87
4.11	Summary	90
Chapter 5.	Defining and Measuring Attitude.....	91
5.1	Overview	91
5.2	Defining attitude	91
5.2.1	Measuring attitude.....	92
5.3	Scaling methods for attitude	92
5.3.1	Thurstone's equally appearing interval scales	93
5.3.2	Likert's procedure	93
5.3.3	Theoretical differences between methods.....	94
5.3.4	Ease of construction and reliability.....	96
5.3.5	Other attitude scales	96
5.3.6	Attitude scale model comparisons	97
5.4	Scaling and the Delphi technique.....	98
5.4.1	Likert scales versus Likert-type scales.....	99
5.4.2	Category labeling	102
5.4.3	The number of Likert scale points	103
5.5	Summary	105
Chapter 6.	The development of first Delphi round	106
6.1	Overview	106
6.2	Item development.....	106
6.2.1	Participant screening questions	107
6.3	The study questionnaire	107
6.3.1	Contributing Factors to FGID	107
6.3.2	The Therapeutic Relationship	108

6.3.3	Areas for Consideration and Possible Measurement.....	108
6.3.4	Multiaxial Assessment and Formulation (Relevance to FGID)	109
6.3.5	Concluding Statements	109
6.4	Web-based Delphi technique	109
6.4.1	Convenience.....	110
6.4.2	Time and cost advantages	110
6.4.3	Administration	111
6.4.4	Web-based Delphi technique limitations	111
6.4.5	Maximising web-based Delphi technique.....	112
6.5	Survey format and design	112
6.5.1	Survey results and analysis facility	114
6.6	The expert panel.....	115
6.6.1	Recruitment of experts	115
6.6.2	Contact of expert population.....	115
6.6.3	Reminders	116
6.6.4	Confidentiality and anonymity.....	116
6.7	Data Analysis	117
6.7.1	Univariate Analysis.....	118
6.7.2	Comparisons between medical specialities	120
6.7.3	Confidence intervals	121
6.7.4	Responders versus non-responders	122
6.7.5	Analysis of Delphi study data	122
6.7.6	Qualitative Analysis.....	123
6.8	Piloting the Questionnaire.....	124
6.9	Paper pilot study	124
6.9.1	Feedback	124
6.10	International Pilot Study	126
6.10.1	Methods.....	126
6.10.2	Feedback	127
6.10.3	Usability.....	127
6.10.4	Content.....	128
6.10.5	Failure to complete.....	129
6.10.6	Pilot data collection and analysis.....	130

6.10.7	Pilot respondent participation in the main Delphi study	130
6.11	Summary	130
Chapter 7.	Results of the international web-based Delphi survey	132
7.1	Overview	132
7.2	From the pilot study to the first round survey	132
7.2.1	Round one	132
7.2.2	Round two	133
7.2.3	Round three	133
7.3	Results	135
7.3.1	Participants	135
7.3.2	Participant total scores	136
7.3.3	Distribution of expert responses to Likert-scale points for each survey round 139	
7.3.4	Participant characteristics	140
7.3.5	Response rates (appendix E)	142
7.3.6	Delphi process findings	144
Chapter 8.	Discussion	166
8.1	Introduction	166
8.2	Expert opinion	166
8.2.1	Risk and contributing factors	167
8.2.2	The therapeutic relationship	172
8.2.3	Areas for measurement	174
8.2.4	Multiaxial assessment	176
8.2.5	Potential FGID assessment criteria framework	177
8.3	Methodological processes and limitations	179
8.3.1	Terminology	179
8.3.2	The online survey	179
8.4	Methodological rigour	180
8.4.1	Trustworthiness	180
8.4.2	The iterative process	185
8.4.3	Survey length and relevance of items	186
8.4.4	The number of Likert scale points	187
8.4.5	Data collection	188
8.4.6	Survey item bias	189

8.5	Participant characteristics	190
8.5.1	Sample size	190
8.5.2	Responder rates	190
8.5.3	Stability of participant response over rounds two and three	191
8.5.4	Sample heterogeneity	192
8.5.5	How important is consensus?.....	193
8.5.6	Responder bias	194
8.5.7	Non-responder bias	195
8.5.8	Maintaining responder rates	197
8.6	Lumping and Splitting FGIDs with other FSSs	197
8.7	The future.....	199
8.8	Summary	200
8.9	Conclusions.....	201
8.10	Recommendations.....	201
8.11	Contributions to knowledge	201
Appendix A: Pilot data summaries		223
Pilot Round (trial table)		223
Individual data sheet example (Likert scores and comments)		231
Appendix B: Delphi round comment sheets		243
Round 1		243
Round 2.....		259
Round 3.....		272
Appendix C: Search protocol flow charts for Delphi survey sections		284
Appendix D: Tables of Articles used in the Systematic Review		288
Risk factors		288
Gender.....		288
Impact of daily life.....		292
Physical risk factors		304
Psychological risk factors		308
Genetic polymorphisms		318
The patient-practitioner relationship.....		324
Domains of measurement		331
Multiaxial assessment criteria.....		343

Appendix E: Response rates	349
Response and non-response rates (1 – responders, 2 – non-responders)	349
Late versus early response rates.....	353
Appendix F: Participant correspondence	357
Research outline letter (paper and electronic).....	357
Invitations (electronic).....	358
Invitation to round one.....	358
Invitation to round two.....	359
Invitation to round three.....	360
Reminders (electronic).....	361
Online Pilot study	361
First reminder (all rounds)	361
Second reminder (all rounds).....	362
Final reminder (all rounds)	362
Thanks for participating letter (electronic)	363
Appendix G: No requirement for ethical approval	364
Appendix H: Rome Foundation Letter of permission for use of research publication	365
An international Delphi study to assess the need for multiaxial criteria in diagnosis and management of functional gastrointestinal disorders”	365
To be used as support for the new Rome IV multi-axial criteria	365
Appendix I: Poster Presentations and peer review article.....	366
British Pain Society ASM, Edinburgh, June 2011	366
Scottish Pain Research Community ASM, Dundee, March 2012	367
American College of Gastroenterology ASM, Las Vegas, Oct 2012	368
Journal of Psychosomatic Research, 2013Aug;75(2): 128-34.....	369

Tables

Table 3-1 The checklist for developing systematic review search strategies for this study (adapted from Naumann 2007)	37
Table 3-2 Data extraction sheet used to locate articles containing data for item development for the Delphi first round survey (National Institute for Health and Clinical Excellence, 2008).	38
Table 3-3 General systematic search protocols to identify population data and methodological criteria of located articles	40
Table 3-4 shows the number of articles found for each type of study design and the geographic spread of study location.....	42
Table 3-5 Keywords and search terms for risk factors for FGIDs	45
Table 3-6 Keywords and search terms for the patient-practitioner relationships in FGIDs..	50
Table 3-7 Keywords and search terms for areas of consideration and possible measurement.	53
Table 3-8 Keywords and search terms for studies researching the use of multiaxial assessment.....	57
Table 4-1 Applications of the Delphi technique during the research process Adapted from (Okoli and Pawlowski, 2004).....	69
Table 7-1 Outline of the round one subject sections from the Delphi survey for future development of MAC for FGIDs.	134
Table 7-2 Countries of experts contacted round one of the Delphi study.	135
Table 7-3 Professional and academic profiles of respondents in first, second and third Delphi rounds.....	141
Table 7-4 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.	146
Table 7-5 . Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.	151
Table 7-6 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.	155
Table 7-7 Overall agreement on the future development of MAC for FGIDs over three survey rounds.	160
Table 7-8 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert	

rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.	161
Table 7-9 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.	164
Table 8-1 IGDA. 7: Standardized multi-axial diagnostic formulation. (World Psychiatric Association, 2003)	178

Figures

Figure 1-1 A model showing the generation and possible use of hypotheses used as a basis for this thesis for the advancement of diagnostic criteria for FGIDs	20
Figure 2-1 William Beaumont (left), a US Army surgeon, known as the “Father of gastric physiology” through his research on human digestion (Bernard Becker Medical Library Archives) and Paul MacLean (right) who first described the limbic system and its effect on behaviour and GI symptoms (Newman and Harris, 2009).	25
Figure 2-2 Overlapping symptoms of GI motility disorders commonly reported by many individuals diagnosed with a specific FGIDs. Abbreviations; chronic constipation (CC), gastroesophageal reflux disease (GERD), irritable bowel syndrome (IBS). Adapted from Baker DE 2005.....	33
Figure 4-1 A diagrammatic representation of the relationship between reliability and validity. The aim of the research question is the centre of the target (Precision Consulting, 2014).....	77
Figure 5-1 A comparison of theoretical item characteristic curves associated with unfolding (squares) and cumulative models (plus signs) in relation to participants true attitudes and scored level of agreement (taken from Roberts et al 1999)	95
Figure 5-2 A Likert item based on the recommendations of Rensis Likert	100
Figure 5-3 An example of a Likert-type item. The choice options are arranged horizontally in a scale. However, the scale is not bipolar, nor does it represent a range from extreme agreement to disagreement	101
Figure 5-4 In this example of a participant scoring matrix using a 5-point Likert scale, “5” represents strongly agree and “1” represents strongly disagree. Participant rating scores of 4 and 5 are added to find the percentage level of agreement. In this matrix, the columns represent group survey scores and the rows, individual survey scores.....	102
Figure 6-1 The Bristol Online Survey: a section from the first round Delphi survey showing a five-point Likert scale	113
Figure 6-2 An example of a results page including a survey overview	114
Figure 6-3 A typical box and whisker plot showing interquartile ranges and extreme markers (SAS Software Solutions, 2015)	120

Figure 7-1 Shows the spread of total expert scores for A round 1, B round 2, C round 3. Data are expressed as box-and-whisker plots showing median, interquartile ranges and extreme scores.....	138
Figure 7-2 Stem and leaf plots showing distribution of participant total scores in rounds one to three with lowest scores at the top and highest scores at the bottom.....	139
Figure 7-3 Shows a stacked bar chart for the percentage of experts responding to each Likert-point in each Delphi rounds 1- 3.....	140
Figure 7-4 Responder-frequency timeline for the Delphi first round survey.....	143
Figure 7-5 Mean percentages of consensus levels over three survey rounds for items relating to A) risk factors, B) the therapeutic relationship, C) domains of measurement and D) Multiaxial assessment criteria (MAC) concerning FGIDs.....	165
Figure 8-1 Structural equation model indicating the latent variables (ovals) and observable variables (boxes). Line thicknesses of the arrows indicate the significance level of the path (Jones et al., 2012)	168
Figure 8-2 The Delphi study audit trail and timescale flow chart.....	185
Figure 8-3 shows the greater range of strong attitude scale points on seven-point scale (7,6,2 and 1), compared to a five-point scale.	188

Glossary

BOS	Bristol Online Survey
CFS	Chronic fatigue syndrome
DSM	Diagnostic Statistical Manual
ED	Eating disorder
FAPS	Functional abdominal pain syndrome
FD	Functional dyspepsia
FGID	Functional gastrointestinal disorder
FMS	Fibromyalgia
FSS	Functional somatic syndrome
GERD	Gastroesophageal reflux disease
GI	Gastrointestinal
GIT	Gastrointestinal tract
H-pylori	Helicobacter pylori
HADS	Hospital Anxiety and Depression Scale
HPA axis	Hypothalamic-pituitary-adrenal axis
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
IGDA	International Guidelines for Diagnostic Assessment
IS	Inquiry system
MAC	Multiaxial assessment criteria
NCCP	Noncardiac chest pain
p	Probability
PRO	Patient reported outcome
PTSD	Post-traumatic stress disorder
RCT	Randomised clinical trial

5-HT	Serotonin
SERT	Serotonin transporter
SF-36	Short form 36
SCL-90	System checklist 90
SSRI	Selective serotonin reuptake inhibitor
PTSD	Post-traumatic stress disorder
TNF α	Tumour necrosis factor alpha
VLPFC	Ventrolateral prefrontal cortex

Chapter 1. Introduction and thesis overview

1.1 Introduction

In this thesis, I examine the opportunity to improve and extend current symptom-based diagnostic criteria for FGIDs by investigating how the systematic evaluation of different specific features or elements of disease (Multiaxial assessment criteria) can improve the current diagnostic process. In the absence of “gold standard” tests, I propose that diagnosis move away from symptom-based criteria to an approach that considers different components of a disorder. Separate measures of each aspect of the condition (each considered as an axis) will thus characterise the entire disorder and indicate the effect of the disorder on the patient. Expert opinion should help to develop FGID diagnostic criteria. This thesis investigated these opinions. I first reviewed the published literature using systematic search procedures to identify risk factors, elements of the patient-practitioner relationship, disorder features and the potential application of MAC to FGIDs. From the review, I made a list of items to be included in a Delphi survey of experts on the concept of MAC as a means of evaluation of FGIDs. Through the course of the Delphi process, I observed the experts’ evolving responses.

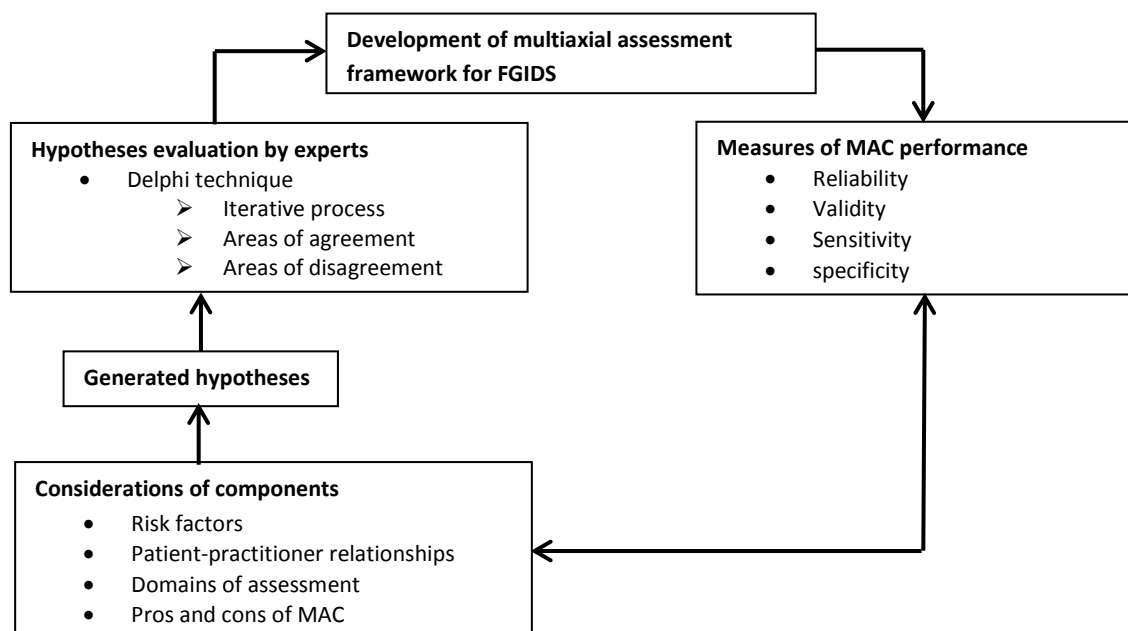
1.2 Study Objectives

My primary aim was to examine attitudes of experts towards the future development of MAC for FGIDs. I also aimed to examine attitudes regarding areas of information that are presently suitable for use in MAC where there is a high level of evidence and those areas where further research is needed due to a lack of reliable evidence. One further aim was to define areas of the patient-practitioner relationship necessary to the management and the well-being of FGID patients.

1.3 Generation and testing of hypotheses

To develop hypotheses for this study I systematically searched and reviewed published articles concerning the phenomena of FGIDs and MAC to a) identify related components relating to the cause, maintenance and evaluation of FGIDs and b) organise these components into a framework for the first round of a Delphi survey. I developed hypotheses supporting an assessment that communicates different components of functional disorders. I used the Delphi technique to evaluate these hypotheses where experts judged items generated from the literature review for levels of agreement and importance over several survey rounds. Results from the Delphi study then established whether the hypotheses were practical and, therefore, appropriate as data to support the future development and testing of MAC for FGIDs. Figure 1.1 shows the steps taken during this study and possible future steps in the development and validation of MAC for FGIDs.

Figure 1-1 A model showing the generation and possible use of hypotheses used as a basis for this thesis for the advancement of diagnostic criteria for FGIDs



1.4 Hypotheses

My hypotheses were that

- 1 “Experts will agree that MAC is a clinically relevant tool for assessing patients with FGIDs”.
- 2 “Experts will agree on elements that are presently suitable for the multiaxial evaluation of FGIDs”.

1.5 Overview of thesis chapters

In Chapter two, I outline a brief history on the recognition of FGIDs and their associations with psychosocial factors. I also review the development of FGID symptom-based evaluation beginning with the work of Manning and colleagues in the late 70s’ to the present Rome III criteria where I discuss confusion surrounding the current definition of FGIDs. I then introduce FGIDs, citing common examples, describing their clinical presentation and prevalence within the general population. I finish with an introduction to MAC, describing their primary purpose of integrating and describing multiple components in functional disorders.

In chapter three, I show my methods for developing and implementing systematic search procedures to locate information for the construction of the first Delphi round. I further show tables of all search terms and keywords while also showing results describing the numbers of accepted articles and their geographical location.

In chapter four I discuss the fundamental systems of inquiry related to theory building, decision-making and consensus. I also discuss methodological considerations and essential features surrounding the Delphi technique and my rationale for using it in this study. I then describe the validity, reliability and trustworthiness of the Delphi technique. Finally, I discuss the nature and size of Delphi panels and their effect on methodological rigour.

In chapter five, I discuss the origins and importance of what defines attitude and its measurement. I then compare Likert and Thurstone's attitude scaling methods where I discuss their reliability, validity, their construction and argue my reasons for selecting Likert scaling. I also discuss differences and difficulties in distinguishing Likert from Likert-type items and scales. Finally, I describe differences in Likert category labelling, the number of scale points and effects of group size and item homogeneity on reliability.

In chapter six, I describe the development of the Delphi first round survey. Here I explain the methods for recruiting and maintaining participation of experts participating in my study. I also discuss the advantages and disadvantages of Web-based Delphi technique regarding convenience, time, cost, administration, and access. I then describe the development of Delphi survey items and the construction of online surveys. Next, I review and discuss quantitative and qualitative data analysis used in the study. Lastly, I present the pen-and-paper and online pilot studies where I discuss participant feedback on questionnaire format and item suitability for the Delphi first round survey.

Chapter seven presents the main online Delphi study. I give a brief overview of its development, followed by a description of the administration over all three Delphi rounds; I then report the features of the experts who took part, their attitude ratings, response rates and the descriptive analysis of responses for all three Delphi rounds.

Finally, in chapter eight, I first discuss the results of the Delphi study. I then review the methods, rigour and limitations of the study. I next show the adherence of my study to trustworthiness criteria, after which I consider the influence of response rate on conclusion. I further consider associations between FGIDs and other functional somatic syndromes (FSSs)

and the relevance to commonly observed symptom overlap. I finish this chapter with my conclusions from the study and recommendations for future inquiry.

Chapter 2. Functional gastrointestinal disorders: origins and perceptions

2.1 Overview

This chapter describes the recognition of FGIDs and the development of symptom-based diagnosis for FGIDs. Summaries of FGIDs and MAC are also described.

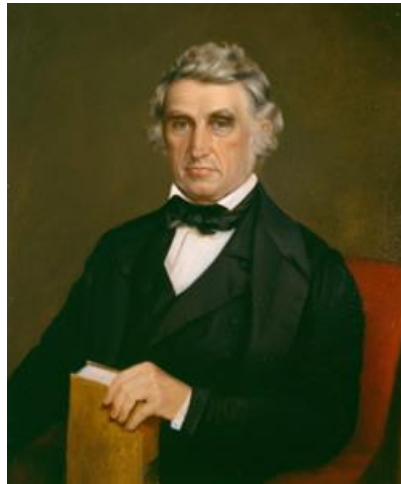
2.2 A brief historical perspective on FGIDs

The earliest articles on disorders such as irritable bowel syndrome (IBS) were published in the 1900s. A few abstracts, opinion papers or case studies describe IBS as a chronic intestinal obstruction disorder (Blacklock, 1965, p8), or a “divided syndrome” presenting as painful “spastic colon” and/or “painless diarrhoea” (No author listed, 1972). Similar articles also describe FD, its treatment, prevalence rates and psychological associations (Fulton, 1907, Sandler and Pollock, 1954). Interestingly, Scott (1933, p521) later described FD as not a physical diagnosis but a name given to a collection of symptoms, a description that at present still remains the only form of diagnosis for all FGIDs.

2.3 Abdominal symptoms and psychosocial factors

It has long been accepted that both emotions and social factors affect the sensorimotor function of the gastrointestinal tract (GIT) with people experiencing changes in gastrointestinal (GI) function during stress or emotional arousal (Van Oudenhove et al., 2010, p201), such as “butterflies in my stomach”. In the early 1800s, an army surgeon, William Beaumont (figure 2.1) had the opportunity to study a patient with a gastric fistula. He observed the effect of “violent passion” on the digestive tract in the presence of bile brought on by anger that is seldom found in the stomach in a healthy state (Beaumont W, 1833).

Figure 2-1 William Beaumont (left), a US Army surgeon, known as the “Father of gastric physiology” through his research on human digestion (Bernard Becker Medical Library Archives) and Paul MacLean (right) who first described the limbic system and its effect on behaviour and GI symptoms (Newman and Harris, 2009).



Remarkably, in the 1880s, an American and a Danish psychologist independently developed similar theories supporting the idea that emotional stimuli automatically induce bodily changes where feedback stimulate further emotional feelings (Van Oudenhove et al., 2010, p203). Research in the early 20th century focused on GI secretion and motor function in patients suffering psychotic disorders such as manic depression and schizophrenia using fluoroscopy after barium meals (Dunbar, 1938). Complete cessation of gastroduodenal peristalsis during suggestion of intake of aversive food was noted, while the stimulation of disgust induced reverse contraction of the stomach (Dunbar, 1938). Later, Paul MacLean (figure 2.1) became the first researcher to communicate a theory on brain mechanisms. His work advanced knowledge by linking emotion and visceral function via the “visceral brain” and later described the role of the limbic system which has recently been shown using functional brain imaging to be involved in visceral sensation and emotion (Newman and Harris, 2009).

2.4 Functional gastrointestinal disorders

FGIDs are a large group of persistent and recurring disorders that occur because of abnormal functioning of the gastrointestinal tract. As FGIDs do not have an identified underlying pathophysiology, they can only be determined using symptom-based diagnostic criteria. FGIDs are associated with and shown to be influenced by psychological conditions (Fortea and Prior, 2013), and organic disease

2.4.1 Symptom-based evaluation of FGIDs

There are currently no known gold standard tests for the diagnosis of FGIDs. Consequently, clinicians base their diagnosis on specific clusters of symptoms rather than an understanding of underlying mechanisms. In the late 1970s, Manning et al. (1978) were the first to develop a questionnaire establishing symptoms thought to be typical of IBS. After the questionnaire was trialled on patients referred to gastroenterological clinics, a review of records 17 – 26 months later revealed four cardinal symptoms significantly more common in IBS patients than those with organic disease. These were abdominal distension, relief of pain on bowel movement, looser and more frequent bowel movements, incomplete evacuations and mucus in stool. Kruis et al. (1984) added new dimensions of symptom duration and negative physical and blood test findings. Not surprisingly, Talley et al. (1990) and Jellema et al. (2009) found that Manning and Kruis criteria were specific in excluding organic disease, but lacked sensitivity. In the same year, Sandler et al. (1984) were the first researchers to consider psychological influences when they showed that IBS patients reported more psychopathology than patients with organic disease showing for the first-time behavioural influences leading to health care seeking in IBS patients.

2.4.2 The Rome process

In 1990, several multinational working teams used the Delphi technique to develop standards aimed to reduce unnecessary diagnostic procedures. These measures were originally designed for research purposes when investigating pathophysiology and treatment responses in comparatively homogeneous populations (Khan and Chang, 2010). These standards became known as the Rome criteria. The most recent set are the Rome III criteria for FGIDs introduced in 2006. These measures are currently used in clinical research and increasingly so in clinical practice and are now available in 23 languages both for adults and children. The Rome criteria are administered by expert panels consisting of 14 investigative committees representing 18 countries worldwide, which at present is considered the accepted FGID diagnostic resource (Chang, 2006, The Rome Foundation, 2006).

Modifications of these symptom-based criteria have aimed to increase discrimination between both individual FGIDs and among healthy people. However, Dang et al. (2012) in a systematic review of validation studies for IBS between 1992 and 2011 using Manning, Kruis and Rome I and II criteria, found that the Manning criteria while performing modestly, were the most valid and diagnostically accurate. Strangely, seven years after the publication of the Rome III criteria, Ford et al. (2013) were the first researchers to validate Rome III for IBS against all iterations of the Rome and the Manning criteria within the same data set. However, while the Rome III criteria introduced further FGID subgroups and frequency thresholds of symptoms, they did not perform better than previous symptom-based criteria. Predictably, when Ford et al. (2014) validated the Rome III criteria for FD in a large cohort of patients with GI symptoms, they found they were no better than the previous Rome definitions. Sood et al. (2014) in their informative review, summarise nicely why these findings are not surprising. They suggest that because symptom-based diagnostic criteria are

generally derived from each other, refinement of pre-existing criteria is unlikely to improve diagnostic accuracy.

2.4.3 The current FGID definition

The current Rome Foundation definition of an FGID is “*a condition where there is disordered functioning motility, visceral sensation, altered mucosal / immune function or brain-gut interactions*” (Drossman et al., 2006, p1378). This definition causes confusion especially concerning the term “functional”. For most, the term means the absence of organic disease while for others it implies a psychiatric problem, which may offend many patients. Furthermore, while not stated, the Rome definition does ***not*** exclude organic co-morbid conditions (Corazziari, 2004) with common cited examples being bacterial and viral infections, asthma, stroke and inflammatory joint and bowel disease (Debley et al., 2006, Halpin and Ford, 2012, Whitehead et al., 2007, Barratt et al., 2011, Rodríguez et al., 2000). To clarify this question and the position of those who develop FGID diagnostic criteria, I discussed this point with the Rome Foundation Chair and two senior board members. They said that the current definition is seen by members to be incorrect and remains controversial (Drossman DA, 2013, Talley N, 2013, Kellow JE, 2103). Further discussion proposed that for future classification purposes, definitions should be made clearer by possibly including; “*in the absence of other diseases that would explain the character of FGID symptoms*” (Drossman DA, 2013). These views are supported by a useful commentary by Corazziari (2004) that cites co-existing conditions where the symptoms of IBS and those of ulcerative colitis or functional dyspepsia and peptic ulcer occur simultaneously.

However, there is debate concerning the symptomatic relationships between IBS and inflammatory bowel disease (IBD). Quigley and Bernstein (2012), legitimately warn that studies reviewing these relationships have not studied IBS symptom clusters, but “symptoms

meeting criteria for IBS” in IBD subjects, where the authors state that every patient with IBD could satisfy diagnostic criteria for IBS. Thus, they recommend that patients in remission from IBD who have IBS symptoms should be regarded as having ongoing IBD. Berrill et al. (2013) and Keohane et al. (2010) examined the prevalence of IBS symptoms in IBD patients in remission and support these comments by stating that regardless of faecal calprotectin levels (a sensitive marker for intestinal inflammation), IBS-type symptoms should be considered as caused by occult inflammation rather than coexisting IBS.

After reviewing the current literature, I conclude that until the introduction of reliable diagnostic modelling for FGIDs, it would be better to combine present symptom-based criteria with appropriate and practical laboratory tests. Unlike well-established pathophysiological diagnostic tests, symptom-based criteria cannot guarantee diagnostic homogeneity and thus creates diagnostic uncertainty. Furthermore, as symptom-based diagnosis leaves little room for exploration of the past (cause) or present contribution; it is also important to obtain a thorough history so that specific FGIDs can be accurately defined. In conclusion, Corazziari (2004) sensibly suggests a more appropriate term of “*disorders of gastrointestinal function*”. This term avoids the term “functional” while also avoiding the diagnosis of separate organic versus non-organic aetiology. Thus, this term also leaves open the consideration of organic pathology at any time-point.

While there are suggestions for use of the Rome criteria in primary care settings, they are at present infrequently used. Only a minority of patients referred to secondary care fulfil the criteria for an FGID or lack alarm symptoms. Improved use of the criteria in primary care in conjunction with available laboratory tests could not only reduce unnecessary endoscopies in patients who attend secondary care, but also reduce the burden on patients and healthcare systems (Kok et al., 2013). Encouragingly, the Rome Foundation recently set up working

committees to collect data for development of the Rome IV multi-axial criteria for FGIDs, which begins in 2014 with an expected publication date in 2016. These new criteria will apply evidence-based knowledge as opposed to expert consensus used in previous versions. In the following sections, I briefly describe different examples of FGIDs.

2.4.4 Irritable bowel syndrome

IBS is a relapsing gastrointestinal disorder typified by recurring abdominal pain and cramping associated with altered defaecation (Halder SLS and Locke GR, 2007), in the absence of detectable organic disease. (Talley NJ and Spiller R, 2002). Longitudinal population-based studies show the prevalence of IBS is constant over time, whereas the severity of symptoms do vary. Additionally, Ringstrom et al. (2007) suggest that IBS symptom severity alone does not explain illness behaviour, but also showed that psychological symptoms and reduced quality of life are most important to the experience of GI symptoms and health care seeking patterns. IBS affects around 11% of the global population with rates varying broadly between populations from 4.7% in France (Dapoigny et al., 2004) to 32% in Nigeria (Okeke et al., 2009). In most populations women report more IBS symptoms irrespective of diagnostic criteria used (Quigley et al., 2006) where rates are approximately two to three times higher than in men (Canavan et al., 2014). Over 50% of all patients with IBS report depression or anxiety (Hamilton et al., 2009) and these patients experience more severe symptoms (Canavan et al., 2014).

2.4.5 Functional Dyspepsia

“Functional dyspepsia” refers to a group of upper GI symptoms that are common in adults which include postprandial fullness, recurrent epigastric pain and epigastric burning in the absence of pathology and other upper GI symptoms such as nausea, vomiting and belching (Talley et al., 1999, Mahadeva and Goh, 2006). Population studies using Rome II criteria

report high prevalences of around 25% in the US, China and Australia (Westbrook and Talley, 2002), with greater prevalence rates in women (Mahadeva and Goh, 2006). Adults with FD also score highly on anxiety scales, but less so on depression scales. Several large population studies also show considerably more psychological morbidity in FD patients than in controls (Van Oudenhove and Aziz, 2013).

2.4.6 Functional abdominal pain syndrome (FAPS)

FAPS is a pain syndrome that has little or no relationship to gut function, is associated with a loss of daily activities, and has been present for at least 6 months (Clouse et al., 2006). The pain is nearly always constant and frequently occurs. Patients typically describe severe pain as covering a large anatomical area for which they persistently seek pain relief and many diagnostic tests. Very few epidemiological studies have investigated FAPS. Here pain is stated to be caused by amplified central perception of normal visceral input rather than enhanced peripheral stimulation from the abdominal viscera (Sperber and Drossman, 2011, p515). Rome III criteria define FAPS as continuous abdominal pain that shows negligible associations with physiological events, and that significantly affects daily life (Rome Foundation, 2006).

2.4.7 Noncardiac chest pain (NCCP)

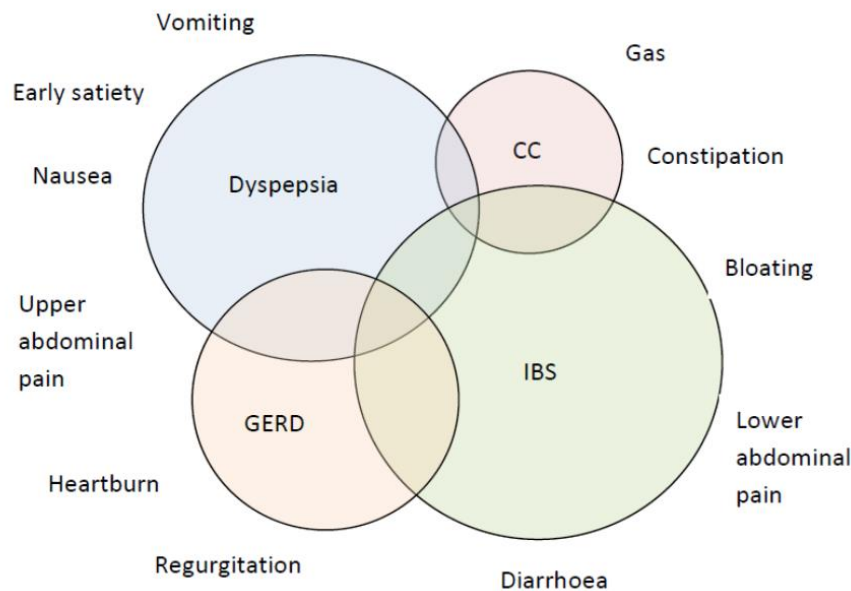
Due to the high morbidity and prevalence of coronary artery disease, chest pain is treated as cardiac in origin until shown otherwise. However, many patients initially considered having cardiac disease are later diagnosed with oesophageal disease. Studies show that around 33% of chest pain cases are diagnosed as NCCP with no other GI (dysphagia, heartburn, acid regurgitation) or psychological symptoms (anxiety, depression) being significantly associated with the condition (Eslick et al., 2003). Although the prevalence in the community is similar, more women than men are referred to tertiary care clinics with NCCP

and are more likely to report anxiety-related symptoms (Carmin et al., 2008, Taylor and Bellumkonda, 2009).

2.4.8 Overlap between FGID conditions

According to the Rome III classification, FGIDs are mutually exclusive disorders that are usually studied independently (Choung, 2012). However, studies show close relationships between different FGIDs, with around 80% of IBS patients reporting symptoms of gastroesophageal reflux disease (GORD) and up to 71% of GORD patients complaining of IBS symptoms (figure 2.2) (Gasiorowska et al., 2009). Debate on the link between these disorders is at present theoretical, however, some studies propose that having one FGID increases the risk for a diagnosis of the other condition due to common sensory and motor dysfunction (Yarandi et al., 2010, Ruigomez et al., 2009). Lee et al. (2009, p200) suggest that these overlaps of diagnosis occur more often in individuals with anxiety disorders.

Figure 2-2 Overlapping symptoms of GI motility disorders commonly reported by many individuals diagnosed with a specific FGIDs. Abbreviations; chronic constipation (CC), gastroesophageal reflux disease (GERD), irritable bowel syndrome (IBS). Adapted from Baker DE 2005



2.5 Multidimensional assessment

The purpose of MAC is to assess and systematically communicate the biological, psychological and social components of a disorder by means of different measure related to various diagnostic features. MAC also provides an overview of the patient's condition that describes its impact on their daily function (Banzato et al., 2009). However, while a biopsychosocial model of health attempts to conceptualise how contributing factors interact in functional somatic syndromes (FSSs); we remain caught in a circular argument that struggles to go beyond the mind-body dichotomy. This presents a challenge for patients and practitioners that is reflected in the number of clinician visits, diagnostic tests and economic loss due to work absence (López-Colombo et al., 2012). Unlike symptom-based diagnostic criteria, MAC aims to integrate the contribution of both peripheral and central factors, past and present that may affect symptom severity and the impact on the patient's health and well-being.

There are no published studies on the formal development or implementation of MAC for FGIDs. Given the nature of FGIDs, separate axes of information could be used to describe features of FGIDs, such as their symptoms, comorbid psychological and physical disorders, adaptive social, physical and occupational functioning and quality of life. Importantly, as FGIDs are common in patients with a history of childhood adversity, the timing of the onset of both physical and psychological symptoms may also help solve the chicken and egg question. As such, I suggest that the inclusion of MAC as part of the Rome criteria should make assessment and treatment more accessible to those who are not experts working in clinical research, and include gastroenterologists and other health care providers who deal with patients with FGIDs.

Chapter 3.A Systematic search for literature relevant to the development of MAC for FGIDs

3.1 Overview

This chapter describes how I developed a search process to gather information to construct items for the first round of the Delphi survey. The details are summarised in tables.

3.2 Introduction

FGIDs are syndromes that are each defined by a collection of symptoms. Research into improving these criteria would be facilitated by agreement on a set of important related features that could be used to better define these conditions. Many people with FGIDs have more anxiety and depression compared with healthy individuals. It is not known whether these and other psychological disorders precede the onset of FGIDs or develop because of symptoms. Furthermore, socioeconomic and cultural circumstances contribute to the severity of symptoms and decreased quality of life. Given the possible causes of these heterogeneous disorders, a single diagnostic test is unlikely to identify people with discrete FGIDs. Therefore, gathering a set of similar features to define different FGIDs would aid research, investigation and treatment, and in particular, allow us to develop guidelines. I conducted a systematic search to collect data for use in the first round of an international Delphi study where experts will make judgements on areas of information related to the application of MAC to FGIDs.

3.2.1 Existing literature

I searched for existing literature researching and reviewing multi-axial assessment for FGIDs in English and non-English articles from 1980 to Sept 2011. I searched MEDLINE and EMBASE using the following terms: functional gastrointestinal disorders, irritable bowel syndrome (IBS),

functional bowel, IBS and FGID. In addition, I used the Boolean search operator, AND, in conjunction with the terms: systematic review, assessment and diagnosis. Concerning MEDLINE, I used the following MeSH terms: “Irritable bowel syndrome”, “Review” and “Diagnostic and Statistical Manual of Mental Disorders”. I searched these terms using the fields, “Text Word”, “Title” and “Title/Abstract”. I refined results by limiting the source type to peer-reviewed journals. Eligible studies had to be systematic reviews concerning the assessment or diagnosis of FGIDs and/or IBS.

I found 104 articles on the diagnosis and management of IBS. Most articles reviewed symptom-based diagnosis as shown in Manning and Rome II and III criteria, e.g. (Furman and Cash, 2011, Moayyedi and Ford, 2011). Others reviewed disease-defining biomarkers related to recent studies supporting low-grade immune activation in FGIDs, e.g., (Spiller, 2011, Barbara and Stanghellini, 2009). I found one article which reviewed and discussed the present requirement of interdisciplinary management of FGIDs (Enck and Martens, 2008). Hence, I found no research or reviews similar to what I had proposed.

3.3 Methods

Systematic reviews search and analyse specific areas of scientific research. In my study, however, there were a large number of subject areas. Thus, in order to develop Delphi items, I used systematic search protocols in MEDLINE and EMBASE using keywords and search terms in areas relating to the cause and maintenance of FGIDs. I followed the Cochrane Haematological Malignancies Group search strategy shown in table 3.1 (Naumann, 2007). Study selection criteria were; publication as a full paper, English and non-English articles (reduce cultural and language bias) and articles that fulfilled high levels of evidence. In many areas of FGIDs research, high levels of evidence such as prospective cohort studies, good systematic reviews and meta-analysis are not available. Thus, I also selected papers that

provided the best-available evidence in these fields. Eligible articles were summarised and compiled using data extraction sheets (figure 3.1) developed by the National Institute for Health and Clinical Excellence (National Institute for Health and Clinical Excellence, 2008). I divided the systematic search into the following sections:

- Risk and contributory factors to FGIDs
- The therapeutic relationship
- Areas of measurement
- Advantages and disadvantages of multiaxial assessment strategies

Table 3-1 The checklist for developing systematic review search strategies for this study (adapted from Naumann 2007)

1	Define text words
2	Determine synonyms for text words
3	Perform test searches – I
4	Identify keywords used for indexing of databases (MeSH & Emtree)
5	Decide on whether to perform “exploded” or “focused” search for keywords
6	Check if all words are spelled correctly
7	Specify the type of search (MEDLINE – advanced search, EMBASE – multi-field search)
8	Specify fields of search (MEDLINE - Title/Abstract, EMBASE – Text Word)
9	Combine logically all search terms (Boolean – AND)
10	Perform test search – II

Table 3-2 Data extraction sheet used to locate articles containing data for item development for the Delphi first round survey (National Institute for Health and Clinical Excellence, 2008).

Study ID, authors	
<i>Type of paper</i>	
<i>Year</i>	
<i>Country</i>	
Population / search from which sample was drawn (inc controls)	
Number	
Age group	
Diagnostic / Expert group	
Study design	
Clinical Setting	
Recruitment of consecutive Patients	
Inclusion/exclusion criteria	
Methods	
Objectives specified in methods section	
Outcomes specified in methods section (inc process, criteria, etc.)	
prospective or retrospective data collection	
Test status: routine; indicated or mixture	
Results	
Number of tests performed	
Results of tests	
Number of 'significant' Abnormal results (?)	
Is test judged to be clinical Useful?	
Main findings	

3.3.1 Search protocols

I designed my search procedures (table 3.3) using a framework as recommended by the Centre for Evidence-Based Medicine (Sackett et al., 1996). I identified the population or patient problem, health status and other demographic information such as age, race, sex or location. I defined a minimum of five out of seven criteria for article inclusion. Information on the rate of outcomes in a comparison group, as well as the intervention or exposed group, was also an inclusion criterion; therefore, valid outcome and data analysis techniques had to be described. As I was investigating analytical, experimental and observational studies, search criteria relating to each study were applied in order to maintain the selection of highest evidence level research relating to each Delphi survey section. Only human-only subjects, clear aims and objectives were also required for inclusion.

Table 3-3 General systematic search protocols to identify population data and methodological criteria of located articles

Definition	Inclusion Criteria
Evidence level	<ul style="list-style-type: none"> 1a – 2b (CEBM) where possible
Language / Text	<ul style="list-style-type: none"> English and Non-English / Full text only
Aims / objectives	<ul style="list-style-type: none"> Clearly described aims or objectives
Population	<ul style="list-style-type: none"> Population definition (minimum of 5) <ol style="list-style-type: none"> Human subjects only Recruitment (where, how, who, consecutive?) exclusion / inclusion criteria age range and mean (in all samples) number of participants matched control groups more than 10 subjects Blinding (open label studies) No subject relatives in control group
Baseline criteria	<ul style="list-style-type: none"> Valid baseline measures for clinical trials Valid inclusion / exclusion criteria for study group
Intervention / Exposure	<ul style="list-style-type: none"> Valid and reliable intervention Valid and reliable outcome measures
Results	<ul style="list-style-type: none"> Appropriate data analysis techniques Statistical validation results (eg., p-values, odds ratios)
Miscellaneous	<ul style="list-style-type: none"> No duplicate reports Access to full text <ul style="list-style-type: none"> U of E database search <ul style="list-style-type: none"> U of E E-Journal search <ul style="list-style-type: none"> ILLIAD hard copy order Google scholar Google search Reviews <ul style="list-style-type: none"> Systematic <ul style="list-style-type: none"> More than one reviewer / database Incomplete methodological explanation

3.3.2 Limitations of systematic search protocols in the present study

Once I had completed the systematic search, my Ph.D. supervisors independently reviewed a random sample of 10% of selected articles using the same search protocols. Due to the large number of articles, I was not able to comprehensively review each article and subject area. As most studies had multiple aims and used large numbers of outcome measures, it was difficult to categorise studies by topic. Therefore, I classified studies according to study design as follows:

- Case-control
- Cross-sectional
- Prospective cohort studies
- Patient characteristics
- Randomised control trials
- Systematic reviews / meta-analysis
- Validation and reliability

Table 3.4 lists the features of articles found according to study type, risk factors, the patient-practitioner relationship, areas of consideration and possible measurement and multi-axial assessment.

Table 3-4 shows the number of articles found for each type of study design and the geographic spread of study location

Study type Subject	Case Control	Patient characteristics	Cross- sectional	Randomised controlled	Systematic reviews	Cohort	Validation	Geographic spread of articles
Risk factors								
Gender	5	9	5	3	2	1	N/A	9 – USA 4 – UK 3 - Canada 2 - Israel 1- France, Italy, Sweden, Japan, S Korea, Iran, India, Mexico
Physical	6	4	11	0	3	3	N/A	8 – USA 7 – UK 3 – Spain 2 – Canada, S Korea 1 – New Zealand, France, Germany, Japan, Israel
Psychological	31	14	10	2	1	2	N/A	25 – USA 7 – UK 4 – Australia 3 – Holland 2 – Germany, Turkey 1 – Norway, Pakistan, Ireland, S Korea, Canada, Peru, Iran, Belgium, Sweden, Finland, China, Spain
Impact of daily life	31	19	18	2	1	5	N/A	29 – USA 8 – UK 6 – Sweden 4 – Australia, Holland 3 – China 2 – Canada, France, Italy, Mexico 1 – Egypt, Iceland, Spain, Norway, Hungary, Croatia, Finland, Pakistan, Greece, Bangladesh

Genetic polymorphism	25	4	0	1	1	0	N/A	12 - USA 4 – Holland 3 – China, S Korea 2 – Sweden, Greece 1 – Turkey, UK, India
Therapeutic relationship	1	4	20	2	2	19	N/A	29 – USA 6 – UK 3 – Sweden 1 – France, Nigeria, Germany, Canada, Greece, Taiwan, Iran, Israel, Australia, Italy, Romania, Holland
Areas of consideration and possible measurement	13	15	12	2	3	7	20	28 - USA 6 – UK 5 – Sweden 4 – France, Iran 3 – Canada, Spain 2 – Australia, Germany, Hungary, Italy, Multinational 1 – China, Israel, Norway, Malaysia, Poland, India, Japan
Multiaxial assessment	0	2	3	1	14 (non-systematic)	0	4	12 - USA 3 – Australia 2 – Spain, Germany 1 – Canada, UK, Holland, Switzerland, Finland

3.4 Risk and contributing factors of FGIDs

3.4.1 Background

FGIDs are disorders where psychological factors can combine with environmental factors such as previous GI infections and food intolerance (Cremonini and Talley, 2005). Genetic predisposition and family aggregation also appear to be potential risk factors (Saito et al., 2005). Early life factors such as trauma, socioeconomic status and early learning of illness behaviour are also strong predictors of adult FGIDs (Chitkara et al., 2008). Evidence further shows a predominance of females presenting with FGIDs with female: male ratios reaching 3:1 (Cremonini and Talley, 2005). Physical risk factors such as previous trauma, truncal surgery and the influence of musculoskeletal disorders have been seldom if at all researched.

3.4.2 Methods

Using keywords and search terms provided in table 3.5. I summarised selected studies describing

- Physical and psychosocial factors shown to initiate and maintain FGIDs
- Inheritable and heritable genetic factors possibly affecting the modulation of GI function, psychopathology and neuro-immune function
- Gender differences in prevalence of FGIDs, the effects of sex hormones on GI sensory and motor function, psychological factors and cultural beliefs
- The impact of FGID symptoms on daily life, illness behaviour, work productivity, family function and QOL

Table 3-5 Keywords and search terms for risk factors for FGIDs

PUBMED	
MeSH Terms Limits <ul style="list-style-type: none"> Boolean Search - AND 	PubMed Keywords Limits <ul style="list-style-type: none"> advanced search – Title & Abstract Boolean search - AND
Gender	
Gender identity AND Colonic diseases Gender identity AND IBS AND prevalence Gender identity AND IBS and culture Gender identity AND IBS and socioeconomic factors Gender identity AND visceral afferents Gender identity AND CNS AND pain Gonadal steroid hormones AND Abdominal pain Gonadal steroid hormone AND IBS Gonadal steroid hormone AND oesophageal motility disorders Gonadal Steroid hormone AND visceral afferents	Gender AND IBS Gender AND Functional gastrointestinal disorders (FGIDs) Gender AND colonic dysfunction Gender AND Visceral pain Gender AND Pain modulation Gender AND IBS AND Prevalence Gender AND FGIDs AND Prevalence Gender AND cultural beliefs AND IBS / FGIDs Gender AND socioeconomic AND IBS / FGIDs Gender AND socioeconomic status AND IBS / FGIDs Gender AND gastric function Gender AND Culture AND IBS Sex hormones AND IBS / FGIDs Sex hormones AND Pain perception Sex hormones AND gastric motility Sex hormones AND colonic motility Sex hormones AND gastric function Sex hormones AND colonic function Sex Hormones AND visceral hypersensitivity
Impact on daily life	
IBS AND Quality of life (QOL) IBS AND acceptance of healthcare (major topics only) IBS AND family relations IBS AND occupational health AND economics	IBS / FGID AND Impact AND QOL IBS / FGID AND Social support IBS / FGID AND healthcare seeking IBS / FGID AND Work productivity IBS/ FGID AND Hypervigilance IBS / FGID AND Altered bowel habit
Physical risk factors	
IBS AND Risk factors IBS AND inflammatory bowel disease (IBD) AND risk factors Gastrointestinal disease AND IBS AND risk factors IBS AND Trauma IBS AND surgery (inc subheading – adverse effects) IBS AND colorectal surgery IBS AND Joint instability IBS AND Nervous system trauma IBS AND brain trauma IBS AND brain injury IBS AND injuries (subheading)	IBS / FGID AND Joint hypermobility IBS / FGID AND Surgery IBS / FGID AND abdominal surgery IBS / FGID AND Myofascial IBS / FGID AND previous gastroenteritis IBS / FGID AND inflammatory bowel disease (IBD) IBS / FGID AND risk factors IBS / FGID AND musculoskeletal dysfunction IBS / FGID AND physical trauma / physical injury / pathological risk factors / traumatic brain injury IBS / FGID AND previous pathology Endometriosis AND myofascial pain syndrome

Gynaecology AND myofascial pain syndrome	IBS / FGID AND somatic dysfunction IBS / FGID AND somatic comorbidity
Genetic polymorphism	
Genetic polymorphism AND Psychology Genetic polymorphism AND IBS	IBS / FGID AND Polymorphism
Psychological risk factors	
IBS AND Child abuse IBS AND Sex offences IBS AND Personality disorders IBS AND mental disorders AND risk factors IBS AND Hypothalamo-hypophyseal system IBS AND eating disorders	IBS / FGID AND Abuse IBS / FGID AND early life IBS / FGID AND autonomic dysfunction IBS / FGID AND altered gut physiology IBS / FGID AND personality traits IBS / FGID AND Psychiatric disorders IBS / FGID AND Eating disorders
EMBASE	
EMTREE limits <ul style="list-style-type: none"> • Exclude Medline journals • Boolean search - AND • All subject heading and subheadings – focus (specificity) • mp (multi-purpose) – free text keyword search <ul style="list-style-type: none"> ○ SH – subject headings 	EMBASE Keywords Limits <ul style="list-style-type: none"> • Exclude Medline journals • Multifield search • Field – Text Word • Boolean search - AND
Gender	
IBS / FGID IBS. mp and focus (irritable colon) / FGID – mp and focus (digestive system function disorder) IBS / FGID AND Gender <ul style="list-style-type: none"> • IBS – focus (irritable colon) / FGID – mp and focus (digestive system function disorder) Gender AND Colonic function <ul style="list-style-type: none"> • <i>SH – Colon, colon motility, diarrhoea, colonic function, colonic, intestine function</i> Gender AND visceral pain <ul style="list-style-type: none"> • mp & visceral pain Gender AND PAIN modulation <ul style="list-style-type: none"> • <i>SH - analgesia, beta endorphin, central nervous system, modulation, morphine, naloxone, nociception, nociceptive receptor, pain, pain modulation, pain threshold</i> Gender AND Prevalence AND IBS / FGID <ul style="list-style-type: none"> • mp & prevalence) Gender AND IBS / FGID AND Cultural Beliefs <ul style="list-style-type: none"> • <i>SH - attitude to health beliefs, cultural, cultural anthropology, cultural beliefs, cultural factor, health behaviour, psychological aspect, religion, women's</i> 	Gender AND IBS Gender AND Functional gastrointestinal disorders (FGIDs) Gender AND colonic dysfunction Gender AND Visceral pain Gender AND Pain modulation Gender AND IBS AND Prevalence Gender AND FGIDs AND Prevalence Gender AND cultural beliefs AND IBS / FGIDs Gender AND socioeconomic AND IBS / FGIDs Gender AND socioeconomic status AND IBS / FGIDs Gender AND gastric function Gender AND Culture AND IBS Sex hormones AND IBS / FGIDs Sex hormones AND Pain perception Sex hormones AND gastric motility Sex hormones AND colonic motility Sex hormones AND gastric function Sex hormones AND colonic function Sex Hormones AND visceral hypersensitivity

<p><i>health</i></p> <p>Gender AND Socioeconomic AND IBS / FGIDs</p> <ul style="list-style-type: none"> • mp & socioeconomics <p>Gender AND socioeconomic status AND IBS / FGID</p> <ul style="list-style-type: none"> • mp & social status <p>Sex hormones AND IBS / FGIDs</p> <ul style="list-style-type: none"> • SH – sex hormones mp or sex hormone <p>Sex hormones AND Pain perception</p> <p>Sex hormones AND Gastric motility</p> <ul style="list-style-type: none"> • mp. Gastric motility, motility, stomach motility <p>Sex hormones AND Colonic motility</p> <ul style="list-style-type: none"> • Mp. Gastric motility, motility, stomach motility <p>Sex hormones AND colonic motility</p> <p>Sex hormones AND IBS / FGID</p> <p>Sex Hormones AND Visceral Hypersensitivity</p> <ul style="list-style-type: none"> • SHs - serotonin antagonist, pain, irritable colon, hyperalgesia, hypersensitivity, gastrointestinal disease, visceral pain, viscera hypersensitivity, neurokinin 1 receptor antagonist 	
Impact on daily life	
<p>IBS / FGID (as / gender) AND QOL AND Impact</p> <p>IBS / FGID AND social support</p> <p>IBS / FGID AND healthcare seeking</p> <ul style="list-style-type: none"> • focus – healthcare seeking behaviour <p>IBS / FGID AND work productivity</p> <p>IBS / FGID AND hypervigilance</p> <ul style="list-style-type: none"> • SH – attention <p>IBS / FGID AND altered bowel habit</p> <ul style="list-style-type: none"> • SH – defecation habit 	<p>IBS / FGID AND Impact AND QOL</p> <p>IBS / FGID AND Social support</p> <p>IBS / FGID AND healthcare seeking</p> <p>IBS / FGID AND Work productivity</p> <p>IBS/ FGID AND Hypervigilance</p> <p>IBS / FGID AND Altered bowel habit</p>
Physical risk factors	
<p>IBS / FGID AND Joint hypermobility</p> <ul style="list-style-type: none"> • SH – joint instability, hypermobility, Marfan’s syndrome <p>IBS / FGID and Surgery</p> <ul style="list-style-type: none"> • All SHs <p>IBS / FGID AND Abdominal surgery</p> <ul style="list-style-type: none"> • All SHs <p>IBS / FGID AND Myofascial pain</p> <ul style="list-style-type: none"> • SH – diagnosis, epidemiology, Aetiology <p>IBS / FGID AND Physical injury</p> <ul style="list-style-type: none"> • Injury, physical disability, physical disease, post-traumatic stress disorder <p>IBS / FGID AND Physical trauma</p> <ul style="list-style-type: none"> • SH – Injury, head injury, risk factor 	<p>IBS / FGID AND Joint hypermobility</p> <p>IBS / FGID AND Surgery</p> <p>IBS / FGID AND abdominal surgery</p> <p>IBS / FGID AND Myofascial</p> <p>IBS / FGID AND previous gastroenteritis</p> <p>IBS / FGID AND inflammatory bowel disease (IBD)</p> <p>IBS / FGID AND risk factors</p> <p>IBS / FGID AND musculoskeletal dysfunction</p> <p>IBS / FGID AND physical trauma / physical injury / pathological risk factors / traumatic brain injury</p> <p>IBS / FGID AND previous pathology</p> <p>Endometriosis AND myofascial pain syndrome</p> <p>IBS / FGID AND somatic dysfunction</p> <p>IBS / FGID AND somatic comorbidity</p>
Genetic polymorphism	
<p>Polymorphism AND IBS / FGID</p> <p>IBS / FGID AND HPA axis</p> <ul style="list-style-type: none"> • mp 	<p>IBS / FGID AND Polymorphism</p>

Psychological risk factors	
IBS / FGID AND Abuse <ul style="list-style-type: none"> • SH – child, child sexual, sexual, human rights abuse IBS / FGID AND Early life <ul style="list-style-type: none"> • mp IBS / FGID AND ANS dysregulation <ul style="list-style-type: none"> • SH – autonomic dysfunction and mp IBS / FGID AND Altered gut physiology <ul style="list-style-type: none"> • mp IBS / FGID AND Personality traits <ul style="list-style-type: none"> • SH – anxiety disorder, personality, personality disorder & mp) IBS / FGID AND Eating disorders <ul style="list-style-type: none"> • All Subheadings 	IBS / FGID AND Abuse IBS / FGID AND early life IBS / FGID AND autonomic dysfunction IBS / FGID AND altered gut physiology IBS / FGID AND personality traits IBS / FGID AND Psychiatric disorders IBS / FGID AND Eating disorders

3.4.3 Results

The search strategy identified 4134 citations. From these, I identified 447 articles that appeared to be relevant to risk and contributing factors. Using systematic search protocols shown in table 3.3, I accepted 219 articles (Appendix D)

3.5 The patient-practitioner relationship

3.5.1 Background

Social support, including the patient-practitioner relationships has been associated with improved health outcomes (Conboy et al., 2010). Patients satisfied with care are more likely to be self-confident, motivated and follow advice resulting from good communication (Greenfield et al., 1988). Conversely, patients unhappy with care are more likely to make repeat visits and change clinicians which weaken the effect of the medical encounter (Conboy et al., 2010). Furthermore, low levels of social support have been linked to altered immune function in both observational and experimental studies (Umberson and Montez, 2010). These factors are important to FGIDs as many patients have psychosocial variables that have been shown to play a significant role in symptom expression (Heitkemper et al., 2012).

3.5.2 Methods

Using keywords and search terms provided in table 3.6, I summarised selected studies describing

- FGID patient health care expectations
- Patient and practitioner perceptions of FGIDs
- Knowledge and awareness of health care professionals to FGIDs
- Patient and clinician-centred education
- Clinician knowledge of co-morbidity associated with FGIDs

Table 3-6 Keywords and search terms for the patient-practitioner relationships in FGIDs.

PubMed	
MeSH Terms Limits Boolean Search - AND	PubMed Keywords Limits <ul style="list-style-type: none"> • advanced search – Title & Abstract • Boolean search - AND
IBS AND clinical competence IBS AND interdisciplinary communication Clinical competence AND Humanism IBS AND cultural characteristics IBS AND ethnology IBS AND cross-cultural comparisons IBS AND ethics IBS AND patient care management IBS AND attitudes to health AND patient IBS AND Health knowledge attitudes practice IBS AND social class IBS AND educational status IBS AND physician-patient relationship	IBS/ FGID AND Clinical competence IBS / FGID AND Clinical experience IBS / FGID AND Clinician / physician experience IBS / FGID AND multidisciplinary approach IBS / FGID AND Multidisciplinary IBS /FGID AND Humanistic burden IBS AND FGID AND Cultural impact IBS / FGID AND ethical consideration IBS / FGID AND ethics IBS / FGID AND symptom experience IBS / FGID AND patient care IBS / FGID AND patient perspective IBS /FGID AND clinician / physician attitudes IBS / FGID AND attitudes IBS / FGID AND Cross-cultural IBS / FGID AND symptom interpretation IBS / FGID AND symptom reporting Emotion AND symptom reporting Patient attitudes AND health care Patient attitudes AND treatment options IBS / FGID AND socioeconomic IBS / FGID AND educational status Patient interview AND techniques IBS / FGID AND Patient expectation Patient-practitioner relationship
EMBASE	
EMTREE limits <ul style="list-style-type: none"> • Exclude Medline journals • Boolean search - AND • All subject heading and subheadings – focus (specificity) • mp (multi-purpose) – free text keyword search SH – subject headings	EMBASE Keywords Limits <ul style="list-style-type: none"> • Exclude Medline journals • Multifield search • Field – Text Word Boolean search - AND
IBS / FGID <ul style="list-style-type: none"> • IBS. mp and focus (irritable colon) / FGID – mp and focus (digestive system function disorder) IBS / FGID AND Patient interview .mp <ul style="list-style-type: none"> • SH – doctor-patient relation, medical education, interview, interpersonal communication 	IBS / FGIDs AND patient interview IBS / FGIDs AND patient expectation IBS / FGID AND physician-patient relationship Clinical reasoning AND self-reflection Cross-cultural AND IBS IBS / FGID AND Symptom expression IBS / FGID AND symptom interpretation IBS / FGID AND Symptom reporting

<p>IBS / FGID AND Patient expectation .mp</p> <ul style="list-style-type: none"> SH – patient satisfaction, patient, patient attitude, expectation <p>IBS / FGID AND Clinical reasoning .mp</p> <ul style="list-style-type: none"> SH – clinical competence, problem-solving, decision making, medical decision making <p>IBS / FGID AND self-reflection .mp</p> <ul style="list-style-type: none"> SH professional competence, self-evaluation, public-relations, thinking <p>IBS / FGID AND symptom reporting. mp</p> <ul style="list-style-type: none"> SH – self-report, symptomatology <p>Emotion .mp (SH – emotion) AND symptom reporting</p> <ul style="list-style-type: none"> SH – self-report, symptomatology <p>Patient attitude .mp</p> <ul style="list-style-type: none"> SH patient attitude AND Healthcare .mp SH – health care <p>IBS / FGID AND socioeconomic .mp</p> <ul style="list-style-type: none"> SH- socioeconomics <p>IBS FGID AND educational status.mp</p> <ul style="list-style-type: none"> SH – educational status <p>IBS / FGID AND Humanistic burden .mp</p> <p>IBS / FGID AND Cultural impact .mp</p> <ul style="list-style-type: none"> SH – cultural anthropology, attitude to health, cultural factors <p>IBS / FGID AND ethical considerations .mp</p> <ul style="list-style-type: none"> SH – medical ethics, ethics <p>IBS / FGID AND patient care .mp</p> <ul style="list-style-type: none"> SH – patient care <p>IBS / FGID AND clinician attitudes .mp</p> <p>IBS / FGID AND patient perspective .mp</p> <p>IBS / FGID AND Attitude .mp</p> <p>SH – social attitude, attitude to life, patient attitude to health, family attitude, attitude to disability, attitude to illness</p>	<p>Patient attitude AND interview</p> <p>Emotion AND symptom reporting</p> <p>IBS / FGID AND Socioeconomic</p> <p>IBS / FGID AND educational status (0)</p> <p>IBS / FGID AND clinical competence</p> <p>IBS / FGID AND clinical experience</p> <p>IBS / FGID AND humanistic burden</p> <p>IBS / FGID AND cultural impact</p> <p>IBS / FGID AND ethical considerations</p> <p>IBS / FGID AND ethics</p> <p>IBS / FGID AND symptom experience</p> <p>IBS / FGID AND patient care</p> <p>IBS / FGID AND patient perspective</p> <p>IBS / FGID AND clinician attitudes</p> <p>IBS / FGID AND Attitudes</p>
--	---

3.5.3 Results

The search strategy identified 927 citations. I identified 148 potential articles concerning the patient-practitioner relationship. Using systematic search protocols shown in table 3.3, I selected 47 articles (appendix D).

3.6 Areas for consideration and possible measurement

3.6.1 Background

Clinicians must rely on patient symptoms to make a diagnosis. The severity of symptoms can be confounded by both poor patient recall and affected by comorbidities, which may also need investigation. FGIDs have a significant impact on physical, emotional and cognitive function, as well as on social and family life and the ability to work. Therefore, it is important not only to elicit the number and severity of disorders but also their comorbidities. Assessment must, therefore, both reflect several areas of information to measure the net effect of FGIDs and their co-morbidities on daily function.

3.6.2 Methods

Using keywords and search terms provided in table 3.7. I summarised selected studies describing outcome measures evaluating:

- GI symptoms
- Psychological symptoms
- Belief systems and their influence on GI symptoms
- Validity and reliability testing of PROs associated with FGIDs
- Validity of English PROs translated into other languages.

Table 3-7 Keywords and search terms for areas of consideration and possible measurement.

PubMed	
MeSH Terms Limits Boolean Search - AND	PubMed Keywords Limits <ul style="list-style-type: none"> • advanced search – Title & Abstract • Boolean search - AND
<ul style="list-style-type: none"> • Personality tests AND IBS • Personality tests AND reproducibility of results AND IBS • Disability evaluation AND IBS Catastrophizing AND IBS • Dysthymic disorder AND IBS • Adaptation AND Psychological AND IBS • Fatigue AND IBS • Aged AND GI motility AND IBS <ul style="list-style-type: none"> • Activities of daily living AND IBS • Disability evaluation AND IBS • Interpersonal relations AND IBS • Occupational function AND IBS Dependence (psychology) AND IBDS <ul style="list-style-type: none"> • Signs and symptoms (digestive) AND diagnosis (subheading) AND IBS subheading (diagnosis) • Pain perception AND IBS 	<ul style="list-style-type: none"> • Anxiety scale AND IBS • Anxiety scale AND FGID • Anxiety rating AND IBS / FGID • Anxiety AND validity AND IBS • Anxiety AND validity AND FGID • Depression AND validity AND IBS • Depression AND validity AND FGID • Depression scale AND IBS • Depression scale AND FGID • Fear avoidance AND IBS / FGID • Kinesiophobia AND IBS / FGID • Fear avoidance AND validity AND measurement • Kinesiophobia AND validity AND measurement • Catastrophizing AND IBS / FGID • Hopelessness AND IBS / FGID • Hopelessness AND validity AND measurement • Coping strategies AND IBS • Coping strategies AND FGID • IBS AND fatigue • FGID AND fatigue • Elderly AND IBS • FGID AND elderly • IBS AND aged • FGID AND aged • Gastrointestinal (GI) motility AND elderly • Manning criteria AND reliability • Rome III criteria AND reliability • Daily activity AND IBS / FGID • Physical disability AND IBS /FGID • Social disability AND IBS / FGID • Interpersonal AND IBS • Interpersonal AND FGID • Dependence AND IBS / FGID • Rome criteria AND symptoms • Abdominal symptoms AND IBS / FGID diagnosis

	<ul style="list-style-type: none"> • Symptomatic diagnosis AND IBS / FGID • Common symptoms AND IBS • Common symptoms AND FGID • Symptom evaluation AND IBS / FGID • IBS AND perception AND symptoms • FGID AND perception AND symptoms • IBS / FGID AND symptom AND description • Alarm symptoms AND IBS • Alarm symptoms AND FGID • FGID / IBS AND Occupational function
EMBASE	
EMTREE limits <ul style="list-style-type: none"> • Exclude Medline journals • Boolean search - AND • All subject heading and subheadings – focus (specificity) • mp (multi-purpose) – free text keyword search SH – subject headings	EMBASE Keywords Limits <ul style="list-style-type: none"> • Exclude Medline journals • Multifield search • Field – Text Word • Boolean search - AND
IBS / FGID IBS. mp and focus (irritable colon) / FGID – mp and focus (digestive system function disorder) IBS AND Anxiety scale mp. <ul style="list-style-type: none"> • (SH – questionnaire, psychometry, anxiety, personality test, rating scale, anxiety disorder) FGID AND Anxiety scale <ul style="list-style-type: none"> • (SH - as for IBS) IBS AND Anxiety rating mp. <ul style="list-style-type: none"> • (SH - anxiety disorder, psychological rating scale, rating scale) FGID AND Anxiety rating mp. <ul style="list-style-type: none"> • (SH – as / IBS) IBS AND Validity mp. <ul style="list-style-type: none"> • (SH - external validity, face validity, construct validity, internal validity, criterion-related validity, concurrent validity, qualitative validity) AND Anxiety (SH – Hospital anxiety, anticipatory anxiety, generalised anxiety disorder, Hamilton anxiety scale, state-trait anxiety Inventory, Self-rating anxiety scale,) FGID AND Validity (as / IBS), AND anxiety (as / IBS) IBS AND validity AND Depression <ul style="list-style-type: none"> • (SH – depression inventory, hospital anxiety and depression scale, Beck depression 	<ul style="list-style-type: none"> • Anxiety scale AND IBS • Anxiety scale AND FGID • Anxiety rating AND IBS / FGID • Anxiety AND validity AND IBS • Anxiety AND validity AND FGID • Depression AND validity AND IBS • Depression AND validity AND FGID • Depression scale AND IBS • Depression scale AND FGID • Fear avoidance AND IBS / FGID • Kinesiophobia AND IBS / FGID • Fear avoidance AND validity AND measurement • Kinesiophobia AND validity AND measurement • Catastrophizing AND IBS / FGID • Hopelessness AND IBS / FGID • Hopelessness AND validity AND measurement • Coping strategies AND IBS • Coping strategies AND FGID • IBS AND fatigue • FGID AND fatigue • Elderly AND IBS • FGID AND elderly • IBS AND aged • FGID AND aged • Gastrointestinal (GI) motility AND

<p>inventory, self-rating depression scale)</p> <p>FGID AND Validity AND Depression</p> <p>IBS AND depression scale mp.</p> <ul style="list-style-type: none"> • (SH – reliability, questionnaire, rating scale, anxiety rating) <p>FGID AND Depression scale</p> <p>IBS AND fear avoidance mp.</p> <ul style="list-style-type: none"> • (SH – disability, avoidance behaviour, fear, psychological aspects) <p>FGID AND fear avoidance</p> <p>IBS / FGID AND Kinesiophobia mp.</p> <ul style="list-style-type: none"> • (SH - fear) Fear avoidance AND Validity AND Measurement mp. <p>Kinesiophobia AND Validity AND Measurement</p> <p>IBS / FGID AND Catastrophizing mp</p> <ul style="list-style-type: none"> • (SH - catastrophizing) <p>IBS / FGID AND Hopelessness mp.</p> <ul style="list-style-type: none"> • (SH – hopelessness and Beck Hopelessness) <p>IBS AND Coping strategies mp.</p> <ul style="list-style-type: none"> • (SH - coping behaviour) <p>FGID AND Coping strategies (IBS AND fatigue mp.</p> <ul style="list-style-type: none"> • (SH –fatigue) <p>FGID AND fatigue</p> <p>IBS and aged mp.</p> <p>FGID AND aged</p> <p>Gastrointestinal motility is mp.</p> <ul style="list-style-type: none"> • (SH – gastrointestinal motility) AND aged <p>Manning Criteria mp. AND reliability mp.</p> <ul style="list-style-type: none"> • (SH – reliability • Rome III criteria mp. AND reliability <p>IBS / FGID AND Daily activity mp</p> <ul style="list-style-type: none"> • (SH – Daily life activity) <p>IBS AND Physical disability mp.</p> <ul style="list-style-type: none"> • (SH – physical disability) <p>FGID and Physical disability</p> <p>IBS / FGID AND Social disability mp.</p> <ul style="list-style-type: none"> • (SH –Social disability) <p>IBS AND interpersonal mp.</p> <ul style="list-style-type: none"> • (SH – interpersonal communication, interpersonal stress) • FGID AND interpersonal <p>Rome III criteria AND symptoms mp.</p> <ul style="list-style-type: none"> • (SH – brief symptom inventory, symptoms) <p>IBS AND Abdominal symptoms mp.</p> <ul style="list-style-type: none"> • (SH – abdomen, gastrointestinal symptoms) AND diagnosis mp. (SH – diagnosis) <p>FGID AND Abdominal symptoms AND diagnosis</p> <p>IBS / FGID AND Symptomatic diagnosis mp.</p> <p>IBS AND Abdominal symptoms AND evaluation mp.</p> <ul style="list-style-type: none"> • (SH – clinical evaluation, evaluation) 	<p>elderly</p> <ul style="list-style-type: none"> • Manning criteria AND reliability • Rome III criteria AND reliability • Daily activity AND IBS / FGID • Physical disability AND IBS /FGID • Social disability AND IBS / FGID • Interpersonal AND IBS • Interpersonal AND FGID • Dependence AND IBS / FGID • Rome criteria AND symptoms • Abdominal symptoms AND IBS / FGID diagnosis • Symptomatic diagnosis AND IBS / FGID • Common symptoms AND IBS • Common symptoms AND FGID • Symptom evaluation AND IBS / FGID • IBS AND perception AND symptoms • FGID AND perception AND symptoms • IBS / FGID AND symptom AND description • Alarm symptoms AND IBS • Alarm symptoms AND FGID <p>FGID / IBS AND Occupational function</p>
--	---

FGID AND abdominal symptoms AND evaluation IBS AND abdominal symptoms AND perception mp. • (SH –perception) FGID AND abdominal symptoms AND perception IBS AND Alarm symptoms mp. FGID AND Alarm symptoms	
--	--

3.6.3 Results

The search strategy identified 2208 citations. I identified 239 articles that appeared to be relevant to risk and contributing factors. Using previously described search criteria, I selected 78 (Appendix D).

3.7 Multiaxial assessment

3.7.1 Background

Patient-centred psychiatry and medicine are emerging as new concepts in response to the recognition of current inadequacies in health care of functional disorders. Using a multidimensional framework allows evaluation of the whole person rather than a particular disease process. This formulation should also provide contextual and standardised description of a clinical condition. (IGDA WORKGROUP, 2003). Although multiaxial assessments such as DSM-IV / V and ICD-10 are not without their problems, recent studies show that the multiaxial approach aids clinical diagnosis and predicts the outcome after treatment (Saavedra et al., 2001).

3.7.2 Methods

Using keywords and search terms provided in table 3.8. I summarised selected studies describing MAC in other clinical fields:

- The strengths and robustness of MAC
- Consensus among experts on uses of MAC
- The validity and implementation of MAC

Table 3-8 Keywords and search terms for studies researching the use of multiaxial assessment.

PubMed	
<p>MeSH Terms Limits Boolean Search – AND SH - subheading</p>	<p>PubMed Keywords Limits</p> <ul style="list-style-type: none"> • advanced search – Title & Abstract (T/A) • All fields (A/F) • Boolean search - AND
<ul style="list-style-type: none"> • Taxonomy AND patient assessment (SH Outcome assessment, health care) • Mental disorders (SH diagnosis, classification , ethnology) 	<ul style="list-style-type: none"> • IBS / FGID AND Multiaxial assessment • IBS / FGID AND Multidimensional assessment • IBS / FGID AND multiaxial diagnostic criteria • IBS / FGID AND multiaxial classification • IBS / FGID AND biopsychosocial assessment • IBS AND Taxonomy • FGID AND Taxonomy • Multiaxial classification AND DSM • Multiaxial classification AND WPA • Multiaxial classification AND ICD (Multiaxial classification AND IASP) • Multiaxial assessment AND ICD • Multiaxial assessment AND DSM Multiaxial assessment AND IASP Multiaxial assessment AND WPA • Multidimensional assessment AND DSM • Multidimensional assessment AND IASP • Multidimensional assessment AND WPA • Multidimensional assessment AND ICD • Multiaxial assessment AND Pain • Multiaxial assessment AND Chronic pain • Multiaxial classification AND Pain • Multiaxial classification AND Chronic pain • Multidimensional assessment AND chronic pain • Multidimensional assessment AND pain • Multiaxial assessment AND development • Multidimensional assessment AND development • Multiaxial classification AND development • Multiaxial assessment AND psychosomatic. • Multidimensional assessment AND

	<p>psychosomatic</p> <ul style="list-style-type: none"> • Multiaxial classification AND psychosomatic • Diagnostic assessment AND Comprehensive • Multidimensional assessment AND Validation • Multidimensional assessment AND reliability • Multiaxial diagnosis AND validation • Multiaxial diagnosis AND reliability • Multiaxial classification AND Validation • Multiaxial classification AND Reliability • Multiaxial taxonomy AND Validation • Multiaxial taxonomy AND reliability • Multiaxial assessment AND Validation • Multiaxial assessment AND Reliability • Multiaxial assessment AND biopsychosocial • person-centred integrative diagnosis • DSM AND Advantages • ICD-10 AND Advantages • Psychiatric diagnosis AND relevance • Multiaxial classification AND Implications • Multiaxial assessment AND implications • Multidimensional assessment (T/A) and implications (AF) • Multiaxial assessment AND relevance • Psychiatric nosology AND relevance • Psychiatric nosology AND advantages
EMBASE	
EMTREE limits <ul style="list-style-type: none"> • Exclude Medline journals • Boolean search - AND • All subject heading and subheadings – focus (specificity) • mp (multi-purpose) – free text keyword search <ul style="list-style-type: none"> ○ SH – subject headings 	EMBASE Keywords Limits <ul style="list-style-type: none"> • Exclude Medline journals • Multifield search • Field – Text Word • Boolean search - AND
IBS / FGID IBS. mp and focus (irritable colon) / FGID – mp and focus (digestive system function disorder) <ul style="list-style-type: none"> • IBS AND Multiaxial assessment - mp. (SH – personality disorder, chronic pain, psychological aspect) 	<ul style="list-style-type: none"> • IBS / FGID AND Multiaxial assessment • IBS / FGID AND Multidimensional assessment • IBS / FGID AND multiaxial diagnostic criteria • IBS / FGID AND multiaxial classification • IBS / FGID AND biopsychosocial assessment • IBS AND Taxonomy • FGID AND Taxonomy

<ul style="list-style-type: none"> • FGID AND Multiaxial assessment • IBS AND Multiaxial classification – mp. (SH depression, comorbidity) • FGID AND Multiaxial classification • IBS AND Multidimensional assessment – mp. (SH – psychological aspect, pain, health status) • FGID AND Multidimensional assessment • IBS AND Biopsychosocial assessment – mp. (SH chronic pain, social psychology, psychological aspects) • IBS AND Taxonomy – mp. (SH – taxonomy) • FGID AND Taxonomy • Multiaxial – mp (SH – disease classification) AND DSM – mp. (SH diagnostic and statistical manual of mental disorders, structured clinical interview for DSM disorders) • Multiaxial AND World psychiatric association – mp. • Multiaxial AND ICD – mp. (SH international classification of diseases) • Multiaxial AND IASP mp. (SH – chronic pain, international association for the study of pain, pain assessment) • Multidimensional mp. (SH – multidimensional scaling) AND DSM • Multidimensional AND World psychiatric association • Multidimensional AND ICD • Multidimensional AND IASP • Multiaxial assessment mp. AND chronic pain mp. • Multiaxial assessment and pain mp (SH - *referred pain/ or *upper abdominal pain/ or *epigastric pain/ or stomach pain/ or *psychogenic pain/ or *lower abdominal pain/ or *chronic pain/ or *visceral pain/ or *abdominal pain/ or *retrosternal pain/ or pain.mp. or gastrointestinal pain/ or *esophagus pain/ or pain/) • Multiaxial classification mp. (SH – comorbidity) AND chronic pain • Multiaxial classification AND pain • Multidimensional assessment mp. AND chronic pain • Multidimensional assessment AND pain • Multiaxial assessment AND Validation mp. (SH – instrument validation, validation study, validation process) 	<ul style="list-style-type: none"> • Multiaxial classification AND DSM • Multiaxial classification AND WPA • Multiaxial classification AND ICD (Multiaxial classification AND IASP) • Multiaxial assessment AND ICD • Multiaxial assessment AND DSM Multiaxial assessment AND IASP Multiaxial assessment AND WPA • Multidimensional assessment AND DSM • Multidimensional assessment AND IASP • Multidimensional assessment AND WPA • Multidimensional assessment AND ICD • Multiaxial assessment AND Pain • Multiaxial assessment AND Chronic pain • Multiaxial classification AND Pain • Multiaxial classification AND Chronic pain • Multidimensional assessment AND chronic pain • Multidimensional assessment AND pain • Multiaxial assessment AND development • Multidimensional assessment AND development • Multiaxial classification AND development • Multiaxial assessment AND psychosomatic. • Multidimensional assessment AND psychosomatic • Multiaxial classification AND psychosomatic • Diagnostic assessment AND Comprehensive • Multidimensional assessment AND Validation • Multidimensional assessment AND reliability • Multiaxial diagnosis AND validation • Multiaxial diagnosis AND reliability • Multiaxial classification AND Validation • Multiaxial classification AND Reliability • Multiaxial taxonomy AND Validation • Multiaxial taxonomy AND reliability • Multiaxial assessment AND Validation • Multiaxial assessment AND Reliability • Multiaxial assessment AND biopsychosocial • person-centred integrative diagnosis • DSM AND Advantages
---	--

<ul style="list-style-type: none"> • Multiaxial assessment AND reliability mp. (SH – inter-rater reliability, inter-rater reliability, test-retest reliability, reliability) • Multiaxial classification AND validation • Multiaxial classification AND reliability • Multidimensional assessment mp. AND validation • Multidimensional assessment AND reliability • Multiaxial assessment AND psychosomatic mp. (SH – psychosomatics) • Multiaxial classification AND psychosomatic • Multidimensional assessment AND psychosomatic • Multiaxial assessment AND biopsychosocial assessment mp. (SH – social psychology, clinical assessment, psychometry, depression) • DSM AND Advantages mp. • ICD-10 - mp. (SH - coding, classification) AND Advantages • Multiaxial AND implications – mp. • Multidimensional AND implications • Multiaxial AND relevance mp. • Multidimensional AND relevance • Psychiatric nosology mp. (SH disease classification, psychiatric diagnosis) • Psychiatric nosology AND advantages 	<ul style="list-style-type: none"> • ICD-10 AND Advantages • Psychiatric diagnosis AND relevance • Multiaxial classification AND Implications • Multiaxial assessment AND implications • Multidimensional assessment and implications • Multiaxial assessment AND relevance • Psychiatric nosology AND relevance • Psychiatric nosology AND advantages
--	--

3.7.3 Results

I identified 1618 citations. From these, I identified 75 articles that appeared to be relevant to risk and contributing factors. Using previously described search protocols, I selected 29 articles (Appendix D).

3.8 Summary

Information and data from accepted studies were used to construct items for the Delphi study first round. Appendix C shows flow charts that describe the search processes for each subject area shown in table 3.4. Appendix D contains tabulated summaries of all selected articles

that show the authors, type of data collected, the population studies, sample size, location of study, type of control group if applicable, and relevant finding.

Chapter 4. Philosophy and Methodology Underlying the Delphi Technique

4.1 Overview

This chapter describes inquiry systems (IS) applicable to Delphi technique, the methodological considerations of the Delphi technique generally, and those specific to my study.

4.2 Introduction

Empirical surveys are challenging. McGrath (1981), Knott et al. (2012) lists the compromises required in design that relate to choices of population and data collection. Inquiry systems differ in application and process. For example, inquiry may begin with an axiom, or with raw data, or with observations which are considered as characteristic of reality (Linstone and Turoff, 2002, p20). However, in order to develop practical concepts, these initial features must be summarised in order to find a “form” that can be presented as information or as models that can be represented by rules such as algorithms or estimates. This processing may be done several times so that decision makers recognise and correctly apply acquired information. Before developing my Delphi study, it was important to distinguish different IS theories, so its design was suited to not only the research question, but also new and existing theories and currently available data.

4.3 Inquiry Systems

Charles Churchman first proposed formal IS. He considered the works of philosophers and developed a classification of rationalism (Leibnizian), empiricism (Lockean), phenomenological (Kantian), dialectical (Hegelian) and pragmatism (Singerian)

(Churchman, 1971, Stevens, 1975). Churchman investigated basic types of information used as the building blocks for inquiry that are then applied to philosophical knowledge systems. Importantly, Churchman also examined the role and issue of the “guarantor” (the component that judges the relevance and competence of the system).

4.3.1 Inquiry systems and the Delphi technique

Depending on the research question, several inquiry systems may be applicable to the Delphi technique. The IS with most associations is Lockean (empiricist) inquiry. Empiricist inquiry operates on the principle that factual questions based on experiential and observational elements can be tested with expert consensus methods. Initial raw data, observations, sensations and questions are addressed by building empirical representations of them through group agreement. Here, data is transformed into summarised information that is more understandable, where knowledge can be used for future actions (Courtney et al., 2007). The guarantor of empiricist IS lies both with the expertise of the panel and the “tightness of agreement” (Mitroff and Turoff, 2002, p21). Mitroff and Turoff (2002, p20) suggest it is hard to find a better example of a Lockean inquirer than the Delphi technique. Firstly, the raw data are either the opinions of the experts (classic Delphi technique) or information from published literature and secondly, the validity of resulting judgements is measured by the degree of consensus and equally important, the tightness of agreement. Although empiricist IS can gather rich sources of experiential data, uncertain expertise leads to a lack of uncertainty, i.e., a lack of trust. Furthermore, agreement can also suppress conflict or debate when they are needed most, therefore; empiricist IS are suited for working with well-defined, singular problems where there is a strong consensual position. Good examples in healthcare are studies examining guideline development for measurement in a particular area of care relating to a single procedure (Knott et al., 2012) or a definable disease (Lodewijckx et al., 2013).

However, I investigated an area of clinical evaluation where there are many areas of information and uncertainty involving experts from many different specialities. Thus, an IS is required that allows the integration of various models where valuable data is produced from many different perspectives. In these cases, Kantian IS is well-suited. Kantian IS relies on input of both theories and empirical observations that are considered by experts or decision-makers from different backgrounds to develop solutions for their problems. Data and theory are regarded as inseparable; thus, Kantian IS are viewed as multimodal where the inquirer can consider many alternative data forms allowing them create a “best-fit” for a presenting problem (Mitroff and Turoff, 2002, Courtney et al., 2007). The guarantor for Kantian IS is the degree of fit between the primary theory(s) and empirical observations from attitudinal data collected from participants. However, due to multiple perspectives from participants, an input is subject to different interpretations where there is no guarantee that the model represents the best solution (Malhotra, 1997, Courtney et al., 2007). Complex, semi-structured problems such as planning and cost-benefit analysis are examples of Kantian enquiry. Delphi studies now use Kantian characteristics where the technique can elicit alternative judgment, producing an overview of an issue where the problem is broader in scope than any one area of expertise, a good healthcare example being international guidelines for pain management (Kumar, 2007).

However, as often occurs with ill-structured problems where there are multiple areas of information, uncertainty and perspectives, there is also conflict. While Kantian inquiry sees participants with different experiences working towards the same goal, Hegelian inquiry relies on constructive conflict to gain data. Hegelian IS operates on the principle that improved results come from the conflict between opposing ideas or plans (Courtney et al., 2007). Hegelian inquiry is more associated with policy Delphi, and its discussion is beyond the scope of the present study. Other forms of inquiry exist. These are Leibnizian inquiry that operates on the principle that results come from formal mathematical modelling processes

and Singerian inquiry that uses a wide range of interchangeable inputs to solve problems by any means possible.

4.3.2 Inquiry systems and the present study

To develop MAC for FGIDs, different areas of knowledge are needed to provide information on these interacting systems. Given the uncertainty of pathophysiological mechanisms and the different contributions from heterogeneous experts, Kantian inquiry is perhaps the best suited. Moreover, as is important to the evaluation of FGIDs, Kantian inquiry places importance on alternative models to gain a comprehensive overview eliciting as many perspectives about the nature of the problem as possible. Therefore, while a Lockean Delphi study is better served for communication within an informed group with similar areas of knowledge, Kantian Delphi studies are designed to allow informed individuals from many different disciplines to provide information to the problem, broader in scope than any one individual possesses. Mitroff and Turoff (2002) summarise this well by stating that the objective is establishing how to fit the pieces of a jigsaw puzzle together or even to determine if there is more than one jigsaw.

4.4 Delphi technique

4.4.1 Overview

The Delphi technique is a structured process for collecting and organising judgements from a group of experts over several survey rounds (Ziglio, 1996, Cho HK et al., 2003). A key feature of the Delphi technique is its use of experts whose responses remain anonymous throughout a series of iterative questionnaire rounds. Controlled feedback (by the researcher) returned to participants between rounds is also an important feature of the Delphi technique. Here, details of collective group opinion allow experts to either retain or change their earlier opinion in light of other experts' views. Thus, strength of the Delphi technique is the

facilitation of decisions for strategies in situations where there is contradictory or little information and limited raw data (Mitroff and Turoff, 2002, Hasson et al., 2000). Some authors refer to Delphi as a methodology (McKenna, 1994, Hasson et al., 2000, Jairath and Weinstein, 1994). Unfortunately, this description is not universal with literature quoting other terms such as survey, technique, process and approach (Suckley, 2102). For my thesis, I chose the singular term “technique” as method, approach and process could involve more than one technique. The Delphi technique is characterised by the following important features: (Dalkey and Helmer, 1962, Mead and Moseley, 2001).

4.4.2 Anonymity of Delphi participants

Anonymity allows free communication without undue social pressures to confirm. Decisions are more likely to be based on the merit of the proposal rather than who made the proposal. Importantly, if judgments turn out to be unsuitable, participants do not lose face. Turoff and Hiltz (1996) point out that anonymity allows experts of high status to produce questionable ideas, and also permits lower status participants to introduce ideas without fear of them being rejected outright. However, anonymity should only extend to identity as a primary influence on the participants is the knowledge that they are communicating with a group of peers and know that other participants may have contributions of equivalent weight. This was especially relevant to my study where leaders in the field of FGIDs were encouraged to participate. For this study to be successful, experts had to be aware that they were sharing opinion and experience with not only fellow experts, but also fellow stakeholders involved in the development of FGID diagnosis.

4.4.3 Iteration and controlled feedback

Iteration is a process of repetitive input that allows interaction among participants over several data collection stages. At the beginning of each stage, the results containing statistical aggregation of participant rating scores and opinions from the previous stage are

summarised and fed back to the group. Participants are then asked to reassess their previous answers in light of what the whole group stated in the previous round (Dalkey, 1968).

Therefore, iteration can allow convergence of opinions **OR** continued disagreement, depending on the topic under investigation. In both cases, median opinion can be measured providing an opportunity for researchers and participants to improve the accuracy of results. However, at this point I would also suggest that as with other forms of feedback, studies might also suffer from “iteration-itis” where aside from participant fatigue; too much feedback from too many experts over too many survey rounds can transform what were originally interesting ideas into well-documented mush.

4.4.4 Application of The Delphi technique

Early Delphi studies were used for long and short range forecasting of future events to gain a consensus opinion or to emphasise differences of opinion to develop alternative future scenarios. Dalkey (1962) and Kaplan, Skogstad and Girshick (1950) showed us that decisions made by groups are more reliable than those made individually. In the 1950s, the intelligence think-tank, (RAND Corporation, 2013) developed Delphi for use in the US military to determine preparedness for a Soviet attack on US soil. Luckily, over 50 years later, it is used for more peaceful purposes, but with the same aims: that is to establish as objectively as possible levels of consensus on complex issues. Linstone, Turoff and Mitroff (2002, p3) accurately describe it as the best and/or only option to a research problem at that time where no accurate information exists.

4.4.5 Theory Building

Okoli and Pawlowski (2004) showed that the Delphi technique may also be used as an inductive approach for theory-building. First, by posing questions to experts, different variables are identified. These variables can be returned to the participants who are then asked to rank them in order of importance. By calculating the mean rank for participants

ranked lists, the researcher can prioritise and choose factors with the strongest effect. Using my research topic as an example, in classic Delphi technique, experts may be asked to list in order of importance, possible mechanisms that might maintain FGIDs. From this ranked list, the researcher could select variables to support their original hypothesis. For example, a factor of interest from the above example may present as “*symptoms of FGIDs persist due to altered central nervous system processing*”. Second, by obtaining information from a heterogeneous group of experts working in the same field, a researcher may extend their own theory or research question across multiple settings. Here, a good example is the multidisciplinary work in the management of chronic pain. Opinions from experts in fields such as anaesthesia, physiotherapy, gastroenterology, general practice, nursing and psychology add unique empirical observations that in turn broaden and consolidate the researcher’s original theory across several specialist areas. Thirdly, by asking experts to justify reasoning for their judgment, a researcher is more able to understand possible underlying relationships between factors necessary for building a theory. Here, an item related to this study may ask an expert to judge the importance of previous acute upper GI infection when assessing people with FD. When asked to justify their answer, an expert may state that in their experience, many patients presenting with FD have been previously diagnosed with an acute upper GIT infection. Thus, the resultant research question or theory may examine the association between the above two variables.

While theory building is not the focus of most Delphi research, it is valuable in confirming or answering research hypotheses. More importantly, I suggest that processes underlying theory building also help to enhance the trustworthiness of Delphi technique (see section 4.9) where justification of opinions from a heterogeneous group help the researcher understand not only a set of judgments, but also the factors behind them. Therefore, the Delphi technique can contribute directly to both theory and practice where theory is consolidated by the design and rigour of the study and where clinicians can have immediate access to ranked

critical factors agreed upon by Delphi participants (Okoli and Pawlowski, 2004). As such, the Delphi technique is a versatile instrument that may be used at many points of research (table 4.1).

Table 4-1 Applications of the Delphi technique during the research process
Adapted from (Okoli and Pawlowski, 2004)

<ul style="list-style-type: none"> • Identification of a research topic • Specification of a research question(s) • Identification of theoretical perspectives • Generation/selection and ranking of variables of interest • Preliminary identification of causal relationships • Definition of theories and generation of the common language of discourse

In health care, the Delphi technique has been used to develop clinical guidelines (Conway et al., 2013, Kumar, 2007), clinical education (Gensichen et al., 2009, Broomfield and Humphris, 2001) and the development of diagnostic criteria (Watkins et al., 2012, Graham et al., 2003b).

4.4.6 Comparisons with other consensus building techniques

Other consensus-building methods were considered for this thesis but were rejected in favour of the Delphi technique: The first, Nominal Group Technique (Delbecq and Van de Ven, 1971) involves a structured face-to-face meeting where experts are requested to submit ideas independently and privately. The ideas are collected and discussed in turn and then ranked in importance. The second, focus groups (Fern, 2001), where participants also meet face-to-face and are asked to provide ideas and information. A moderator then limits discussion to the areas of importance. Here, the quality of the information presented depends on the skill of the moderator (Gallagher et al., 1993).

With applied fact-finding problems with no known solutions, Nominal Group Technique and the Delphi technique are equally effective (Ven and Delbecq, 1974, Graefe and Armstrong, 2011). However, focus groups and Nominal Group Technique require face-to-face meetings that for this study would be impossible due to the geographical spread of experts involved in FGID diagnosis. Several studies demonstrate the difficulties of face-to-face interaction (Dalkey, 1969). The most serious is the influence of dominant individuals who talk most, with little correlation between status of speech and knowledge (Dalkey, 1968, p7). Another unfortunate influence is “noise”; in the semantic sense, where communication relates more to individual and group interest than with actual problem-solving. Additionally, group pressure for members to conform can also cause significant distortion in individual judgement. (Asch, 1951). Not surprisingly, studies show that, after face-to-face discussion, group response is not the same as the median of individual opinion without discussion (Dalkey, 1969). Thus, I chose the Delphi technique because it allows anonymity, asynchronous interaction and participation of experts who are geographically spread and where negative influence has the least effect.

4.5 Methodological considerations

Mullen (2003) lists over 20 variations of the Delphi technique where researchers modify the technique to suit their needs. There are no standardised guidelines relating to the conduct of the Delphi technique. Quantitative researchers consider the method lacks scientific rigour. While some Delphi protagonists argue that Delphi design is situational and guided by the research question as opposed to methodological requirement, I would argue (as discussed in sections 4.9 and 8.4) this observation should only relate to the research question and not to how the technique is carried out.

Generally, the Delphi technique avoids the need to bring experts together, reduces costs and saving experts' time. refer to the ability to progress without experts needing to meet or take part at the same time as "asynchronous". The same authors also show that the effectiveness of group communication is increased by introducing Web-based communications in place of paper and pen style methods. Importantly, Delphi techniques never produce decisions, but evaluate potential decisions that can then be offered to aid decision-making processes. Geist (2010) and Brandon (1998) propose that stakeholders should always be included for participation, suggesting that this increases both the attention paid to findings and the validity of findings. However, I would caution that when developing a Delphi study, researchers should account for stakeholder-related response bias, as they usually have a vested interest in study outcomes. I discuss this issue in more detail in section 4.10.

4.6 Establishing rigour in qualitative research

The conventional rationalistic (quantitative) model assumes a single reality (situation or condition) on which inquiry can converge and where reality is separated into independent parts (variables). In contrast, the naturalistic (qualitative) model assumes that there are multiple realities based on human interpretation. So, for a given reality, an inquiry will diverge rather than converge as more is known and where all parts of reality are interrelated (Guba, 1981). Importantly, the term "qualitative research" is generic and describes a wide range of contrasting research methods based on the naturalistic model. Hence, there is controversy over the value of qualitative research and how far it should be subjected to the same rigours as quantitative research. Guba (1981, p88) suggests that qualitative research is so diverse that a single set of criteria cannot encompass all non-quantitative methods. Furthermore, Cutcliffe and McKenna (1999) in a searching review argue that some qualitative concepts are so abstract or built around 'hunches' or 'felt sense' it would be impossible to investigate them using empirical measurements. In these cases, Rolfe (2006,

p304) reasonably suggests that a generic framework for evaluating the quality of qualitative research methods should be abandoned in favour of individual judgments of specific studies.

Nonetheless, attempts have been made to describe differences between different qualitative research paradigms and how best to judge their quality (Rolfe, 2006, Guba, 1981, Sandelowski, 1993). Authors argue that mixing quantitative and qualitative models provide options for examining complex research questions and improving analytic power of studies (Driscoll et al., 2007, Sandelowski, 2000). Insightfully, Rolfe (2006, p306) suggests that if the terms “quantitative” and “qualitative” refer directly to data collection where qualitative research uses verbal and textual data and quantitative research uses numerical data, then there is little at issue with mixed methodology studies. However, he further argues that if the terms have deeper epistemological or ontological significance where one is judging knowledge, then genuine philosophical problems arise when attempting to combine positivism with interpretivism. Here, as discussed by Cutcliffe and McKenna (1999) and Sandelowski and Barroso (2002), the epistemological scope of qualitative methodologies are simply too broad to represent a single set of criteria. Given these areas of uncertainty, Rolfe (2006), as stated above argues that there is no qualitative paradigm at all where each research methodology is appraised on its own merits. Generally, there are conflicts of opinion where, some authors argue that the same validity criteria be used as for quantitative research, while others have sought to identify specific frameworks or criteria explicit to qualitative research, one of which refers to the term “trustworthiness” (section 4.9). Further discussion surrounding the issues of rigour in qualitative research exploring human experience is beyond the scope of my thesis. I shall, therefore, narrow the discussion to methodological rigour surrounding Delphi technique. Interested readers are referred to (Lincoln and Guba, 1985b), (Guba, 1981), (Sandelowski, 1993) as well as (Rolfe, 2006) who nicely summarises these confusing and often conflicting concepts.

4.7 Methodological rigour in Delphi research

In conventional science, criteria such as accuracy, repeatability, and validity are used to gauge the rigour of a process: for interpretative studies, trustworthiness is a more frequent feature. The Delphi technique is often chosen because there is no “first best” method available to answer particular research questions (Mead and Moseley, 2001, p10). Subsequently, consensus is ranked low in traditional hierarchies of evidence defining expert opinion as low (level 5) evidence (Strauss et al., 2011). Researchers using the Delphi technique must show that it is an appropriate method to use, where other means are not appropriate or ineffective.

Reaching consensus does not seek to provide new knowledge or truth. The aim is to make best use of available information and give a snapshot of knowledge at a point in time (Everett, 1993, Black et al., 1999). Evidence-based research methods including clinicians’ judgement and experience must give the best answer to a clinical question (Sackett et al., 1996). Thus, rigorous sampling criteria are required to ensure some form of uniformity in the level of knowledge and understanding of participants. Although knowledge is important, it is not necessary to have highly developed scientific structure, as items of common sense are considered equally reliable.

The Delphi technique is used to achieve consensus and to support decisions by discussing alternate views of a “preferred future” as is the case with Policy Delphi (Hasson and Keeney, 2011, p6). In the present study, I aimed to gather data on areas of agreement and disagreement where there is little knowledge in subject fields important to FGIDs. These two aims had equal importance as items gaining consensus may be developed for future use in MAC, while those items where a lack of credible research was cited, as a reason for disagreement can be further investigated. The following sections review both quantitative and qualitative procedures for assessing methodological rigour in the Delphi technique.

4.8 Reliability and validity

4.8.1 Reliability

Data from Delphi studies are subjective and qualitative in nature. A different expert panel could produce different results when given the same Delphi survey, so, critics have concerns about reliability (Goodman, 1987, Sackman, 1974). Some would argue that it is not possible to determine reliability as each Delphi survey round involves the creation of a new measuring instrument (Rowe et al., 1991, Engels and Powell Kennedy, 2007). Nevertheless, Duffield (1993) found 93% inter-rater agreement in the same survey, between two separate panels that differed marginally in size and composition. Interestingly, Delphi studies have also shown good test-retest reliability with one study being conducted 16 years apart (Ono and Wedemeyer, 1994). However, testing any type of questionnaire for reliability is difficult (Brace, 2010b, Hasson and Keeney, 2011). Opinions differ on reliability concerning Delphi panel size, the use of open first rounds, the interactive nature of Delphi technique and the avoidance of group bias (Linstone and Turoff, 2002, Okoli and Pawlowski, 2004, Hasson and Keeney, 2011). The following sub-sections summarise attempts to “best fit” the Delphi technique into common forms of reliability used in psychometric research.

4.8.1.1 Test-retest reliability

The usual means of testing reliability of questionnaires is to administer them twice to the same group to establish whether they give consistent answers. Williams (2003, p249) suggests a gap of two or three weeks between surveys, however, Brace (2010a) highlights difficulties in knowing which time is optimal. Too soon, and respondents remember their responses and may change them, or alternatively they may deliberately try to be consistent with their response. Test-retest methods with the Delphi technique are considered inappropriate due to the assumption that no changes have been made to the construct of the questionnaire, where each survey round may differ in the number of items (Keeney et al., 2011c, Okoli and Pawlowski, 2004). However, I propose a novel suggestion that test-re-test

reliability could be established in a pilot round or final survey round by administering it twice over a given time period. This would ensure that each survey retest would have the same number of items each with the same wording and given to the same participants. The potential benefit of re-administering a final survey round could allow investigators to check both the reliability and participant justification of attitude ratings.

4.8.1.2 Inter-rater reliability

Inter-rater reliability assesses the degree to which different raters give consistent scores to the same scale. Findings suggest that the Delphi technique shows good inter-rater reliability both simultaneously (Woudenberg, 1991, Welty, 1972) and at different time points (Ament, 1970). However, like many other types of survey, the Delphi technique relies on judgements. Here, results can be influenced by personal and situational bias where people from different backgrounds may have different opinions (Kahneman et al., 1982). Furthermore, participants must consider many cognitive factors both in the understanding of an item and their interpretation of scale categories when making a single judgment on a rating scale item (see section 5.4.2). Moreover, distraction between raters may also affect the reliability. These issues can also affect the association of ratings by different participants. One potential way to improve this form of reliability is to issue instructions or training to participants on how to rate or estimate rating scales. While not researched in the Delphi technique, rating instruction increases rater competency and result-consistency in non-Delphi surveys (Waninge et al., 2011, Oremus et al., 2012, Castorr et al., 1990). Thus, similar rating instructions for each survey round may reduce the number of confounding factors and increase inter-rater reliability over the rounds of a Delphi study.

4.8.1.3 Parallel form reliability

Parallel form reliability is obtained by administering different versions of a test to the same group by changing or modifying the order and wording. Scores are evaluated for consistency between tests. This type of reliability may be applied to each item of the Delphi first round

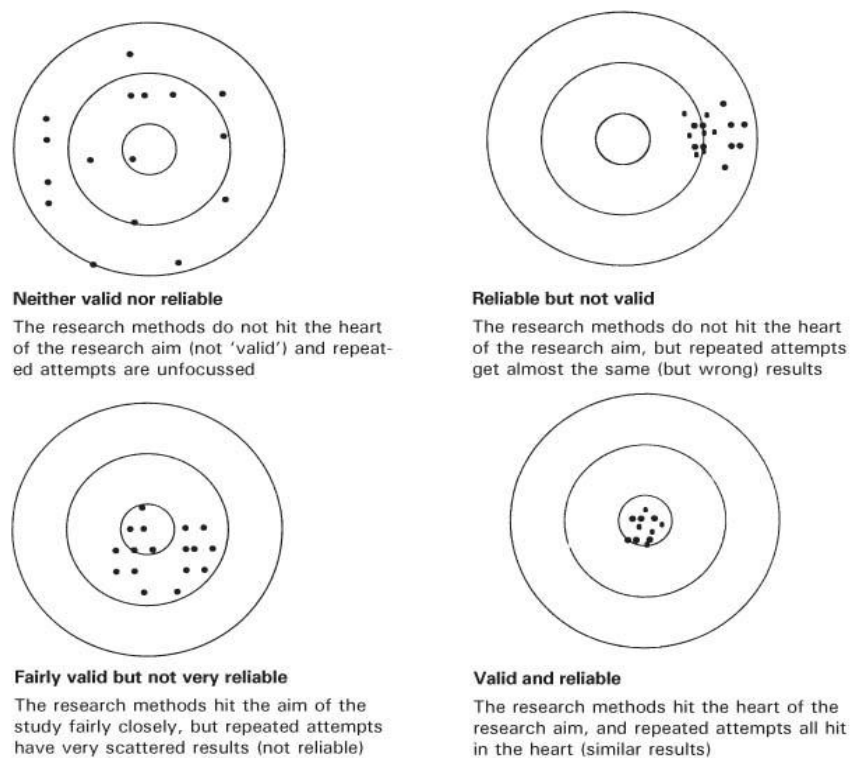
(Hasson and Keeney (2011). However, subtle changes in wording can influence participant responses. Gordon (1994, p8) suggests participants who undertake a Delphi survey in its original layout, should attend focus groups to investigate respondents' interpretations of each item. Unfortunately, due to the geographic spread of the experts and limited resources, these face-to-face follow-ups were impractical.

4.8.2 Validity

The validity of a questionnaire relies in part on its reliability with consistency of results across a given population. However, although test results may be consistent, they may not necessarily accurately measure the construct intended as nicely shown in figure 4.1.

It is unclear how the validity of Delphi results can be established. Some argue that if experts are well matched, results will convey a degree of validity (Williams and Webb, 1994, Goodman, 1987). However, others argue that non-experts provide judgments indistinguishable from experts' judgements. (Sackman, 1974, Welty, 1972). Importantly, as with reliability, these observations depend on the specificity of the research question, who and how many are able to answer it, the inquiry system and the rigour of expert selection. Therefore, stating the details of methodological and analytical decisions made during a Delphi study is strongly recommended to reduce criticism and more importantly increase methodological rigour. Some authors suggest results of Delphi studies be checked and/or improved through follow-up research using further surveys or interviews (Ono and Wedemeyer, 1994, Skulmoski et al., 2007). Below are summaries of the different forms of validity, which may apply to the Delphi technique.

Figure 4-1 A diagrammatic representation of the relationship between reliability and validity. The aim of the research question is the centre of the target (Precision Consulting, 2014)



4.8.2.1 Content validity

Content validity is the estimate of how much a measure represents each element of a construct. Content validity of the Delphi technique is determined by expert evaluation of whether test items assess predefined content. Authors claim the Delphi technique provides content validity based on the assumptions that firstly, results come from group judgment and not a single person and secondly that expert judgment provides empirical confirmation of “reality”. However, Delphi groups may have features that influence results such as the number of experts taking part, the level of knowledge the experts possess, participant self-interest and ambiguous imprecise statements (Bollen, 2002). Non-participant experts (including investigators) should assess content validity of items suggested by participants in the first round of a classic Delphi study or items generated from a literature review. However, as I later discuss in section 4.9.1, if Delphi is a theory-building process, and no

theory exists at the beginning of a study, what defines an expert in these cases and will any exist?

4.8.2.2 Construct validity

Construct validity should follow content validity where test items are not only relevant and representative of a construct, but also that the test does not measure other non-intended constructs. A construct may be simple such as weight or height; however, a test may be intended evaluate a condition or an attitude where a construct may be made up of many components. Here, each component would have to be defined to determine which feature should and should not be included in a given construct. In Classic Delphi development, construct validity is determined by the researcher's endorsement of items given by experts for inclusion in the following round. Okoli and Pawlowski (2004) propose that as the level of group expertise not anonymous to the researcher, the construct validity of approved items is acceptable. This being the case, I suggest that not only the level, but also the area(s) of expertise are necessary if construct validity is to exist, highlighting the importance of rigorous purposive sampling for Delphi study participants.

4.8.2.3 Concurrent validity

Concurrent validity is measured by administering a new test simultaneously against a measure that has previously been validated. Several studies suggest that successive rounds of a Delphi questionnaire contribute to concurrent validity (Keeney et al., 2011c, Sharkey and Sharples, 2001). However, Keeney et al. (2011c) question the establishment of concurrent validity citing a lack of accountability for views expressed, high dropout rates and the influence of group thinking leading to a bandwagon effect. However, I suggest that if strict criteria for purposive sampling are followed, then the opinions of all experts should be accounted for, where the researcher is confident that opinions come from both knowledge and informed insight as opposed to wishful-thinking or speculation.

4.8.2.4 Predictive validity

Predictive validity indicates that scores from a test make accurate predictions about the construct they represent (e.g., achievement, intelligence). Similarly, in consensus studies, researchers assess predictive validity by examining whether a predefined forecast comes true (Von der Gracht, 2008, p67). Researchers have attempted to demonstrate both congruence of prediction across independent Delphi studies and substantiate results of older studies by observing the accuracy of previous forecasts (Ono and Wedemeyer, 1994). Keeney et al. (2011c) cites de Meyrick (2003) warns that experts typically hold a position of power and, therefore, may try to shape results. For example, if treatment guidelines for a particular condition are being considered, participants from different areas of speciality involved in its treatment may only suggest or agree with items relating to their area of expertise. As such, results from such a study may not be accurate, or may not help in determining an overall solution. Interestingly, Linstone and Turoff (2002) expand this discussion by suggesting that depending on the duration of a forecast or plan of action, long-range forecasts tend to be pessimistic, whereas short-range forecasts are more optimistic. As such, predictive validity is difficult to apply especially in situations where predictions reflect the interests of the participants rather than the research problem under investigation.

4.9 Trustworthiness

The Delphi technique combines both quantitative (psychometric scale scores) and qualitative (themes based on expert opinion) research processes. (Hasson and Keeney, 2011, Zimmermann et al., 2012). However, qualitative criteria cannot be used in a quantitative context without significant adjustment. Therefore, trustworthiness has become an accepted gauge for the thoroughness of many qualitative research methods that now includes the Delphi technique. Guba (1981) and later Lincoln and Guba (1985a) devised criteria that parallel those of conventional quantitative paradigms: internal validity, external validity, reliability and objectivity. These criteria respond to four fundamental questions concerned

with truth-value, applicability, consistency and neutrality. Most literature discussing trustworthiness with the Delphi technique give overviews describing analogies to the above questions (Day and Bobeva, 2005, Zimmermann et al., 2012, Hasson and Keeney, 2011) without describing in detail the application of each criterion to the Delphi technique. Given that, many Delphi studies now use components of trustworthiness. I suggest that in order to obtain the most reliable and valid attitude ratings from a group of experts, each element of trustworthiness and its application to the Delphi technique must be defined by the researcher(s) or by the development of guidelines. This would allow investigators to apply the elements of trustworthiness more easily to a Delphi study, where methodological rigour is based on criteria that provide transparency, consistency and reproducibility. In the following subsections I suggest where possible, how the Delphi technique may “best fit” the components of each criterion. I further note that these ideas are not recommendations for future use but only initial suggestions for future consideration.

4.9.1 Credibility (Internal validity)

In order to understand a research situation, it is important to both identify and document recurring patterns, ideas and ethical considerations. Credibility involves sufficient commitment to a research setting for recurrent considerations or patterns to be identified and verified (Lincoln and Guba, 1985a).

4.9.1.1 Prolonged engagement

Prolonged engagement with a topic allows the researcher to understand it and grasp the features and context. The researcher may detect and adjust for possible biases while also recognising their own biases about the research setting. In a Delphi study, prolonged engagement may relate to immersion into the research topic of investigation through the literature and regular contact with participants. Hence, familiarity would allow increased understanding of the subject. Here, an investigator can describe the study to participants

prior to the first round survey with contact continuing between and after all the survey rounds with updates and feedback on data analysis. However, the danger exists that researchers become so immersed that they become complacent and biased toward the interests of the participants, and/or miss obvious ideas.

4.9.1.2 Persistent observations

While prolonged engagement allows breadth of understanding of a topic, persistent observation allows an in-depth pursuit of characteristics and important features found through prolonged engagement. Here, the research may pursue features relevant to the research question and discard those not relevant. Regular interviews, meetings and observations over the course of a study allow adequate revision of context and checks for misinformation (Johnson and Saville-Troike, 1992, p603). Concerning the Delphi technique, the iterative process of ongoing survey rounds may be viewed as persistent observation. By observing participant agreement levels, researchers are able to identify important elements of both consensus and disagreement on matter requiring further investigation or revision.

4.9.1.3 Triangulation

Triangulation or crosschecking of data uses different sources, different investigators, different theories and different methods that allow the consistency of findings to be checked. In such cases, the ‘validity’ of a study can be examined by comparing it with other kinds of evidence. This point is expanded by Porter (2007) in an interesting opinion piece. He suggests that by using different investigators (other readers), judgement on a researcher’s report will provide a point of mediation that will help both parties to agree on the most appropriate interpretive tools to use in making claims. Several Delphi studies have used triangulation using multiple methods such as mixing Delphi survey rounds and face to face interviews with participants and non-participating experts (Di Rezze et al., 2014). Other Delphi studies have used literature reviews on study subject material extracted from databases for questionnaire development coupled with a classic Delphi study (Golchin et al.,

2010). Su et al. (2010, p696) suggest that using participants from different professions in a Delphi also warrants investigator triangulation. However, I would point out that participants are not investigators. While they can give a judgement on a given topic, they do not have investigator responsibilities or may not have the necessary training to perform research procedures. Therefore, Guba (1981, p85) appropriately suggests obtaining perceptions and analysis of several “disinterested” qualified investigators. However, triangulation poses other issues.

As discussed in section 4.8.2.1, Cutcliffe and McKenna (1999) argue that if there is an absence of an initial theory relating to the phenomenon being studied, it is difficult to know how to define an expert in the field or if researchers familiar with the phenomenon even exist. Equally, given that the primary researcher is “immersed” in the research question, and another “experienced” colleague is not, it is unlikely these two people will interpret data in the same way. Fortunately, for the present study, there are abundant observational data concerning patient characteristics, prevalence and symptom clusters as well as current research examining possible markers for FGIDs. Thus, data triangulation was applied by using data from the Delphi study literature review, evaluation and guidance from other recognised experts in the same field and thesis supervisors throughout the course of the study.

4.9.1.4 Peer debriefing

Peer debriefing requires the researcher to work together with one or more colleagues whose views are impartial to the study. These colleagues examine the investigator’s transcripts, general methodology, results and conclusions where their feedback to the researcher enhances credibility. Peer debriefing is seldom discussed in Delphi research. I suggest that investigators using the Delphi technique could submit themselves to searching questions or data reanalysis by faculty colleagues. Peers may identify problems concerning item content

and ambiguity in relation to both relevant literature and opinions/ideas expressed by experts in classic Delphi studies. Peers may also detect errors in the data and biases or assumptions made by the researcher. I would note however that the same issues regarding disinterested investigators or faculty colleagues arising with triangulation could also negatively affect peer debriefing and other elements of trustworthiness.

4.9.1.5 Member checks

Member checks require that data and the researcher's interpretations are checked by returning to participants following analysis or peer checking, using experts to re-analyse the data (Rolfe, 2006). The crucial issue for this exercise is how far the researcher's interpretation of data corresponds with that of the members (participants) involved in the research. Lincoln and Guba (1985a, p22) see member checks as a way to examine and verify data and describe it as the most important action inquirers can take. However, Sandelowski (1993, p30) justly argues that participants look out for themselves and their own reality in researchers' accounts, whereas researchers are trying to present multiple realities in a way that remains close to each member's reality. Therefore, if reality is assumed to be unique to the individual, repeatability is not practicable as one cannot expect other researchers or respondents to respond in the same way. While rarely discussed in Delphi methodology, ongoing iteration and feedback given to participants may also be viewed as member checks (Keeney et al., 2011b). However, Paraskevas and Saunders (2012, p10) suggest that member checking in Delphi studies is best served by participant interviews where transcripts are returned to participants for review of the accuracy of their own response.

4.9.2 Transferability (external validity)

Qualitative researchers believe that interpretations of data for a given context are unique and only relevant to a particular situation. For this reason; purposive sampling is preferred in Delphi studies as it should be specific to a given context and is neither intended to be typical

nor representative of the general population (Guba, 1981). However, Lincoln and Guba (1985a, p17) suggest “thick description” as a way of achieving transferability. Thick description refers to the detailed description of a phenomenon or participant experiences that take into account not only immediate behaviours/opinions (scale scores), but also the circumstantial and experiential understanding of such behaviours/opinions (comments). The researcher then formulates clear patterns of cultural, social or experiential relationships and puts them into an overall context. In this way, other inquirers can then consider a degree of fit to situations or findings elsewhere. I suggest that Delphi researchers ask participants to justify their responses. Here, researchers are able to gather data concerning not only opinion, but also experiences and thoughts of participants behind their opinions in sufficient detail so that one may evaluate the extent to which these responses are transferable to other research settings.

4.9.3 Confirmability and Dependability (objectivity and reliability)

Confirmability indicates the degree to which outcomes are the product of focus of an inquiry and not of the biases of the researcher. Here, an audit trail that includes notes on the methods used, raw data, analysis techniques and pilot forms should allow an auditor to determine if conclusions, interpretations and recommendations are supported by the inquiry (Egger et al., 1997, Guba, 1981). Alternatively, dependability relies on the same detailed description of the research process that allows the reader (or external auditor) to track, and if necessary replicate the research process concerning data collection and analysis. However, while audit feedback may help to assess and summarise findings, it may also lead to confusion where the auditor disagrees with the researcher’s interpretations. Perhaps in such cases auditors and researchers could follow the advice of Porter (2007) and mediate via discussion on common ground for both methods and interpretation of results. Skulmoski et al. (2007) and Keeney et al. (2011c) advocate the use of audit trails of critical theoretical and methodological

decisions to demonstrate trustworthiness in a Delphi study, while Engels and Powell Kennedy (2007, p436) further suggest that confirmability and dependability, depend on:

- the tracking of participant selection and participation of panel members,
- the construction and administration of the questionnaires and criteria for identification of consensus
- the divergence of opinion amongst panel members

4.10 Delphi panels

4.10.1 Does size matter?

The Delphi technique combines both quantitative (numerical rating scores) and qualitative (verbal & textual opinion) data. However, the Delphi technique is used in a variety of ways relating either to quantitative (deductive) approaches where pre-determined hypotheses are tested to answer “what” or to more qualitative (inductive) approaches that aim to gain information to provide understanding on “how” or “why”. Sample sizes and sampling methods must be determined by the research question and not the preference of the investigator. As such, researchers require a sample that can answer a specific research question. Thus, purposive sampling is chosen over convenience and random sampling methods. For instance, choosing a random person to answer a specific question is akin to asking a bystander how to renovate a house, rather than ask a qualified builder. However, as discussed further in this section, purposive sampling for gaining opinion can also have its issues. Dalkey (1969) nicely describes the technology for dealing with opinion using the adage “Two heads are better than one” or more generally, “n heads are better than one.” However, it is important to note that for relevant information in “n heads”, there may also be at least as much distortion. Hence, if there are no exact techniques for extracting information from n heads to form reliable opinion, then heads rule depends on how the heads are used

(Dalkey, 1969, p6). Additionally, there is no way of asserting that one expert is more knowledgeable than another.

In general, confusion regarding Delphi sample size is because there are no methodologically established sampling criteria. (Akins et al., 2005a). Current literature presents only empirical choices of Delphi expert sample sizes made by individual researchers. Many published Delphi studies have used panels consisting of between 10-100 or more panellists. Other Delphi studies have used only small samples, for example, only five experts were used to identify serious drug interactions most likely to occur in ambulatory pharmacy settings (Malone et al., 2004). Conversely, 1142 out of 2865 participants returned questionnaires in a study investigating the methodology of response rates and effects of feedback in large-scale Delphi surveys (Barnette et al., 1978). Nevertheless, several “rule of thumb” figures are suggested with Ziglio (1996) generally proposing between 10 to 15 participants. More specifically, Clayton (1997, p378) proposes between 15 to 30 for homogeneous groups (same discipline). However, for heterogeneous groups (experts with different perspectives/disciplines), Hoffman (1958, p31) suggests that only five to 10 experts are needed as they give a higher proportion of high-quality solutions than homogeneous groups (see section 4.9.2.1).

While there may be differences in decision-making abilities between homogeneous and heterogeneous groups, group size recommendations in both cases appear arbitrary, where group size is determined by individual researchers as opposed to standardised guidelines. Another potential issue regarding group size relates to the research question and the number of recognised experts able to answer it. For example, if a Delphi study aims to obtain judgment among general practitioners on improving an area of prescription practice, the pool of potential experts would be large. However, if one aimed to obtain judgment on diagnostic guidelines for a rare disease, the number of experts in this field may be no more than five or

six. Thus, the stability of response characteristics is also applicable to who is sampled rather than the number of participants. Akins et al. (2005b) show that in fields where experts are limited, those with similar training and understanding in the field of interest allow for effective and reliable use of a small sample. However, I suggest that there may be situations where for each recognised expert there may also be an unrecognised one. Purposive sampling methods recruit recognised experts who are leaders in their field, publish regular articles and work in academic, clinical and recognised expert groups. However, there may be unrecognised experts outside these areas, who have valuable experience and insight. Such experts might be found in non-research-based tertiary care settings. Thus, experts selected for participation in a Delphi study may only represent a small and special sample of a larger and partly unknown population whose clinical experience and expertise are unrecognised.

4.10.2 Group expertise

The composition of an expert panel is central to the outcomes of Delphi methods of the subject being investigated and the subject being examined (Mead and Moseley, 2001, p10) as it is their opinions that form the data on which future decisions may be made. However, developing criteria for expertise is difficult as general definitions, although citing knowledge or skill-sets in particular areas, do not state how these should be measured. Mead and Moseley (2001, p10) in their summary review define expertise within healthcare as a position within a clinical and/or academic hierarchy or by reference to particular experiences such as patients undergoing specific treatments. However, Baker et al. (2006) when discussing the definition of an expert for the Delphi technique, unhelpfully concluded that there is no real answer with responsibility lying with researchers to choose and defend the most appropriate group of experts.

4.10.2.1 Homogeneous or heterogeneous groups

Although authors state that heterogeneous groups produce a greater range of perspectives and a higher proportion of acceptable solutions than homogeneous groups, (Delbecq et al., 1975, Hoffman and Maier, 1961, Hoffman, 1958), they refer to varying personality and perspectives rather than differing specialities. Some authors suggest that experts should be drawn from varied backgrounds in order to guarantee a wide knowledge base (Skulmoski et al., 2007, p8, Mead and Moseley, 2001). However, heterogeneity can be a double-edged sword where diversity may also bring less integration and as such the potential for disapproval with the content of items and opinions of other panellists. Jones and Hunter (1995) support the use of more homogeneous panels for studies concerned with speciality-specific clinical intervention. Oddly though, the same authors then recommend the inclusion of other clinicians such as general practitioners who may provide alternative clinical views particularly when study results could have an impact outside a specialist field.

The desire of the participant to participate is also an important factor. This can be source of bias when considering the forecasting of future events. Here, homogeneous groups tend to be optimistic in their predictions on subsequent actions such as benefits, future costs and associated risks (Lovallo and Kahneman, 2003). However, the anonymity, feedback and iterative processes of the Delphi method may improve forecast accuracy and reduce prediction error (Ecken et al., 2011). In addition, Wright et al. (1996) show that a heterogeneous group of experts in combination with the Delphi procedure further reduces overconfidence.

Two studies tested whether homogeneous subgroups for the same Delphi survey altered agreement in large heterogeneous organisations. In the first, authors used personnel from various US Air Force organisations, to examine how closely the organisational opinions agreed (Jones, 2001). The average level of agreement obtained within a subgroup varied

significantly over three rounds for some items where information would not have been obtained had one Delphi study been run with a single larger heterogeneous group. The overall number of participants was 61 but data for the subgroups was not given. In the second study, Okoli and Pawlowski (2004), aimed to provide a more rigorous approach to selecting appropriate experts for a Delphi study. The authors selected separate stakeholder groups related to e-commerce in Sub-Saharan Africa. They asked each expert group to list relevant factors for infrastructure and expediency lists. They then requested each expert to select at least 10 given factors for each list, followed by further request for each expert to rank each list in order of importance. These ranked lists were then returned to the experts for re-ranking in light of rankings given by other experts. However, while this study was carefully designed using homogeneous expert groups to compare different perspectives, the authors supply no data to supporting their conclusions.

I sampled experts who were all involved in research of the diagnosis of FGIDs. Although it would have been helpful to conduct separate Delphi surveys with each speciality subgroup, there were in some cases only one or two experts from any given field. Even given that some Delphi participant groups are small, as determined by the research question, I considered these numbers too small. Many scientific papers analyse data inappropriately, using samples that are too small. Such results are unreliable and irreproducible (Button et al., 2013).

Previous studies using the Delphi technique contain such examples. While I could have recruited other specialists in each field to increase subgroup size, this would not meet the strict purposive sampling criteria proposed for this study. Furthermore, all experts involved in this study were currently working together towards the same goal where their judgments have an impact across specialist fields and different levels of health care. I, therefore, chose one Delphi study to survey one heterogeneous group.

4.11 Summary

The Delphi technique is well suited for the development of theories or hypotheses directed toward building process models (several processes of a similar nature classified together) from a factor-based framework where there is incomplete knowledge about a problem. Concerning my study, the process model is MAC, and the framework is based on existing data and expert judgment on FGID evaluation. Multidimensional assessment for the evaluation of FGIDs is not a single or well-understood problem, but a set of many complex problems that are not well understood. This Delphi study, therefore, required a structure that allowed informed experts in different specialities to contribute varied and valid judgments to the concept of MAC of FGIDs.

Chapter 5. Defining and Measuring Attitude

5.1 Overview

This chapter considers the concepts of attitude, opinion, and their application to different types of rating scale. I then discuss questions and arguments surrounding Likert scale construction.

5.2 Defining attitude

Attitude is surprisingly hard to define. The central concept is that attitude will affect how an individual chooses one action over another. Early attempts concentrated on direct relationships between a given attitude and its associated behavioural response. However, attitude measurements were found to poorly reflect subsequent behaviour. From further studies, it became clear that many other factors affect how a person responds in a specific situation (Wicker, 1969, Edwards, 1957). If behaviour varies, then additional factors must influence what would otherwise be predicted by a “known” attitude (Tittle and Hill, 1967). Tittle and Hill (1967) showed a greater correspondence between measured attitude and behaviour is found when the behaviour criterion incorporates a wide range of activity with respect to the attitude object.

However, Likert (1932, p9), adds to this debate by suggesting that attitude, in the form of opinion is a verbal substitute and an indirect method of measuring an individual’s “dispositions”. Silk (1969) further describes attitudes as not being directly observable but rather inferred as they cannot be identified from any one particular act or response, but rather from a large number of similar acts or responses. As such, both positive and negative variables such as salience have to be considered. Therefore, when measuring levels of attitude, it is important that researchers develop accurate attitude constructs that will assess their questions.

5.2.1 Measuring attitude

To measure group attitude, consistency or covariation among responses is fundamental to all methods. Therefore, patterns of intercorrelation among responses are evidence used to identify the attitude of a group (Silk, 1969, Likert, 1932). Thus, the participant's thoughts and feelings must be expressed orally or written in the form of an opinion (Thurstone, 1928, p531). Attitude can be measured using observation, self-reporting methods and graded or binary scale responses to statements which are then collected and analysed (Roberts et al., 1999). Such approaches provide valid measures of attitude where a participant's score reflects their real attitude. Most attitude scales are unidimensional with simple anchors such as like-dislike, or important-unimportant. This is not to say that attitudes are not multidimensional; rather they are easier to understand.

Analysis of attitudes in each round of the Delphi procedure has several objectives. Firstly, analysis of subjective judgments can generate a range of views or considerations. Secondly, analysis detects disagreements and judgmental biases that require further clarification or investigation. Thirdly, analysis may find patterns of information and attitude clusters among subgroups. Turoff and Hiltz (1996).

5.3 Scaling methods for attitude

Attitude scales provide quantitative measurement of attitudes, opinion or values by summarising numerical scores given by researchers to gain peoples' responses to sets of statements (Payne and Payne, 2004). In this section, I broadly discuss and compare Thurstone's equally appearing interval scale and Likert's summated scale in terms of their design, function and application generally and specifically for the Delphi technique. I also briefly discuss item response theory models and their application to attitude-scaling techniques. However, it is beyond the scope of this thesis to discuss the mathematical modelling involved in item response theory. These topics are addressed in full by Andrich (1988), Andrich (1996), Roberts et al. (2000).

5.3.1 Thurstone's equally appearing interval scales

Developed by Thurstone and Chave (1929), equal appearing interval scales involve two process stages. Firstly, an informed group spanning an entire range of possible opinions write a large number of attitude statements on a given subject. A separate judgment group then independently rate each statement on a scale numbered 1-to-11 that represent an evenly graduated series of attitudes from least favourable in category 1 to most favourable in category 11 with all neutral statements in category 6 (Thurstone and Chave, 1929). Ambiguous items or those showing considerable variance in judgment are eliminated. Secondly, the median value and interquartile range is computed for the “tightness of judgment” for each statement. Statements with the smallest interquartile range are then allocated at equal intervals with the median of each statement representing a point on the 1-to-11 scale (Trochim, 2006). Therefore, the final Thurstone scale is limited to only “relevant” items with scale values determined by the judges’ classification that are distributed across an attitude continuum.

The questionnaire is then presented to participants who are asked to agree or disagree to a set of statements. The median of the scale values of the agreed items is taken as a measure of the strength of participant attitudes. Thurstone termed this response function as a “single-peaked” response where people are expected to agree with statements close to their own position and disagree with statements not close in either direction (Andrich, 1996, p348). Roberts et al. (2000) followed the work of (Coombs, 1964) and (Andrich, 1988) to develop single peaked response theory as an unfolding model of proximity. This theory predicts item scores and total scores are based on the distance between a participant’s preferred choice and that of an each item.

5.3.2 Likert's procedure

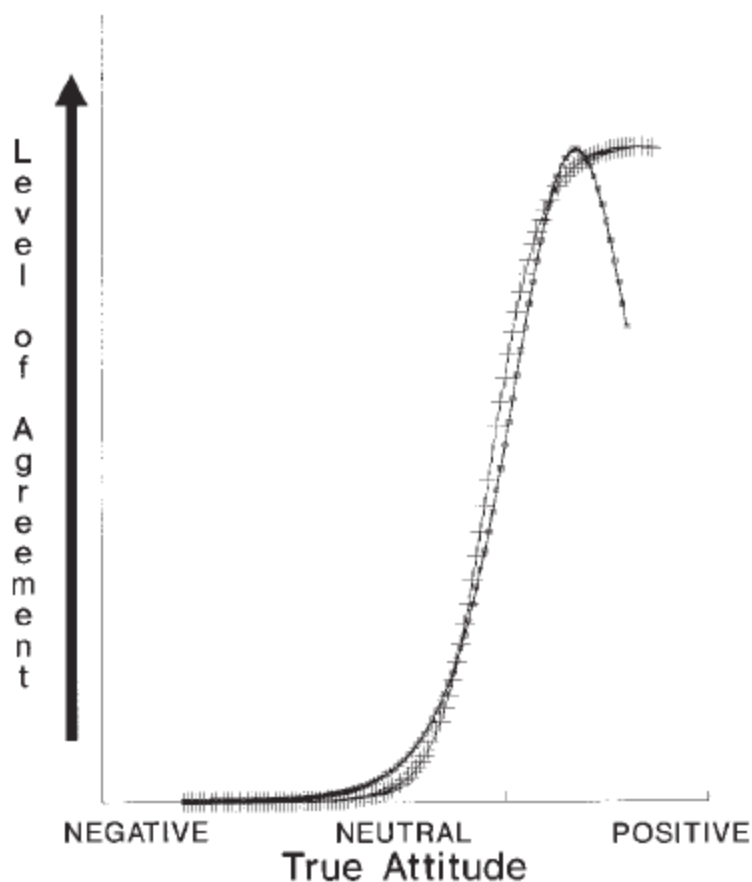
Likert (1932) introduced his technique deliberately to remove the time-consuming task of first locating statements on a scale. Instead, the Likert procedure consists of a series of

statements similar to those of Thurstone but requires participants to indicate degrees of attitude also using a neutral category (Likert, 1932, p46). Likert used statements expressing positive and negative attitudes, the latter which have to be reversed scored (Roberts et al., 1999). Each statement consists of several response categories that are numerically weighted. The responses are simply summed across all items to locate each person (in this case, degree of attitude). It is imperative that all included items are homogeneous, internally consistent and measure a single common factor. Otherwise, it would not make sense to sum the individual item scores to attain a person's scale score. Items are then analysed in various ways such as examining the correlation between responses to an item and the total score for the whole scale while other tests may measure levels of inter-item correlation (de Winter and Dodou, 2012, Gadermann et al., 2012).

5.3.3 Theoretical differences between methods

Although Likert never provided a definitive theoretical model for his methods, item response theorists propose Likert scaling as a cumulative response function rather than a single-peak function implied by Thurstone's procedure. There are several useful definitions of cumulative response function (Andrich, 1996, Roberts et al., 1999). They generally state that the probability of item agreement increases as the distance between the individual and the item on the attitude scale increases in a positive direction (i.e., an individual's attitude is more positive than the content expressed in the item) and decreases in a negative direction. Figure 5.1 shows a graph comparing the theoretical item characteristic curves associated with both unfolding and cumulative item response models.

Figure 5-1 A comparison of theoretical item characteristic curves associated with unfolding (squares) and cumulative models (plus signs) in relation to participants true attitudes and scored level of agreement (taken from Roberts et al 1999)



Given this increase item agreement in a positive direction, cumulative modelling may lead to decreased validity when extreme attitudes are measured. However, the Likert procedure generally performs well and only produces conflicting estimates in a small minority of studies (Silk, 1969). The term “cumulative” while similar in theory should not be confused with cumulative (Guttman) scales (see section 5.3.5.1), where agreement of one extreme scale item means the endorsement of all other items that take a less extreme position (Guttman, 1954).

5.3.4 Ease of construction and reliability

When Likert developed his summated scale, Thurstone's procedure was already well-established. One reason the Likert procedure became more popular was that it avoided the task of using judgment groups to construct the scale while retaining similar reliability and results (Edwards and Kenney, 1946). At the time, there was concern that the judging group could influence the scale values. However, this has been disproved: in most cases the attitudes of the judges are unimportant. (Hinckley, 1963, Hinckley, 1932).

Early studies show that scales can be constructed using the Likert technique more quickly and with less labour than the construction of a Thurstone scale (Barclay and Weaver, 1962, Edwards and Kenney, 1946). Surprisingly, there are no recent studies comparing reliability for Thurstone and Likert scales. However, previous investigations do show favour toward Likert scales. A possible explanation could be that participants answer every statement on a Likert scale, whereas they may only be required to check attitude statements they agree or disagree with using Thurstone's scale. As such, Likert scales have more chance of providing data covering a larger proportion of attitude compared to Thurstone scales (Kothari CR, 2004). Unfortunately, earlier comparative studies fail to discuss the fact that Likert procedure does have several limitations, the most notable being that data only determines whether participants are more or less favorable to a topic. Unlike Thurstone's equal appearing interval scale, it cannot indicate how much, or less agreement differs as there is no theoretical basis showing equal spacing between each item category meaning that data are ordinal.

5.3.5 Other attitude scales

5.3.5.1 Guttman scales

The Guttman Scale was developed in the 1940's in order to establish relationships between groups of attitude items. Respondents are asked to agree or disagree with items, but more typically presented in a yes/no dichotomous format.(Trochim, 2006). Questions in these scales gradually increase in specificity and require the respondent to agree with all

statements up to a point and then stop agreeing. The scale is generally used to determine the extremity of a specific viewpoint with successive statements presenting increasingly extreme attitudes (Trochim, 2006).

5.3.5.2 Semantic differential scaling

The Semantic Differential scale is a technique also developed in the 1940s as a quantitative measure of the meaning of language where words may have different meaning to different individuals as a function of their experiences (e.g., “the experience of hardship” is viewed differently by eight and 80-year-olds or by rich and homeless). The semantic differential scale offers a bipolar pair of adjectives between which respondents must select a point of agreement on that scale (Page-Bucci H, 2003, McLver, 2004). Whilst this scale is comparatively easy for the respondent to complete, it has more relevance to market research on how customers feel about certain material products (Dalton et al., 2008).

5.3.6 Attitude scale model comparisons

Very few studies have compared the performance of different types of rating scale. However, Tittle and Hill did study Thurstone’s successive interval technique, the semantic differential procedure, and the Likert and Guttman scales to determine the degree of correlation between measured attitude and the capacity of the four techniques to predict behaviour. Analysis showed that when scales are of equal length, Likert scaling is the best predictor of behaviour while also exhibiting the greatest reliability. Surprisingly, Thurstone’s technique was the poorest predictor and the least reliable (Tittle and Hill, 1967). These findings clash with methodological thinking concerning attitude measurement with Thurstone’s technique being considered as the standard against which other attitude measures are compared. Tittle and Hill (1967) suggest that Thurstone’s technique may be negatively influenced by the fact that respondents can only agree or disagree with each item so that it does not allow for an indication of degrees of attitudes.

Guttman scales are used far less frequently than Likert scales, partly because they are more difficult to construct, and have reduced reliability and validity due to their succinctness (Gothwal et al., 2009, p4496). Tittle and Hill (1967) found that although 10-item Guttman scales predicted behaviour well, they were less reliable than the 10-item Likert scale. They also found that the semantic differential scale as a measure of attitude suffers a serious disadvantage where participants tend to respond in sets, as desirable category points appear on one side, and undesirable appear on the other. However, Friberg et al. (2006) does suggest that the bipolar nature of semantic differential scaling reduces the tendency for respondents to agree with statements regardless of content (acquiescence bias) without lowering psychometric quality.

5.4 Scaling and the Delphi technique

I found no studies examining or comparing different scales for use with the Delphi technique. However, as the Delphi technique aims to assess levels of participant attitude on a given topic and because Likert scales measure the degree of attitude of respondents to a set of statements, Keeney et al. (2011a, p77) state unequivocally that the Likert scale “*is a perfect scale to use with Delphi technique*”. I used Likert scaling for my Delphi study because it measures degrees of attitude and is easy to construct. Although Thurstone’s equally appearing interval scale was a worthy option, limited funding and facilities prevented me from forming the statement judging panels. Furthermore, Thurstone’s procedure only determines if participants either agree or disagree with a statement and this is unsuitable for rating agreement. However, one potential use of Thurstone’s scale has now been suggested for use in a Delphi study with many items. Mead (2013) suggests that a Thurstone-like procedure may be used in a literature review based first round where experts agree only with those statements important to the development of MAC for FGIDs. The chosen statements would then be used as Likert scale items to gain consensus and/or degrees of agreement on a subsequent series of Delphi rounds. Furthermore, requesting experts to rank items in order of

importance for each Delphi survey section as suggested by Okoli and Pawlowski (2004) may also help to prioritise items which are considered by the expert panel as the most important.

5.4.1 Likert scales versus Likert-type scales

Likert never published a definitive account of his methods. Hence, there is uncertainty as to the number of scale points and types of attitude to be measured. Thus, some authors describe Likert scale categories in terms of agreement and others such as importance (Garland, 1991). Higher education websites further advocate many attitude categories for use in Likert scale construction (Siegle, 2010). It is, therefore, understandable that the term “Likert-scale” may be used inappropriately and indiscriminately.

Likert in his original paper used approval and disapproval categories (Likert, 1932, p14). It is interesting to note that the terms; agree and approve appear very similar. However, they differ where the term “approve” (of) suggests a position of authority where the rater gives their consent (or judgment) on an issue that they may not necessarily agree with. This is opposed to the term, “agree” (with) which refers to parties having the same view or opinion. It is possible that the use of “approve” categories may be seen as a sign that one seeks approval rather than agreement from a high-ranking expert. However, participants in Likert’s first study were not experts, but students. Here the term “approval” may be seen as a sign of the times given the period, the rater population and the topic under investigation (attitudes of students attending prestigious universities rating “Internationalism and Negro scales” in the 1929).

Defining a true Likert scale is not easy. Researchers use the term “Likert-scale” in different ways to apply to groups of items and single items where in both cases there is uncertainty as to what format to apply. Methods of construction described in Likert’s original article state that all statements be “expressions of desired behaviour” and not statements of fact (Likert, 1932, p44). Furthermore, each expression of behaviour should apply to the present, or I

suggest in the case of the Delphi technique, the future. Statements should also be unambiguous and worded so the model reaction is approximately in the middle of possible responses (Uebersax, 2006). From Likert's original article, scale characteristics can be stated as:

1. The scale contains several items
2. Response levels are arranged horizontally
3. Response levels are anchored with consecutive integers
4. Response levels are anchored with verbal labels that indicate more-or-less evenly spaced graduations
5. Verbal labels of attitude are bipolar and symmetrical about a neutral middle integer

As there is no "definitive" Likert-scale or verbal measures of attitude, the above criteria should separate a Likert scale from a Likert-type scale. In the below example (figure 5.2), all criteria are met where the attitude term "approval" is used.

Figure 5-2 A Likert item based on the recommendations of Rensis Likert

How do you feel about your local MP's performance on lowering outpatients waiting times?				
Strongly approve	Somewhat approve	Neutral	Somewhat disapprove	Strongly disapprove
5	4	3	2	1

Here, the statement expresses a behaviour, rather than fact and the anchor labels are distinctly bi-directional (bipolar) and symmetrical with a middle neutral point. If response labels are not bipolar or symmetrical, but the scale still demonstrates anchored and consecutive integers that are arranged horizontally (figure 5.3), it may be considered a Likert-type item as in the example seen below. Here we see a statement of fact. There is no exact opposite of "very often", and the opposite of "never" should be "always".

Figure 5-3 An example of a Likert-type item. The choice options are arranged horizontally in a scale. However, the scale is not bipolar, nor does it represent a range from extreme agreement to disagreement

How often do you think about seeking medical attention?				
Very often	Often	Average	Sometimes	Never
5	4	3	2	1

Clason and Dormody (1994) describe differences between Likert scales and Likert-type items in a different manner. They suggest that Likert-type items are individual questions or groups of single questions that use Likert scale properties where no attempt is made by the researcher to combine responses from the items into a composite scale. Alternatively, Likert scales are composed of a series of homogeneous items that represent an attitude where scores are combined to gain a total that represents a character of personality trait. However, when applying Likert scales to Delphi surveys, researchers are aiming to measure attitudes of the respondents for each survey item. The scores of all the respondents are summarised for each item to find the respondents' attitude to that element of the study. Here, there is no value in summarising the responses of each participant to all items of the study, which more indicates the individual participant's character. I followed Likert scale protocols for scale development but followed Delphi protocols of calculating the percentage of "agree" categories across each item to gain group agreement levels. However, I also calculated the median and distribution of the total scores to show the distribution and stability in overall agreement levels of the expert panel over all three rounds. For this study, I used a combination of "agree" and "important" verbal labels depending on the content of each set of items. Figure 5.4 below shows an example of a 5-point Likert-scoring matrix with four items scored by four participants'. Here I show that both individual participants' character (rows) and percentage level of group agreement for each item (columns) can be simply calculated and displayed.

Figure 5-4 In this example of a participant scoring matrix using a 5-point Likert scale, “5” represents strongly agree and “1” represents strongly disagree. Participant rating scores of 4 and 5 are added to find the percentage level of agreement. In this matrix, the columns represent group survey scores and the rows, individual survey scores.

	Item 1	Item 2	Item 3	Item 4	Total individual score
Participant 1	4	4	2	5	15
Participant 2	4	3	2	5	14
Participant 3	3	3	3	4	13
Participant 4	4	2	3	5	14
% agreement	75	25	0	100	

As can be seen above, defining a true Likert scale is not straightforward. Given the different approaches offered, and that Likert did not publish any definitive guidelines, common-sense criteria appear the most obvious alternative. Here, statements describe behaviour and not fact, where verbal categories define an attitude most relevant to the content of the statement and that the scale consists of a set of homogeneous items representing the same construct.

5.4.2 Category labeling

When participants respond to a question on a rating scale, they match the end points of their cognitive image to the end categories of the rating scales, where equally distanced categories in between divide this image into equal measures (Menold et al., 2014, p22). However, paying sufficient attention to all the verbal categories can become complicated or tiring, as respondents are faced with increasing amounts of information to consider. Churchill and Peter (1984) found no differences in reliability between Likert scales with each response clearly defined and those with only anchor points labeled. Not surprisingly however, most studies show respondents being more satisfied when each point is verbally labeled (Dickinson and Zellinger, 1980), and that clearly labeled categories yield higher test-retest and inter-rater reliability than those with only end points labeled (Weng, 2004, Menold et al.,

2014, Peters and McCormick, 1966, Bendig, 1953). Understandably, clearly stating each label also enhances the interpretation of central tendency measurement (Weng, 2004). Thus, I selected verbal labels for each category, because features such as participant satisfaction, study reliability and participant understanding were essential to the validity and final results of this study.

5.4.3 The number of Likert scale points

Reports on effects of the number of Likert scale points on reliability are conflicting. This is due to the effect of latent variables. The obvious criterion for choosing the number of scale categories is the ability of subjects to discriminate between categories (Komorita and Graham, 1965). Here, a scale with few categories may not allow the participant to make full use of their capacity to discriminate, whereas, a scale with a large number of categories may be beyond the participant's ability to accurately rate the correct category. In both situations there is a potential for an increase in measurement error (Symonds, 1924). Scales using a two-point scale such as agree-disagree are much shorter, more convenient to administer and score (Komorita and Graham, 1965). So why do so many investigators recommend the use of five or seven-point scales?

5.4.3.1 Reliability measurement

It is generally thought that large numbers of categories increase a scale's reliability. However, there are different measures of reliability. Internal consistency measures the degree of inter-relatedness among individual items, whereas test-retest reliability assesses the consistency of scale scores across time using the same participants. Previous findings on the relationship between the number of response categories and internal consistency (coefficient α) are conflicting. Here, some studies show that the number of response categories has no effect on coefficient α (Aiken, 1983) while others find it to be affected by the number of scale points (Matell and Jacoby, 1971, Weng, 2004). Concerning test-retest reliability, research is equally unconvincing where studies incorporate a range of numbers of response

categories (Bendig, 1953). As such, results from studies examining the effects of the number of scale points on rating scale reliability show that in some situations only two or three categories are required (Matell and Jacoby, 1971), while in other conditions more than seven (Preston and Colman, 2000) and even up to 18 categories have been justified (Champney and Marshall, 1939).

5.4.3.2 Item homogeneity

Item homogeneity is the likely explanation for these contradictory findings. Komorita and Graham (1965) found that while increasing the number of Likert item categories increased internal consistency; they did so only in certain situations. Consequently, they next studied the relationship between the number of item categories and the internal consistency reliability of scales under varying degrees of item homogeneity. Their study, later replicated by Weng (2004) showed that reliability of homogeneous questionnaires was independent of the number of response categories used while the reliability of heterogeneous questionnaires was improved with the addition of categories. These findings suggest that scales measuring one or two constructs have fewer latent variables and as such fewer measurable constructs. In such cases, the relevance of each item to the scale is a greater indicator of reliability than the number of item categories. Alternatively, heterogeneous scales measuring many constructs have more latent variables and therefore more measurable concepts. Thus, increasing the numbers of item response categories permit more attitude responses that reflect the concepts of the scale, thus reducing random errors and increasing reliability.

5.4.3.3 Diversity of opinion

The influence of diversity of opinion on the number of categories is another factor that is rarely discussed. I found one study that tested two questionnaires covering different topics; both with a homogeneous set of items that were scaled with 2, 3, 4, 5, 6 and 7 agree-disagree categories. Masters (1974) interestingly found that where opinion was diverse, the number of categories had only little effect on internal consistency. Strangely, however, where there was little discrimination between participant opinions, reliability increased with the number of

categories. Since this type of study has not been repeated, it is difficult to know if this is likely to be a consistent finding of consistency and as such any meaning. Repeated measures in this instance would help not only to confirm reliability, but also to strengthen validity, something that is rarely considered in most studies evaluating Likert scales. However, although reliability may increase, this could be an artifact caused by response biases. Therefore, as stated by Lei Chang (1994), both reliability and validity must be considered where both the respondents' attitudes (reliability) and attitudes reflected by the items (construct validity) determine the responses on a Likert scale.

5.5 Summary

To analyse degrees of attitude from experts involved in the development of FGID diagnostic criteria, it was necessary to understand the concept of attitudes before researching suitable scales to measure them. I chose a five-point Likert scale for its ease of construction. Using Delphi technique with a Likert scale also gathers data showing degrees of attitude on a set of items over a series of rounds where participants have a chance to review and change their responses in light of other participant responses.

Chapter 6. The development of first Delphi round

6.1 Overview

This chapter discusses the development of the Delphi survey, starting with construction of the first round survey items. I then present the development of the web-based Delphi survey, participant recruiting procedures and data analysis techniques. Feedback from the pen-and-paper and web-based Delphi pilot studies are then presented.

6.2 Item development

I aimed to make the survey relatively short while maintaining an adequate number of questions to gather sufficient attitudinal data on the development of a multi-axial assessment for FGIDs. Therefore, before drafting the first round survey the following questions were considered:

- What am I trying to analyse and does this question and/or item help?
- Who is the intended audience and is this item appropriate?
- Is the information returned useful or just “nice to know”?
- Should an item be mandatory or optional?
- If it is a mandatory item, can all respondents answer it?
- Are possible responses to the items consistent? i.e. don’t mix verbal categories

Firstly, I developed the questions on paper. This was edited several times adding items that were more relevant or dismissing items where information gained would only be “nice to know”. I set all Likert scaling points as mandatory but added comment boxes next to each item and at the end of each section. Here, experts were able to justify their decision and offer

further opinion they felt was important to the content of an item. Feedback between rounds also gave participants the opportunity to check discussion provided by other panellists.

6.2.1 Participant screening questions

It was not known exactly which area of expertise participants were presently working, the number of years they had consistently worked in their field or their present geographical location. These screening questions confirmed the eligibility of each participant and made available data that could be analysed inferentially if required later.

6.3 The study questionnaire

Brace (2010c, p174) suggests that a survey begins with broad themes and ends with specific topics. I began with a set of introductory items relating to the biological and psychosocial features of FGIDs. I then divided the questionnaire into the following sections

- Contributing factors to FGID
- The therapeutic relationship
- Areas of assessment
- Multiaxial assessment and formulation relevance to FGID

6.3.1 Contributing Factors to FGID

In this section, items focused on factors that may initiate, exacerbate or maintain FGID symptoms. As FGIDs are considered as a set of multifactorial disorders, it was necessary to explore biological, psychological and social features associated with the evaluation and diagnosis of FGID. Thus, statements on areas of possible contribution were divided into the following sections:

- Physical origin
- Psychological origin

- Gender differences
- The impact of symptoms on daily life
- Genetic polymorphism

6.3.2 The Therapeutic Relationship

Clinicians should be skilled in many areas of communication, reflective practice and clinical competence. They must also have insight into the relative influences of different factors on FGIDs. Clinicians must also be aware of skills required to both elicit and assess the patient's symptoms, and their beliefs, concerns, expectations, and misunderstandings about their complaint. Additionally, clinicians are expected to offer education and discuss customised treatment strategies. Thus, items concerning the therapeutic relationship were divided into the following subsections:

- Clinician qualities and experience
- The patient-clinician relationship
- Cultural factors

6.3.3 Areas for Consideration and Possible Measurement

Due to a lack of gold standard testing, FGID diagnosis is often one of exclusion following comprehensive investigations. FGIDs have a significant impact on physical, emotional, social, occupational and cognitive function. Thus, considering the biopsychosocial impact of FGIDs on quality of life, various domains and factors require consideration. I divided items into the following subsections:

- Abdominal Symptoms
- Patient description of abdominal symptoms
- Emotional function
- Social function
- Physical function

6.3.4 Multiaxial Assessment and Formulation (Relevance to FGID)

Multiaxial assessments evaluate different areas of information believed to be significantly relevant to the presenting patient. It was important to gain opinion on advantages and disadvantages of MAC so experts could review and judge if these factors either hinder or enhance FGID assessment. Furthermore, as neuroscience and genetic research advances data from these areas may also help identify risk factors which may then be targeted by novel therapies targeting these specific disorders that make up the FGID spectrum (Hyman, 2002, Farmer et al., 2010). I divided statements on MAC and future research into the following subsections:

- Advantages
- Disadvantages
- Future research: identifying risk factors and replication of findings that may benefit MAC in FGID patients

6.3.5 Concluding Statements

I developed these statements to gain opinion on current and future benefits of MAC when evaluating FGIDs patients based on information provided through the survey. Finally, I asked experts' to comment on the content of the survey and any areas of information they felt had not been explored.

6.4 Web-based Delphi technique

The Internet offers a promising and rapidly developing instrument for Delphi research (Donohoe et al., 2012, Turoff and Hiltz, 1996). Web-based Delphi technique computerises and automates the Delphi technique in order to optimise its ability to reach widespread and diverse groups. Web-based Delphi studies rely on an Internet-based platform that provides a

convenient and acceptable alternative for organising, controlling and enabling communications between the researcher and expert panel (Donohoe et al., 2012, Deshpande et al., 2005).

6.4.1 Convenience

A researcher can login to a virtual test centre at any time to design, adjust or monitor any number of activities. Researchers can also select different levels of anonymity ranging from entirely anonymous to the disclosure of user profiles such as location and area of expertise. Investigators control access to information where participants either see only their own responses or they can see the contributions from other experts, which is fundamental to the Delphi technique. Both researchers and participants can also access a web-based Delphi study anytime or anywhere when it is convenient for them to do so thereby allowing a “real-time” process where the researcher and participants are connected as the Delphi process develops (Cole et al., 2013, p10).

6.4.2 Time and cost advantages

Savings in time and cost are the biggest advantage of Internet-based research (Donohoe et al., 2012, p40). Web-based Delphi technique eliminates long delays between Delphi rounds as with post mail communications. The time associated with conventional Delphi is seen both as a cause of attrition and as a deterrent to Delphi research for both the researcher and participants. Donohoe et al. (2012) and De Villiers et al. (2005, p642) note that the major advantage of web-based Delphi studies are the cutting of costs, time and effort where an expert panel can be canvassed rapidly and inexpensively.

6.4.3 Administration

Web-based survey tools and statistical software assist the organisation and storage of large amounts of data, reduce the risk of error and increase the transparency of analytical processes. This primarily applies to the Delphi technique where large amounts of quantitative and qualitative data are gathered over several survey rounds. Additionally, web-based tools used to develop Delphi studies provide helpline access, survey management (data storage, statistical summaries) and survey tool updates. Researchers can also export data to Microsoft Excel™, SPSS™ and other software programs for further analysis. Participants and other interested parties can also access statistical reports throughout the course of a web-based Delphi study through either an established web portal or via email request (Cole et al., 2013, p11).

6.4.4 Web-based Delphi technique limitations

While authors recognise the advantages of the Internet for Delphi research, they also warn about challenges concerning the design, implementation and evolution of web-based Delphi research. Obvious limitations lie with Internet accessibility, technical difficulties and inconvenience of entering data using computer data screens compared with the convenience of hard copy page-turning (also reported as an advantage) (Donohoe et al., 2012, p42).

6.4.4.1 Access

Internet access depends on two fundamental properties. Firstly, the economy and infrastructure of a region may hinder access in developing countries. However, such countries are catching up, with China now passing the United States with the most Internet users. Nearly 40% of the world's population now have Internet access; a rise of over 30% in the past 10 years (Miniwatts Marketing Group, 2103). However, unreliable Internet access presents challenges for web-based Delphi researchers where dial-up access is interrupted by weather or controlled by political interventions. Donohoe et al. (2012) highlight this as an

issue when conducting health research in rural, indigenous or disadvantaged communities where access remains relatively expensive.

6.4.4.2 Research control

Web-based Delphi technique is “virtual” regardless of the type of communication due to the anonymity required as part of the methodology. However, as the Delphi technique requires purposive sampling where participants are selected for their level and specificity of expertise, methods must ensure that expertise is assessed thoroughly (section 6.7.1). If not, anonymity itself can present concerns related to representation where a degree of uncertainty exists concerning the knowledge of each participant identity.

6.4.5 Maximising web-based Delphi technique

Web-based Delphi technique best suits this study as potential participants are geographically dispersed. All prospective participants were contactable through their clinical and academic email addresses. Thus, experts who accepted could receive up-to-date communication. Moreover, those who did not initially accept could be sent reminders to participate. Unfortunately, one form of technology not available at the time of study period was the use of mobile and smartphone software. This would have been an ideal platform for communication and data collection as virtually all potential experts travel, thus allowing considerable ease in participation via either the Internet or mobile communications.

6.5 Survey format and design

The Bristol Online Survey (BOS) is an online tool that can develop, launch and analyse web-based surveys. The BOS contains a number of pre-defined question types, pre-populated templates and established question styles. I developed a survey using a range of pre-defined question styles including multiple choice, single-line answer and multiple answer question styles (figure 6.1).

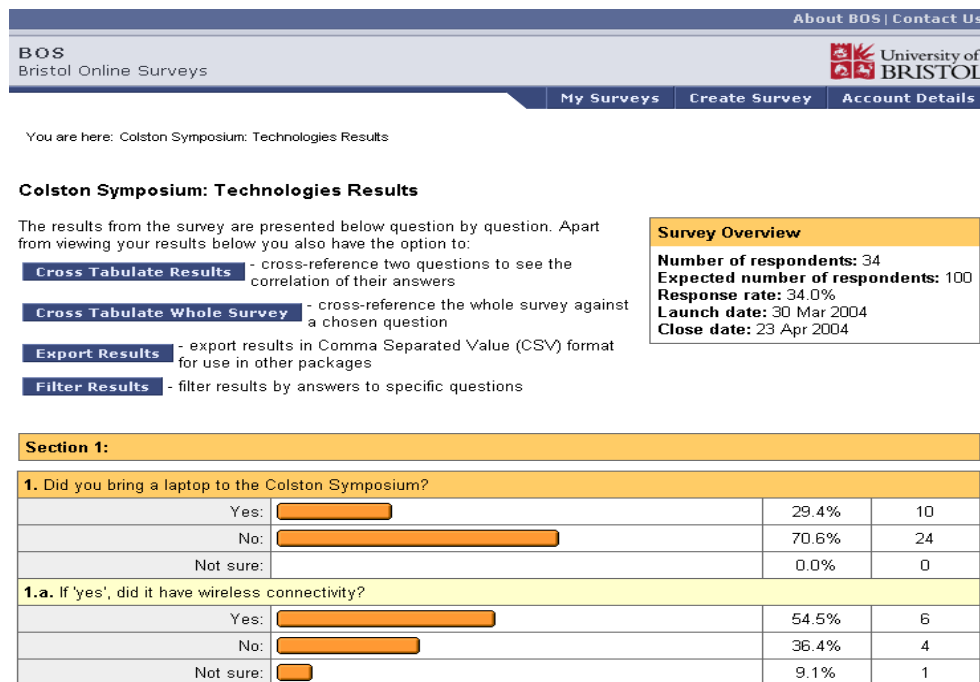
Figure 6-1 The Bristol Online Survey: a section from the first round Delphi survey showing a five-point Likert scale

9. Impact of symptoms on daily life.						
	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Comments (Optional)
a. Reduced health-related quality of life.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
b. Increased healthcare seeking behaviour.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
c. Help from significant others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
d. Increased symptom vigilance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
e. Effect of current medication.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
f. Altered daily toilet habits	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
g. Altered daily occupational functioning.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
h. Altered daily family functioning.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
i. Other.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

6.5.1 Survey results and analysis facility

BOS supply a “real time” survey summary that displays respondent numbers, response rates, launch dates, closing dates and a breakdown of data including the number of responses for each question. Tools for further analyses are also displayed that includes a cross-tabulation tab allowing cross-referencing and correlation of results between surveys. Additionally, filter tools allow inclusion or exclusion of questions, an extra statistics tab that show the mean and standard deviation for each question and an export results tab that allows the exportation of results to other software packages (figure 6.2).

Figure 6-2 An example of a results page including a survey overview



6.6 The expert panel

The selection of participants is critical to Delphi research, as their opinions form the primary data. To reduce sampling and response bias, rigorous procedures for selection ensure the identification of suitable experts (Okoli and Pawlowski, 2004, p6). The Delphi technique does not use random sampling: it relies on the experiences and knowledge of each participant regarding the subject being investigated. Sampling in qualitative research in relation to this study is discussed in more detail in chapter eight (section 8.5.1.).

6.6.1 Recruitment of experts

I sampled the entire population of experts that were involved in the development of diagnostic criteria for FGIDs. I chose this sampling method because the size of the population with the above characteristics is very small. I defined the population characteristic as experts currently assigned to committees and advisory boards of The Rome Foundation and The International Foundation of Functional Gastrointestinal Disorders. I targeted this group as they are responsible for the development and implementation of internationally recognised international diagnostic criteria for FGIDs. This panel was predominantly made up of clinical GI consultants but also included members from other specialities involved in patient treatment (e.g., clinical psychiatrists and pain specialists). Participants were also required to have published research relating to FGID diagnosis in the previous 5 years and be consistently employed in their field of expertise. Experts were chosen for their expertise in relation to the research question in addition to them being potential stakeholders in the outcome of the study due to their involvement in the development of diagnostic criteria.

6.6.2 Contact of expert population

The invitation that was sent to potential participants in round one contained information on the aims of the study, how to use the online survey tool, the expected time to complete, their

expertise in the subject being investigated and information about how data would be handled throughout the study period. This letter also had a hyperlink to the survey and my contact details if they had any questions. Consent was implied when experts responded to the email to inform the author of their agreed participation or after completion of the first Delphi round (Harper et al., 2012, Alahlafi and Burge, 2005). Subsequent invitations to later rounds included thanks for participation in the previous round, attached copies of personal and group data, a brief written update of the study and link to the next survey round.

6.6.3 Reminders

Reminders were sent two weeks after the initial invitation, one week before the end of the round and one day before the end of the round. Each reminder stated the time the survey had been open, the percentage of experts who had returned their survey so far and a link to the survey. Additionally, in rounds two and three, I asked participants in all reminder letters if they wanted extra time to complete due to their heavy workloads.

6.6.4 Confidentiality and anonymity

Experts were directed to the Bristol Online Survey website via a hyperlink embedded in a personalised email as this prevents the risk of computer viruses which can occur when downloading material from the web (Duffy, 2002, p87). Experts were also contacted through their research work email address from the author's university email address in order to avoid the message being treated as unsolicited mail by spam filters. Individual responses and participation were kept confidential. Experts were also guaranteed anonymity throughout the full Delphi study, report writing and subsequent publications. I protected identity of the experts by corresponding with each expert individually. Additionally, all computers used during this study were password protected and stored in a lockable room when not in use.

6.7 Data Analysis

Data gathered from Likert scales are considered ordinal as responses are ranked by a set of numbers expressing levels of attitude assumed not to have equal distances between categories (Hildebrand et al., 1977). Categories for an ordinal data set have a natural order, in the case of this study, an attitudinal rating scale of 5 to 1. However, the interval between each category on the scale may not be equivalent in magnitude and as such, there is no measure of the distance between two scale values. In other words, Likert scale categories represent relatively “more or less of something”, such as pain intensity scores or participant attitudes. Furthermore, all verbal categories are mutually exclusive and ideally exhaustive.

When analysing nominal and ordinal data, hypotheses are evaluated using non-parametric tests because they do not require the data to fit an assumed distribution. The assumed variance may not be homogeneous, and these tests make no assumptions that numerical values are equally spaced. For instance, in a test-retest reliability trial for a five-point Likert scale for satisfaction, a participant scores ‘2’ at time point one and scores ‘4’ at time point two. Under ordinal conditions, these data suggest that the participant’s satisfaction has increased from second to the fourth position on the scale. However, under interval conditions, these data imply that the participant experiences twice as much satisfaction at time point two than at one, which, as a result, may affect internal consistency. In other words, nonparametric tests rank outcome variables from low to high or from best to worst (Motulsky, 1995). However, if the data is not assumed to be drawn from a normally distributed population even interval and ratio data may require non-parametric tests. For these reasons, data analysis is driven by the research question, i.e., the specific objectives which preserve the meaning and the characteristic of the scale.

6.7.1 Univariate Analysis

Univariate analysis involves the evaluation across cases of one variable at a time. Three major features of a single variable are necessary to assess:

- Distribution
- Central tendency
- Dispersion

6.7.1.1 Distribution

The basic feature of data in my study was to define levels of attitude using descriptive statistics that summarise a data set without employing a probabilistic formulation. In contrast, inferential statistics attempt to reach a conclusion beyond the immediate data alone. For example, inferential statistics may infer from sample data what a given population may resemble, or assess the probability that the observed difference between groups may have occurred by chance in a particular study, on the basis of a null hypothesis (Trochim, 2006)

6.7.1.2 Central Tendency

The central tendency of a distribution is an estimate of the centre of distribution of values.

There are three main types of estimates of central tendency:

- Mean
- Median
- Mode

The mean or the average is the most commonly used method to describe central tendency and is given in over half mainstream medical research papers (Harris and Taylor, 2009).

However, as data collected from this study are ordinal, parametric methods with calculations based upon mean are invalid as these tests are designed for continuous and normally

distributed data (with equal variance) and are not easy to interpret with regard to subjective ratings (Jakobsson, 2007).

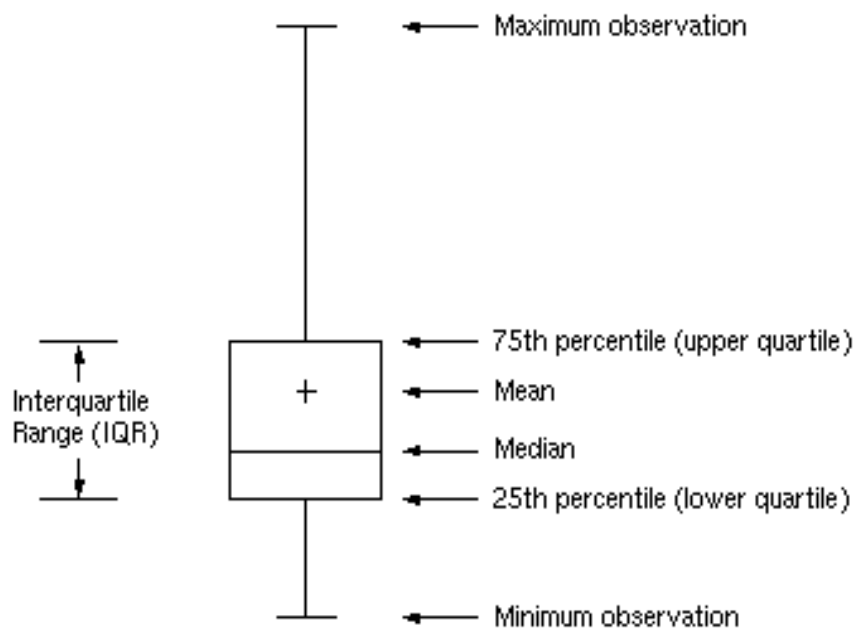
The median is the score found exactly in the middle of an ordered set of values and used to represent the average when the data are not symmetrical (distorted) and is more robust to outliers (Cottrell and McKenzie, 2010). Therefore, median values are typically used with non-parametric tests.

The mode is useful when data are asymmetric or skewed and simply labels the most frequently occurring event. This makes Delphi questionnaire results much easier for the analyst and the experts receiving statistical feedback to interpret. Concerning this study, the median will show which Likert scale points have most responses, which are converted to percentages of experts who agree or disagree.

6.7.1.3 Dispersion

Dispersion shows the spread of values around the central tendency (Trochim, 2006). There are two common measures of non-parametric dispersion, those being the range and quartiles. The range describes the difference between the largest and smallest values; however, this gives minimal information regarding the spread of the data. The interquartile range encompasses the 25% - 75% of values that better describes the distribution of the data. Put differently, the interquartile range defines the range between the first and third quartile and contains exactly 50% of the data within the distribution (Lohninger, 1999). Box-and-whisker plots are a simple and straightforward way to display these values (figure 6.3):

Figure 6-3 A typical box and whisker plot showing interquartile ranges and extreme markers (SAS Software Solutions, 2015)



Standard deviation is used for data which are normally distributed, to present information on how much the data differ around their mean (Harris and Taylor, 2009). Normal distribution of data suggests that most variables in a data set are close to the “average” while relatively few variables lean to one extreme, or another. However, standard deviation, like mean, is not appropriate when ordinal data are considered. Additionally, Likert-derived data often produce skewed or polarised distribution (most participants agree or disagree) and is, therefore, not considered to be “normal distribution”, to which standard deviation is suitably applied (Jamieson, 2004; 1218).

6.7.2 Comparisons between medical specialities

In order to determine if differences in opinion exist between gastroenterologists and “other” health care professions on the content of Delphi items, I used the Mann-Whitney U test. This is a null hypothesis test that combines and ranks combined data from two independent

samples from the highest to lowest values. The average of ranks in each group is then found and then shown as two averages. If the means of the ranks in the two groups are very different, the p value will be small, meaning the null hypothesis is less likely. (Cheung and Klotz, 1997).

6.7.2.1 Multiple group comparisons

I considered comparing attitudes of experts from each of the nine specialist fields responding to my study. The Kruskal-Wallis test is an extension of the Mann-Whitney U test that allows the comparison of more than two independent groups. This test is appropriate when dependent variables are measured at the ordinal or interval level such as Likert scales and when the independent variable consists of two or more independent groups (Lund and Lund, 2013). However, the Kruskal-Wallis test also assumes that individuals must be randomly selected from a given population, while experts for this study were purposively sampled for their particular area of expertise and experience, regardless of location. Furthermore, sample sizes must be as equal as possible and be large enough to generate reliable probability (p) values, where in this study, numbers of experts in each group varied considerably and where most groups contained between one and three subjects. All but one sample was very small, therefore, as recommended by (Cumming, 2008) I considered them too low and unreliable as a strength of evidence.

6.7.3 Confidence intervals

Kline RB (2004) and Cumming (2008) recommend using confidence intervals when analysing attitudinal and behavioural data, I followed the convention of using 95% CI of the median which shows the range centred around the median point of each item and where the width either side of the median is called the margin of error. Cumming (2008) argues that the p -values from small single studies give only vague information about what is likely to happen in a replicate study. In contrast, a confidence interval improves representation and

gives a better sense of a whole set, including inherent uncertainty. Button et al. (2013, p367) also state that small sample studies have limited statistical power, poor positive predictive value and a greater likelihood of type I error. If the result does happen to be positive, these studies generally exaggerate the effect size, (that is, they suggest a greater difference between two variables or the strength of an association between two variables than is actually present in the overall populations sampled).

6.7.4 Responders versus non-responders

I tested response bias by comparing responders with non-responders and by comparing early with late responders, checking if late responders had similar traits to non-responders. For this, I used Pearson's Chi-square test that determines the goodness of fit between theoretical and experimental/observational data where random samples are taken from defined populations. The resultant tables are described as unconditional because neither the row nor the column marginal totals are fixed in advance. Thus, the hypothesis is concerned with "goodness-of-fit" (Ludbrook, 2013). Chi-square tests are not appropriate for analysis of subject performance scores (0-100); however, Chi-square is appropriate if subjects are categorised into "pass" or "fail" groups or in the case of this study, "respond" or "non-respond". Additionally, all variables must be mutually exclusive, i.e., participation in one category excludes participation in another (Ling and Fall, 2008). For response versus non-response and early versus later response, I tested four traits that were academic title (three variables), job speciality (five variables), gender (two variables) and work location (three variables).

6.7.5 Analysis of Delphi study data

I chose Cronbach's alpha (α) to quantify the reliability of responses of participants. The internal consistency of the Delphi panel for each item would be expected to reflect the level of consensus within the group for the importance of that item (Graham et al., 2003a, p 1153).

If the reliabilities of each rater for item importance were known to be equal, then Spearman-Brown formula could be used to estimate reliability of the sum of the raters (Bravo and Potvin, 1991). However, as the reliability of each participant is unknown, I chose Cronbach's α because the sums of the participants' responses are more important than the sum of reliabilities.

6.7.6 Qualitative Analysis

To analyse comments and suggestions given by the participants, I examined data using thematic analysis. I categorised information by locating themes relating to different areas of knowledge contained in the survey. Thematic analysis in my study relates to the grouping of similar occurrences of data under an umbrella term. This grouping process allowed accurate placement of statements submitted by the participants into the appropriate section or sub-section of the second round survey.

Thematic analysis was broken down into the following phases:

- **Familiarisation with the data** (reading and noting initial themes of information)
- **Generating terms** (collating data into relevant terms relating to sections of the survey)
- **Searching for themes** (gathering all data relevant to each theme from sentences and/or paragraphs submitted by participants)
- **Reviewing themes** (check all themes relate to a term)
- **Defining the themes** (refinement of themes into the form of statements that may be clearly understood by the participants taking part in round two)
- **Producing the round two survey** (input the refined statements into the relevant sections and subsections)

This type of qualitative approach was used, because the intent of analysis was not to report patterns and themes within a long, complex narrative, but to analyse phrases and short paragraphs written in sections already categorised in the survey.

6.8 Piloting the Questionnaire

I piloted the first round survey using paper and online versions in order to establish the intended meaning of each item and to determine whether respondents interpret items as intended by the researcher (Bowden et al., 2002). Another aim was to estimate the time taken to complete a survey. Additionally, the questionnaire was checked for errors in grammar or format. Piloting a questionnaire should check four areas: validity, reliability, error testing and trustworthiness (Brace, 2010b).

6.9 Paper pilot study

Prior to the online pilot survey, paper versions were given to fellow members of the gastro-network research group at Karolinska Institutet. Specialities included a gastroenterologist, an orthopaedic surgeon researching spinal influences on visceral pain, a neurobiological scientist and an integrative medicine researcher. Each test subject was asked to give feedback on the time taken to complete the survey, the survey's readability, usability, clarity of instruction, and an understanding of each question and format.

6.9.1 Feedback

6.9.1.1 Areas of expertise:

There were several queries concerning the areas of expertise. There was confusion regarding two options for "Neurogastroenterology" that were "basic science" and "physiology & motility/sensation" as partitioned by the Rome Foundation committees. Participants queried the separation of basic science and physiology. Consequently, I combined this category and dropped the basic science and physiology label as the two labels are comparable.

6.9.1.2 Description of participant nationality

Participants were also uncertain if “description of nationality” meant place of birth or country of residence and practice. For the purposes of this study, it was important to elicit the country of residence and practice as I wanted to gain opinion on the population status of FGID patients.

6.9.1.3 Fixed periods of health worker experience

I asked how many years of experience the participants had in their area of expertise. The format for this question was originally described in blocks of years (e.g., 5-10 years, 20-30 years etc.). One participant pointed out that open format questioning is far more flexible in terms of statistical possibilities when analysing data if the need for comparative analysis were to arise. While these statistical possibilities are of more use to larger samples, the term “consistently” was added to the question in order to clarify full-time work as oppose to inconsistent periods over an occupational duration. This style of question removes the possibilities of ignoring time away from their field of work during their career.

6.9.1.4 FGID questions

Generally, participants questioned the wording of the question topic prior to each list of statements. For example, in question five, statements relating to the physical origins of FGIDs were originally worded as “previous and/or ongoing visceral disease”. It was suggested that the terms “ongoing” and “previous” be split into separate questions as experts may agree with one feature but not with another. While I aimed not to include ambiguous items, I made some initial errors when trying to reduce the length of the survey. Additionally, question eight was drafted as “Impact of symptom related outcomes”. The author was referring to the reduced quality of life due to the effect of symptoms on the patient. This question was re-worded as “Impact of symptoms on daily life”.

6.9.1.5 Timing

The time taken to complete the survey was the biggest obstacle to maintaining the interest of the participant. I aimed to achieve a completion time of approximately 20 minutes. However, during early pilot trials, the time taken by some was around 35 -40 minutes. Therefore, a section on a potential MAC model was omitted. I also removed ambiguous and repetitive questions highlighted by the experts. Furthermore, the section examining FGIDs in the elderly was omitted and added as a single statement in question 17 (physical functioning). Before the international pilot study, the paper pilot study was distributed again, and time taken to complete the questionnaire was timed between 19 and 23 minutes. The native tongue of many participants was not English. As this is an international study, great care was taken in making survey instruction and items as clear and as simple as possible to understand without reducing their impact. Two Swedish supervisors also helped make the items clear for non-English speaking experts. Overall, the feedback was positive, the main point being that the survey was comprehensive concerning the issues of FGID assessment. Additionally, the items were described as “good”, “interesting” and “informative”.

6.10 International Pilot Study

6.10.1 Methods

I piloted the online survey on educational staff from the Pain Management Research Institute (PMRI) at the University of Sydney, and one research fellow from Karolinska Institutet in Stockholm was asked to participate. These included the director of PMRI, an epidemiologist, three pain medicine specialists, a psychologist, a physiotherapist specialising in chronic pain, and an integrative medicine expert. No participants worked with FGIDs, but they all had many years of experience in multidisciplinary medicine and functional pain disorders. Experts in the field of FGID were excluded from the development studies to allow a large cohort for the definitive Delphi study.

Participants were asked whether they understood each question and if they could answer them in an intelligible way. They were also asked to give “frank” feedback on the format of the survey concerning its appearance, ease of use, time taken to complete the survey and any features or content they considered relevant that was missing from the questionnaire. The pilot participants were also requested to complete the questionnaire. There were 132 questions divided into four sections:

- Risk and contributing factors to FGIDs
- The therapeutic relationship
- Areas for consideration and possible measurement
- Multiaxial Assessment and formulation relevant to FGIDs

Responses to each statement were described by frequency, median, range and percentage responses using Statistical Package for the Social Sciences (SPSS) Version 19.

6.10.2 Feedback

The online pilot survey showed the ease of accessing international samples of experts who were geographically dispersed. Levels of agreement were very similar throughout the survey highlighting the potential for consensus to be generated in a shorter period. However, there were some methodological limitations, which are discussed below.

6.10.3 Usability

Three participants completed the survey, with all three completing the survey in approximately 20 minutes. All the participants liked the format, believed the survey to be comprehensive and found the questionnaire easy to use.

6.10.4 Content

For additional content, participants suggested the inclusion of:

- “Other functional visceral syndromes” (e.g., irritable bladder) to question five (physical origin)
- “Genetic factors involved in the regulation of pain modulatory pathways” to question 9 (Genetic polymorphism)
- “Ability to work in a multi-disciplinary team to question 10 (Clinician qualities and experience)

I inserted the above suggestions, however due to the number of questions, and the issue of time, question 5f (previous myofascial dysfunction) was omitted due to its similarity with question 5e (previous physical trauma).

Other recommendations proposed by participants were as follows:

- Recruitment of FGID patients as participants
- More focus on integration and facilitation of “team-work” between healthcare providers and patients.
- Insert more “disagree” statements as most were “agree” type.

I did consider the inclusion of patients as participants. However, these participants would have to have significant knowledge concerning FGIDs to complete the survey. The theme of teamwork between the healthcare practitioner and patient is covered in question 10 and 11 whose statements examine the patient-practitioner relationship. Furthermore, I reduced the number of “agree” questions in order to keep the completion time at a level. I also included five further “disagree” items in order to prevent pattern answering

6.10.5 Failure to complete

Two participants declined, stating lack of available time to give to the pilot survey while another did not reply. Of the remaining six participants, three failed to complete the survey. The first, a physiotherapist stated that their area of expertise did not extend to the survey topic. The participant concerned had no background FGIDs knowledge and was included as a chronic pain specialist working in a multidisciplinary pain research centre. Their selection was an attempt to represent all aspects of the FGIDs; concerning chronic visceral pain and physical disability associated with FGIDs. The second test participant failed to complete the survey due a misunderstanding on how to save their answers with the intention of returning later to complete the questionnaire. While there are instructions at the end of the online questionnaire, the participant was not aware of the “finish later” tab at the end of the survey. Here they could follow instructions on the various ways of saving data for use later. The participant was unwilling to start the survey again. In light of this issue, I inserted a highlighted bullet point explaining the “finish later” instructions at the beginning of the survey.

After agreeing to take part, the third participant was sent the questionnaire; however, no correspondence could be made under the survey period with two reminder letters developed for this study. However, six weeks after the cut-off date, the participant made contact stating computer problems as the reason for the failure to open their email account. They asked if they could still help, however I informed them that round one was due to commence and thanked them for their time. I further stated that while they did not complete the survey, the issues of the reliance on functioning computers, and late participation was useful as a discussion point in the development of the Delphi study.

6.10.6 Pilot data collection and analysis

While simple data parametric analysis was available on the Bristol Online Survey results page, it was necessary to export the results to SPSS for further non-parametric data analysis. The instructions to export data files in Comma Separated Value format (CSV) appeared to be straightforward. The implementation of this process was less so as exportation was primarily designed for MS Excel. Once in MS Excel, data were then exported to SPSS as a comma-separated values (CSV) file. Likert scale codes were then reversed as SPSS scaling sequencing only recognise the first Likert point as the lowest number.

6.10.7 Pilot respondent participation in the main Delphi study

One online pilot participant requested to take part in the main Delphi study. Their reasons were that of wanting to contribute to the main study due to their interest as a participant. Several factors had to be considered; firstly, having been exposed to the survey, the respondent may have become more adept at using the online questionnaire compared to the experts sampled for the main study. Secondly, and more negatively, the participant may also show a decline in following protocols because the survey is no longer novel. However, because of pilot participant feedback, several items were omitted, and new ones included while a complete section was removed. Therefore, as the survey format had changed adequately together with a “run in” period before the main study of three months, I considered participant fatigue and response bias not to be a significant contributing factor. I followed literature recommendations and conducted frequency analysis with and without the above participant in order to assess the extent of the influence of possible contamination (van Teijlingen and Hundley, 2002).

6.11 Summary

Due to the specific knowledge required and relatively small numbers of experts involved in developing FGID diagnostic criteria, purposive sampling methods were applied to select

potential participants regardless of job location. I chose Web-based Delphi technique over the pen, paper and mail version due to minimal cost and its convenience in organising quick communications between the researcher and participants. I designed the order of first round survey items to reflect the order of the evaluation and diagnosis of FGIDs. I also selected data analysis models that reflect the qualitative and non-parametric data gathered as well as the size and sampling of participants. Finally, I ran two pilot studies (pen/paper and Internet) to evaluate the competency, the length, completion time and the format of the round one survey. Modifications as recommended by the pilot participants were implemented for the following round one survey.

Chapter 7. Results of the international web-based Delphi survey

7.1 Overview

In this chapter, I first give an overview of how round one was finalised and distributed, after which I report and tabulate the results of the main online Delphi survey.

7.2 From the pilot study to the first round survey

Using recommendations given by experts who took part in the pilot studies, I re-drafted the survey for the main Delphi first round. The round one survey contained 132 items divided into sections outlined in table 7.1. I employed three rounds of data collection and set the consensus level of 75% for each item by combining the number of experts scoring ‘strongly agree’ and ‘agree’ on a Likert rating scale. I then asked participants to re-rate and further comment on items in light of their own previous rating and overall panel response on items that gained less than 75% of consensus in the previous round.

7.2.1 Round one

One month before the first round survey, I contacted experts by standard mail using University of Edinburgh headed paper. Two weeks later, I sent an email and a week later, the first round survey. I allowed participants four weeks to respond to the questionnaire. I additionally sent electronic reminders marked as “priority” after two and three weeks and then one day before the survey closed. I also placed “read receipts” to observe the number of emails opened and/or deleted. In five cases, experts opened the emails up to 15 months after the cut-off date. Table 7.2 shows the countries of experts contacted to participate in round one of the Delphi study.

7.2.2 Round two

Items not reaching the consensus level of 75% in round one were included in round two as were additional items developed from expert feedback using thematic analysis described in chapter six. I continued to include comment boxes with each item, but not at the end of each section. I asked participants to re-rate if they so wished and to comment on items in light of their own previous rating and overall panel response. The deadline was set at four weeks. However, because of the busy schedules of many of the experts and their willingness to participate, I extended deadlines in order to collect as much data as possible. I waited until all participants requesting extra time returned their surveys before results were analysed.

7.2.3 Round three

I applied the same protocols in round three as for round two but did not apply thematic analysis. I mailed reminders at the same times as for rounds one and two over the four-week period but allowed participants to respond according to their own deadline. I sent a further reminder one week before each participant's deadline. Finally, I tabulated all data, which was then returned to participants, thanking them for participation, and stating that the study was closed.

Table 7-1 Outline of the round one subject sections from the Delphi survey for future development of MAC for FGIDs.

Survey section	Subsection
FGID related comorbidity	
Risk and contributing factors	Physical Psychosocial Impact on daily life Gender differences Genetic polymorphism
The therapeutic relationship	Clinician qualities and experience The patient-clinician relationship Cultural factors
Areas of possible measurement	Abdominal symptoms (in the absence of pathology) Patient description of abdominal symptoms Emotional function Social function Physical function
Multiaxial assessment criteria	Advantages Disadvantages
Areas of future research (brain-gut axis)	Identification of risk factors Replication of findings

Table 7-2 Countries of experts contacted round one of the Delphi study.

Europe	North America	Asia / Australia
Belgium (n=2)	Canada (n=5)	China (n=2)
Denmark (n=3)	Mexico (n=1)	India (n=1)
Eire (n=1)	USA (n=43)	Japan (n=2)
France (n=1)		Singapore (n=1)
Italy (n=3)		Australia (n=4)
Norway (n=1)		
Poland (n=1)		
Romania (n=1)		
Spain (n=2)		
Sweden (n=3)		
UK (n=13)		

7.3 Results

7.3.1 Participants

I invited 90 experts to participate. In round one, 40% (n = 36) of experts returned completed surveys. In round two, 86% (n = 31) of round one participants returned completed surveys and 100% (n = 31) in round three, yielding a 77% response rate across survey rounds. Some experts responded to the invitation explaining their ineligibility. There were several reasons for non-participation. Most cited heavy work load; however, they also expressed interest and encouragement toward success of the study. Several research scientists cited their non-clinical background as their reason for exclusion, while one expert tersely stated that they were not used to being told what to do and what's going to happen and were therefore not

interested. Additionally, some email addresses were not recognised and returned by the email server. Participation from Asia was reduced to one expert in India. Japanese participants were willing to take part, but were unable due to the 2011 tsunami. Email addresses in China were returned, and one expert from Singapore refused the survey due to lack of time and lack of “bandwidth”.

Respondents had an average of 20 years working experience in the field of FGIDs (table 7.4). In round one, ninety-six out of 132 items gained consensus. Thirty-three additional items were generated for round two using qualitative feedback from round one. In round two, 19 out of 69 items gained consensus, and in round three, nine items out of 50 gained consensus. Experts contributed substantial comment that covered all sections and rounds of the survey (appendix B). Themes from participants’ comments relating to a particular item and / or items in other sections were transcribed into the correct survey section related category for round two.

7.3.2 Participant total scores

The following figures and tables show the spread, distribution and median of participants’ total scores for each round. Total scores were calculated by summing all Likert scores for each participant in each round. Calculating participant total scores provide a simple overview of the agreement distribution between participants for each round and changes in overall panel agreement distribution between each round. Furthermore, plotting total score distributions for each round shows both symmetry and stability in score distributions for each round and changes in distribution symmetry between rounds.

Figure 7.1 shows interquartile ranges, median and extreme values for total scores in each round. Note that while ranges and quartile distributions appear to differ over each round, they do so due to the change in the number of items in each round. However, when the

number of items in each survey round was considered, the percentage spread of IQR and ranges remained relatively constant over all three survey rounds. Stem and leaf plots in table 7.3 summarise the change in distribution of total agreement scores for each round.

Figure 7-1 Shows the spread of total expert scores for **A** round 1, **B** round 2, **C** round 3. Data are expressed as box-and-whisker plots showing median, interquartile ranges and extreme scores

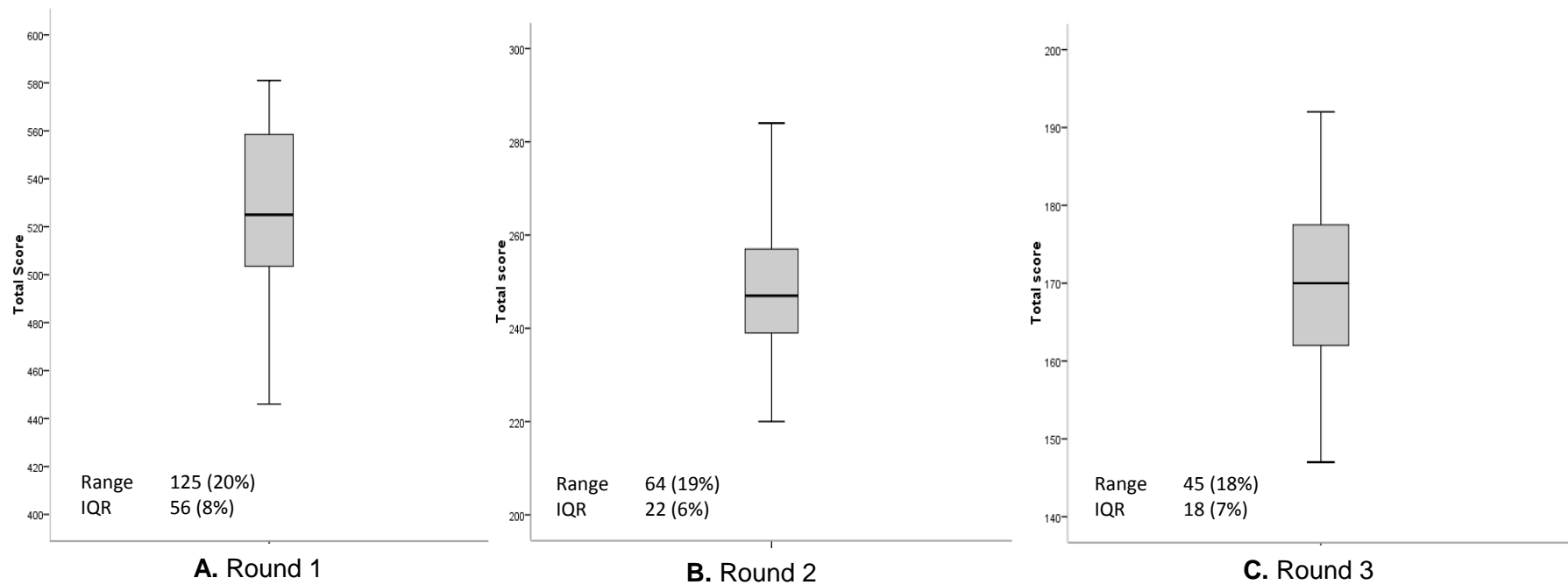


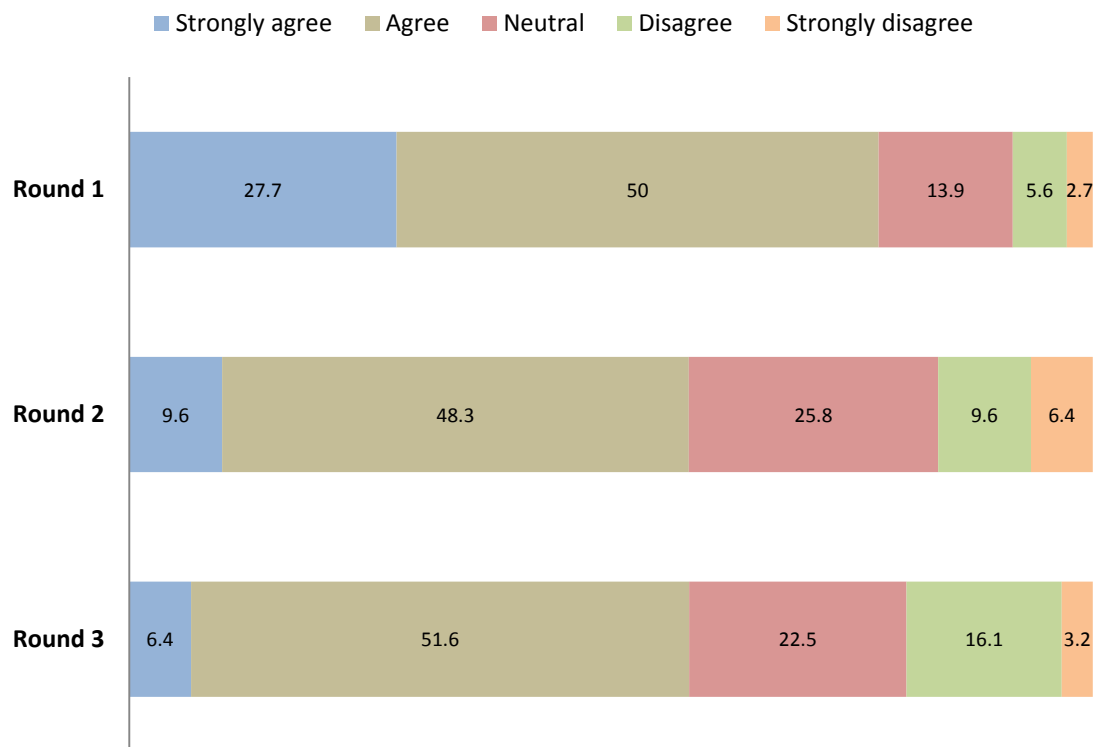
Figure 7-2 Stem and leaf plots showing distribution of participant total scores in rounds one to three with lowest scores at the top and highest scores at the bottom

Round 1		Round 2		Round 3	
Stem	Leaf	Stem	Leaf	Stem	Leaf
44	6	22	0478	14	78
46	6	23	14599	15	4589
47	09	24	0011267788	16	004567779
49	49	25	0033	17	0013446788
50	1234788	26	1378	18	1167
51	278	27	156	19	12
52	04688	28	4		
53	237				
54	08				
55	89				
56	156				
57	579				
58	11				

7.3.3 Distribution of expert responses to Likert-scale points for each survey round

Figure 7.2 shows the percentage of experts responding to the 5-point Likert scale used in this study. The figure shows that the ‘agree’ Likert category was the most frequently scored. The ‘agree’ point was also the most stable with only a 3% variation in percentage response over all three rounds. This figure further indicates the possibility of increased uncertainty in response to remaining items through the survey rounds where nearly 30% of experts score the ‘strongly agree’ category in round one compared to on 6.4% in round three. This uncertainty may also be shown with the increased scoring of ‘disagree’ categories from 8.3% in round one to 19.3% in round three.

Figure 7-3 Shows a stacked bar chart for the percentage of experts responding to each Likert-point in each Delphi rounds 1- 3



7.3.4 Participant characteristics

Self-reported professional background and experience among participants who completed all three round ($n = 36 / 31$) confirmed that the sample met the goal of reaching clinical and research experts known to have extensive knowledge of FGIDs (table 7.3). Participants were primarily heads of research departments, and all had authored, or co-authored peer-reviewed publications investigating phenomena related to the diagnosis of FGIDs. In examining the characteristics of responders versus non-responders, only job location was significant ($p = 0.0001$) with 61% of participants based in Europe, 33% from North America and 14% from Asia and Australia. There were no significant differences between academic title ($p = 0.768$), gender ($p = 0.278$) and speciality ($p = 0.204$).

There were minimal differences in agreement between gastroenterologists and other specialities showing differences in only 10% of items in round one, 7% in round two and 6%

in round three. Significant differences in judgement related to the use of non-directive interview techniques ($p = 0.005$), engaging with both the patient and their family when managing FGIDs ($p = 0.006$) with gastroenterologists marking higher agreement scores for the above items. Cronbach's α for round one was 0.956 indicating excellent internal consistency of participant opinion. However, internal consistency decreased through rounds two (0.853) and three (0.749). While maintaining good levels, the decrease in score shows less consistent agreement with reduced numbers of items in later rounds.

Table 7-3 Professional and academic profiles of respondents in first, second and third Delphi rounds.

Variable		Round 1 N (%)	Rounds 2 & 3 (%)
Academic title	Professor	26 (72)	24 (77)
	Associate Professor	2 (6)	1 (3)
	Senior research fellow	8 (22)	6 (19)
Primary occupational setting	Gastroenterology	19 (53)	15 (48)
	Psychiatry	3 (8)	2 (7)
	Paediatrics	3 (8)	3 (10)
	Primary care	2 (6)	2 (7)
	General surgery	2 (6)	2 (7)
	Pain medicine	2 (6)	2 (7)
	Physiology	2 (6)	2 (7)
	Integrative medicine	1 (3)	1 (3)
	Nursing	1 (3)	1 (3)
	Pharmacology	1 (3)	1 (3)
Consecutive years' experience with FGIDs	Median: 20 years		
	Range: 4 – 30 years (26)		
Job location	Europe	22 (61)	20 (64)
	North America	12 (33)	10 (32)
	Asia & Australia	2 (14)	1 (3)

7.3.5 Response rates (appendix E)

The overall response rate was low with 40% (n=36) of the total sample and 53% of the eligible sample. Expert sampling was broad in both location and speciality, but relatively narrow concerning academic title. Due to some regions being poorly represented, I grouped areas by continent. The same representation issues occurred with expert speciality. I compared gastroenterologists with all other specialist areas, as no other area of expertise had more than four representatives. Generally, there was no significant non-responder bias within or between groups with the exception of geographical location where responder rates in Europe were significantly higher than other continents. Chi-square analysis comparing gender, academic rank, geographical location and specialist field with responder rates are shown in appendix C.

7.3.5.1 Location

There were significant geographic differences in responder and non-responder rates ($p = 0.0001$). Sixty-one % of experts from Europe (n=22) responded to round one compared to 14% from Asia and Australia (n=2) and 26% in North and Central America (n=13).

7.3.5.2 Speciality

There were some demographic differences between responders and non-responders concerning individual specialties, however, overall they were not considered to be significant ($p = 0.204$). Fifty percent of paediatricians (n=3) and 61% of “other specialties” (n=11) had high responder rates compared to 0% psychologists (n=0), 25% of physiologists (n=2) and 39% of gastroenterologists (n=19).

7.3.5.3 Academic title

Overall, there were no significant differences between experts holding different academic titles ($p = 0.768$). Responses from associate professors were less (29%) than professors (40%) and research fellows (44%).

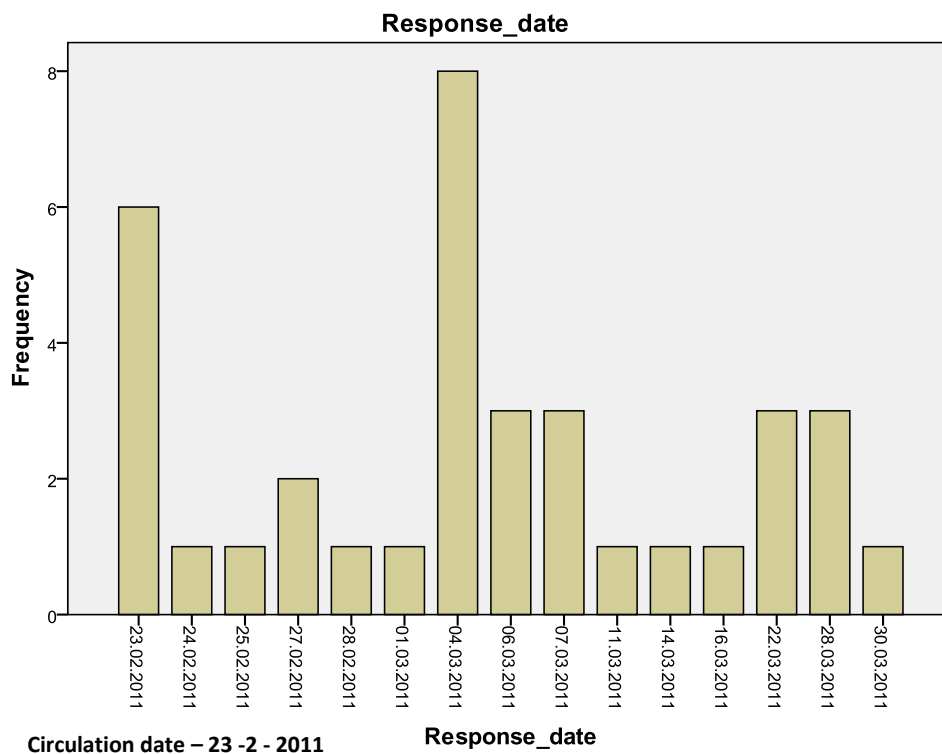
7.3.5.4 Gender

There were minor difference in response between male and female experts with 44% of male and 33 % of female experts responding to the survey.

7.3.5.5 Early versus late responders

Another strategy to assess non-responder bias is to compare early respondents with later respondents on the assumption that later respondents are more similar to non-respondents. This was done by counting the first 10 days as an early response and the remainder as a late response (figure 7.2). This assumption proved not to be the case; there were no significant differences in responder versus non-responder rates in geographical location. However, there were significant differences between late response and gender, with female experts responding significantly later than male counterparts do ($p = 0.017$) (see appendix E).

Figure 7-4 Responder-frequency timeline for the Delphi first round survey.



7.3.5.6 Reasons for non-response to the round one survey

Experts were not requested to respond to the invitation. However, 42 replied showing either willingness or reluctance to participate. Explanations for non-participation were,

- Heavy work commitments
- Not interested in the study
- Ineligibility for the study due to non-clinical expertise
- No longer working in the field of FGIDs
- Disagreement with the Delphi as a method of gaining consensus
- Disagreement with the style and content of the invitation letter

Three email addresses were returned and marked as “failure to deliver”. Alternative email addresses were sought, but none were found. One basic scientist was persuaded to participate while a further three physiologists nominated senior gastroenterologists working in their research groups.

7.3.6 Delphi process findings

7.3.6.1 Contributing and risk factors to FGIDs

Consensus was gained on items relating to the impact of symptoms on daily life, psychological risk factors and the contribution of previous GI infections to the onset and maintenance of FGIDs (table 7.4). Experts disagreed on the impact of physical trauma, previous surgery (both 53% in round three) and connective tissue disorders such as joint hypermobility syndrome (47% in round three) as being significant contributing factors to the onset of FGIDs. Most experts did not view physical trauma as a relevant risk factor in most patients. Experts also regarded surgery as an intervention in response to abdominal symptoms, or health-care seeking behaviour, while others suggested miss-diagnosis due to surgeons’ unawareness of FGIDs. In contrast to physically identified causes of FGIDs, experts agreed that ongoing myofascial dysfunction such as FMS (94%) do contribute to the

onset of FGIDs as they viewed these conditions as having similar underlying central mechanisms.

Experts agreed that the prevalence of FGIDs was greater in the female population (92%). However, opinion differed over valid explanations for these relationships. Experts agreed in principle that increased perception of symptoms may be related to central nervous system processing of visceral stimuli (83%). However, disagreement continued over likely pathophysiological mechanisms behind the effect of sex hormones on GI sensitivity and motility. Only one item gained consensus regarding contribution of genetic factors to the mediation of GI function and psychological disorders. Experts commented that while they agreed in principle with the gender and genetic-based items, a lack of human data and the need to translate present diagnostic criteria into endophenotypic (heritable phenotype associated with illness in the population) analysis needed to be more rigorously scrutinised.

Table 7-4 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.

Risk and contributing factors to FGIDs	Median	Range	Agreement %
	R1	R1	R1
	R2	R2	R2
	R3	R3	R3
Physical Origin			
a. Previous truncal surgery such as cholecystectomy and bowel surgery (<i>i.e., the effect of local injury repair mechanisms and/or the stress response on hypothalamic-pituitary-adrenal axis dysregulation</i>)	4	3	53
	4	3	62
	4	3	53
b. Previous visceral conditions	4	3	89
c. Ongoing visceral conditions	4	2	86
d. Previous neuropathic conditions	4	3	78
e. Previous physical trauma (<i>i.e., the effect of local injury repair mechanisms and/or the stress response on hypothalamic-pituitary-adrenal axis dysregulation</i>)	4	3	67
	4	3	42
	4	3	53
f. Ongoing myofascial dysfunction	4	2	94
g. Brain injury / condition (effect on CNS - emotion, cognition, personality)	4	3	69
	4	4	52
	4	3	58
h. Ongoing connective tissue disorders such as Marfan's and Ehlers-Danlos syndromes (<i>possible resultant abnormalities such as external and hiatus herniae, intestinal diverticula and rectal prolapse</i>)	-	-	-
	3	3	39
	4	3	47
i. Previous gastrointestinal infection/infestation (<i>e.g., bacterial, viral, protozoal</i>)	-	-	-

	4	4	97
j. Aberrant enteric microbiota (<i>i.e., due to previous infection, dietary change or drugs</i>)	- 4	- 4	- 84
Psychological Origin			
a. Early life events (<i>eg., maternal separation</i>)	4	2	89
b. Effects of life stress resulting in autonomic dysregulation and susceptibility to changes in GI physiology	4	1	100
c. Abuse (<i>physical, psychological, sexual</i>)	4	2	92
d. Psychiatric disorders (<i>effect on FGID experience and behaviour</i>)	4	3	92
e. Personality trait (<i>effect on FGID experience and behaviour</i>)	4	2	94
f. Dietary habit (<i>regularity, nutritional value and side effects</i>)	4 4	3 3	67 97
g. Significant life events (<i>family bereavement, family/partner separation</i>)	- 4	- 3	- 97
h. There is no proof that FGID are psychological in origin	- 3 2	- 4 4	- 45 32
Gender Differences			
a. Perception of pain.	4	2	78
b. Sex hormone effect on GI sensitivity, function and motility (<i>i.e., alterations in GI transit/colonic permeability during follicular and luteal phases of the menstrual cycle</i>)	4 4 4	2 3 2	62 74 71
c. Sex hormone effect on nociceptive processing (<i>i.e., modulation of visceral pain via oestrogen receptors expressed in the dorsal root ganglion</i>)	4 4 4	2 3 3	64 74 65
d. Central nervous system processing of visceral stimuli	4	2	83
e. Cultural values and beliefs	4	2	81

f. Socioeconomic status (<i>trait anxiety in relation to family, job security and lack of medical insurance</i>)	3 4 4	4 3 3	47 65 68
g. Increased prevalence of FGID in the female population	4	3	92
h. There is insufficient evidence to make claims on statements a and b	- 2 3	- 4 3	- 29 29
i. There are gender differences in response to pharmacological treatments	- 4 4	- 3 4	- 70 71
Impact of Symptoms on Daily Life			
a. Reduced health-related quality of life	5	1	100
b. Increased healthcare seeking behaviour	4	2	91
c. Help from significant others (<i>aid in everyday activity from family or friends</i>)	4 4 3	2 3 3	72 56 48
d. Increased symptom vigilance	4	2	94
e. Effect of current medication	4	2	78
f. Altered daily toilet habits	4	2	97
g. Altered daily occupational functioning	4	2	97
h. Altered daily family functioning	4	1	100
i. Reduced school attendance in children and adolescents	- 4	- 2	- 77
Genetic Polymorphism			
a. Contribution of genetic factors to the mediation of psychological disorders (<i>e.g., reduced function polymorphisms in the serotonin reuptake pump in conditions such as anxiety and depression</i>)	4 4	2 3	61 71

	4	3	
b. Contribution of genetic factors to the mediation of gastrointestinal sensory and motor function (<i>e.g., polymorphisms of enteric serotonin transporter genes and alpha-2 adrenoceptors</i>)	4	2	52
	4	3	65
	4	3	81
c. Contribution of genetic factors to pain modulatory pathways (<i>e.g., polymorphisms of serotonin receptors in the dorsal root ganglion</i>)	4	2	58
	4	3	68
	3	3	71
d. Polymorphism in genes that modulate immune and/or neuro-immune functions (<i>i.e., possible contribution to the onset of symptoms in the presence of other exogenous stressors</i>)	4	2	64
	4	3	69
		2	68
e. Epigenetics (<i>heritable changes in phenotype appearance or gene expression caused by mechanisms other than changes in the underlying DNA sequence</i>) is an area of importance in several areas of FGID expertise	-	-	-
	4	3	55
	4	4	65

7.3.6.2 The therapeutic relationship

Eighty-two percent of items relating to the therapeutic relationship obtained consensus (table 7.5). Most comments stressed the importance of continued training in FGID related areas of knowledge and the ability to listen to the patient, believing it is their reality. However, while consensus was achieved on the above items, several comments suggested that the reliance on evidence-based medicine with FGID patients is a weakness because it is based on an atheoretical approach not suitable for complex disorders. Participants disagreed on how clinicians share information with patients and with the negative effects of the socioeconomic environment (job security, medical insurance situation, etc.), educational status (lower educational levels and beliefs that symptoms are a signal of harm unrelated to emotional experiences) and spiritual attitudes on health (interpretation of symptoms and attitudes toward medical treatment). Participants cited a lack of epidemiological data in these areas with FGID patients. Additionally, some experts commented that educated people with little medical knowledge could also misinterpret symptoms and signs because they may have more access to information.

The use of non-directive interview technique gained consensus (83%). Respondents noted that this method allows emphasis on how patient interprets their symptoms while not concentrating on the symptoms alone. The item on structured interview technique failed to gain consensus (65% in round three), but many experts commented that both types of interview are complementary, especially in complex cases, or when information is required for research purposes. Experts noted that structured interviews lead to over-reliance on criteria leading to all-or-nothing diagnoses that may be different depending on the criteria chosen.

Table 7-5 . Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.

The therapeutic relationship	Median	Range	Agreement %
	R1	R1	R1
	R2	R2	R2
	R3	R3	R3
Clinicians working with FGID patients need the following qualities and experience			
a. Scientific competence in evidence-based assessment and management procedures)	4	2	97
b. The number of years working with FGID patients (<i>i.e., does the number of clinical working years necessarily mean continued professional and reflective development?</i>)	4	3	70
	4	3	55
	4	3	71
c. The ability to work in a multi-disciplinary team	4	2	94
d. Concern (empathy on the role of the patient as an individual)	5	2	97
e. Awareness of ethical issues (social, economic, legal and cultural)	4	3	89
f. Awareness that symptoms are "real"	5	3	97
g. Engagement with the patient and their family in selecting and monitoring a given treatment plan	4	2	97
h. Continuity of consultation.	4.5	2	94
i. Healthcare professionals working with FGID patients should be trained in the spirit of the biopsychosocial model	-	-	-
	4	2	90
j. Clinicians working with FGID patients must have a thorough understanding of the pathophysiology of FGID	-	-	-
	5	2	97
The patient-practitioner relationship			
a. Patient-practitioner relationship (patterns of communication)	5	2	94
b. Non-directive interview (inviting the patients to talk about their own experienced problems, concerns,	4	2	83

etc.)			
c. Structured interview using direct questions to elicit information about the patient's presentation	4	3	69
	4	3	65
	4	4	65
d. Role of family (support, influence and corroboration)	4	3	97
e. Developmental history (significant events in the patient's life)	4	3	89
f. Consideration of family illness history	4	2	97
g. Integration of medical and social history	4	1	100
h. Utilisation of past clinical records	4	2	86
i. . FGID patient judgement on past clinicians concerning the progression of symptoms over a given management period (<i>i.e., medical specialty, complementary or alternative therapies</i>)	4	3	56
	3	2	48
	4	3	58
j. Patient expectation	4	2	94
k. The availability of clinician ("regular work hours at a given clinic")	-	-	-
	4	4	74
	4	2	92
Cultural Factors			
a. The patient's cultural background and attitudes toward abdominal symptoms (ingrained habits such as stoicism and expectation of sympathy)	4	2	94
b. Interpretation of abdominal symptoms (thresholds and description of character)	4	2	86
c. Expression of reporting symptoms (verbal style and level of emotion)	4	2	86
d. Attitudes toward health practitioners	4	2	83
e. Favoured type of practitioner (medical, complementary, alternative)	4	2	78
f. Favoured type of clinical approach (e.g., pharmacological, physical, psychological)	4	2	92
g. Socio-economic status (family & job security, lack of medical insurance, etc.)	4	3	67
	4	4	65

	4	2	58
h. Educational status (<i>i.e., FGID patients with lower educational levels may have greater beliefs that pain is a "signal of harm" unrelated to emotional experiences</i>)	4	3	58
	4	3	58
	3	4	48
i. Spiritual and religious attitudes (<i>interpretation of symptoms & attitudes towards medical treatment</i>)	4	3	39
	3	4	45
	3	4	42
j. Patient gender	4	3	72
d. The clinician's outlook on homosexuality and other gender issues	-	-	-
	3	3	13
	3	3	10
e. The patient's understanding of FGID terminology (<i>e.g., the Chinese language has no good terminology for "heartburn"</i>)	-	-	-
	4	4	68
	4	3	78

7.3.6.3 Areas for possible measurement

Seventy-five percent of items gained consensus relating to specific abdominal symptoms.

Many experts commented that each symptom must be considered individually with organic causes ruled out for each. Additionally, for items relating to rare symptoms or those not part of the Rome process, comments suggested that while a specific symptom may not be common or not a component of criteria, they must be considered as important as may be equally disabling (e.g., increased flatulence).

Full consensus was gained on items referring to the patient's description of symptoms and with factors relating to emotional and social functioning. However, experts only agreed with 57% of items relating to measurement of physical function (table 7.6). While consensus was achieved on functional physical impairment, experts disagreed on the measurement of physical disability such as fear of movement caused by feeling vulnerable to symptoms such as incontinence or flatulence. Most commented that they had not observed or considered the importance of physical function. However, in round three, comments from some experts acknowledged that lack of exercise might have important implications for digestive function and that attitudes toward physical activity can change due to fear passing gas or incontinence during activity.

Experts disagreed on the consideration of old age in FGID assessment. Some commented that this group is seen as important only in ruling out organic disease as FGIDs are associated with younger populations while others commented that if a thorough case history is not taken, long-standing symptoms are often mistakenly interpreted.

Table 7-6 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.

Areas for Consideration and Possible Measurement	Median	Range	Agreement %
	R1	R1	R1
	R2	R2	R2
	R3	R3	R3
Abdominal Symptoms			
a. Abdominal pain	5	1	100
b. Altered stool character	5	2	92
c. Functional abdominal bloating	5	2	97
d. Increased flatulence.	4	2	72
	4	3	78
e. Functional heartburn	4	2	92
f. Aerophagia (<i>often observed in gastroesophageal reflux disease patients- Rome III criteria B2a</i>)	4	3	70
	4	4	61
	4	4	68
g. Functional nausea / vomiting	4	3	83
h. Functional constipation	5	2	97
i. Functional diarrhoea	5	2	94
j. Functional faecal incontinence	4	3	86
k. Changes in appetite (<i>e.g., decreased appetite accompanying functional constipation - Rome III criterion G7</i>)	4	3	64
	4	3	74
	4	2	81
l. Dysphagia	-	-	-
	4	3	81

m. Epigastric pain	- 4	- 1	- 100
n. Postprandial fullness	- 4	- 1	- 100
o. Obstructive defecation	- 4	- 4	- 87
p. Flatulence odour	- 3 3	- 3 4	- 45 48
q. Polyuria	- 3	- 3	- 39 42
r. Dysuria (<i>described as an extra-colonic symptom of IBS</i>)	- 3 4	- 3 4	- 45 68
Patient description of abdominal symptoms			
a. Location	4	2	80.5
b. Duration	4	1	100
c. Progression	4	2	97
d. Severity	4	3	97
e. Character	4	2	86
f. Frequency	4	3	97
g. Exacerbating factors	5	1	100
h. Relieving factors	5	1	100
i. Associated symptoms (bodily regions)	4	2	97
j. Interference of symptoms on social and household functioning	4	2	86
k. Patient control of symptoms (e.g., medication, physical therapy, distraction techniques, relaxation)	4	2	92

l. Patient's interpretation of abdominal symptoms	4	2	86
m. Patient's previous experiences of similar symptoms	- 4	- 2	- 90
n. Patient's view of inter-relationships between more than one symptom (<i>e.g., bloating and constipation or abdominal pain and referred musculoskeletal pain</i>)	- 4	- 2	- 94
Emotional Functioning			
a. Anxiety (difficulty in controlling worry, muscle tension, altered bowel habit and sleep disturbance)	5	1	100
b. Depression and sadness (subjective report or observed by others)	5	2	97
c. Symptom-related fear-avoidance beliefs (avoidance of situations associated with abdominal symptoms)	4	2	89
d. Sense of hopelessness (none or minimal expectation for symptom improvement)	4	2	92
e. Diminished interest or pleasure in activities	4	2	92
f. Anger and irritability (hostility)	4	3	83
g. Emotive coping strategies (The patient's adjustments to their symptoms, i.e., diet and toilet habit)	4	2	92
h. Catastrophic misinterpretation of FGID related pain (fear of movement resulting in hypervigilance)	4	2	89
i. Fatigue - inertia and inability to concentrate	4	2	78
Social Functioning			
a. Occupational functioning (<i>change in job status due to physical deconditioning, psychosocial dysfunction and / or any resultant interpersonal conflict</i>)	4	2	97
b. Familial functioning (<i>altered interpersonal relationships due to conflict and lack of cohesion between the patient and significant others</i>)	4	1	100
c. Avoidant behaviour (<i>fear of being away from toilet facilities</i>)	4	3	88
d. Reliance on medication and / or significant others (<i>instead of the patient managing their own symptoms</i>). (ITEM SPLIT FOR R2 – SEE BELOW)	4	2	86
e. Reliance on significant others (<i>instead of patient managing their own symptoms</i>) (ERROR, Not submitted for round 3)	4	3	68

f. Reliance on medication. (<i>relief of physical and psychological symptoms</i>)	4	2	93.6
Physical Functioning			
a. Functional impairment (assistance with one or more personal care tasks)	4	3	86
b. Kinesiophobia (excessive, irrational fear of physical movement due to the feeling of vulnerability to painful injury or re-injury)	3 3 3	3 3 4	36 29 36
c. Symptom associated disability (disproportionate restriction of physical daily activities for observable dysfunction)	4	3	89
d. Previous physical ability (<i>compared to present</i>)	4 4 4	3 3 3	66 71 81
e. The patient's past attitudes towards physical activity	4 4 4	3 3 3	56 61 55
f. The patient's present attitudes towards physical activity	4 4 4	3 2 3	70 71 84
g. Old Age (<i>under-recognised due to associations of younger age groups with FGID symptoms and secondary non-colonic symptoms including lethargy and backache commonly linked with the elderly</i>)	3 3 4	3 3 3	33 39 52

7.3.6.4 Multiaxial assessment criteria for FGIDs

Experts agreed on the current need for MAC and FGIDs with 75% of participants agreeing that MAC should be developed now regardless of gold standard diagnostic testing (table 7.7). Experts agreed with 78% of items exploring the advantages of MAC (table 7.8) especially regarding the benefit of a systematic approach to physiological and psychological components of FGIDs (92%). Experts disagreed on the benefit of reductions in reliance on clinical judgement (55% in round three) and commented that while MAC may improve classification of FGID subtypes, it does not replace the importance of clinical judgement. However, many participants commented that presently MAC might be the only diagnostic approach available for FGID where no gold standard tests exist, and as such should be considered. Comment continued on MAC promoting “all or nothing” diagnoses in individuals with FGIDs (32% agreement in round three), with some experts stating that this issue presently occurs when using Rome questionnaires in that if patients don’t meet the criteria they will not be diagnosed with an FGID. However, concerning the item relating to the term “diagnosis” as implying a distinct illness with no relevance to FGIDs (19% agreement in round three), several experts suggest a need to move beyond symptom-based systems toward the description of multidimensional endophenotypes that are quantifiable factors in gene-to-sensorimotor / behaviour pathways.

Idiographic evaluation reflecting the individuality of the patient caused confusion with three participants commenting that they did not understand the concept or context of the item. However, over the three survey rounds agreement increased from 61% to 71%, just short of the 75% consensus level. Finally, most agreement concerning disadvantages of MAC related to time required to implement MAC to FGID patients (65% in round three).

Table 7-7 Overall agreement on the future development of MAC for FGIDs over three survey rounds.

Future development of multiaxial assessment criteria	Median Likert score	Agree%
Multiaxial assessment criteria will benefit the management of FGIDs	4	75
Multiaxial assessment criteria may benefit the management of FGIDs (depending on results from further clinical research)	4	83
Multiaxial assessment criteria will only be valuable when further "gold standard" diagnostic testing is available	3	28
Multiaxial assessment criteria will never benefit the management of FGIDs	2	8

Table 7-8 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.

Multiaxial Assessment and Formulation (Relevance to FGIDs)	Median	Range	Agreement %
	R1	R1	R1
	R2	R2	R2
	R3	R3	R3
Advantages			
a. It expands from single-item diagnosis to several axes that provide additional "domains" of information of high clinical value	4	3	89
b. Reduces reliance on clinical judgement for diagnosis and therefore reduce clinical subjectivity	3.5	4	50
	4	4	55
	3	3	55
c. Allows users to systematically approach both physiological and psychological components of FGIDs	4	2	92
d. It can be applied in conjunction with laboratory testing (e.g., functional MRI), well-validated psychological tests and self-reporting criteria in FGID patients	4	4	83
e. Conveys large amounts of information related to disorders in the form of clinical shorthand that are otherwise difficult to communicate	4	3	83
f. Promotes structured clinical dialogue based on standardised criteria, compared to self-reporting questionnaires or loosely structured interviews	4	3	78
g. Allows for quantitative rating of a person's mood, cognition and behaviour, which may create a profile of functioning	4	4	81
h. Should encompass not only multiaxial evaluation, but also personal idiography that reflect their individual strengths and weaknesses	4	3	61
	4	2	61
	4	2	71

i. Can often validate the patient's own experience by informing them that others have similar experiences	4	3	78
Disadvantage			
a. Is only applicable to psychiatric diagnosis and therefore not appropriate for FGIDs	2	4	25
b. Is a time-consuming exercise and is of little value to the management of FGIDs (<i>split into two items for round 2</i>)	2	3	17
	4	3	65
	4	2	65
c. MAC is of little value to the management of FGIDS (round 2)	2	3	07
	2	3	03
d. The term "diagnosis" implies a distinct illness that is, therefore, not relevant in many cases of FGIDs	2	3	14
	2	3	19
	2	3	16
e. Multiaxial diagnostic criteria often lack clear distinctions between normal and abnormal &, therefore, do not avoid diagnostic consideration of ordinary problems of daily living	3	4	11
	3	3	16
	2	3	36
f. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability	3	3	19
	3	3	42
	3	2	32
g. Tends to promote "all or nothing" diagnoses when considering an individual's problem (i.e., how many symptoms from a list are required before action is taken)	3	3	61
	3	3	29
	3	3	32

7.3.6.5 Future research

Experts agreed that future incorporation of well-replicated neuroscientific (80%) and genetic and pharmacogenomic information (87%) would provide stronger bases for diagnosis and therapies targeting specific FGIDs (table 7.9). However, participants commented that while much has been learnt from functional imaging techniques, they comment that this approach is based on the assumption that FGIDs are disorders of sensation and central processing which leaves little room for the exploration of the “past” (i.e., is disturbed altered motility the cause or the symptom of a particular mechanism?). Experts remarked that gene-wide association studies have provided little information about predisposing genes and could only work if we obtain accurate phenotypic definitions of more homogeneous and stable FGID subgroups.

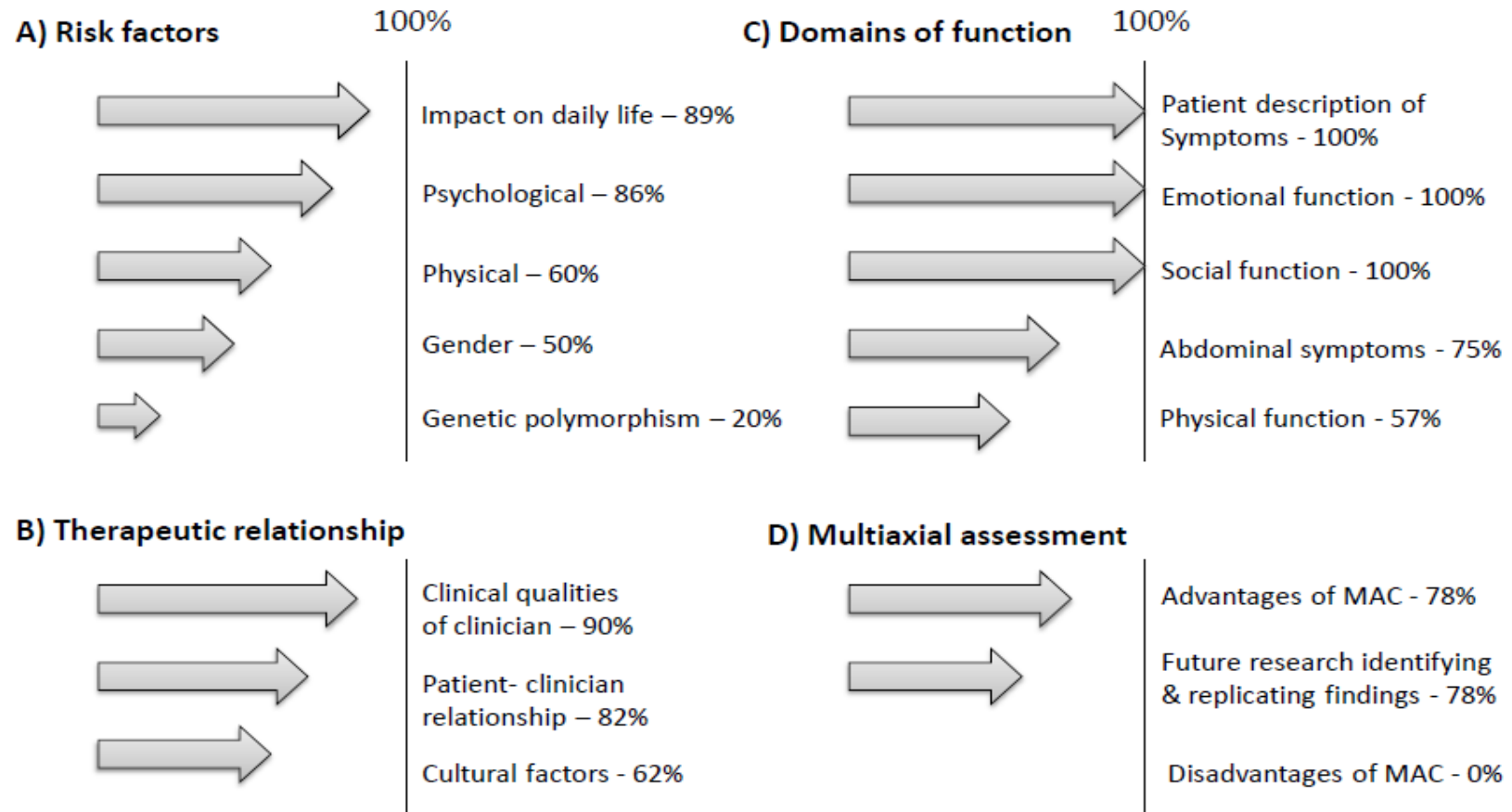
Consensus was gained on the incorporation of pharmacogenomics relating to therapies targeting specific disorders (89%) and epigenetics (84%) when considering changes in gene expression caused by mechanisms other than altered motility and sensation. However, consensus came with the overriding opinion that while these questions are important, more studies are needed where future information would not be used in clinical practice in the foreseeable future.

Figure 7.5 below summarises the mean percentages of agreement for each survey section over three Delphi rounds.

Table 7-9 Frequency tables containing Item content, Likert scale ratings and agreement ratings for risk and contributing factors to FGIDs over three rounds. Note that each row of results represents one survey round; while each range figure denotes the spread of Likert rating scores around the median. Rows shaded in grey represent items gained by thematic analysis in round one and yellow highlights gained consensus.

Future research Identifying Risk Factors and the Replication of Findings (brain-gut-axis)	Median	Range	Agreement %
	R1	R1	R1
	R2	R2	R2
	R3	R3	R3
a. Incorporation of well-replicated neuroscientific data providing bases for diagnosis	4	3	81
b. Incorporating genetic information in relation to psychological & visceral conditions (<i>e.g., polymorphism in genes that encode opioidergic or serotonergic receptors</i>)	4	3	64
	4	4	87
c. Incorporating pharmacogenomic research relating to therapies targeting of specific diseases (<i>i.e., distinct FGID may have different underlying genetic influences, pathological mechanisms and, therefore, personalised drug strengths and combinations?</i>) 1 Missing item – Valid %	4	3	71
	4	3	89
d. Incorporating data on immunological and microbiome-gut-brain interactions (<i>potential for specific modulation of enteric microbiota as a strategy for modulating co-morbid aspects of FGID</i>)	-	-	-
	4	2	94
e. Incorporating epigenetics when considering changes in gene expression caused by mechanisms other than alterations in DNA sequencing	-	-	-
	4	3	74
	4	4	84

Figure 7-5 The arrows represent the magnitude of consensus over three survey rounds for items relating to A) risk factors, B) the therapeutic relationship, C) domains of measurement and D) Multiaxial assessment criteria (MAC) concerning FGIDs



Chapter 8. Discussion

8.1 Introduction

The key finding from my study shows that in the absence of gold standard testing, FGID experts agree that MAC is required for the evaluation of FGIDs. The experts agreed on 1) domains of information that may benefit multiaxial assessment of FGIDs and 2) domains, which may be valuable to FGID evaluation and diagnosis, but require further human research. I obtained both quantitative and qualitative data on subjects highlighting information that may be used in the development of MAC for FGIDs. Many observational, epidemiological and physiological data exist regarding cause and maintenance of FGIDs. However, these data are at present only recognised as separate areas of information. Thus, collective considerations of these distinct areas of information are required to increase the systematic evaluation of FGID patients. This study accessed geographically spread experts participating in the development of FGID diagnostic criteria. Data from this study represents expert opinion and not fact. Therefore, while I do not show that MAC will benefit FGID diagnosis, participating experts evaluated and agreed with my hypotheses, enabling the possibility for the process to continue toward development and testing of such evaluative criteria.

8.2 Expert opinion

In this section, I discuss the analysis of expert judgement of items generated from the literature review and those generated by experts in round one. This part of the discussion also compares expert opinion and levels of agreement from this study with findings in current literature.

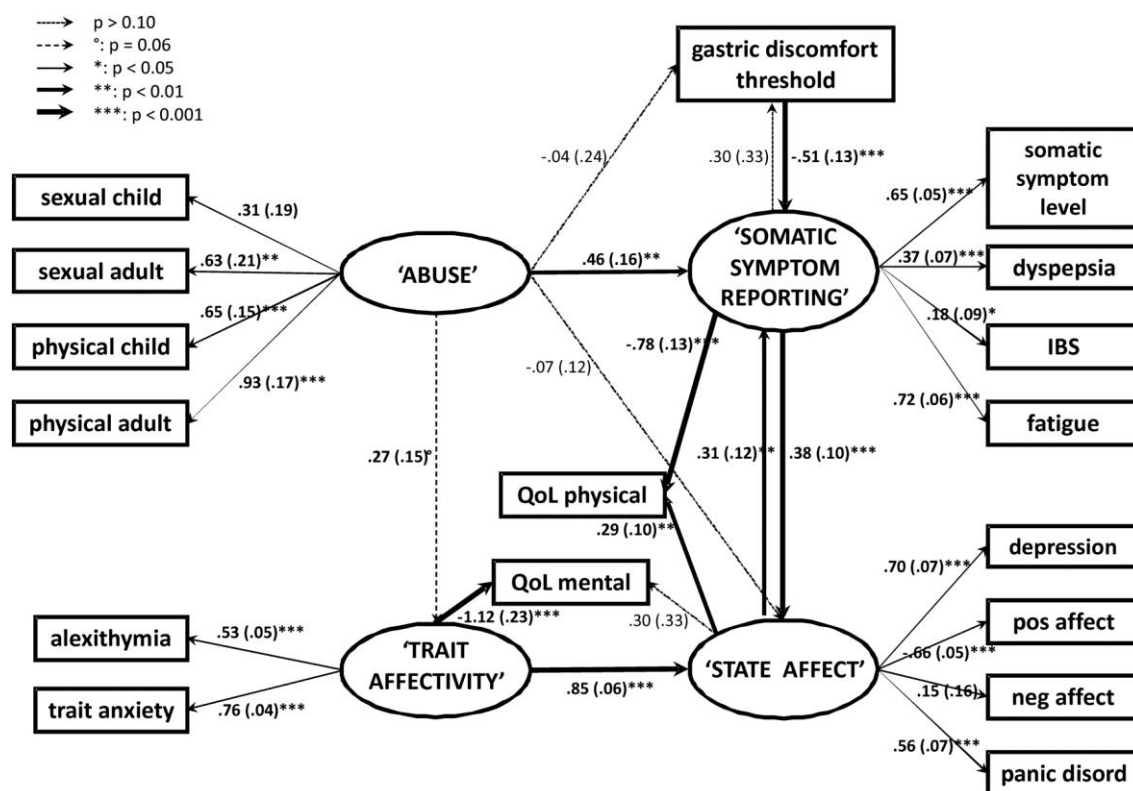
8.2.1 Risk and contributing factors

8.2.1.1 Psychological

By far the strongest area of agreement lay with psychological contribution and the impact of symptoms on daily life to people with FGIDs. These results support many studies that show evidence for contribution of anxiety, depression and personality traits to the onset and maintenance of FGIDs (Lackner et al., 2004, Riedl et al., 2009). Studies also show how co-morbid conditions affect well-being, health care seeking behaviour and how people report symptoms in clinical settings (Lee et al., 2010, Aro et al., 2011, Kaji et al., 2010). However, the link between psychological risk factors and FGIDs is not straightforward. Current data shows differing levels of FGID prevalence, symptom intensity and co-morbid psychology depending on the level and type of observable and latent variables.

Recently, Jones et al. (2012) provided the first comprehensive model of the complex interactions that are thought to play a role in abdominal symptom generation (figure 8.1). Although only FD patients were selected for this study, results show that over 60% of subjects have comorbid IBS symptoms and 40% have chronic fatigue-like symptoms. However, while the authors only used PROs, these findings help to show associations with both specific psychological traits and other functional disorders. Due to the issue of possible recall bias, the authors appropriately recommend further longitudinal studies to replicate and develop data where the course of a disorder is measured over time. Although anxiety and depression are considered as reliable indicators for FGIDs, most attention has been paid to IBS (Mikocka-Walus et al., 2008, Bouchoucha et al., 2013). Further studies are, therefore, required to a) examine associations between psychological traits, gender and other FGIDs and b) examine physiological mechanisms associated with both FGIDs symptoms and their associated psychological traits.

Figure 8-1 Structural equation model indicating the latent variables (ovals) and observable variables (boxes). Line thicknesses of the arrows indicate the significance level of the path (Jones et al., 2012)



Recent studies have shown strong associations between anxiety, hypervigilance and catastrophising, but not depression among patients presenting with pain features associated with central sensitisation. Hence, while there are no gold standard tests for pathophysiological mechanisms associated with FSSs, both physical and psychological components shown to be typical of such mechanisms can discriminate between these and other mechanisms. Good examples are shown with the recent validation of criteria for that discriminate between nociceptive, neuropathic and centrally mediated pain (Smart et al., 2011) and features strongly associated with disorders commonly associated with central sensitisation (Neblett et al., 2014).

8.2.1.2 Physical risk factors

Experts agree that only aberrant enteric microbiota due to previous GI infection were significant physical risk factors to the onset and maintenance of FGIDs. These findings support current literature that shows post-infections IBS to be the only reproducible model of FGID pathogenesis (Marshall et al., 2006) with other authors proposing that enteric infections as a risk factor for FGIDs are equal to anxiety, and greater than depression (Spiller and Lam, 2012, p259). Concerning the effects of truncal surgery, evidence suggests that only preoperative psychological variables predict symptom development and post-surgical dissatisfaction (Sperber et al., 2008, Favreau et al., 2012). These findings are consistent with several follow-up studies investigating other FSSs that show high prevalence rates of chronic pain after physical trauma and surgery, especially those with psychological factors such as comorbid stress and PTSD symptoms in the aftermath of an accident or operation (Jenewein et al., 2009, Roth et al., 2008, Althaus et al., 2012).

The alternative explanation that surgery causes FGID symptoms has been seldom addressed where investigation show little to advance current knowledge. Heaton et al. (1993) describes postoperative diarrhoea and shorter gut transit times after cholecystectomy without complications. However, there are no data showing that cholecystectomy causes any other symptoms such as abdominal pain or altered bowel habit, that are mostly required to diagnose many FGIDs. Another study shows changes in bowel function (constipation and anal incontinence) after hysterectomy (Altman et al., 2004). However, the study was retrospective in design and therefore susceptible to recall bias. At present, we do not know if appendectomy or other laparoscopic abdominal surgery initiate the onset of FGIDs. Nikolajsen et al. (2004) show that persistent pain is typical after caesarean section and seen in around 6% of patients, with Kainu et al. (2010) corroborating these results finding that persistent pain is significantly more

common one year after caesarean section than after vaginal birth. Although these studies discuss the likelihood of psychological susceptibility to pain, no measures were used to confirm this.

Concerning other physical risk factors, experts disagreed on the effects of non-abusive trauma, and physical injury, with many participants viewing these factors as irrelevant to FGID practice. However, given the many similarities between FGIDs and other FSSs and reports of increased rates of other FSS such as FMS following physical injury (Buskila et al., 1997, Buskila and Mader, 2011, Berglund et al., 2001), pilot research is needed to explore possible relationships between physical risk factors and the onset of FGIDs. Importantly, if psychological factors are critical to post trauma or surgical pain, investigations into centrally mediated pathophysiological mechanisms of pre, peri and post-operative or post-traumatic patient characteristics are also required.

8.2.1.3 Gender risk factors

Experts agreed on observations of increased prevalence of FGIDs in females. However, current evidence on the role of gender is inconsistent. Some studies show female subjects reporting increased severity of GI symptoms, however, evidence also shows that males and females report different GI symptoms rather than one particular group reporting more or increased symptom severity than the other (Chang et al., 2009, Smith et al., 1991). Recent meta-analysis demonstrates that women with IBS were considerably more likely to exhibit constipation type symptoms while men were more likely to report diarrhoea-related symptoms (Lovell and Ford, 2012). Studies examining the role of gender and other FGIDs are uncommon. However, one study investigating FD shows that while men show a higher rate of positive (13)C-urea breath tests for *Helicobacter pylori* (H-pylori), there is a highly significant increased mean (13)C-urea breath test values in females of all age groups compared to age-matched males. This study also

shows that anti-H-pylori therapy is also significantly less successful in women than in men. Possible explanations may lie with animal study data where Ohtani et al. (2007) show the adverse effect of ovarian-dependent female hormone on H-pylori induced gastric cancer in mice, while Crabtree et al. (2004) demonstrated differences in the extent of gastric cytokine responses to H-pylori in gerbils. These findings suggest that mechanism-based studies are required to show mechanisms explaining gender differences in prevalence and symptom patterns of FGIDs. It is important that future studies recruit more male subjects to demonstrate if more females consult a clinician with their symptoms than men or that there is a greater overall prevalence of FGIDs in females within the general population.

8.2.1.4 Genetic risk factors

Only one item relating to the influence of genetic polymorphisms gained consensus. Disagreement was attributed by experts to a lack of credible human research rather than the concept. While many experts commented on the likelihood of genetic contribution to factors such as altered motility, visceral sensation and even health care seeking behaviour, they recognise existing complications of heterogeneity in both identifying phenotypes and the link to genetics. Saito et al. (2010) highlight these complications where different SERT polymorphisms have been shown in both FGIDs and disorders such as anxiety and depression where each may have one or more polymorphism. Present data suggests there can be no single candidate gene for these complex traits where many genes and the influence of environment lead to unique endophenotypic presentations. Lembo et al. (2009) managed to validate a set of biomarkers for differentiating IBS patients from healthy volunteers. Following a review of the published literature, 60,000 biomarkers with potential relationships to the pathological processes of IBS were identified. When only those that were serum-based and had a viable commercial assay were considered, the number decreased to 140 and of these 10 were chosen; among them IL-1 β ,

anti-tissue transglutaminase and anti-neutrophil cytoplasmic antibody. However, sensitivity and specificity levels (50% and 88% respectively) appear no better than current symptom-based criteria. Jones et al. (2014) in a more recent study, added another 24 biomarkers, but again showed only modest levels of sensitivity and specificity (81% and 64 respectively) with anti-tissue transglutaminase being shown as the only reliable biomarker between these studies.

The heterogeneity of individual FGIDs represents the greatest limitation to human polymorphism studies. Progress in this field requires a combination of approaches that incorporate reliable genome sequencing. This may allow researchers to identify individual genes and unique phenotypes reflecting a person's environment that may differentiate people with FGIDs from healthy individuals and more importantly from those with organic GI disease.

8.2.2 The therapeutic relationship

8.2.2.1 The patient-practitioner relationship

My findings show that experts recognise only some areas of the patient-practitioner relationship.

Experts recognise the effects of FGIDs on well-being and the need to avoid treating FGIDs as a diagnosis of exclusion (Casiday et al., 2009, Spiegel et al., 2010). However, they did not consider socio-economic and educational status to be relevant to FGID patient assessment.

Lower levels economic and educational status are frequently observed in FGID and other FSS populations (Suarez and Ford, 2011, Bytzer et al., 2001, Johannes et al., 2010, Gundel et al., 2002). Job security, low levels of social support, lack of medical insurance, the number of household rooms, religious attitudes toward health and beliefs that symptoms are a "signal of harm" unrelated to the emotional experience are all important during evaluation (Hoy et al., 2010, Bussing et al., 2005). Most research examining socioeconomic status is cross-sectional where it is unclear if demonstrated socioeconomic associations are the cause or effect of FGIDs.

Thus, longitudinal studies may help to determine if FGIDs are 1) the cause of low socioeconomic status, 2) the result of low socioeconomic status or 3) as a result of confounding factors such as psychological comorbidity. Experts working with FGID patients in tertiary care settings claim there is value in the patient-practitioner relationship. However, published studies show that FGID patients experience dissatisfaction and negative attitudes from health care providers (Chassany et al., 2006, Harris and Roberts, 2008). Halpert and Godena (2011) show that the patient-practitioner relationship is central to the patient's illness experience where at present, patients have major concerns about being heard and receiving empathy. Thus, future studies should investigate education that suit both clinicians and patients concerning clinical, socio-economic and cultural factors associated with FGID presentation.

8.2.2.2 Religion and spiritual beliefs

Experts disagree with the importance of spiritual and religious beliefs and how they influence patients' symptoms and care management. Although no research exists concerning religion and spirituality in FGID patients, research has shown their significance in chronic pain conditions and other chronic diseases such as cancer and multiple sclerosis (Bussing et al., 2007, Bussing et al., 2009, Rippentrop et al., 2005, Harrison et al., 2005). Virtually all studies show that patients rely on and trust spirituality/religion as a form of active coping. Importantly, studies show positive attitudes and internal coping styles to be significantly lower in patients with nonreligious/spiritual views. Moreover, these patients do not regard their illness as a chance to reflect and reappraise lifestyle (Bussing et al., 2005).

The "analgesic effects" of religious and spiritual belief systems have also been investigated. Using functional magnetic resonance imaging, Wiech et al. (2008) in a fascinating study found that religious believers can modulate their pain experience when shown religious images.

Imaging revealed increased activity of the ventrolateral prefrontal cortex (VLPFC). However, while these results are encouraging, several other factors must be considered and investigated. First, research is needed to identify how the VLPFC contributes to the analgesic effects. Second, investigation is required to see whether pain modulation initiated at the VLPFC or driven by other areas of the prefrontal cortex, and third, do religious beliefs have a distinctive role in analgesic effects or can similar effects be observed using stimuli that lack religious connotations, but have similar cultural influence on a subject group. Results from this and other studies (Baetz and Bowen, 2008, Rippentrop et al., 2005) show strong associations between an individual's spirituality and/or religion with positive psychological and physical coping mechanisms. Thus, asking a patient about their beliefs may help to identify positive or negative forms of coping that may otherwise go unnoticed.

8.2.3 Areas for measurement

There was strong agreement for the evaluation of emotional and social functioning together with the importance of the patient's symptom description. This agreement supports a wide range of studies using well-validated PROs described throughout this study that measure psycho-emotional and GI symptoms and their intensity. These observations indicate a set of illnesses that in many patients significantly impair quality of life with relatively little in terms of objective correlates. Many studies show anxiety and depression to be independent predictors for FGIDs with associations of reduced ability to cope with daily living (Farndale and Roberts, 2011, Ferreira et al., 2012). Surprisingly, no studies address other beliefs such as fear-avoidance and kinesiphobia in people with FGIDs. However, at the time of writing this thesis, some studies have started to report correlations between anxiety, hypervigilance and catastrophisation in IBS patients (McKinnon et al., 2013, Ng and Chow, 2012, Labus et al., 2013).

While experts agree that symptom associated disability restricts daily activity, they disagreed on items referring to areas of measurement that determine disability. I suggest there is a lack of awareness both clinically and in research where fear avoidance, kinesiophobia and work-related disabilities are rarely considered. Expert comment over the three Delphi rounds highlight this unawareness, with some experts stating having never seen or read about such beliefs. However, studies show high levels of pain catastrophising and fear avoidance in chronic pain patients versus controls (Picavet et al., 2002, Leeuw et al., 2007, Verbunt et al., 2005). These cognitive and emotional factors also need to be considered in FGID patients where fear avoidance may not necessarily be due to pain but that of embarrassment related to flatulence or incontinence. I suggest that recognition of these cognitive factors would enhance prevention programs developed for FGID and related disability.

Many experts commented that the evaluation of each abdominal symptom is necessary particularly when ruling out organic disease. Experts also suggest that all symptoms are relevant, but not all in the same way and not all in the same patient. Furthermore, depending on a particular FGID, other abdominal symptoms and general somatic symptoms (fatigue, headache and sleep disturbance) are commonly reported by FGID patients but are not included in an FGID definition. Therefore, as a lead into the discussion on the relevance of MAC for FGIDs, we need to consider the following:

- Are comorbid conditions part or distinct from FGIDs?
- Do overlapping abdominal and somatic symptoms confound an FGID definition?
- Are extra-abdominal symptoms and/or syndromes part of the same clinical presentation?

- Are non-diagnostic abdominal symptoms to a specific FGID important to the overall patient presentation?

8.2.4 Multiaxial assessment

Results show consensus for the current requirement for MAC, irrespective of gold standard tests.

Current evidence for the use of MAC in FGID assessment is minimal, with most information coming in literature review form, advocating its development due to the biopsychosocial nature of FGIDs. However, recent observational studies using well-validated PROs and sensory testing methods have consolidated the value of multidimensional assessment for FGIDs, mainly relating to abdominal symptom severity, sensitivity, abuse history, ‘state and trait’ psychological factors, somatic symptom reporting and quality of life (Jones et al., 2012, Lackner et al., 2013).

However, with the use of multiple domain PROs for FSSs including FGIDs there is the problem of aggregation bias. For example, pathophysiological mechanisms that influence some symptoms may differ from mechanisms, which affect other symptoms that make up an individual symptom-based diagnosis. This bias may be masked during evaluation, but can become evident when pharmacological treatment is successful for abdominal pain, but has limited, if any effect on constipation (Lackner et al., 2013). Therefore, when applying MAC, latent variables may account for unique variance, suggesting that while MAC evaluates the patient as a whole, each component should also be considered individually. This is especially important when considering comorbid disease where the current FGID definition does not exclude organic disease including IBD. Thus, attention is needed when separating symptoms specific to FGIDs and symptoms of IBD that may have the same anatomical location, but may differ in character and associated symptoms. Please see section 2.4.2 for a detailed discussion.

Multiaxial assessment considers different components, peripheral and central, current and past, that can contribute to symptom severity and impact on well-being. Multiaxial evaluation must systematically approach physiological and psychological components both with PROs in conjunction with physiological testing. Due to the potential presence of extra-intestinal symptoms, any MAC model developed for FGIDs must also be generalisable to both FGIDs and FSSs. Such a model could help shed light on whether these disorders are due to common underlying aetiopathological processes or are indeed separate entities.

8.2.5 Potential FGID assessment criteria framework

Features of the World Psychiatric Association's International Guidelines for Diagnostic Assessment (IGDA) formulation are well suited to the nature of FGIDs and with some adaptation, would be an appropriate model of assessment. This approach considers a wide range of areas that can be tailored toward the individuality of the patient. A "stepwise approach" (Gatchel, 2001) using biopsychosocial assessments can fit the order of steps to meet the needs of a specific patient, depending on the importance of each axis (table 8.1).

Table 8-1 IGDA. 7: Standardized multi-axial diagnostic formulation. (World Psychiatric Association, 2003)

Axis	Factors
I	Clinical Disorders (mental and general medical conditions)
II	Disabilities (in personal care, occupational functioning, functioning with family, and broader social functioning)
III	Contextual factors (interpersonal and other psychosocial and environmental problems)
IV	Quality of life (primarily reflecting the patients self-perceptions)

Diagnostic evaluation is an essential feature of clinical care. It involves the gathering of information to describe and understand the patient's clinical condition and to manage effective care (Mezzich, 2002). The care of FGID patients should involve starting with an explicit therapeutic goal and engaging both the patient and their family using a comprehensive range of standardised, reliable and valid assessment tools that reflect multiple aspects of the FGID experience. This requires the clinician to be scientifically competent and empathetic in their approach to FGIDs. This can only occur if when data becomes available concerning both identifiable risk factors, and replicated findings. Unfortunately, symptom-only based labels do not consider pathophysiological mechanisms and lead to redundant diagnoses like "functional vomiting syndrome".

8.3 Methodological processes and limitations

8.3.1 Terminology

The original research question for my thesis examined the assessment of chronic functional visceral pain. Information for the literature review focused not only on the GIT, but also on also other known visceral nociceptive areas such as mesenteric tissues including those surrounding solid abdominal organs. The terminology of functional visceral pain was an area of concern as “functional abdominal pain syndrome is already classified as a distinct FGID concerning pain arising from the GIT. I wanted to include a description of symptom duration and developed the term “chronic functional visceral pain” (CFVP) thus excluding chronic visceral pain related to observable pathology. However, during email discussions with the chair of the Rome Foundation for a request to contact their committee members for study participation, he suggested that as the term CFVP was unique to this study, experts would not recognise the term and, therefore, less likely to participate in the study. Furthermore, while pain is the most apparent and intolerable of abdominal symptoms, it is not the only one. Therefore as nearly all functional abdominal complaints present as a collection of different symptoms such as bloating, constipation and pain, I decided that the term for this project be changed to the well-established “functional gastrointestinal disorders”.

8.3.2 The online survey

I chose the Bristol Online Survey tool for several reasons. Firstly, like other web-survey tools it has many types of question that capture data such as multiple choice, comment boxes or rating scales. Additionally, unlike other surveys bar Ostrakon™ it can use, re-use and share any survey as a template. Furthermore, BOS is able to generate results prior to completion, browse individual results and export results to statistical software packages. The BOS is also the survey tool administered by the University of Edinburgh. The construction of the survey was

straightforward. However, the method by which data is exported to SPSS is not. The BOS online support was not able to help directly, but was able to refer users to a University of Durham document (University of Durham Information Technology Service, 2006) describing the procedure more clearly. While MS Excel performs well in exporting data to statistical software programs, it would be helpful to design steps to allow direct data export to commonly used data analysis programs in addition to MS Excel.

When formatting the second and third Delphi survey rounds, I intended to include items that gained consensus in the previous round keeping the items the same in order to inform the expert while participating. I was not able to format extra question columns to show previous results while automatic numbering also included participant information questions. Therefore, I was not able to label the first Delphi question, as “1”. BOS support was not able to offer solutions to these issues and stated that they were aware of the limitations in the current version.

8.4 Methodological rigour

Following my suggestions for the applicability of trustworthiness criteria to the Delphi technique in section 4.9, I next show how it was applied in testing the methodological rigour in my study.

8.4.1 Trustworthiness

8.4.1.1 Credibility (internal validity)

Credibility criteria are implemented to ensure that a study measures or tests what is actually intended (Shenton AK, 2004; p64). First, Delphi technique was seen as the “correct” measure for gathering judgment data from geographically spread respondents who remain anonymous and can participate in their own time. I maintained prolonged engagement first by reviewing, systematically, subject matter to be used for item development. Second, in order to gain trust, I

engaged with all participants by explaining the outline of the study before it began, while corresponding with those who had any queries throughout the study period. I also updated respondents individually between each round with data analysis from the previous round. I continued contact after the study, supplying participants with final data analysis and notification of subsequent publications. I further continued prolonged engagement through updates of current literature throughout the study and thesis write-up period.

I followed persistent observations with checks for misinformation in the content and relevance of Delphi items throughout the study using two pilot studies, the iterative process over the main study period and contact with non-participating experts. I crosschecked judgements (triangulation) using data from both the initial literature review and regular database search updates throughout the study and write-up periods. Although peer debriefings are rarely utilised in the Delphi methods, I was able to obtain feedback from some experts and impartial colleagues before, during and after the study. Additionally, expert judgment, the application of methods, results and conclusions were examined and checked for errors by colleagues and doctoral supervisors. I informally applied member checks between survey rounds and after the study with several participants and non-participant experts who were involved in the development of both FGID diagnostic criteria and of this study via email correspondence and/or via face-to-face meetings at conferences. Feedback confirmed individual opinion on the study in general and its importance in the re-evaluation of current FGID diagnostic criteria. However, many participants did not respond to email correspondence between or after the study, and it was, therefore not possible to carry out a full member checks concerning complete re-analysis of the final data.

8.4.1.2 Transferability (external validity)

I purposively sampled participants for their expertise in the development FGID diagnostic criteria. Therefore, this study was bound to a specific research area. However, given the heterogeneous nature of the expert panel and similarities between FGIDs and other FSSs, it is feasible that many Delphi items concerning risk factors, the patient-practitioner relationship, domains of assessment/measurement and MAC may be applicable to other research fields. This supports Lincoln and Guba (1985a) observations that studies should provide sufficient contextual information using thick description to enable the reader to compare observations described in a study with those emerging in their situations. Experiential and circumstantial information was gained by offering the chance for experts to comment on each item. This aspect of the study was successful where the number of comments per statement increased through each survey round (*1.5/statement in round one, 2.2/statement in round two and 3.6/statement in round three*) giving a much richer pool of information that not only justified their judgments, but also generated items for round two based on clinical experiences not available in the literature. The importance of justification is highlighted by Green et al. (1999) who in a candid critique of their own classic Delphi study discuss negative consequences of iteration without expert justification. Experts taking part in their study became concerned that their responses included as statements in subsequent rounds appeared generalised and failed to show any contextual reasoning. While it is not clear how each round was developed, simple instruction to justify each statement could have enhanced incoming qualitative data. Additionally, Green et al. (1999) further quote Goodman (1987, p 731) who state “*that Delphi technique is not sensitive enough to differentiate between those who grade a topic low because they believe insufficient research is available, and those who place it low because they do not think it is important*”. I disagree with this statement and suggest that instruction to participants to justify their level of attitude for Delphi items

would show sufficient sensitivity to differentiate the above factors, while also, as observed in my study, highlight a lack of awareness on a given subject.

8.4.1.3 Dependability (reliability)

The below points show that my study could be repeated, in the same context, with the same methods and the same participants

Chapter 1: describes the development of the research question and the process of generating my hypotheses

Chapter 3: describes the development of systematic protocols to search relevant literature (Appendix D)

Chapter 4: describes the Delphi technique as a methodological approach

Chapter 5: describes the theory of attitude rating scales and their relevance to this study

Chapter 6: describes item and questionnaire development, sampling methods, data analysis and pilot studies

Chapter 7: describes and tabulates results of the international online Delphi study

Chapter 8: discusses, evaluates and draws conclusions on the study results and full research process

This study also showed high levels of internal consistency in item agreement throughout the survey. However, internal consistency reliability decreased slightly over each survey round.

There are several reasons why this may have occurred. First, there is a positive correlation between the number of items and levels of internal consistency. Second; internal consistency reliability becomes less accurate as sample size decreases (Javali, 2011). In my study, the drop in internal consistency reliability is most likely due to the decrease in items over each round from 132 to 50. Sample size is a less likely reason as while the sample size decreased between the first

and second rounds, it remained the same for round three. Other latent variables may have also contributed, such as participant boredom and fatigue. While the effects of these variables are often discussed in Delphi studies (Boulkedid et al., 2011, Whitehead, 2008), they are done so only as warning not to continue either past three survey rounds or when agreement levels have not changed significantly.

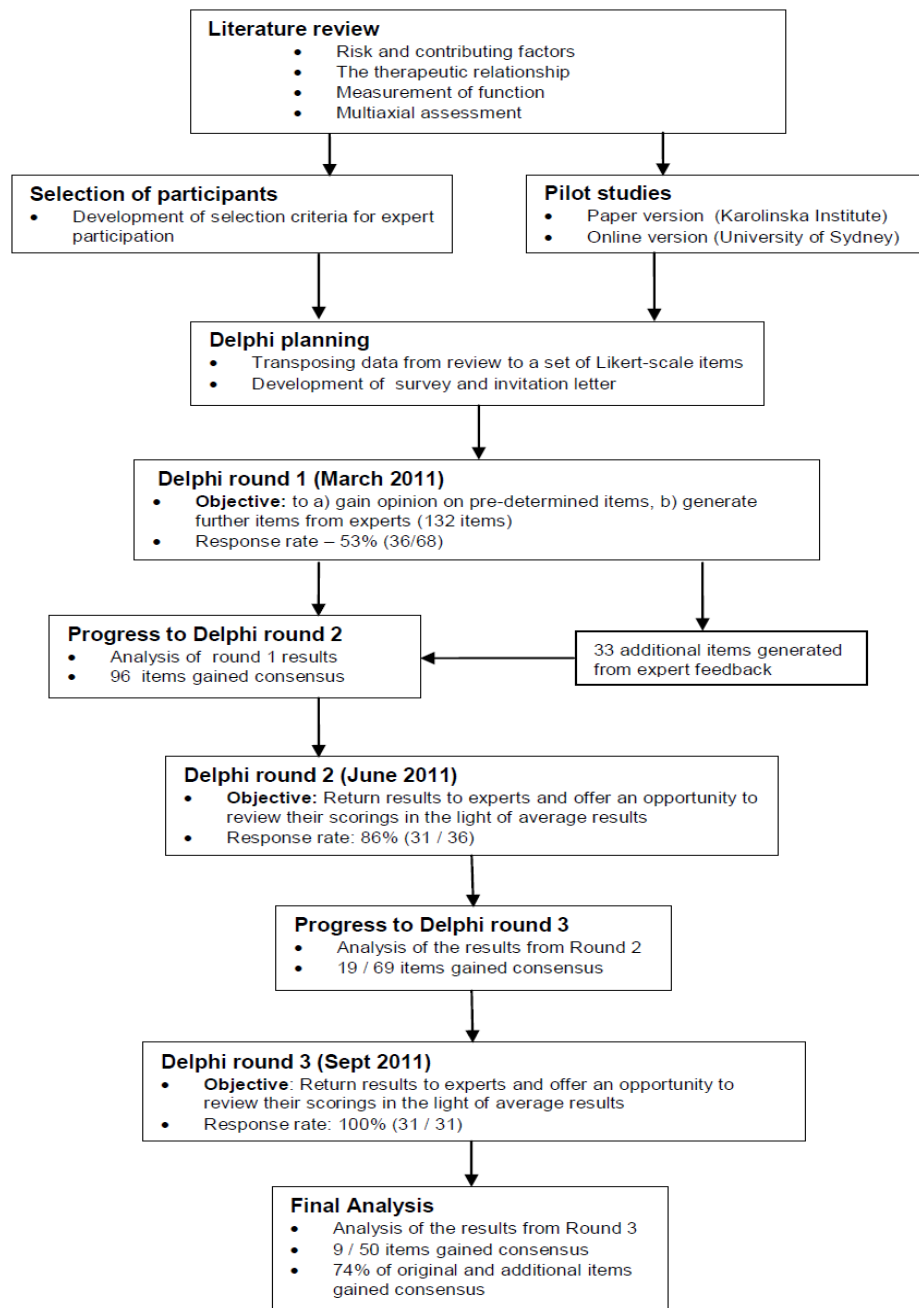
8.4.1.4 Confirmability (objectivity)

In order to show my findings are based on unique attitudes and experiences of the respondents rather than my own bias, I describe how and why each methodological decision was made over the duration of the study. I further crosschecked data (triangulation) and systematically collected, analysed, tabulated and made available all survey data. Thus, readers should be able to track and verify each methodological stage of my study while also confirming that my study conclusions are supported by material in this audit trail. Additionally, all participants had the opportunity to give their own feedback throughout the survey. These procedures allowed for transparency over the course of the study between the participants, other stakeholders in the Rome process and doctoral supervisors and myself. Figure 8.2 shows a flow chart representing an overview of the audit trail and time scale for my study.

8.4.1.5 Summary

While the concept of trustworthiness is not without its shortcomings, it is valuable when evaluating the worth of qualitative research. Thus, I followed trustworthiness criteria in order to systematically and accurately describe the research question, the quality of current literature, the number of experts, and the level of expertise. In doing so, I made explicit my initial hypotheses, my methodological approach establishing credibility of the data as well as their interpretations. It is also my intention that readers regard the findings of this study as meaningful and applicable to both the present research question and other situations reflecting their own experience.

Figure 8-2 The Delphi study audit trail and timescale flow chart



8.4.2 The iterative process

The iterative process allowed participants to review their original judgment at the beginning of each survey round against the judgments of other participants. While some revised their score on

reconsideration, most experts retained their original score even when it was significantly different from the median. In many cases, participants wrote statements justifying their original score. This observation supports the suggestion of Farmer et al. (2010, p479) that cooperation between specialities concerning future research and patient management will be assisted by identifying such areas of disagreement. Comments provided by experts in the survey and email correspondence suggested that most respondents found the process rewarding. Several participants stated that they found the process useful and thought provoking, regarding the improvement FGID evaluation. Interestingly, Hsu and Sandford (2007) suggest that through multiple iterations, participants become more problem-solving oriented where ratings and opinions become more insightful. However, in this study, iteration through participant disagreement showed not so much problem-solving, but usefully highlighted problems in areas that require further investigation to warrant future consideration for evaluation of FGIDs.

8.4.3 Survey length and relevance of items

Due to the large amount of subject matter relating to FGID assessment, I developed many items for the round one survey. This limited the number of experts willing to take part in the study. Moreover, the length of the survey may have also caused respondent fatigue and subsequent acquiescence bias. If I were to repeat a similar study I would acknowledge items gaining dissensus. Here all items gaining 25% agreement levels and less would be regarded as having reached an equivalent cut-off level of disagreement and as such discarded from subsequent survey rounds. By following this procedure, only those items where true uncertainty existed could be rated and commented on, saving both unnecessary rating and comment and participant time taking part in the study.

Alternative methods may have also included the ranking of items by experts for importance and inclusion for round two (Okoli and Pawlowski, 2004). Items may also be ranked by the researchers using composite scores of all scale points whereby the higher the composite score, the stronger the importance of the item (Smart et al., 2010). As such, I could have identified not only the most important factors, but also which ones were viewed as more important than others.

8.4.4 The number of Likert scale points

I chose five Likert scale points to represent levels of agreement and importance in this Delphi study. While evidence shows only minor differences in reliability between five and seven-point scales, a seven-point scale would have improved the quality of the incoming data. Seven-point scales provide a more accurate and sensitive measure of a participants' true attitude as it allows for a greater precision of ratings between extreme scale points. Additionally, a more recent study also shows participants using a five-point scale are more likely to try place their rating between scale points compared to seven point scales, indicating a possible lack of scale sensitivity (Finstad, 2010). Thus, while I elected to use a five-point scale for this Delphi study, I should, in hindsight given its superior sensitivity chosen a seven-point scale.

8.4.4.1 The five and seven-point argument

Choosing the number of scale points depends on the concept being measured and the cognitive and discriminating ability of the target population (Weng, 2004, Komorita and Graham, 1965, Menold et al., 2014). Many studies find that an increase in the number of scale points is associated with an increase in reliability (Weng, 2004, Champney and Marshall, 1939, Alwin, 1997). However, the reliability of homogeneous scales have also been shown to decrease after five or seven-point scales (Colman et al., 1997, Lissitz and Green, 1975). Colman et al. (1997) using a number of statistical methods showed strong correlations between five-point and seven-

point scales ($r = .921, p < .001$). However, while five and seven-point scales appear to show similar reliability, five-point scales do not have two more extreme options of “very strongly agree” or “very strongly disagree”. This creates a potential risk to reliability where with five-point scales there may be a reluctance of some respondents to check extreme scale points as seen with acquiescence bias and not state their true attitude. With more points, the seven-point scale allows participants to choose scale points indicating strong attitudes without necessarily checking the outer extreme points as seen below in Figure 8.3.

Figure 8-3 shows the greater range of strong attitude scale points on seven-point scale (7,6,2 and 1), compared to a five-point scale.

7	6	5	4	3	2	1
Completely agree	Mostly agree	Somewhat agree	Neither agree or disagree	Somewhat disagree	Mostly disagree	Completely disagree

5	4	3	2	1
Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree

Finally, no studies have investigated how the number of points on a Likert-scale affects the cumulative response function (section 5.3.3). Given that the Likert procedure may falter for participants who hold extreme attitudinal positions and that individuals with extreme attitudes may form an important segment of a given study, research investigating these variables may also yield informative results.

8.4.5 Data collection

All data collection strategies were described from the outset where rating and expert comment data was presented to all participants and thesis supervisors after each round. The ease and speed of using the online survey was an important factor for both the participants and researchers. The

participants had quick and easy access to the survey using only a link given to them on the email invitations at the beginning of each round. However, the most significant decision concerning data collection process was to relax the time of the survey closing dates. This allowed nearly 25% of experts taking part in rounds two and three to return data that otherwise would have been lost. Compared to postal surveys, online survey distribution, participant email correspondence and instant notification of survey completion make significant differences in both costs and return rates.

8.4.6 Survey item bias

I developed round one items from the literature review to reduce content bias and the potential for ambiguous broad statements, both of which are common in traditional Delphi surveys (Hasson and Keeney, 2011, Hsu, 2007). However, this form of first round survey did bias response. The bias arose due to too many items declaring FGIDs as part of the spectrum of psychiatric illnesses especially with terms such as “functional” and “psychosocial”. Furthermore, efforts to limit the time taken to complete the survey gave rise to an uneven balance between positive and negative statements. Two experts commented on this issue during round one. Other terms such as “biopsychosocial”, while all-encompassing, were seen by some participants as being weighted towards psychosocial dysfunction. In future studies, I would use a more neutral term “multidimensional” that describes the involvement of several dimensions without specifying what types. Several items also contained conditional words such as “may” and “can” and therefore reduced the assertiveness of some statements. Additionally in round one, participants described some items, especially concerning physical contribution as too vague. I addressed this by inserting bracketed examples taken from research articles found during the review process.

8.5 Participant characteristics

8.5.1 Sample size

I aimed to recruit only the most qualified experts who work continually in clinical and research areas and who have responsibility for decisions relating to the development of diagnostic criteria for FGIDs. I purposively sampled the total population of experts who met the predefined sampling criteria of those assigned to positions within the Rome Foundation and the International Foundation of Functional GI disorders. Total population sampling makes it possible to get a deep insight into phenomena of interest. Okoli and Pawlowski (2004, p6) add that Delphi sample size does not depend on statistical power like traditional surveys, but on the knowledge and experience participants bring to a study as well as group dynamic for arriving at consensus.

However, while obtaining lists of the above populations was straightforward, a large proportion of FGID experts could not be reached or who decided not to take part. As such it was difficult to make analytical generalisations about the FGID expert group as a whole and subgroups within the sampled population. In this study only 36 experts from 90 sampled chose or were able to participate, I considered this number low but acceptable, given the research question, the purposive sampling of the group and the large number of comments based on expert knowledge and experience given during each survey round.

8.5.2 Responder rates

Response to the first round of my Delphi study was low with 36 out of 90 experts completing the first round survey. This study sought participants worldwide where many were far removed both in location and in acquaintance with me as a researcher. If the study had been local to Edinburgh, United Kingdom or Europe, the response rate would have been much higher as was

shown with 61% of European-based participants responding compared to significantly lower response rates in other continents.

There are no response rate recommendations for Delphi surveys. However, (Keeney et al., 2011b, p53) suggest that in order to preserve methodological rigour, a 70% response rate should be maintained. Face-to-face interviews before or in the first round are also shown to increase response rates (McKenna, 1994, Keeney et al., 2011b). As previously discussed, financial restraints made this option impossible. However, I did arrange face-to-face meetings with individual experts at conferences. While informal, these meetings helped to maintain interest during and after the study period. Financial limitations also prevented the use of follow-up discussion using focus groups with participants on how to best use the study data. However, a publication from this study was examined and used by the Rome Foundation Board as supporting data for development of the new Rome IV Diagnostic Criteria (Austin et al., 2013).

8.5.3 Stability of participant response over rounds two and three

While consensus was not reached for many items, the stability of participant responses over rounds two and three were stable. For nearly all items, there were only very small differences in score with 36 out of 50 items showing a 5% and less change and 45 showing a 10% and less change in scores from all 31 participants. In following Scheibe et al. (1975) recommendation, a 15% change or lower in any two distributions was considered as stable. These observations suggest that while experts did not necessarily agree with items, consensus was stable for both 'neutral' and 'disagree' categories in the final two rounds. These figures are encouraging as good response stability is typically associated with small homogeneous groups working in the same field (Akins et al., 2005b). Given that, there is presently no general standard of how to measure consensus, analysing the stability of Delphi results are now being viewed as a valid

measure consensus and to stop a Delphi survey (von der Gracht, 2012). This is especially relevant when investigating the opinions of heterogeneous groups where ‘dissensus’ may be considered to be of greater relevance. In my study, while there was no significant convergence of opinion for many items, the results indicate useful trends where disagreement may stimulate the investigation of areas where uncertainty exists.

8.5.4 Sample heterogeneity

I sampled a spectrum of expertise in order to gain opinion and judgement from all areas of research relating to the investigation and clinical management of FGIDs. Literature suggests that heterogeneous groups reduce prediction error due to over-optimism often observed in groups that are more homogeneous (Ecken et al., 2011, Delbecq et al., 1975). However, my experience with this study showed that while there were no differences in item response between specialities, issues arose because of a mixed sample. Namely, some experts involved in basic physiology, nursing and layperson employment with patient groups felt that despite being involved in work surrounding the assessment of FGIDs, the survey was either not directly related to their area of expertise or the survey items were beyond their level of expertise. Therefore, future Delphi studies of this nature should specify the type of heterogeneity and consider not only differing areas of expertise, but also differing levels of academic and clinical training.

Table 7.3, shows how most of the experts in this survey were gastroenterologists. The numbers in other specialities were small. Only one or two respondents in such groups cannot influence the overall outcome and I can only suggest that responses from these subgroups may be mixed. As they are all working towards the same diagnostic goals, it would be expected that they would agree, in principle with questions raised in this study. I am aware that experts from these

subgroups may rate items differently compared to gastroenterologists in subject areas not reviewed or developed by the FGID diagnostic committees.

8.5.5 How important is consensus?

Typically, Delphi studies aim for a high degree of convergence through survey rounds to gain consensus on items relating to future decisions or events. Therefore, studies that exhibit a high degree of convergence are often accepted for future use, while those that show wide differences of opinion after the final survey round are considered unfeasible. Importantly, Linstone and Turoff (2002; p73) suggests that the suppression of uncertainty can mask the real significance of Delphi results. Furthermore, the ability to expose uncertainty and divergent views is an inherent strength of the Delphi process where disagreement and misunderstandings can be resolved.

Setting a high pre-determined percentage level of consensus may help to show areas of disagreement that may otherwise be hidden by a lower consensus level. Like other areas of Delphi methodology, numerical levels of consensus are set somewhat arbitrarily by researchers with levels ranging from 51% (Loughlin and Moore, 1979) to 90% (Herdman et al., 2002).

While the Delphi technique uses anonymity and iteration to eliminate many issues associated with the pressure to conform to group agreement, low consensus levels allow the increased probability of ‘silence’ in the form of neutral agreement rankings to mask participants’ actual attitudes. In such cases, a group can make a decision that is to the preference of very few group members. This issue, known as the Abilene paradox, was aptly coined by the late Israeli orator and diplomat Abba Eban, who cynically stated “*A consensus means that everyone agrees to say collectively what no one believes individually*”.

8.5.6 Responder bias

My aim was not to compare judgments between different areas of expertise involved in FGID diagnosis, but to request opinions from those involved in the same mission. The diagnosis of FGIDs is a problem that is much broader in scope than any one set of experts possess. Thus, I viewed the participant panel as experts working toward the same clinical and research endpoints. To confirm or deny this assumption, it was necessary to test for differences in item response between the different groups. However, all specialities bar gastroenterologists numbered between one and four experts. As shown by Cumming (2008) and Button et al. (2013) small samples provide paltry statistical power and inadequate positive predictive value. I, therefore, chose to compare only responses of gastroenterologists against other specialities. As explained in chapter seven, there were no significant differences in item response between the two groups in each survey round.

8.5.6.1 Pilot respondent participation

I conducted frequency analysis with and without one expert who participated in the online pilot and main study. I wanted to assess the extent of the influence of possible contamination as recommended by (van Teijlingen and Hundley, 2002). Analysis showed no difference in the number of items gaining consensus overall survey rounds. However, my choice of allowing a pilot participant into the main survey had the potential to bias results at any point of the three survey rounds. Therefore, in order to rule out this type of bias, only experts not participating in a pilot study would in future be eligible for a main Delphi study. However, I found only one study (van Teijlingen and Hundley, 2002) addressing the question of subjects participating in both pilot and main studies. Surprisingly, this review gave only a short paragraph of recommendations without any reference to evidence or other opinion.

8.5.7 Non-responder bias

I tested four demographic dimensions (location, academic rank, specialist field and gender).

Firstly, I tested respondents versus non-respondents where I found only geographic location to be significantly different between experts from Europe compared to North America and Asia and Australia. Secondly, I tested early versus late responders to observe if there were any similarities between late and non-responders. Analysis showed no differences between geographic locations, academic rank or specialist field, but did show significant differences in gender response timing with 71% female experts responding late compared to 24% of male experts. While research shows how to minimize non-response in the Delphi technique (Hsu, 2007), I did not find any studies examining non-response bias as part of a Delphi study methodology. Instead, I reviewed the literature that investigated response rates, non-response bias and response error in population-based mail and web surveys (Johnson and Wislar, 2012, Smith, 2008, Sheehan, 2001).

Studies reviewing difference in geographic response rates are conflicting with some showing response rates as being higher in Asia than in North America and vice-versa (Kriger and Esther, 1992, Harzing, 1997). Others show North America having higher response rates than European countries (Harzing, 1997). However, as suggested by Harzing et al. (2012, p22), it is more likely, in the case of my study that higher response rates were achieved because respondents were geographically and culturally closer to the research project's originating country.

There were no significant differences in response rates between specialities. However, some areas of expertise responded less than others. Of the low responders, most physiologists replied to the initial invitation explaining that the survey was clinically based and therefore beyond their expertise. Psychologists were the only group that neither responded nor participated in the

survey. The highest response rates came from areas of expertise related to multidisciplinary practice (primary care, psychiatry, pain medicine). Little investigation exists concerning survey response rates of different specialist field in health care; however, those experts who responded most came from areas of health care with experience in the field of multidisciplinary patient management.

Concerning academic ranking, there were no significant differences between the responders versus non-responders, associate professors responded less (29%) compared to both professors and research fellows (40% and 44%). I found no studies comparing attitudes at different levels of academia. I suggest that the reduced response was due to the high workload of research, teaching and department responsibility, whereas the job description of the professor and a research fellow are more that of responsibility or teaching/research and not both. While I found no studies to validate my assumptions, faculty policy literature is available on the different workloads and responsibilities within individual faculties (Misra et al., 2010, University of Maryland, 2012).

There were significant differences between gender with early and late response rates, but no differences in response and non-response. In round one, 43 % of male experts responded compared to 30% of female experts. I found only one study reviewing differences in gender response rates where Smith (2008) argues that links between gender and response behaviour relate to male “separative” characteristics versus female “empathetic closeness”. As this assumption opposes observations from this study, and no other studies were found, no obvious conclusions can be made.

8.5.8 Maintaining responder rates

I considered the ongoing management of participants as one of the most important factors of this study. I was able to persuade some participants in doubt or with very little spare time at the beginning and between rounds to participate and continue in the study. Therefore, while initial response was low at 45%, excellent stability prevailed in rounds two and three with only five experts dropping out in round two (86% response rate) and no dropouts in round three (100% response rate).

8.6 Lumping and Splitting FGIDs with other FSSs

While terms such as “functional” and “psychosocial” attempt to broaden the approach to patients with GI disorders with no observable pathology, the same approach applies to many other chronic disorders with the only difference potentially being the clinician’s bias as to the type of functional disorder or to what is organic and what is not. These questions are approached by (Wessely et al., 1999) who in a provoking article discusses the limited value of existing definitions to syndromes in terms of specific symptoms where substantial overlap exists. This opinion piece prompted debate of whether to lump and split FSSs.

Recent observations have strengthened both camps with lumpers’ showing that FGIDs like other FSSs are epidemiologically linked and have the same premorbid risk markers. These include childhood adversity, suffering with anxiety and depressive related disorders, similar abnormalities in the HPA axis and autonomic nervous system as well as sensitisation in the central nervous system (White, 2013, Warren et al., 2013, Whorwell et al., 1986). Research also shows that patients with one FSS frequently meet diagnostic criteria for other syndromes with IBS being linked to conditions such as FMS, CFS, tension headaches, NCCP and hyperventilation syndrome (Whorwell et al., 1986, Wessely et al., 1999). One can argue that as

the number of FSSs increases so does the possible number of possible combinations of associated processes, however, the question of which came first, the process or the syndrome has not been answered for any of the above associations.

Splitters suggest that each FSS is heterogeneous. Studies show heterogeneity in conditions such as CFS where findings have shown different phenotypes and genotypes (Warren et al., 2013, Aslakson et al., 2009). Moreover, a prospective study also found that different aetiological factors precipitate different FSSs. Their results showed that the odds of developing IBS were significantly higher post campylobacter than post-infectious mononucleosis. In contrast, the odds of developing CFS were significantly greater after post-infectious mononucleosis than campylobacter. Interestingly, the authors also noted that anxiety and depression were the strongest predictors for CFS whereas the nature of an infection was the strongest predictor of IBS (Moss-Morris and Spence, 2006), indicating the importance of consideration to physical risk factors with FGIDs.

White summarises the need to both split and lump FSSs together and move away from considering symptoms themselves and instead focus on differentiating FSS sub-phenotypes using already available biomarkers in order to reveal underlying pathophysiology and underlying endophenotypes associated with an individual FSS (White, 2013). As a final thought, symptom-based labels do not describe any underlying process and lead to nomenclature like “functional vomiting syndrome”, or “failed back surgery syndrome”. How do we explain these diagnoses to patients who are in constant pain or throwing up and have come to the end of their care strategy?

8.7 The future

Clinicians need diagnostic tests that not only discriminate between people with FGIDs and healthy individuals, but that also discriminate between FGIDs from other organic diseases. Recently, several biomarkers relating to visceral hypersensitivity (Ludidi et al., 2012), altered pain perception (Chacaltana Mendoza et al., 2012), gene expression (Jones et al., 2014) and faecal metabolites (Ahmed et al., 2013) have been studied. While Jones et al. (2014) differentiated between people with IBS and healthy volunteers using a combination of 34 biomarkers and psychological measures, success was modest with only four biomarkers showing discriminative value. However, it may be more relevant to clinical practice to investigate biomarkers that discriminate between functional disorders and organic disease. Ahmed et al. (2013) in a small study were the first to show the promise of faecal volatile organic metabolites (VOMS) in the differentiation of IBS-diarrhoea from active IBD with sensitivity and specificity of 94% and 82% respectively. However, this type of study needs to be replicated using larger samples where other demographic data such as age, gender in conjunction with other upper and lower abdominal symptoms are considered.

Concerning more centrally mediated symptoms, neuroimaging has greatly improved understanding of brain mechanisms involved in processing and perception of visceral information but has not been sufficiently integrated with epidemiological or behavioural evidence (Van Oudenhove, 2011). Encouragingly, the somatic pain research field have made progress investigating changes in brain structure associated with FSSs using novel neuroimaging techniques such as voxel-based morphometry and diffusion tensor imaging (Bandettini, 2009). Neural mechanisms of emotional modulation (anxiety, sadness) and cognition (attention, expectation) have also been identified using fMRI showing distinct spinal and supraspinal mechanisms (Roy et al., 2009). Interestingly, somatic pain studies are now measuring dynamic

changes in brain activation during spontaneous or fluctuating levels of pain using an fMRI technique called arterial spin labelling that is well suited to longer stimulus duration (Van Oudenhove, 2011, p295). These psychological process investigations are important but should not be done in isolation, as is often the case, but together with equally important studies investigating peripheral mechanisms that also result in symptoms of FGIDs.

8.8 Summary

Presently, there are no methods able to show the true presence of FGIDs. This is most likely due to their multifactorial aetiology. However, the combination of biomarkers and psychological markers has recently shown promise where Jones et al. (2014) was able to distinguish between patients with IBS and healthy controls with modestly improved sensitivity from 81% to 85% and specificity from 64% to 88% when compared with biomarkers alone. However, it is essential that further studies include larger sample sizes, inclusion criteria that allow other FGIDs and biomarkers that may differentiate between FGIDs and organic GI disease. Unfortunately, FGIDs are a set of complex disorders for which there are no unifying explanations. Clinicians want to define FGIDs but are not able to using present symptom-based criteria.

Thus, given the present uncertainty over both cause and underlying pathophysiological mechanisms, I set out in this thesis to examine the opportunity to improve and extend current symptom-based diagnostic criteria for FGIDs, investigating how systematic evaluation of different specific features or elements of a disorder (MAC) can improve current criteria. My aim was not to prove beyond doubt, the existence of specific relationships between these elements or that MAC is the best diagnostic process, so much as to describe and gain expert judgment on the

interdependence of many more-or-less proven features that may be valuable for the diagnosis of FGIDs in the future.

8.9 Conclusions

The main finding from this study are that experts agree that evaluation of FGIDs requires multi-axis assessment. They agree that psychosocial factors, the impact of symptoms on daily life, and physical disability caused by these factors should be considered in developing MAC. Experts disagree on the influence of genetic and gender-based risk factors, but consider them important but require more research.

8.10 Recommendations

I recommend that the planned Rome IV criteria move away from consensus-driven measures towards criteria based upon high-quality evidence in areas where science has advanced.

Additionally, future diagnostic criteria should include assessment of psychological and social function while also evaluating physical disability. I suggest that cultural background be considered during patient evaluation. Hence, the Rome IV criteria should define different predisposing factors where diagnostic features are confirmed using physiological and psychosocial clinical markers. The Rome IV criteria should be incorporated into a diagnostic process that is relevant to primary care and can easily be used by primary care physicians

8.11 Contributions to knowledge

- This study encouraged the Rome Foundation to develop new diagnostic criteria that will include a multidimensional clinical profile.

- We now know that experts agree that multidimensional assessment using symptoms, available biomarkers and psychological markers will improve the diagnose FGIDs
- The primary publication from this study was requested by the Rome Foundation for use as a supporting document for the development of the Rome IV diagnostic criteria for FGIDs.

References

The Sage encyclopedia of social science research methods, Sage.

- AHMED, I., GREENWOOD, R., COSTELLO, B. D. L., RATCLIFFE, N. M. & PROBERT, C. S. 2013. An Investigation of Fecal Volatile Organic Metabolites in Irritable Bowel Syndrome. *PLoS ONE*, 8, e58204.
- AIKEN, L. R. 1983. Number of Response Categories and Statistics on a Teacher Rating Scale. *Educational and Psychological Measurement*, 43, 397-401.
- AKINS, R., TOLSON H & COLE B 2005a. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Medical Research Methodology*, 5, 37.
- AKINS, R., TOLSON, H. & COLE, B. 2005b. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Medical Research Methodology*, 5, 37.
- ALAHIAFI, A. & BURGE, S. 2005. What should undergraduate medical students know about psoriasis? Involving patients in curriculum development: modified Delphi technique. *Bmj*, 330, 633-6.
- ALTHAUS, A., HINRICHS-ROCKER, A., CHAPMAN, R., ARRANZ BECKER, O., LEFERING, R., SIMANSKI, C., WEBER, F., MOSER, K. H., JOPPICH, R., TROJAN, S., GUTZEIT, N. & NEUGEBAUER, E. 2012. Development of a risk index for the prediction of chronic post-surgical pain. *Eur J Pain*, 16, 901-10.
- ALTMAN, D., ZETTERSTROM, J., LOPEZ, A., POLLACK, J., NORDENSTAM, J. & MELLGREN, A. 2004. Effect of hysterectomy on bowel function. *Dis Colon Rectum*, 47, 502-8; discussion 508-9.
- ALWIN, D. 1997. Feeling Thermometers Versus 7-Point Scales: Which are Better? *Sociological Methods & Research*, 25, 318-340.
- AMENT, R. H. 1970. Comparison of delphi forecasting studies in 1964 and 1969. *Futures*, 2, 35-44.
- ANDRICH, D. 1988. The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude. *Applied Psychological Measurement*, 12, 33-51.
- ANDRICH, D. 1996. A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies.
- ARO, P., TALLEY, N. J., AGREUS, L., JOHANSSON, S. E., BOLLING-STERNEVALD, E., STORSKRUBB, T. & RONKAINEN, J. 2011. Functional dyspepsia impairs quality of life in the adult population. *Aliment Pharmacol Ther*, 33, 1215-24.
- ASCH, S. E. 1951. Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership and men; research in human relations*. Oxford, England: Carnegie Press.
- ASLAKSON, E., VOLLMER-CONNA, U., REEVES, W. C. & WHITE, P. D. 2009. Replication of an empirical approach to delineate the heterogeneity of chronic unexplained fatigue. *Popul Health Metr*, 7, 17.
- AUSTIN, P., HENDERSON S, POWER I, JIRWE M & ALANDER T 2013. An international Delphi study to assess the need for multiaxial criteria in diagnosis and management of functional gastrointestinal disorders. *J Psychosom Res*, 75, 128-34.
- BAETZ, M. & BOWEN, R. 2008. Chronic pain and fatigue: Associations with religion and spirituality. *Pain Res Manag*, 13, 383-8.
- BAKER, J., LOVELL, K. & HARRIS, N. 2006. How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. *Nurse Res*, 14, 59-70.

- BANDETTINI, P. A. 2009. What's New in Neuroimaging Methods? *Annals of the New York Academy of Sciences*, 1156, 260-293.
- BANZATO, C., JORGE MR & KASTRUP M 2009. Multiaxial Schemas for Psychiatric Diagnosis. In: SALLOUM IM & MEZZICH JE (eds.) *Psychiatric Diagnosis: Challenges and Prospects*. Wiley.
- BARBARA, G. & STANGHELLINI, V. 2009. Biomarkers in IBS: when will they replace symptoms for diagnosis and management? *Gut*, 58, 1571-1575.
- BARCLAY, J. E. & WEAVER, H. B. 1962. Comparative Reliabilities and Ease of Construction of Thurstone and Likert Attitude Scales. *The Journal of Social Psychology*, 58, 109-120.
- BARNETTE, J., DANIELSON LC & ALGOZZINE RF 1978. Delphi Methodology: An Empirical Investigation. *Educational Research Quarterly*, 3, 67-73.
- BARRATT, S. M., LEEDS, J. S., ROBINSON, K., LOBO, A. J., MCALINDON, M. E. & SANDERS, D. S. 2011. Prodromal irritable bowel syndrome may be responsible for delays in diagnosis in patients presenting with unrecognized Crohn's disease and celiac disease, but not ulcerative colitis. *Dig Dis Sci*, 56, 3270-5.
- BEAUMONT W 1833. *Experiments and Observations on the Gastric Juice and the Physiology of Digestion*, Plattsburgh FP Allen.
- BENDIG, A. W. 1953. The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *Journal of Applied Psychology*, 37, 38-41.
- BERGLUND, A., ALFREDSSON, L., JENSEN, I., CASSIDY, J. D. & NYGREN, A. 2001. The association between exposure to a rear-end collision and future health complaints. *J Clin Epidemiol*, 54, 851-6.
- BERNARD BECKER MEDICAL LIBRARY ARCHIVES William Beaumont Papers, 1807-1853. In: WASHINGTON UNIVERSITY SCHOOL OF MEDICINE (ed.). Washington.
- BERRILL, J. W., GREEN, J. T., HOOD, K. & CAMPBELL, A. K. 2013. Symptoms of irritable bowel syndrome in patients with inflammatory bowel disease: examining the role of sub-clinical inflammation and the impact on clinical assessment of disease activity. *Aliment Pharmacol Ther*, 38, 44-51.
- BLACK, N., MURPHY, M., LAMPING, D., MCKEE, M., SANDERSON, C., ASKHAM, J. & MARTEAU, T. 1999. Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy*, 4, 236-48.
- BLACKLOCK, J. B. 1965. CHRONIC INTESTINAL OBSTRUCTION: A TRUE "IRRITABLE BOWEL" SYNDROME. *Ann Surg*, 161, 819-23.
- BOLLEN, K. A. 2002. LATENT VARIABLES IN PSYCHOLOGY AND THE SOCIAL SCIENCES. *Annual Review of Psychology*, 53, 605-634.
- BOUCHOUCHA, M., HEJNAR, M., DEVROEDE, G., BABBA, T., BON, C. & BENAMOUZIG, R. 2013. Anxiety and depression as markers of multiplicity of sites of functional gastrointestinal disorders: a gender issue? *Clin Res Hepatol Gastroenterol*, 37, 422-30.
- BOULKEDID, R., ABDOUL, H., LOUSTAU, M., SIBONY, O. & ALBERTI, C. 2011. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS ONE*, 6, e20476.
- BOWDEN, A., FOX-RUSHBY, J. A., NYANDIEKA, L. & WANJAU, J. 2002. Methods for pre-testing and piloting survey questions: illustrations from the KENQOL survey of health-related quality of life. *Health Policy Plan*, 17, 322-30.

- BRACE, I. 2010a. Objectives in writing a questionnaire. In: BARR D & BIRN RJ (eds.) *Questionnaire Design*. 2nd ed. London: Kogan Page.
- BRACE, I. 2010b. Piloting the Questionnaire. In: BIRN RJ & BARR D (eds.) *Questionnaire Design*. 2nd ed. London: Kogen-Page.
- BRACE, I. 2010c. Planning the Questionnaire. In: BARR D & BIRN RJ (eds.) *Questionnaire Design*. 2nd ed. London: Kogan Page Ltd.
- BRANDON, P. R. 1998. Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *The American Journal of Evaluation*, 19, 325-337.
- BRAVO, G. & POTVIN, L. 1991. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *J Clin Epidemiol*, 44, 381-90.
- BROOMFIELD, D. & HUMPHRIS, G. M. 2001. Using the Delphi technique to identify the cancer education requirements of general practitioners. *Med Educ*, 35, 928-37.
- BUSKILA, D. & MADER, R. 2011. Trauma and work-related pain syndromes: risk factors, clinical picture, insurance and law interventions. *Best Pract Res Clin Rheumatol*, 25, 199-207.
- BUSKILA, D., NEUMANN, L., VAISBERG, G., ALKALAY, D. & WOLFE, F. 1997. Increased rates of fibromyalgia following cervical spine injury. A Controlled study of 161 cases of traumatic injury. *Arthritis & Rheumatism*, 40, 446-452.
- BUSSING, A., MATTHIESSEN, P. F. & OSTERMANN, T. 2005. Engagement of patients in religious and spiritual practices: confirmatory results with the SpREUK-P 1.1 questionnaire as a tool of quality of life research. *Health Qual Life Outcomes*, 3, 53.
- BUSSING, A., MICHALSEN, A., BALZAT, H. J., GRUNTHNER, R. A., OSTERMANN, T., NEUGEBAUER, E. A. & MATTHIESSEN, P. F. 2009. Are spirituality and religiosity resources for patients with chronic pain conditions? *Pain Med*, 10, 327-39.
- BUSSING, A., OSTERMANN, T. & KOENIG, H. G. 2007. Relevance of religion and spirituality in German patients with chronic diseases. *Int J Psychiatry Med*, 37, 39-57.
- BUTTON, K. S., IOANNIDIS, J. P., MOKRYSZ, C., NOSEK, B. A., FLINT, J., ROBINSON, E. S. & MUNAFO, M. R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*, 14, 365-76.
- BYTZER, P., HOWELL, S., LEEMON, M., YOUNG, L. J., JONES, M. P. & TALLEY, N. J. 2001. Low socioeconomic class is a risk factor for upper and lower gastrointestinal symptoms: a population based study in 15 000 Australian adults. *Gut*, 49, 66-72.
- CANAVAN, C., WEST, J. & CARD T 2014. The epidemiology of irritable bowel syndrome. *Clinical Epidemiology*, 6, 71-80.
- CARMIN, C. N., OWNBY, R. L., WIEGARTZ, P. S. & KONDOS, G. T. 2008. Women and non-cardiac chest pain: gender differences in symptom presentation. *Archives of Womens Mental Health*, 11, 287-293.
- CASIDAY, R. E., HUNGIN, A. P., CORNFORD, C. S., DE WIT, N. J. & BLELL, M. T. 2009. GPs' explanatory models for irritable bowel syndrome: a mismatch with patient models? *Fam Pract*, 26, 34-9.
- CASTORR, A. H., THOMPSON, K. O., RYAN, J. W., PHILLIPS, C. Y., PRESCOTT, P. A. & SOEKEN, K. L. 1990. The process of rater training for observational instruments: implications for interrater reliability. *Res Nurs Health*, 13, 311-8.
- CHACALTANA MENDOZA, A., DIAZ RIOS, R., ALVA ALVA, E., VASQUEZ VALVERDE, N. & CELESTINO MORALES, C. 2012. [Perception of pain during sigmoidoscopy flexible as an

- additional diagnostic method for irritable bowel syndrome]. *Rev Gastroenterol Peru*, 32, 178-83.
- CHAMPNEY, H. & MARSHALL, H. 1939. Optimal refinement of the rating scale. *Journal of Applied Psychology*, 23, 323-331.
- CHANG, J. Y., LOCKE, G. R., 3RD, SCHLECK, C. D., ZINSMEISTER, A. R. & TALLEY, N. J. 2009. Risk factors for chronic diarrhoea in the community in the absence of irritable bowel syndrome. *Neurogastroenterol Motil*, 21, 1060-e87.
- CHANG, L. From Rome to Los Angeles -- The Rome III Criteria for the Functional GI Disorders. Digestive Diseases Week, 2006.
- CHASSANY, O., LE-JEUNNE, P., DURACINSKY, M., SCHWALM, M. S. & MATHIEU, M. 2006. Discrepancies between patient-reported outcomes and clinician-reported outcomes in chronic venous disease, irritable bowel syndrome, and peripheral arterial occlusive disease. *Value Health*, 9, 39-46.
- CHEUNG, Y. & KLOTZ, J. 1997. The Mann Whitney Wilcoxon Distribution using linked lists. *Statistica Sinica*, 7, 805-813.
- CHITKARA, D. K., VAN TILBURG, M. A., BLOIS-MARTIN, N. & WHITEHEAD, W. E. 2008. Early life risk factors that contribute to irritable bowel syndrome in adults: a systematic review. *Am J Gastroenterol*, 103, 765-74; quiz 775.
- CHO HK, TUROFF M & HILTZ SR. The Impacts of Delphi Communication Structure on Small and Medium Sized
- Asynchronous Groups: Preliminary Results. 38th International Conference on System Sciences, 2003 Hawaii. 10.
- CHOUNG, R. S. 2012. [Natural history and overlap of functional gastrointestinal disorders]. *Korean J Gastroenterol*, 60, 345-8.
- CHURCHILL, G. A., JR. & PETER, J. P. 1984. Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research*, 21, 360-375.
- CHURCHMAN, C. W. 1971. *The design of inquiring systems: Basic concepts of systems and organization*, Basic Books New York.
- CLASON, D. & DORMODY, T. 1994. Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, 35, 31-35.
- CLAYTON, M. J. 1997. Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, 17, 373-386.
- CLOUSE, R. E., MAYER, E. A., AZIZ, Q., DROSSMAN, D. A., DUMITRASCU, D. L., MONNIKES, H. & NALIBOFF, B. D. 2006. Functional abdominal pain syndrome. *Gastroenterology*, 130, 1492-1497.
- COLE, Z. D., DONOHOE, H. M. & STELLEFSON, M. L. 2013. Internet-based Delphi research: case based discussion. *Environ Manage*, 51, 511-23.
- COLMAN, A., NORRIS CE & PRESTON CC 1997. Comparing Rating Scales of Different Lengths: Equivalence of Scores From 5-Point and 7-Point. *Psychological Reports* 80, 355-362.
- CONBOY, L. A., MACKLIN, E., KELLEY, J., KOKKOTOU, E., LEMBO, A. & KAPTCHUK, T. 2010. Which patients improve: characteristics increasing sensitivity to a supportive patient-practitioner relationship. *Soc Sci Med*, 70, 479-84.
- CONWAY, A., ROLLEY, J., PAGE, K. & FULBROOK, P. 2013. Issues and challenges associated with nurse-administered procedural sedation and analgesia in the cardiac catheterisation laboratory: a qualitative study. *J Clin Nurs*.

- COOMBS, C. H. 1964. *A theory of data*, Oxford, England, Wiley.
- CORAZZIARI, E. 2004. Definition and epidemiology of functional gastrointestinal disorders. *Best Practice & Research Clinical Gastroenterology*, 18, 613-631.
- COTTRELL, R. & MCKENZIE, J. 2010. Instrumentation. *Health Promotion and Education and Research Methods*. 2nd ed. Sudbury: Jones and Bartlett Publishers.
- COURTNEY, J., CROSSDELL D & PARADICE D 2007. Inquiry organizations. In: COURTNEY J (ed.). Texas: Texas A&M University.
- CRABTREE, J. E., COURT, M., ABOSHKIWA, M. A., JEREMY, A. H., DIXON, M. F. & ROBINSON, P. A. 2004. Gastric mucosal cytokine and epithelial cell responses to *Helicobacter pylori* infection in Mongolian gerbils. *J Pathol*, 202, 197-207.
- CREMONINI, F. & TALLEY, N. J. 2005. Irritable bowel syndrome: epidemiology, natural history, health care seeking and emerging risk factors. *Gastroenterol Clin North Am*, 34, 189-204.
- CUMMING, G. 2008. Replication and p Intervals: p Values Predict the Future Only Vaguely, but Confidence Intervals Do Much Better. *Perspectives on Psychological Science*, 3, 286-300.
- CUTCLIFFE, J. R. & MCKENNA, H. P. 1999. Establishing the credibility of qualitative research findings: the plot thickens. *Journal of Advanced Nursing*, 30, 374-380.
- DALKEY, N. 1968. Predicting the Future. *RAND Corporation*. Santa Monica
- DALKEY, N. 1969. The Delphi method: an experimental study of group opinion. *RM-5888-PR*. Santa Monica: Rand Corporation.
- DALKEY, N. & HELMER, O. 1962. *An Experimental Application of the Delphi Method to the Use of Experts*, Santa Monica, RAND Corporation.
- DALTON, P., MAUTE, C., OSHIDA, A., HIKICHI, S. & IZUMI, Y. 2008. The Use of Semantic Differential Scaling to Define the Multi-Dimensional Representation of Odors. *J Sens Stud*, 23, 485-497.
- DANG, J., ARDILA-HANI, A., AMICHAH, M. M., CHUA, K. & PIMENTEL, M. 2012. Systematic review of diagnostic criteria for IBS demonstrates poor validity and utilization of Rome III. *Neurogastroenterol Motil*, 24, 853-e397.
- DAPOIGNY, M., BELLANGER, J., BONAZ, B., BRULEY DES VARANNES, S., BUENO, L., COFFIN, B., DUCROTTE, P., FLOURIE, B., LEMANN, M., LEPICARD, A. & REIGNEAU, O. 2004. Irritable bowel syndrome in France: a common, debilitating and costly disorder. *Eur J Gastroenterol Hepatol*, 16, 995-1001.
- DAY, J. & BOBEVA, M. 2005. A Generic Toolkit for the Successful Management of Delphi Studies. *Electronic Journal of Business Research Methodology*, 3, 103-116.
- DE MEYRICK, J. 2003. The Delphi method and health research. *Health Education*, 103, 7-16.
- DE VILLIERS, M. R., DE VILLIERS, P. J. T. & KENT, A. P. 2005. The Delphi technique in health sciences education research. *Medical Teacher*, 27, 639-643.
- DE WINTER, J. & DODOU, D. 2012. Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon *Practical Assessment and Research Evaluation*, 15, 1-16.
- DEBLEY, J. S., CARTER, E. R. & REDDING, G. J. 2006. Prevalence and impact of gastroesophageal reflux in adolescents with asthma: a population-based study. *Pediatr Pulmonol*, 41, 475-81.
- DELBECQ, A., VAN DE VEN, A. & GUSTAFSON DH 1975. Group Decision Making in Modern Organisations. *Group Techniques for Program Planning*

- DELBECQ, A. L. & VAN DE VEN, A. H. 1971. A Group Process Model for Problem Identification and Program Planning. *The Journal of Applied Behavioral Science*, 7, 466-492.
- DESHPANDE, A. M., SHIFFMAN, R. N. & NADKARNI, P. M. 2005. Metadata-driven Delphi rating on the Internet. *Comput Methods Programs Biomed*, 77, 49-56.
- DI REZZE, B., LAW, M., EVA, K., POLLOCK, N. & GORTER, J. W. 2014. Therapy behaviours in paediatric rehabilitation: essential attributes for intervention with children with physical disabilities. *Disability and Rehabilitation*, 36, 16-22.
- DICKINSON, T. L. & ZELLINGER, P. M. 1980. A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology*, 65, 147-154.
- DONOHUE, H., STELLEFSON M & TENNANT B 2012. Advantages and Limitations of the e-Delphi Technique:: Implications for Health Education Researchers. *American Journal of Health Education*, 43, 39-46.
- DRISCOLL, D., YEBOAH, A., SALIB, P. & RUPERT, D. 2007. Merging Qualitative and Quantitative Data in Mixed Methods Research: How To and Why Not. *Ecological and Environmental Anthropology*, 3, 19-27.
- DROSSMAN, D., CORAZZIARI E, DELVAUX M, SPILLER R, TALLEY N, GRANT W, THOMPSON W, WHITEHEAD WE & EDS 2006. *Rome III: The Functional Gastrointestinal Disorders*, McLean VA, Degnon Associates.
- DROSSMAN DA. November 18, 2013 2013. *RE: The term "functional"*. Type to AUSTIN, P.
- DUFFIELD, C. 1993. The Delphi technique: a comparison of results obtained using two expert panels. *Int J Nurs Stud*, 30, 227-37.
- DUFFY, M. E. 2002. Methodological issues in Web-based research. *J Nurs Scholarsh*, 34, 83-8.
- DUNBAR, F. H. 1938. *Emotions And Bodily Changes*, New York, Columbia Press.
- ECKEN, P., GNATZY, T. & VON DER GRACHT, H. A. 2011. Desirability bias in foresight: Consequences for decision quality based on Delphi results. *Technological Forecasting and Social Change*, 78, 1654-1670.
- EDWARDS, A. L. 1957. *Techniques of attitude scale construction / by Allen L. Edwards*, New York : Appleton-Century-Crofts, [1957].
- EDWARDS, A. L. & KENNEY, K. C. 1946. A comparison of the Thurstone and Likert techniques of attitude scale construction. *J Appl Psychol*, 30, 72-83.
- EGGER, M., ZELLWEGER-ZÄHNER, T., SCHNEIDER, M., JUNKER, C., LENGELER, C. & ANTES, G. 1997. Language bias in randomised controlled trials published in English and German. *The Lancet*, 350, 326-329.
- ENCK, P. & MARTENS, U. 2008. Der nächste Konsensus zum Reizdarmsyndrom muss interdisziplinär sein. *Z Gastroenterol*, 46, 211,215.
- ENGELS, T. C. E. & POWELL KENNEDY, H. 2007. Enhancing a Delphi study on family-focused prevention. *Technological Forecasting and Social Change*, 74, 433-451.
- ESLICK, G. D., JONES, M. P. & TALLEY, N. J. 2003. Non-cardiac chest pain: prevalence, risk factors, impact and consulting - a population-based study. *Alimentary Pharmacology & Therapeutics*, 17, 1115-1124.
- EVERETT, A. 1993. Piercing the veil of the future. A review of the Delphi method of research. *Prof Nurse*, 9, 181-5.

- FARMER, A., AZIZ, Q., TACK, J. & VAN OUDENHOV, E. L. 2010. The future of neuroscientific research in functional gastrointestinal disorders: Integration towards multidimensional (visceral) pain endophenotypes? *Journal of Psychosomatic Research*, 68, 475-481.
- FARNDAL, R. & ROBERTS, L. 2011. Long-term impact of irritable bowel syndrome: a qualitative study. *Prim Health Care Res Dev*, 12, 52-67.
- FAVREAU, C., SIPROUDHIS, L., ELEOUET, M., BOUGUEN, G. & BRETAGNE, J. F. 2012. Underlying functional bowel disorder may explain patient dissatisfaction after haemorrhoidal surgery. *Colorectal Dis*, 14, 356-61.
- FERN, E. 2001. The Focus Group Moderator. *Advanced Focus Group Research*. Sage Publications.
- FERREIRA, N. B., EUGENICOS, M. P., MORRIS, P. G. & GILLANDERS, D. T. 2012. Measuring acceptance in irritable bowel syndrome: preliminary validation of an adapted scale and construct utility. *Qual Life Res*.
- FINSTAD, K. 2010. Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales. *Journal of Usability Studies*, 5, 104-110.
- FORD, A. C., BERCIK, P., MORGAN, D. G., BOLINO, C., PINTOS-SANCHEZ, M. I. & MOAYYEDI, P. 2013. Validation of the Rome III criteria for the diagnosis of irritable bowel syndrome in secondary care. *Gastroenterology*, 145, 1262-70.e1.
- FORD, A. C., BERCIK, P., MORGAN, D. G., BOLINO, C., PINTOS-SANCHEZ, M. I. & MOAYYEDI, P. 2014. The Rome III criteria for the diagnosis of functional dyspepsia in secondary care are not superior to previous definitions. *Gastroenterology*, 146, 932-40; quiz e14-5.
- FORTEA, J. & PRIOR, M. 2013. Irritable bowel syndrome with constipation: a European-focused systematic literature review of disease burden. *J Med Econ*, 16, 329-41.
- FRIBORG, O., MARTINUSSEN, M. & ROSENVINGE, J. H. 2006. Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, 40, 873-884.
- FULTON, D. 1907. Differential Diagnosis of Organic from Functional Dyspepsia. *Cal State J Med*, 5, 172-5.
- FURMAN, D. L. & CASH, B. D. 2011. The Role of Diagnostic Testing in Irritable Bowel Syndrome. *Gastroenterology Clinics of North America*, 40, 105-119.
- GADERMANN, A., GUHN M & ZUMBO D 2012. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide *Practical Assessment and Research Evaluation*, 17, 1-9.
- GALLAGHER, M., HARES T, SPENCER J, BRADSHAW C & WEBB I 1993. The Nominal Group Technique: A Research Tool for General Practice? *Fam. Pract.*, 10, 76-81.
- GARLAND, R. 1991. The Mid-Point on a Rating Scale: Is it Desirable. *Marketing Bulletin*, 66-70.
- GASIOROWSKA, A., POH, C. H. & FASS, R. 2009. Gastroesophageal Reflux Disease (GERD) and Irritable Bowel Syndrome (IBS)-Is It One Disease or an Overlap of Two Disorders? *Digestive Diseases and Sciences*, 54, 1829-1834.
- GATCHEL, R. J. 2001. A biopsychosocial overview of pretreatment screening of patients with pain. *Clin J Pain*, 17, 192-9.
- GEIST, M. R. 2010. Using the Delphi method to engage stakeholders: A comparison of two studies. *Evaluation and Program Planning*, 33, 147-154.

- GENSICHEN, J., VOLLMAR, H. C., SONNICHSEN, A., WALDMANN, U. M. & SANDARS, J. 2009. E-learning for education in primary healthcare--turning the hype into reality: a Delphi study. *Eur J Gen Pract*, 15, 11-4.
- GOLCHIN, M., HEIDARI, H., ZIAIE, S. & SALEHI, S. 2010. Creating national care standards for neonatal intensive care units in 2007. *Iran J Nurs Midwifery Res*, 15, 54-9.
- GOODMAN, C. M. 1987. The Delphi technique: a critique. *J Adv Nurs*, 12, 729-34.
- GORDON, T. 1994. The Delphi Method. *Futures Research Methodology*, 3, 29.
- GOTHWAL, V. K., WRIGHT, T. A., LAMOUREUX, E. L. & PESUDOV, K. 2009. Guttman Scale Analysis of the Distance Vision Scale. *Investigative Ophthalmology & Visual Science*, 50, 4496-4501.
- GRAEFE, A. & ARMSTRONG, J. S. 2011. Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting*, 27, 183-195.
- GRAHAM, B., REGEHR, G. & WRIGHT, J. G. 2003a. Delphi as a method to establish consensus for diagnostic criteria. *Journal of Clinical Epidemiology*, 56, 1150-1156.
- GRAHAM, B., REGEHR, G. & WRIGHT, J. G. 2003b. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol*, 56, 1150-6.
- GREEN, B., JONES, M., HUGHES, D. & WILLIAMS, A. 1999. Applying the Delphi technique in a study of GPs' information requirements. *Health Soc Care Community*, 7, 198-205.
- GREENFIELD, S., KAPLAN, S. H., WARE, J. E., JR., YANO, E. M. & FRANK, H. J. 1988. Patients' participation in medical care: effects on blood sugar control and quality of life in diabetes. *J Gen Intern Med*, 3, 448-57.
- GUBA, E. 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*, 29, 75-91.
- GUNDEL, H., LADWIG, K. H., WOLOWSKI, A., FISCHER, A., GRUBL, A., MARTEN-MITTAG, B., KOLK, A., SCHEUTZEL, P., HAMMES, M. & NEFF, A. 2002. [Psychic and somatic findings in jaw or facial pain of unclear origin. Comparison of patients with severe and mild symptoms]. *Schmerz*, 16, 285-93.
- GUTTMAN, L. 1954. An Outline of Some New Methodology for Social Research. *Public Opinion Quarterly*, 18, 395-404.
- HALDER SLS & LOCKE GR 2007. Epidemiology and Socioeconomic Impact of Visceral and Abdominal Pain Syndromes. In: PASRICHA PJ, W. W., GEBHART GF (ed.) *Chronic Abdominal and Visceral Pain: Theory and Practice*. New York, London: Inroma Healthcare.
- HALPERT, A. & GODENA, E. 2011. Irritable bowel syndrome patients' perspectives on their relationships with healthcare providers. *Scand J Gastroenterol*, 46, 823-30.
- HALPIN, S. J. & FORD, A. C. 2012. Prevalence of symptoms meeting criteria for irritable bowel syndrome in inflammatory bowel disease: systematic review and meta-analysis. *Am J Gastroenterol*, 107, 1474-82.
- HAMILTON, W. T., GALLAGHER, A. M., THOMAS, J. M. & WHITE, P. D. 2009. Risk markers for both chronic fatigue and irritable bowel syndromes: a prospective case-control study in primary care. *Psychol Med*, 39, 1913-21.
- HARPER, M. G., ASSELIN, M. E., KURTZ, A. C., MACARTHUR, S. K. & PERRON, S. 2012. Research priorities for nursing professional development: a modified e-Delphi study. *J Nurses Staff Dev*, 28, 137-42.

- HARRIS, L. R. & ROBERTS, L. 2008. Treatments for irritable bowel syndrome: patients' attitudes and acceptability. *BMC Complement Altern Med*, 8, 65.
- HARRIS, M. & TAYLOR, G. 2009. *Medical Statistics Made Easy*, Bloxham, Scion Publishing Ltd.
- HARRISON, M. O., EDWARDS, C. L., KOENIG, H. G., BOSWORTH, H. B., DECASTRO, L. & WOOD, M. 2005. Religiosity/spirituality and pain in patients with sickle cell disease. *J Nerv Ment Dis*, 193, 250-7.
- HARZING, A.-W. 1997. Response rates in international mail surveys: Results of a 22-country study. *International Business Review*, 6, 641-665.
- HARZING, A., RIEICHE S & PUDELKO M 2012. Challenges in International; Survey Research: A Review with Illustrations and Suggested Solutions for Best Practice. *European Journal of International Management*. Barcelona: IESE Business School.
- HASSON, F. & KEENEY, S. 2011. Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, 78, 1695-1704.
- HASSON, F., KEENEY, S. & MCKENNA, H. 2000. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32, 1008-1015.
- HEATON, K. W., PARKER, D. & CRIPPS, H. 1993. Bowel function and irritable bowel symptoms after hysterectomy and cholecystectomy--a population based study. *Gut*, 34, 1108-11.
- HEITKEMPER, M. M., CAIN, K. C., DEECHAKAWAN, W., POPPE, A., JUN, S. E., BURR, R. L. & JARRETT, M. E. 2012. Anticipation of public speaking and sleep and the hypothalamic-pituitary-adrenal axis in women with irritable bowel syndrome. *Neurogastroenterol Motil*, 24, 626-31, e270-1.
- HERDMAN, M., RAJMIL, L., RAVENS-SIEBERER, U., BULLINGER, M., POWER, M. & ALONSO, J. 2002. Expert consensus in the development of a European health-related quality of life measure for children and adolescents: a Delphi study. *Acta Paediatr*, 91, 1385-90.
- HILDEBRAND, D., LAING JD & ROSENTHAL H 1977. *Analysis of Ordinal Data*, Thousand Oaks, SAGE Publications Inc.
- HINCKLEY, E. D. 1932. The Influence of Individual Opinion on Construction of an Attitude Scale. *The Journal of Social Psychology*, 3, 283-296.
- HINCKLEY, E. D. 1963. A follow-up study influence of individual opinion on the construction of an attitude scale. *The Journal of Abnormal and Social Psychology*, 67, 290-292.
- HOFFMAN, L. 1958. Homogeneity of member personality and its effect on group problem-solving. *Journal of Abnormal and Social Psychology*, 54, 51-55.
- HOFFMAN, L. R. & MAIER, N. R. 1961. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *J Abnorm Soc Psychol*, 62, 401-7.
- HOY, D., BROOKS, P., BLYTH, F. & BUCHBINDER, R. 2010. The Epidemiology of low back pain. *Best Pract Res Clin Rheumatol*, 24, 769-81.
- HSU, C. 2007. Minimizing Non-Response in The Delphi Process: How to Respond to Non-Response. *Practical Assessment, Rresearch & Evaluation*, 12.
- HSU, C. & SANDFORD, B. 2007. The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research and Evaluation*, 12, 1-8.
- HYMAN, S. E. 2002. Neuroscience, genetics, and the future of psychiatric diagnosis. *Psychopathology*, 35, 139-44.
- IGDA WORKGROUP, W. 2003. IGDA. 7: Standardised multi-axial diagnostic formulation. *The British Journal of Psychiatry*, 182, s52-s54.
- JAIRATH, N. & WEINSTEIN, J. 1994. The Delphi methodology (Part one): A useful administrative approach. *Can J Nurs Adm*, 7, 29-42.

- JAKOBSSON, U. 2007. Statistical Presentation and Analysis of Ordinal Data in Nursing Research. *Scand J Caring Sci*, 18, 437-440.
- JAMIESON, S. 2004. Likert scales: how to (ab)use them. *medical Education*, 38, 1217-1218
- JAVALI, S. 2011. Effect of Varying Sample Size in Estimation of Coefficients of Internal Consistency. *Webmed Central*, 1-8.
- JELLEMA, P., VAN DER WINDT, D., SCHELLEVIS, G. & VAN DER HORST HHE 2009. Systematic review: accuracy of symptom-based criteria for diagnosis of irritable bowel syndrome in primary care. *Alimentary Pharmacology & Therapeutics*, 30, 695-706.
- JENEWEIN, J., MOERGELI, H., WITTMANN, L., BUCHI, S., KRAEMER, B. & SCHNYDER, U. 2009. Development of chronic pain following severe accidental injury. Results of a 3-year follow-up study. *J Psychosom Res*, 66, 119-26.
- JOHANNES, C. B., LE, T. K., ZHOU, X., JOHNSTON, J. A. & DWORKIN, R. H. 2010. The prevalence of chronic pain in United States adults: results of an Internet-based survey. *J Pain*, 11, 1230-9.
- JOHNSON, D. M. & SAVILLE-TROIKE, M. 1992. Validity and Reliability in Qualitative Research on Second Language Acquisition and Teaching. Two Researchers Comment. *TESOL Quarterly*, 26, 602-605.
- JOHNSON, T. & WISLAR, J. 2012. Response rates and nonresponse errors in surveys. *JAMA*, 307, 1805-1806.
- JONES, C. 2001. Delphi Evaluation of Agreement between Organizations. In: LINSTONE HA & TUROFF M (eds.) *The Delphi Method Techniques and Applications* reading MA: Addison-Wesely.
- JONES, J. & HUNTER, D. 1995. Qualitative Research: Consensus methods for medical and health services research. *BMJ*, 311, 376-380.
- JONES, M. P., CHEY, W. D., SINGH, S., GONG, H., SHRINGARPURE, R., HOE, N., CHUANG, E. & TALLEY, N. J. 2014. A biomarker panel and psychological morbidity differentiates the irritable bowel syndrome from health and provides novel pathophysiological leads. *Aliment Pharmacol Ther*, 39, 426-37.
- JONES, M. P., COPPENS, E., VOS, R., HOLVOET, L., LUYTEN, P., TACK, J. & VAN OUDENHOVE, L. 2012. A multidimensional model of psychobiological interactions in functional dyspepsia: a structural equation modelling approach. *Gut*.
- KAHNEMAN, D., SLOVIC P & TVERSKY A 1982. *Judgment under Uncertainty: Heuristic and Biases*, Cambridge, Cambridge University Press.
- KAINU, J. P., SARVELA, J., TIIPPANA, E., HALMESMAKI, E. & KORTTILA, K. T. 2010. Persistent pain after caesarean section and vaginal birth: a cohort study. *Int J Obstet Anesth*, 19, 4-9.
- KAJI, M., FUJIWARA, Y., SHIBA, M., KOHATA, Y., YAMAGAMI, H., TANIGAWA, T., WATANABE, K., WATANABE, T., TOMINAGA, K. & ARAKAWA, T. 2010. Prevalence of overlaps between GERD, FD and IBS and impact on health-related quality of life. *J Gastroenterol Hepatol*, 25, 1151-6.
- KEENEY, S., HASSON, F. & MCKENNA H 2011a. Conducting the Research Using the Delphi Technique. *The Delphi Technique in Nursing and Health Research*
- Oxford: Wiley Blackwell.

- KEENEY, S., HASSON, F. & MCKENNA H 2011b. The Delphi Technique. *Delphi Technique in Nursing and Health Research*. Oxford: Wiley Blackwell.
- KEENEY, S., HASSON, F. & MCKENNA. H 2011c. *Reliability and Validity*, Wiley-Blackwell.
- KELLOW JE. 2103. *RE: The term functional*. Type to AUSTIN PD.
- KEOHANE, J., O'MAHONY, C., O'MAHONY, L., O'MAHONY, S., QUIGLEY, E. M. & SHANAHAN, F. 2010. Irritable bowel syndrome-type symptoms in patients with inflammatory bowel disease: a real association or reflection of occult inflammation? *Am J Gastroenterol*, 105, 1788, 1789-94; quiz 1795.
- KHAN, S. & CHANG, L. 2010. Irritable Bowel Syndrome (IBS): Diagnosis and Management: Diagnosis of IBS. *Medscape Education* [Online]. Available: http://www.medscape.org/viewarticle/729760_2.
- KLINE RB 2004. *Beyond significance testing: Reforming data analysis methods in behavioral research*, Washington DC, American Psychological Association.
- KNOTT, A., PATHAK, S., MCGRATH, J. S., KENNEDY, R., HORGAN, A., MYTHEN, M., CARTER, F. & FRANCIS, N. K. 2012. Consensus views on implementation and measurement of enhanced recovery after surgery in England: Delphi study. *BMJ Open*, 2.
- KOK, L., ELIAS, S. G., WITTEMAN, B. J., GOEDHARD, J. G., ROMBERG-CAMPS, M. J., MURIS, J. W., MOONS, K. G. & DE WIT, N. J. 2013. Application of the Rome III criteria is not likely to reduce the number of unnecessary referrals for colonoscopy in primary care. *Eur J Gastroenterol Hepatol*, 25, 568-74.
- KOMORITA, S. S. & GRAHAM, W. K. 1965. Number of scale points and the reliability of scales. *Educational and Psychological Measurement*, 25, 987-995.
- KOTHARI CR 2004. Measurement and scaling techniques. *Research Methodology-Methods and Techniques*. 2 ed. New Delhi: New Age International.
- KRIGER, M. & ESTHER, E. 1992. Strategic Mindsets and Decision-Making Autonomy in U.S. and Japanese MNCs. *MIR: Management International Review*, 32, 327-343.
- KRUIS, W., THIEME, C., WEINZIERL, M., SCHUSSLER, P., HOLL, J. & PAULUS, W. 1984. A diagnostic score for the irritable bowel syndrome. Its value in the exclusion of organic disease. *Gastroenterology*, 87, 1-7.
- KUMAR, N. 2007. WHO Normative Guidelines on Pain Management. *In: WORLD HEALTH ORGANISATION* (ed.). Geneva: WHO,.
- LABUS, J., GUPTA, A., GILL, H. K., POSSERUD, I., MAYER, M., RAEEN, H., BOLUS, R., SIMREN, M., NALIBOFF, B. D. & MAYER, E. A. 2013. Randomised clinical trial: symptoms of the irritable bowel syndrome are improved by a psycho-education group intervention. *Aliment Pharmacol Ther*, 37, 304-15.
- LACKNER, J. M., JACCARD, J. & BAUM, C. 2013. Multidomain Patient-Reported Outcomes of Irritable Bowel Syndrome: Exploring Person-Centered Perspectives to Better Understand Symptom Severity Scores. *Value in Health*, 16, 97-103.
- LACKNER, J. M., QUIGLEY, B. M. & BLANCHARD, E. B. 2004. Depression and abdominal pain in IBS patients: the mediating role of catastrophizing. *Psychosom Med*, 66, 435-41.
- LEE, H. J., LEE, S. Y., KIM, J. H., SUNG, I. K., PARK, H. S., JIN, C. J., KANG, S. G., YOON, H. & CHUN, H. J. 2010. Depressive mood and quality of life in functional gastrointestinal disorders: differences between functional dyspepsia, irritable bowel syndrome and overlap syndrome. *Gen Hosp Psychiatry*, 32, 499-502.

- LEE, S. Y., LEE, K. J., KIM, S. J. & CHO, S. W. 2009. Prevalence and Risk Factors for Overlaps between Gastroesophageal Reflux Disease, Dyspepsia, and Irritable Bowel Syndrome: A Population-Based Study. *Digestion*, 79, 196-201.
- LEEuw, M., HOUBEN, R. M., SEVEREIJNS, R., PICALET, H. S., SCHOUTEN, E. G. & VLAEYEN, J. W. 2007. Pain-related fear in low back pain: a prospective study in the general population. *Eur J Pain*, 11, 256-66.
- LEI CHANG 1994. A Psychometric Evaluation of 4-Point and 6-Point Likert-Type Scales in Relation to Reliability and Validity. *Applied Psychological Measurement*, 18, 205-215.
- LEMBO, A. J., NERI, B., TOLLEY, J., BARKEN, D., CARROLL, S. & PAN, H. 2009. Use of serum biomarkers in a diagnostic test for irritable bowel syndrome. *Aliment Pharmacol Ther*, 29, 834-42.
- LIKERT, R. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140, 55.
- LINCOLN, Y. & GUBA, E. 1985a. Establishing Trustworthiness. *Naturalistic Inquiry*. California: Sage Publications.
- LINCOLN, Y. & GUBA, E. 1985b. *Naturalistic Inquiry*, Newbury Park, California Sage Publications.
- LING & FALL. 2008. *Tutorial: Pearson's Chi-square Test for Independence* [Online]. Philadelphia: University of Pennsylvania. Available: <http://www.ling.upenn.edu/~clight/chisquared.htm> 2013].
- LINSTONE, H. & TUROFF, M. 2002. *The Delphi method: Techniques and applications.*, Reading, MA, Addison-Wesley Publishing Company.
- LISSITZ, R. W. & GREEN, S. B. 1975. Effect of the Number of Scale Points on Reliability A Monte Carlo Approach. *Journal of Applied Psychology*, 60, 10-13.
- LODEWIJCKX, C., SERMEUS, W., PANELLA, M., DENECKERE, S., LEIGHEB, F., TROOSTERS, T., BOTO, P. A., MENDES, R. V., DECRAMER, M. & VANHAECHE, K. 2013. Quality indicators for in-hospital management of exacerbation of chronic obstructive pulmonary disease: results of an international Delphi study. *J Adv Nurs*, 69, 348-62.
- LOHNINGER, H. 1999. *Teach Me Data Analysis.*, Berlin New York Tokyo: Springer.
- LÓPEZ-COLOMBO, A., MORGAN, D., BRAVO-GONZÁLEZ, D., MONTIEL-JARQUÍN, A., MÉNDEZ-MARTÍNEZ, S. & SCHMULSON, M. 2012. The Epidemiology of Functional Gastrointestinal Disorders in Mexico: A Population-Based Study. *Gastroenterology Research and Practice*, 2012, 8.
- LOUGHLIN, K. G. & MOORE, L. F. 1979. Using Delphi to achieve congruent objectives and activities in a pediatrics department. *Academic Medicine*, 54, 101-6.
- LOVALLO, D. & KAHNEMAN, D. 2003. Delusions of success. How optimism undermines executives' decisions. *Harv Bus Rev*, 81, 56-63, 117.
- LOVELL, R. M. & FORD, A. C. 2012. Effect of gender on prevalence of irritable bowel syndrome in the community: systematic review and meta-analysis. *Am J Gastroenterol*, 107, 991-1000.
- LUDBROOK, J. 2013. Analysing 2 x 2 contingency tables: which test is best? *Clin Exp Pharmacol Physiol*, 40, 177-80.
- LUDIDI, S., CONCHILLO, J. M., KESZTHELYI, D., VAN AVESAAT, M., KRUIJEL, J. W., JONKERS, D. M. & MASCLÉE, A. A. 2012. Rectal hypersensitivity as hallmark for irritable bowel syndrome: defining the optimal cutoff. *Neurogastroenterol Motil*, 24, 729-33, e345-6.
- LUND, M. & LUND, A. 2013. *Kruskal-Wallis H Test using SPSS* [Online]. Derby: Laird Research Ltd. Available: <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php> [Accessed 17 Jan 2014].

- MAHADEVA, S. & GOH, K. 2006. Epidemiology of functional dyspepsia: A global perspective. *World J Gastroenterol*, 12, 2661-6.
- MALHOTRA, Y. 1997. Knowledge Management in Inquiring Organizations. *3rd Americas Conference on Information Systems*. Indianapolis: Philosophy of Information Systems Mini-track.
- MALONE, D. C., ABARCA, J., HANSTEN, P. D., GRIZZLE, A. J., ARMSTRONG, E. P., VAN BERGEN, R. C., DUNCAN-EDGAR, B. S., SOLOMON, S. L. & LIPTON, R. B. 2004. Identification of serious drug-drug interactions: results of the partnership to prevent drug-drug interactions. *J Am Pharm Assoc (2003)*, 44, 142-51.
- MANNING, A. P., THOMPSON, W. G., HEATON, K. W. & MORRIS, A. F. 1978. Towards positive diagnosis of the irritable bowel. *Br Med J*, 2, 653-4.
- MARSHALL, J. K., THABANE, M., GARG, A. X., CLARK, W. F., SALVADORI, M. & COLLINS, S. M. 2006. Incidence and Epidemiology of Irritable Bowel Syndrome After a Large Waterborne Outbreak of Bacterial Dysentery. *Gastroenterology*, 131, 445-450.
- MASTERS, E. R. 1974. THE RELATIONSHIP BETWEEN NUMBER OF RESPONSE CATEGORIES AND RELIABILITY OF LIKERT-TYPE QUESTIONNAIRES¹. *Journal of Educational Measurement*, 11, 49-53.
- MATELL, M. S. & JACOBY, J. 1971. Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity. *Educational and Psychological Measurement*, 31, 657-674.
- MCGRATH, J. E. 1981. Dilemmatics: The Study of Research Choices and Dilemmas. *American Behavioral Scientist*, 25, 179.
- MCKENNA, H. P. 1994. The Delphi technique: a worthwhile research approach for nursing? *J Adv Nurs*, 19, 1221-5.
- MCKINNON, A. C., VAN OUDENHOVE, L., TACK, J. & JONES, M. 2013. The association of personality, appraisal, catastrophising and vigilance with gastrointestinal symptom-specific anxiety. *Journal of Health Psychology*.
- MCLVER, J. 2004. Semantic Differential Scale. In: LEWIS-BECK MS, BRYMAN A & LIAO TF (eds.) *The SAGE Encyclopedia of Social Science Research Methods*. Sage publications.
- MEAD, D. & MOSELEY, L. 2001. The use of the Delphi as a research approach. *Nurse Res*, 8, 4-23.
- MENOLD, N., KACZMIREK, L., LENZNER, T. & NEUSAR, A. 2014. How Do Respondents Attend to Verbal Labels in Rating Scales? *Field Methods*, 26, 21-39.
- MEZZICH, J. 2002. The WPA International Guidelines for Diagnostic Assessment. *World Psychiatry*, 1, 36-9.
- MIKOCCA-WALUS, A., TURNBULL, D., MOULDING, N., WILSON, I., ANDREWS, J. M. & HOLTSMANN, G. 2008. Psychological comorbidity and complexity of gastrointestinal symptoms in clinically diagnosed irritable bowel syndrome patients. *J Gastroenterol Hepatol*, 23, 1137-43.
- MINIWATTS MARKETING GROUP. 2103. *Internet World Stats* [Online]. Columbia Miniwatts Marketing Group. Available: <http://www.internetworldstats.com/emarketing.htm> [Accessed Jan 15 2013].
- MISRA, J., LUNDQUIST, J., DAHLBERG HOLMES E & AGIOMAVRTIS S. 2010. *Associate Professors and Gendered Barriers to Advancement Executive Summary* [Online]. Massachusetts: University of Massachusetts. Available:

- http://people.umass.edu/misra/Joya_Misra/Work-Life_Research_files/Associate%20Professors%20Report%20Executive%20Summary.pdf.
- MITROFF, I. & TUROFF, M. 2002. Philosophical and Methodological Foundations of Delphi. In: LINSTONE HA & TUROFF M (eds.) *The Delphi Method: Techniques and Applications*. Reading MA: Addison-Wesley.
- MOAYYEDI, P. & FORD, A. C. 2011. Symptom-Based Diagnostic Criteria for Irritable Bowel Syndrome: the More Things Change, the More They Stay the Same. *Gastroenterology Clinics of North America*, 40, 87-103.
- MOSS-MORRIS, R. & SPENCE, M. 2006. To "lump" or to "split" the functional somatic syndromes: can infectious and emotional risk factors differentiate between the onset of chronic fatigue syndrome and irritable bowel syndrome? *Psychosom Med*, 68, 463-9.
- MOTULSKY, H. 1995. Intuitive Biostatistics *Choosing a statistical test*. Oxford: Oxford University press.
- MULLEN, P. M. 2003. Delphi: myths and reality. *J Health Organ Manag*, 17, 37-52.
- NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE 2008. Irritable Bowel Syndrome: Costing Report, Implementing NICE Guidelines. *NICE clinical guideline 61*. London: National Institute for Health and Clinical Excellence.
- NAUMANN, F. 2007. How to develop a search strategy for a Cochrane Review. In: COLLABORATION, C. (ed.). Cologne: Cochrane Haematological Malignancies Group.
- NEBLETT, R., HARTZELL, M. M., COHEN, H., MAYER, T. G., WILLIAMS, M., CHOI, Y. & GATCHEL, R. J. 2014. Ability of the Central Sensitization Inventory to Identify Central Sensitivity Syndromes in an Outpatient Chronic Pain Sample. *Clin J Pain*.
- NEWMAN, J. D. & HARRIS, J. C. 2009. The scientific contributions of Paul D. MacLean (1913-2007). *J Nerv Ment Dis*, 197, 3-5.
- NG, S. M. & CHOW, K. W. 2012. Symptoms catastrophizing versus social hypervigilance in irritable bowel syndrome patients. *Soc Work Health Care*, 51, 743-56.
- NIKOLAISEN, L., SORENSEN, H. C., JENSEN, T. S. & KEHLET, H. 2004. Chronic pain following Caesarean section. *Acta Anaesthesiol Scand*, 48, 111-6.
- NO AUTHOR LISTED 1972. Irritable bowel syndrome. *Br Med J*, 1, 197-8.
- OHTANI, M., GARCIA, A., ROGERS, A. B., GE, Z., TAYLOR, N. S., XU, S., WATANABE, K., MARINI, R. P., WHARY, M. T., WANG, T. C. & FOX, J. G. 2007. Protective role of 17 beta -estradiol against the development of Helicobacter pylori-induced gastric cancer in INS-GAS mice. *Carcinogenesis*, 28, 2597-604.
- OKEKE, E. N., LADEP, N. G., ADAH, S., BUPWATDA, P. W., AGABA, E. I. & MALU, A. O. 2009. Prevalence of irritable bowel syndrome: a community survey in an African population. *Ann Afr Med*, 8, 177-80.
- OKOLI, C. & PAWLOWSKI, S. D. 2004. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42, 15-29.
- ONO, R. & WEDEMEYER, D. J. 1994. Assessing the validity of the Delphi technique. *Futures*, 26, 289-304.
- OREMUS, M., OREMUS, C., HALL, G. B. C., MCKINNON, M. C., ECT & TEAM, C. S. R. 2012. Inter-rater and test-retest reliability of quality assessments by novice student raters using the Jadad and Newcastle-Ottawa Scales. *BMJ Open*, 2.
- PAGE-BUCCI H 2003. The value of Likert scales in measuring attitudes of online learners. In: HKA DESIGNS (ed.).

- PARASKEVAS, A. & SAUNDERS, M. 2012. Beyond consensus: an alternative use of Delphi enquiry in hospitality research. *International Journal of Contemporary Hospitality Management*, 24, 907-924.
- PAYNE, G. & PAYNE, J. 2004. Key Concepts in Social Research. *SAGE Key Concepts series*. London: SAGE Publications Ltd.
- PETERS, D. L. & MCCORMICK, E. J. 1966. Comparative reliability of numerically anchored versus job-task anchored rating scales. *J Appl Psychol*, 50, 92-6.
- PICAVET, H. S. J., VLAHEYEN, J. W. S. & SCHOUTEN, J. S. A. G. 2002. Pain Catastrophizing and Kinesiophobia: Predictors of Chronic Low Back Pain. *American Journal of Epidemiology*, 156, 1028-1034.
- PORTER, S. 2007. Validity, trustworthiness and rigour: reasserting realism in qualitative research. *J Adv Nurs*, 60, 79-86.
- PRECISION CONSULTING. 2014. *Survey Development and Analysis* [Online]. New York: Precision Colsultinh LLC. Available: <http://precisionconsultingcompany.com/survey-development-and-analysis.shtml> [2014].
- PRESTON, C. C. & COLMAN, A. M. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol (Amst)*, 104, 1-15.
- QUIGLEY, E. M. & BERNSTEIN, C. N. 2012. "Irritable bowel symptoms" in inflammatory bowel disease: diagnostic uncertainty meets pathological reality. *Am J Gastroenterol*, 107, 1483-5.
- QUIGLEY, E. M., BYTZER, P., JONES, R. & MEARIN, F. 2006. Irritable bowel syndrome: the burden and unmet needs in Europe. *Dig Liver Dis*, 38, 717-23.
- RAND CORPORATION. 2013. *History and Mission* [Online]. Santa Monica: RAND Corporation. Available: <http://www.rand.org/about/history.html>.
- RIEDL, A., MAASS, J., FLIEGE, H., STENGEL, A., SCHMIDTMANN, M., KLAPP, B. F. & MONNIKES, H. 2009. Subjective theories of illness and clinical and psychological outcomes in patients with irritable bowel syndrome. *J Psychosom Res*, 67, 449-55.
- RINGSTROM, G., ABRAHAMSSON, H., STRID, H. & SIMREN, M. 2007. Why do subjects with irritable bowel syndrome seek health care for their symptoms? *Scand J Gastroenterol*, 42, 1194-203.
- RIPPENTROP, E. A., ALTMAIER, E. M., CHEN, J. J., FOUND, E. M. & KEFFALA, V. J. 2005. The relationship between religion/spirituality and physical health, mental health, and pain in a chronic pain population. *Pain*, 116, 311-21.
- ROBERTS, J. S., DONOGHUE, J. R. & LAUGHLIN, J. E. 2000. A General Item Response Theory Model for Unfolding Unidimensional Polytomous Responses. *Applied Psychological Measurement*, 24, 3-32.
- ROBERTS, J. S., LAUGHLIN, J. E. & WEDELL, D. H. 1999. Validity Issues in the Likert and Thurstone Approaches to Attitude Measurement. *Educational and Psychological Measurement*, 59, 211-233.
- RODRÍGUEZ, L. A. G., RUIGÓMEZ, A., WALLANDER, M. A., JOHANSSON, S. & OLBE, L. 2000. Detection of Colorectal Tumor and Inflammatory Bowel Disease during Follow-up of Patients with Initial Diagnosis of Irritable Bowel Syndrome. *Scandinavian Journal of Gastroenterology*, 35, 306-311.
- ROLFE, G. 2006. Validity, trustworthiness and rigour: quality and the idea of qualitative research. *J Adv Nurs*, 53, 304-10.

- ROME FOUNDATION 2006. Rome III Diagnostic Criteria for Functional Gastrointestinal Disorders. *Gastroenterology*, 20, 885-898.
- ROTH, R. S., GEISSER, M. E. & BATES, R. 2008. The relation of post-traumatic stress symptoms to depression and pain in patients with accident-related chronic pain. *J Pain*, 9, 588-96.
- ROWE, G., WRIGHT G & F, B. 1991. Delphi: A reevaluation of research and theory. *Technological Forecasting and Social Change*, 39, 235-251.
- ROY, M., PICHE, M., CHEN, J. I., PERETZ, I. & RAINVILLE, P. 2009. Cerebral and spinal modulation of pain by emotions. *Proc Natl Acad Sci U S A*, 106, 20900-5.
- RUIGOMEZ, A., WALLANDER, M. A., JOHANSSON, S. & RODRIGUEZ, L. 2009. Irritable Bowel Syndrome and Gastroesophageal Reflux Disease in Primary Care: Is There a Link? *Digestive Diseases and Sciences*, 54, 1079-1086.
- SAAVEDRA, J. E., MEZZICH, J. E., SALLOUM, I. M. & KIRISCI, L. 2001. Predictive validity of the physical disorders axis of the DSM multiaxial diagnostic system. *J Nerv Ment Dis*, 189, 435-41.
- SACKETT, D. L., ROSENBERG, W. M. C., GRAY, J. A. M., HAYNES, R. B. & RICHARDSON, W. S. 1996. Evidence based medicine: what it is and what it isn't. *BMJ*, 312, 71-72.
- SACKMAN, H. 1974. Delphi assessment: expert opinion, forecasting and group process. *Rand Corp*, R-1283-PR.
- SAITO, Y., MITRA, N. & MAYER, E. 2010. Genetic approaches to functional gastrointestinal disorders. *Gastroenterology*, 138, 1276-85.
- SAITO, Y. A., PETERSEN, G. M., LOCKE, G. R., 3RD & TALLEY, N. J. 2005. The genetics of irritable bowel syndrome. *Clin Gastroenterol Hepatol*, 3, 1057-65.
- SANDELOWSKI, M. 1993. Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *ANS Adv Nurs Sci*, 16, 1-8.
- SANDELOWSKI, M. 2000. Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Res Nurs Health*, 23, 246-55.
- SANDELOWSKI, M. & BARROSO, J. 2002. Finding the Findings in Qualitative Studies. *Journal of Nursing Scholarship*, 34, 213-219.
- SANDLER, J. & POLLOCK, A. B. 1954. Studies in psychopathology using a self-assessment inventory. II. Some neurotic gastro-intestinal symptoms: functional dyspepsia in men. *Br J Med Psychol*, 27, 146-57.
- SANDLER, R., DROSSMAN, D., NATHAN HP & MCKEE DC 1984. Symptom complaints and health care seeking behavior in subjects with bowel dysfunction. *Gastroenterol*, 87, 314-8.
- SAS SOFTWARE SOLUTIONS. 2015. *BOXCHART Statement: SHEWHART Procedure* [Online]. Cary NC, USA: SAS. Available: http://support.sas.com/documentation/cdl/en/qcug/65562/HTML/default/viewer.htm#qcug_shewhart_sect027.htm 2015].
- SCHEIBE, M., SKUTSCH, M. & SCHAFER, J. 1975. Experiments in Delphi Methodology. In: LINSTONE, H. & TUROFF, M. (eds.) *The Delphi Method: Techniques and Applications*. Reading Mass: Addison-Wesley
- SCHMULSON, M., ADEYEMO, M., GUTIERREZ-REYES, G., CHARUA-GUINDIC, L., FARFAN-LABONNE, B., OSTROSKY-SOLIS, F., DIAZ-ANZALDUA, A., MEDINA, L. & CHANG, L. 2010. Differences in gastrointestinal symptoms according to gender in Rome II positive IBS and dyspepsia in a Latin American population. *Am J Gastroenterol*, 105, 925-32.
- SCOTT, J. W. 1933. FUNCTIONAL DYSPEPSIA. *Can Med Assoc J*, 28, 518-21.

- SHARKEY, S. B. & SHARPLES, A. Y. 2001. An approach to consensus building using the Delphi technique: developing a learning resource in mental health. *Nurse Educ Today*, 21, 398-408.
- SHEEHAN, K. B. 2001. E-mail Survey Response Rates: A Review. *Journal of Computer-Mediated Communication*, 6, 0-0.
- SHENTON AK 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-74.
- SIEGLE, D. 2010. *Likert Scale* [Online]. Neag School of Education: University of Connecticut. Available: <http://www.gifted.uconn.edu/siegle/research/instrument%20reliability%20and%20validity/likert.html> [Accessed 16 Dec 2013].
- SILK, A. 1969. *NOTES ON THE APPLICATION OF ATTITUDE MEASUREMENT AND SCALING TECHNIQUES IN MARKETING RESEARCH* M.I.T., Cambridge M.I.T.
- SKULMOSKI, G., HARTMAN, F. & KRAHN J 2007. The Delphi Method for Graduate Research. *Journal of Information Technology Education*, 6, 1-21.
- SMART, K., BLAKE C, STAINES A & DOODY C 2010. Clinical indicators of 'nociceptive', 'peripheral neuropathic' and 'central' mechanisms of musculoskeletal pain. A Delphi survey of expert clinicians. *Manual Therapy*, 15, 80-87.
- SMART, K. M., BLAKE, C., STAINES, A. & DOODY, C. 2011. The Discriminative validity of "nociceptive," "peripheral neuropathic," and "central sensitization" as mechanisms-based classifications of musculoskeletal pain. *Clin J Pain*, 27, 655-63.
- SMITH, R. C., GREENBAUM, D. S., VANCOUVER, J. B., HENRY, R. C., REINHART, M. A., GREENBAUM, R. B., DEAN, H. A. & MAYLE, J. E. 1991. Gender differences in Manning criteria in the irritable bowel syndrome. *Gastroenterology*, 100, 591-5.
- SMITH, W. 2008. Does Gender Influence Online Survey Participation?: A Record-linkage Analysis of University Faculty Online Survey Response Behavior *Educational Resources Information Centre*. 2008 ed.: US Department of Education.
- SOOD, R., LAW, G. R. & FORD, A. C. 2014. Diagnosis of IBS: symptoms, symptom-based criteria, biomarkers or 'psychomarkers'? *Nat Rev Gastroenterol Hepatol*.
- SPERBER, A. D. & DROSSMAN, D. A. 2011. Review article: the functional abdominal pain syndrome. *Aliment Pharmacol Ther*, 33, 514-24.
- SPERBER, A. D., MORRIS, C. B., GREEMBERG, L., BANGDIWALA, S. I., GOLDSTEIN, D., SHEINER, E., RUSABROV, Y., HU, Y., KATZ, M., FREUD, T., NEVILLE, A. & DROSSMAN, D. A. 2008. Development of abdominal pain and IBS following gynecological surgery: a prospective, controlled study. *Gastroenterology*, 134, 75-84.
- SPIEGEL, B. M., FARID, M., ESRAILIAN, E., TALLEY, J. & CHANG, L. 2010. Is irritable bowel syndrome a diagnosis of exclusion?: a survey of primary care providers, gastroenterologists, and IBS experts. *Am J Gastroenterol*, 105, 848-58.
- SPILLER, R. 2011. Potential Biomarkers. *Gastroenterology Clinics of North America*, 40, 121-139.
- SPILLER, R. & LAM, C. 2012. An Update on Post-infectious Irritable Bowel Syndrome: Role of Genetics, Immune Activation, Serotonin and Altered Microbiome. *J Neurogastroenterol Motil*, 18, 258-68.
- STEVENS, E. 1975. *American Educational Research Journal*, 12, 94-96.
- STRAUSS, S., RICHARDSON, W., GLASZIOU P & HAYNES RB 2011. *Evidence-based medicine: How to practice and teach EBM*, Churchill Livingstone.

- SU, Y., WIN, K. & CHUNG TC 2010. Applying the modified delphi method to identify the Taiwanese electronic health record systems evaluation framework and instrument. *14th Pacific Asia Conference on Information Systems Taiwan: PACIS*
- SUARES, N. C. & FORD, A. C. 2011. Prevalence of, and risk factors for, chronic idiopathic constipation in the community: systematic review and meta-analysis. *Am J Gastroenterol*, 106, 1582-91; quiz 1581, 1592.
- SUCKLEY, J. 2102. *CORE CLINICAL COMPETENCIES FOR EXTENDED-SCOPE PHYSIOTHERAPISTS WORKING IN MUSCULOSKELETAL (MSK) INTERFACE CLINICS BASED IN PRIMARY CARE: A DELPHI CONSENSUS STUDY*. PhD Doctoral, University of Salford.
- SYMONDS, P. M. 1924. On the Loss of Reliability in Ratings Due to Coarseness of the Scale. *Journal of Experimental Psychology*, 7, 456-461.
- TALLEY N. November 15 2013. *RE: The term "functional"*. Type to AUSTIN, P.
- TALLEY NJ & SPILLER R 2002. Irritable bowel syndrome: a little understood organic bowel disease ? *Lancet*, 360, 165-77.
- TALLEY, N. J., PHILLIPS, S. F., MELTON, L. J., MULVIHILL, C., WILTGEN, C. & ZINSMEISTER, A. R. 1990. Diagnostic value of the Manning criteria in irritable bowel syndrome. *Gut*, 31, 77-81.
- TALLEY, N. J., STANGHELLINI, V., HEADING, R. C., KOCH, K. L., MALAGELADA, J. R. & TYTGAT, G. N. 1999. Functional gastroduodenal disorders. *Gut*, 45 Suppl 2, li37-42.
- TAYLOR, A. L. & BELLUMKONDA, L. 2009. Minority Women and Cardiovascular Disease. *Cardiovascular Disease in Racial and Ethnic Minorities*. Humana Press Inc.
- THE ROME FOUNDATION. 2006. *Rome III Disorders and Criteria* [Online]. Mclean VA,: The Rome Foundation. [Accessed Oct 19th 2010].
- THURSTONE, L. & CHAVE, E. 1929. *The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the church*, Chicago, University of Chicago Press.
- THURSTONE, L. L. 1928. Attitudes Can Be Measured. *American Journal of Sociology*, 33, 529-554.
- TITTLE, C. R. & HILL, R. J. 1967. Attitude measurement and prediction of behavior: an evaluation of conditions and measurement techniques. *Sociometry*, 30, 199-213.
- TROCHIM, W. 2006. Descriptive Statistics. In: TROCHIM, W. (ed.) *Research Methods Knowledge Base*.
- TUROFF, M. & HILTZ, S. 1996. Computer based Delphi Processes. In: ADLER, M. & ZIGLIO, E. (eds.) *Gazing into the oracle: the Delphi method and its application to social policy and public health*. Bristol: Kingsley Publishers.
- UEBERSAX, J. 2006. *Likert scales: dispelling the confusion* [Online]. Available: <http://john-uebersax.com/stat/likert.htm> [Accessed Dec 15 2013].
- UMBERSON, D. & MONTEZ, J. K. 2010. Social relationships and health: a flashpoint for health policy. *J Health Soc Behav*, 51 Suppl, S54-66.
- UNIVERSITY OF DURHAM INFORMATION TECHNOLOGY SERVICE. 2006. *Bristol Online Survey (BOS) User Guide* [Online]. Durham. [Accessed Jan 23 2013].
- UNIVERSITY OF MARYLAND. 2012. *II-1.25(A) UMCP Policy on Full-Time Faculty Workload* [Online]. Maryland. Available: <http://www.president.umd.edu/policies/ii125a.html>.
- VAN OUDENHOVE, L. 2011. Understanding gut-brain interactions in gastrointestinal pain by neuroimaging: lessons from somatic pain studies. *Neurogastroenterology & Motility*, 23, 292-302.

- VAN OUDENHOVE, L. & AZIZ, Q. 2013. The role of psychosocial factors and psychiatric disorders in functional dyspepsia. *Nature Reviews Gastroenterology & Hepatology*, 10, 158-167.
- VAN OUDENHOVE, L., VANDENBERGHE, J., DEMYTTENAERE, K. & TACK, J. 2010. Psychosocial factors, psychiatric illness and functional gastrointestinal disorders: a historical perspective. *Digestion*, 82, 201-10.
- VAN TEIJLINGEN, E. & HUNDLEY, V. 2002. The importance of pilot studies. *Nurs Stand*, 16, 33-6.
- VEN, A. H. V. D. & DELBECQ, A. L. 1974. The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *Academy of Management Journal*, 17, 605-621.
- VERBUNT, J. A., SIEBEN, J. M., SEELEN, H. A., VLAEYEN, J. W., BOUSEMA, E. J., VAN DER HEIJDEN, G. J. & KNOTTNERUS, J. A. 2005. Decline in physical activity, disability and pain-related fear in sub-acute low back pain. *Eur J Pain*, 9, 417-25.
- VON DER GRACHT, H. A. 2008. The Delphi Technique for Future Research. In: JAHNS C (ed.) *The Future of Logisitics: Scenarios for 2025*. Wiesbaden: Gabler-Verlag.
- VON DER GRACHT, H. A. 2012. Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change*, 79, 1525-1536.
- WANINGE, A., ROOK, R. A., DIJKHUIZEN, A., GIELEN, E. & VAN DER SCHANS, C. P. 2011. Feasibility, test-retest reliability, and interrater reliability of the Modified Ashworth Scale and Modified Tardieu Scale in persons with profound intellectual and multiple disabilities. *Res Dev Disabil*, 32, 613-20.
- WARREN, J. W., LANGENBERG, P. & CLAUW, D. J. 2013. The number of existing functional somatic syndromes (FSSs) is an important risk factor for new, different FSSs. *J Psychosom Res*, 74, 12-7.
- WATKINS, R. E., ELLIOTT, E. J., MUTCH, R. C., PAYNE, J. M., JONES, H. M., LATIMER, J., RUSSELL, E., FITZPATRICK, J. P., HAYES, L., BURNS, L., HALLIDAY, J., D'ANTOINE, H. A., WILKINS, A., PEADON, E., MIERS, S., CARTER, M., O'LEARY, C. M., MCKENZIE, A. & BOWER, C. 2012. Consensus diagnostic criteria for fetal alcohol spectrum disorders in Australia: a modified Delphi study. *BMJ Open*, 2.
- WELTY, G. 1972. Problems of Selecting Experts for Delphi Exercises. *The Academy of Management Journal*, 15, 121-124.
- WENG, L.-J. 2004. Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability. *Educational and Psychological Measurement*, 64, 956-972.
- WESSELY, S., NIMNUAN, C. & SHARPE, M. 1999. Functional somatic syndromes: one or many? *Lancet*, 354, 936-9.
- WESTBROOK, J. I. & TALLEY, N. J. 2002. Empiric Clustering of Dyspepsia into Symptom Subgroups: a Population-based Study. *Scandinavian Journal of Gastroenterology*, 37, 917-923.
- WHITE, P. D. 2013. Functional somatic syndromes may be either "polysyndromic" or "monosyndromic". *Journal of psychosomatic research*, 74, 2-3.
- WHITEHEAD, D. 2008. An international Delphi study examining health promotion and health education in nursing practice, education and policy. *J Clin Nurs*, 17, 891-900.
- WHITEHEAD, W. E., PALSSON, O. S., LEVY, R. R., FELD, A. D., TURNER, M. & VON KORFF, M. 2007. Comorbidity in irritable bowel syndrome. *Am J Gastroenterol*, 102, 2767-76.
- WHORWELL, P. J., MCCALLUM, M., CREED, F. H. & ROBERTS, C. T. 1986. Non-colonic features of irritable bowel syndrome. *Gut*, 27, 37-40.

- WICKER, A. W. 1969. Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects. *Journal of Social Issues*, 25, 41-78.
- WIECH, K., FARIAS, M., KAHANE, G., SHACKEL, N., TIEDE, W. & TRACEY, I. 2008. An fMRI study measuring analgesia enhanced by religion as a belief system. *Pain*, 139, 467-76.
- WILLIAMS, A. 2003. How to ... Write and analyse a questionnaire. *J. Orthod.*, 30, 245-252.
- WILLIAMS, P. & WEBB, C. 1994. The Delphi technique: a methodological discussion. *J Adv Nurs*, 19, 180 - 6.
- WORLD PSYCHIATRIC ASSOCIATION 2003. IGDA. 7: Standardised multi-axial diagnostic formulation. *The British Journal of Psychiatry*, 182, s52-54.
- WOUDENBERG, F. 1991. An evaluation of Delphi. *Technological Forecasting and Social Change*, 40, 131-150.
- WRIGHT, G., LAWRENCE, M. J. & COLLOPY, F. 1996. The role and validity of judgment in forecasting. *International Journal of Forecasting*, 12, 1-8.
- YARANDI, S. S., NASSERI-MOGHADDAM, S., MOSTAJABI, P. & MALEKZADEH, R. 2010. Overlapping gastroesophageal reflux disease and irritable bowel syndrome: Increased dysfunctional symptoms. *World Journal of Gastroenterology*, 16, 1232-1238.
- ZIGLIO, E. 1996. The Delphi method and its contribution to decision making. *Gazing into the Oracle: the Delphi method and its application to social policy and public health*, 3 - 33.
- ZIMMERMANN, M., DARKOW, I.-L. & VON DER GRACHT, H. A. 2012. Integrating Delphi and participatory backcasting in pursuit of trustworthiness — The case of electric mobility in Germany. *Technological Forecasting and Social Change*, 79, 1605-1621.

Appendix A: Pilot data summaries

• Pilot Round (trial table)

Statement	Frequency and (%)					Agreement	
	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Median	Range
1. A biopsychosocial understanding of FGID							
a. FGID may represent an accumulation of biological, psychological, social and environmental contributing factors	1(33)	2 (67)				4	1
b. FGIDs can be associated with long-term comorbidity such as depression and anxiety	1 (33)	2 (67)				4	1
c. Comorbidity should be seriously considered with FGID patients	3 (100)					5	0
d. Comorbid psychiatric conditions increase the likelihood of health seeking behaviour in FGID patients		1 (33)	2 (67)			3	1
e. Comorbid psychiatric disorders negatively affect emotional stress through visceral symptoms		2 (67)	1 (33)			4	1
f. The presence of additional disease can complicate or interfere with treatment	1 (33)	2 (67)				4	1
g. Awareness of possible comorbidities may modify the assessment, diagnosis and subsequent management of FGID	1 (33)	2 (66)				4	1
Contributors to Functional Gastrointestinal Disorders.							
2. Physical Origin							
a. Previous truncal surgery			3 (100)			3	0
b. Previous visceral disease		2 (67)	1 (33)			4	1
c. Previous neuropathic disease			3 (100)			3	0

d. Previous physical trauma			2 (67)	1 (33)		3	1
e. Previous myofascial dysfunction		1 (33)	2 (67)			3	1
f. Ongoing myofascial dysfunction		1 (33)	2(67)			3	1
g. Brain injury / condition (effect on CNS - emotion, cognition, personality).		1 (33)	2 (67)			3	1
3. Psychological Origin							
a. Adverse early life events (eg. maternal separation)		1 (33)	2 (67)			3	1
b. Effects of life stress (autonomic dysregulation and susceptibility to changes in GI physiology)		3 (100)				4	0
c. Abuse (physical, psychological, sexual)	1 (33)	1 (33)	1 (33)			4	2
d. Psychiatric disorders (effect on FGID experience and behaviour)		3 (100)				4	0
e. Personality trait (effect on FGID experience and behaviour)		2 (67)	1 (33)			4	1
f. Dietary habit (regularity, nutritional value and side effects)			3 (100)			3	0
4. Gender differences							
a. Perception of pain		2 (67)	1 (33)			4	1
b. Sex hormone effect on GI sensitivity, function and transit time		1 (33)	2 (67)			3	1
c. Sex hormone effect on nociceptive processing		1 (33)	2 (67)			3	1
d. Central nervous system processing of visceral stimuli	1 (33)	1 (33)	1 (33)			4	2
e. Cultural values and beliefs		2 (67)	1 (33)			4	1
f. Socioeconomic status		1 (33)	1 (33)	1 (33)		3	2
g. Increased prevalence of FGID in the female population	1 (33)	2 (67)				4	1
5. Impact of daily living on symptoms.							

a. Reduced health-related quality of life	1 (33)	2 (67)				4	1
b. Increased healthcare seeking behaviour	1 (33)	2 (67)				4	1
c. Increased symptom vigilance	1 (33)	1 (33)	1 (33)			4	2
d. Effect of current medication		2 (67)	1 (33)			4	1
e. Altered daily toilet habits		3 (100)				4	0
f. Altered daily occupational functioning		2 (67)	1 (33)			4	1
g. Altered daily family functioning		2 (67)	1 (33)			4	1
6. Genetic Polymorphism							
a. Contribution of genetic factors to the mediation of psychological disorders		2 (67)	1 (33)			4	1
b. Contribution of genetic factors to the mediation of gastrointestinal sensory and motor function		2 (67)	1 (33)			4	1
c. Polymorphism in genes that modulate immune and/or neuro-immune functions may contribute to the onset of symptoms in the presence of other exogenous stressors		2 (67)	1 (33)			4	1
The Therapeutic Relationship							
7. Clinicians working with FGID patients need the following qualities and experience							
a. Scientific competence	1 (33)	1 (33)	1 (33)			4	2
b. Employment of scientific objectivity (evidence-based assessment and management procedures)		3 (100)				4	0
c. Humanistic concern (empathy on the role of the patient as an individual)	2 (67)	1 (33)				5	1
d. Awareness of ethical issues (social, economic, legal and cultural)		3 (100)				4	0
e. Awareness that symptoms are "real"	1 (33)	2 (67)				4	1
f. Engagement with the patient and their family in selecting and monitoring a given treatment plan		3 (100)				4	0

g. Continuity of consultation.		3 (100)				4	0
8. The patient-practitioner relationship							
a. Patient-practitioner relationships (patterns of communication)	1 (33)	2 (67)				4	1
b. Open-ended questions (establishment of symptom content and emotional context)		3 (100)				4	0
c. Non-directive interview (inviting the patient to talk about their own experienced problems, concerns etc)	1 (33)	1 (33)		1 (33)		4	3
d. Role of family (support, influence and corroboration)		3 (100)				4	0
e. Developmental history (significant events in the patient's life)	1 (33)	2 (67)				4	1
f. Family illness history		3 (100)				4	0
g. Symptom evaluation	2 (67)	1 (33)				5	1
h. Integration of medical and social history	2 (67)	1 (33)				5	1
i. Past clinical records		3 (100)				4	0
j. Past clinicians		2 967)	1 (33)			4	1
k. Patient expectation	1 (33)	2 (67)				4	1
9. Cultural Factors							
The patient's present geographical location			2 (67)		1 (33)	3	2
Origin of the patient's family			2 (67)	1 (33)		3	1
Attitudes and beliefs toward pain (ingrained habits such as stoicism and expectation of sympathy)		2 (67)	1 (33)			4	1
Interpretation of pain (thresholds and description of character)	1 (33)	2(67)				4	1
Expression of reporting symptoms (verbal style and level of emotion)		3 (100)				4	0

Attitudes toward health practitioners		1 (33)	2 (67)			3	1
Favoured type of practitioner (medical, complementary, alternative)		2 (67)		1 (33)		4	2
Favoured type of clinical approach (e.g., pharmacological, physical, psychological)		2 (67)		1 (33)		4	2
Socio-economic status (family & job security, lack of medical insurance etc.)		1 (33)	1 (33)	1 (33)		3	2
Educational status		1 (33)	1 (33)	1 (33)		3	2
Spiritual and religious attitudes			3 (100)			3	0
Patient gender		3 (100)				4	0
Areas for Consideration and Possible Measurement							
10. Abdominal Symptoms							
a. Abdominal pain	2 (67)	1 (33)				5	1
b. Altered stool character	2 (67)	1 (33)				5	1
c. Functional abdominal bloating		3 (100)				4	0
d. Increased flatulence		2 (67)	1 (33)			4	1
e. Functional heartburn		2 (67)	1 (33)			4	1
f. Aerophagia		2 (67)	1 (33)			4	1
g. Functional nausea / vomiting		2 (67)	1 (33)			4	1
h. Functional constipation		3 (100)				4	0
i. Functional diarrhoea		3 (100)				4	0
j. Functional faecal incontinence		2 (67)	1 (33)			4	1

k. Changes in appetite		1 (33)	2 (67)			4	1
11. Patient Description of Abdominal Symptoms							
Location	1 (33)	2 (67)				4	1
Duration	1 (33)	2 (67)				4	1
Progression	1 (33)	2 (67)				4	1
Severity	1 (33)	1 (33)		1 (33)		4	3
Character	1 (33)	2 (67)				4	1
Frequency	1 (33)	2 (67)				4	1
Exacerbating factors	2 (67)	1 (33)				5	1
Relieving factors	1 (33)	2 (67)				4	1
Associated symptoms (bodily regions)		3 (100)				4	0
Interference of symptoms on social and household functioning		3 (100)				4	0
Patient control of symptoms (e.g., medication, physical therapy, distraction techniques, relaxation)	1 (33)	2 (67)				4	1
12. Emotional Functioning							
a. Anxiety (difficulty in controlling worry, muscle tension, altered bowel habit and sleep disturbance)	2 (67)	1 (33)				5	1
b. Depression and sadness (subjective report or observed by others)		3 (100)				4	0
c. Pain-related fear avoidance beliefs (avoidance of situations associated with visceral pain)		3 (100)				4	0
d. Sense of hopelessness (none or minimal expectation for symptom improvement)		2 (67)	1 (33)			4	1
e. Diminished interest or pleasure in activities		3 (100)				4	0

f. Anger and irritability (hostility)		2 (67)	1 (33)			4	1
g. Emotive coping strategies (The patient's adjustments to their symptoms)		3 (100)				4	0
h. Significant increase or decrease in bodyweight (eg. a change of more than 5% in a month)		1 (33)	2 (67)			3	1
i. Catastrophic misinterpretation of pain (fear of movement and recurrent pain resulting in hypervigilance)		2 (67)	1 (33)			4	1
j. Fatigue - inertia and inability to concentrate		2 (67)	1 (33)			4	1
13. Social Functioning							
a. Occupational functioning (change in job status due to physical deconditioning, psychosocial dysfunction and / or any resultant interpersonal conflict)		2 (67)	1 (33)			4	1
b. Familial functioning (altered interpersonal relationships due to conflict and lack of cohesion between the patient and significant others)		2 (67)	1 (33)			4	1
c. Avoidant behaviour (fear of being away from toilet facilities)		3 (100)				4	0
d. Reliance on medication and / or significant others (instead of the patient managing their own pain)		2 (67)	1 (33)			4	1
14. Physical Functioning							
a. Functional impairment (assistance with one or more personal care tasks)		3 (100)				4	0
b. Kinesiophobia (excessive, irrational fear of physical movement due to the feeling of vulnerability to painful injury or re-injury)		2 (67)	1 (33)			4	1
c. Symptom associated disability (disproportionate restriction of physical daily activities for observable dysfunction)		2 (67)	1 (33)			4	1
d. Previous physical ability		1 (33)	2 (67)			3	1
e. The patient's past attitudes towards physical activity		1 (33)	2 (67)			3	1
f. The patient's present attitudes towards physical activity		3 (100)				4	0
g. Old Age (eg. decreased visceral sensitivity, impaired communication)		2 (67)	1 (33)			4	1
Multiaxial Assessment and Formulation (Relevance to FGID)							

15. Pros							
a. It expands from single-item diagnosis to several axes that provide additional "domains" of information of high clinical value	1 (33)	2 (67)				4	1
b. Reduces reliance on clinical judgement for diagnosis and therefore reduces clinical subjectivity		1 (33)	2 (67)			3	1
c. Allows users to systematically approach both physiological and psychological components of FGID	1 (33)	2 (67)				4	1
d. It can be applied in conjunction with laboratory testing (e.g., functional MRI), well validated psychological tests and self-reporting criteria		3 (100)				4	0
e. Conveys large amounts of information related to disorders in the form of clinical shorthand that are otherwise difficult to communicate	1 (33)	2 (67)				4	1
f. Promotes structured clinical dialogue based on standardised criteria, compared to self-reporting questionnaires or loosely structured interviews	1 (33)	2 (67)				4	1
g. Allows for quantitative rating of a person's mood, cognition and behaviour, which may create a profile of functioning		3 (100)				4	0
h. Should encompass not only multiaxial evaluation but personal idiography that reflect their individual strengths and weaknesses		2 (67)	1 (33)			4	1
i. Can often validate the patient's own experience by informing them that others have similar experiences		3 (100)				4	0
j.							
16. Cons							
a. Is only applicable to psychiatric diagnosis and therefore not appropriate for FGID				2 (67)	1 (33)	2	1
b. Is a time consuming exercise		1 (33)		2 (67)		2	2
c. The term "diagnosis" implies a distinct illness that is therefore not relevant in many cases of FGID			1 933)	1 (33)	1 (33)	2	2
d. Multiaxial diagnostic criteria often lack clear distinctions between normal and abnormal & therefore do not avoid diagnostic consideration of ordinary problems of daily living			2 (67)	1 (33)		3	1
e. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability			1 (33)	2 (67)		2	1
f. Tends to promote "all or nothing" diagnoses when considering an individual's problem (i.e., how many symptoms from a list are required before action is taken)				3 (100)		2	0
17. Multiaxial assessment systems: future development							

a. Multiaxial diagnostic formulation will benefit the management of FGID		3 (100)				4	0
b. Multiaxial diagnostic formulation may benefit the management of FGID depending on results from further clinical research and consensus		1 (33)		2 (67)		2	0
c. Multiaxial diagnostic formulation will only benefit the management of FGID when further clinical "gold standard" diagnostic testing becomes available				2 (67)	1 (33)	2	1
d. Multiaxial diagnostic formulation will never benefit the management of FGID				2 (67)	1 (33)	2	1

• **Individual data sheet example (Likert scores and comments)**

Statements	Response
5. A biopsychosocial understanding of FGID	
a. FGID may represent an accumulation of biological, psychological, social and environmental contributing factors.	Strongly agree
b. FGID is a single item diagnosis with no relevance to psychological disorders	Strongly disagree
c. FGIDs can be associated with long-term comorbidity such as depression and anxiety	Strongly agree
d. Comorbid psychiatric conditions increase the likelihood of health seeking behaviour in FGID patients	Strongly agree
e. Psychiatric disorders have minimal impact on FGID	Strongly disagree
f. The presence of additional disease can complicate or interfere with treatment	Strongly agree
g. Awareness of possible comorbidities may modify the assessment, diagnosis and subsequent management of FGID	Strongly agree
h. Other. (optional) - (Comment - <i>The section was a bit redundant. FGID represent somatoform manifestations of psychiatric problems. However, they fit into a wide spectrum of problems integrating biological, psychosocial and environmental components. This spectrum includes 'biologically' defined illness (e.g., Crohn's disease), where we learn yet one more time</i>	Strongly agree

<i>that biological markers of disease do not sufficiently explain clinical manifestations)</i>	
Contributors to Functional Gastrointestinal Disorders	
6. Physical Origin	
a. Previous truncal surgery. (Comment - <i>surgery is more often a consequence of medical intervention in response to symptoms</i>)	Neutral
b. Previous visceral conditions (Comment - <i>relevant for a subgroup (e.g., post-infectious IBS), but not major cause</i>)	Agree
c. Ongoing visceral conditions (Comment - <i>ongoing disease and experience of/concerns related to disease can trigger relevant disease mechanisms (e.g., hypervigilance, sensitization); IBD may serve as example; inflammation, perhaps increase in intestinal permeability, can secondarily influence affect (nicely shown in animal experiments by Steve Collins and in human data on Crohn's)</i>)	Agree
d. Previous neuropathic conditions (Comment - <i>Minor role only</i>)	Agree
e. Previous physical trauma. (Comment - <i>Minor role (perhaps mediated via anxiety/PTSD)</i>)	Agree
f. Ongoing myofascial dysfunction. (Comment - <i>This is likely NOT causal in a sense of FM=>FGID, but rather both being linked to underlying central mechanisms</i>)	Agree
g. Brain injury / condition (effect on CNS - emotion, cognition, personality). (Comment - <i>not injury, but affect, stress tolerance...</i>)	Strongly agree
h. Other (optional)	No response
7. Psychological Origin	
a. Early life events (eg. maternal separation) (Comment - <i>overlap with the psychiatric confounders</i>)	Agree
b. Effects of life stress resulting in autonomic dysregulation and susceptibility to changes in GI physiology	Agree
c. Abuse (physical, psychological, sexual)	Agree

d. Psychiatric disorders (effect on FGID experience and behaviour)	Strongly agree
e. Personality trait (effect on FGID experience and behaviour)	Strongly agree
f. Dietary habit (regularity, nutritional value and side effects). (Comment - <i>This is a HUGE topic and goes way beyond the conventional "it's all about fibre."</i>)	Strongly agree
g. Other. (optional)	No response
8. Gender Differences	
a. Perception of pain. (Comment - <i>Quite complex as the gender specific difference depend on modality and testing approach</i>)	Agree
b. Sex hormone effect on GI sensitivity, function and motility	Agree
c. Sex hormone effect on nociceptive processing	Neutral
d. Central nervous system processing of visceral stimuli	Strongly agree
e. Cultural values and beliefs	Agree
f. Socioeconomic status	Neutral
g. Increased prevalence of FGID in the female population. (Comment - <i>in western populations</i>)	Strongly agree
h. Other. (optional)	No response
i. Impact of Symptoms on Daily Life	
a. Reduced health-related quality of life	Strongly agree
b. Increased healthcare seeking behaviour	Strongly agree
c. Help from significant others	Agree
d. Increased symptom vigilance	Strongly agree

e. Effect of current medication. (Comment - <i>while pharmacotherapy is common, the cost-benefit ratio is less impressive</i>)	Agree
f.. Altered daily toilet habits (Comment - <i>unclear what you mean, as it could refer to encopresis, patients with faecal incontinence or functional dyspepsia (the answer might be strikingly different)</i>)	Neutral
g. Altered daily occupational functioning	Agree
H. Altered daily family functioning	Agree
i. Other. (optional)	No response
10. Genetic Polymorphism	
a. Contribution of genetic factors to the mediation of psychological disorders. (Comment - <i>there will be genetic contribution to factors from altered motility to sensation or even healthcare-seeking behaviour. However, this will be difficult to identify, as the ROME criteria as common standard are atheoretical and do not translate into endophenotypes that can even be examined in some meaningful way; with the limited ability to identify or use phenotypes, linking this information to genotypes will be an exercise in futility (see SERT in IBS or also anxiety/depression). As is true for complex traits, there will not be a single candidate gene; many genes and the environment will interact to lead to a phenotype. The genetic variance does matter (see twin studies)</i>)	Agree
b. Contribution of genetic factors to the mediation of gastrointestinal sensory and motor function	Agree
c. Contribution of genetic factors to pain modulatory pathways	Agree
d. Polymorphism in genes that modulate immune and/or neuro-immune functions may contribute to the onset of symptoms in the presence of other exogenous stressors. (Comment - <i>suggested with some emerging data</i>)	Agree
e. Other. (optional)	No response
The Therapeutic Relationship	
11. Clinicians working with FGID patients need the following qualities and experience	
a. Scientific competence. (evidence-based assessment and management procedures) (Comment - <i>the weakness of evidence-based medicine is its basis on an atheoretical approach in a complex phenotype</i>)	Strongly agree

b. The number of years working with FGID patients	Neutral
c. The ability to work in a multi-disciplinary team	Strongly agree
d. Humanistic concern (empathy on the role of the patient as an individual). (Comment - <i>see Tony Lembo's article in the BMJ</i>)	Strongly agree
e. Awareness of ethical issues (social, economic, legal and cultural)	Strongly agree
f. Awareness that symptoms are "real"	Strongly agree
g. Engagement with the patient and their family in selecting and monitoring a given treatment plan	Strongly agree
h. Continuity of consultation	Strongly agree
i. Other. (optional)	No response
12. The patient-practitioner relationship	
a. Patient-practitioner relationships (patterns of communication)	Strongly agree
b. Non-directive interview (inviting the patient to talk about their own experienced problems, concerns etc.)	Strongly agree
c. Structured interview using direct questions to elicit information about the patient's presentation. (Comment - <i>B&C are complementary</i>)	Strongly agree
d. Role of family (support, influence and corroboration)	Agree
e. Developmental history (significant events in the patient's life)	Agree
f. Consideration of family illness history	Strongly agree
g. Integration of medical and social history	Agree
h. Utilisation of past clinical records	Agree
i. Past clinicians	Neutral

j. Patient expectation	Agree
k. Other. (optional)	No response
13. Cultural Factors	
a. The patient's cultural background and attitudes toward abdominal symptoms (ingrained habits such as stoicism and expectation of sympathy)	Agree
b. Interpretation of abdominal symptoms (thresholds and description of character)	Strongly agree
c. Expression of reporting symptoms (verbal style and level of emotion)	Agree
d. Attitudes toward health practitioners	Agree
e. Favoured type of practitioner (medical, complementary, alternative)	Agree
f. Favoured type of clinical approach (e.g., pharmacological, physical, psychological)	Strongly agree
g. Socio-economic status (family & job security, lack of medical insurance etc.)	Neutral
h. Educational status	Neutral
i. Spiritual and religious attitudes	Agree
j. Patient gender	Neutral
k. Other. (optional)	No response
Areas for Consideration and Possible Measurement	
14. Abdominal Symptoms	
a. Abdominal pain	Strongly agree
b. Altered stool character	Strongly agree

c. Functional abdominal bloating	Strongly agree
d. Increased flatulence	Strongly agree
e. Functional heartburn	Strongly agree
f. Aerophagia	Strongly agree
g. Functional nausea / vomiting	Strongly agree
h. Functional constipation	Strongly agree
i. Functional diarrhoea	Strongly agree
j. Functional faecal incontinence	Strongly agree
k. Changes in appetite	Strongly agree
i. Other. (optional) - (Comment - <i>all of them are relevant but not all the same way and all in the same patient; if patients have more than one symptom I typically ask them to give me a sense of priority for their different problems</i>)	Agree
15. Patient Description of Abdominal Symptoms	
a. Location	Strongly agree
b. Duration	Strongly agree
c. Progression	Strongly agree
d. Severity	Strongly agree
e. Character	Strongly agree
f. Frequency	Strongly agree

g. Exacerbating factors	Strongly agree
h. Relieving factors	Strongly agree
i. Associated symptoms (bodily regions)	Strongly agree
j. Interference of symptoms on social and household functioning	Strongly agree
k. Patient control of symptoms (e.g., medication, physical therapy, distraction techniques, relaxation)	Strongly agree
l. Patient's interpretation of abdominal symptoms (Comment - <i>ultimately, we are dealing with an illness that in some significantly impairs quality of life with relatively little in terms of objective correlate; anxiety is a driver; cognitive appraisal of symptoms is related to anxiety; the effect of CBT is largely due to focusing on this point</i>)	Strongly agree
m. Other. (optional)	No response
16. Emotional Functioning	
A. Anxiety (difficulty in controlling worry, muscle tension, altered bowel habit and sleep disturbance)	Strongly agree
b. Depression and sadness (subjective report or observed by others)	Strongly agree
c. Symptom-related fear-avoidance beliefs (avoidance of situations associated with abdominal symptoms)	Strongly agree
d. Sense of hopelessness (none or minimal expectation for symptom improvement)	Strongly agree
e. Diminished interest or pleasure in activities	Strongly agree
f. Anger and irritability (hostility)	Strongly agree
g. Emotive coping strategies (The patient's adjustments to their symptoms, i.e., diet and toilet habit)	Strongly agree
h. Catastrophic misinterpretation of FGID related pain (fear of movement resulting in hypervigilance)	Strongly agree
i. Fatigue - inertia and inability to concentrate	Strongly agree
j. Other. (optional)	No response

17. Social Functioning	
a. Occupational functioning (change in job status due to physical deconditioning, psychosocial dysfunction and / or any resultant interpersonal conflict)	Strongly agree
b. Familial functioning (altered interpersonal relationships due to conflict and lack of cohesion between the patient and significant others)	Strongly agree
c. Avoidant behaviour (fear of being away from toilet facilities)	Strongly agree
d. Reliance on medication and / or significant others (instead of the patient managing their own symptoms)	Strongly agree
e. Other. (optional)	No response
18. Physical Functioning	
a. Functional impairment (assistance with one or more personal care tasks)	Strongly agree
b. Kinesiophobia (excessive, irrational fear of physical movement due to the feeling of vulnerability to painful injury or re-injury)	Strongly agree
c. Symptom associated disability (disproportionate restriction of physical daily activities for observable dysfunction)	Strongly agree
d. Previous physical ability	Strongly agree
e. The patient's past attitudes towards physical activity	Strongly agree
f. The patient's present attitudes towards physical activity	Strongly agree
g. Old Age (eg. decreased visceral sensitivity, impaired communication)	Agree
h. Other. (optional)	Neutral
Multiaxial Assessment and Formulation (Relevance to FGID)	

19. Pros	
a. It expands from single-item diagnosis to several axes that provide additional "domains" of information of high clinical value. (Comment - <i>While I do this integrated into clinical practice, it is obviously not formalized such as in psychiatry; I anticipate at least a potentially better and more comprehensive description of disease in its context; however, I do not know whether we will get closer to the 'endophenotypes' that may provide mechanistic understanding/treatment in some subgroups</i>)	Neutral
b. Reduces reliance on clinical judgement for diagnosis and therefore reduces clinical subjectivity	Neutral
c. Allows users to systematically approach both physiological and psychological components of FGID	Agree
d. It can be applied in conjunction with laboratory testing (e.g., functional MRI), well validated psychological tests and self-reporting criteria	Agree
e. Conveys large amounts of information related to disorders in the form of clinical shorthand that is otherwise difficult to communicate	Neutral
f. Promotes structured clinical dialogue based on standardised criteria, compared to self-reporting questionnaires or loosely structured interviews	Neutral
g. Allows for quantitative rating of a person's mood, cognition and behaviour, which may create a profile of functioning	Agree
h. Should encompass not only multiaxial evaluation but personal idiography that reflect their individual strengths and weaknesses	Neutral
i. Can often validate the patient's own experience by informing them that others have similar experiences	Agree
j. Other. (optional)	No response
20. Cons	
a. Is only applicable to psychiatric diagnosis and therefore not appropriate for FGID	Disagree
b. Is a time consuming exercise and is of little value to the management of FGID	Neutral

c. The term "diagnosis" implies a distinct illness that is therefore not relevant in many cases of FGID. (Comment - <i>Psychiatry obviously would be close to non-existent with such an attitude; however, by focusing on syndromic definitions only and perhaps expanding them, we stay in the atheoretical domain that does not look for cause/mechanism</i>)	Disagree
d. Multiaxial diagnostic criteria often lack clear distinctions between normal and abnormal & therefore do not avoid diagnostic consideration of ordinary problems of daily living. (Comment - <i>that statement is true for any phenomenon that is not truly dichotomous</i>)	Disagree
e. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability	Disagree
f. Tends to promote "all or nothing" diagnoses when considering an individual's problem (i.e., how many symptoms from a list are required before action is taken)	Disagree
g. Other. (optional)	No response
21. Future Research: Identifying Risk Factors and the Replication of Findings (brain-gut-axis).	
a. Incorporation of well-replicated neuroscientific data providing bases for diagnosis. (Comment - <i>we have obviously learned quite a lot with functional neuroimaging; however, this approach is based on the assumption that FGID is largely disorders of sensation/central processing; it leaves little room for the explanations of the 'past' (i.e., disturbed motility is the cause) or some of the emerging views related to unique subgroups (i.e., subclinical inflammation in post-infectious disease)</i>)	Agree
b. Incorporating genetic information in relation to psychological & visceral conditions. (Comment - <i>in even fairly well defined illnesses (e.g., Crohn's disease), GWAS have demonstrated relatively little in my eyes; while we now know about several predisposing genes, genetic mechanisms generally account for a small fraction of the variance only. We may eventually learn about convergent pathways and thus be able to infer/target mechanisms. The problem in FGID is different from IBD or comparable illnesses, as phenotypes are less distinct, often overlap with other GI or non-GI disorder and/or shift over time; thus, genetic studies can only work if we have true phenotypic definition of more homogeneous and hopefully stable subgroups; the recent studies using populations exposed to a waterborne illness show some promise in this context (in this unique case, the only caveat is that the genetics may uncover risk factors for more severe bacterial GI infections as these correlate with higher risk of PI-IBS)</i>)	Agree

c. Incorporating pharmacogenomic research relating to therapies targeting of specific diseases (<i>clearly promising, especially if we had medications that are more promising</i>)	Agree
d. Other	No response
22. Multiaxial assessment systems: future development	
a. Multiaxial diagnostic formulation will benefit the management of FGID	Agree
b. Multiaxial diagnostic formulation may benefit the management of FGID depending on results from further clinical research and consensus	Agree
c. Multiaxial diagnostic formulation will only benefit the management of FGID when further clinical "gold standard" diagnostic testing becomes available. (Comment - <i>before 'gold standard testing' we need mechanistic understanding, assuming that you refer to biomarker when you mention test; the Rome criteria (or comparable approaches) are not tests</i>)	Disagree
d. Multiaxial diagnostic formulation will never benefit the management of FGID	Disagree
e. Other. (optional)	No response
23. What recommendations would you make for future development in any of the areas covered by this survey? (Optional)	
24. Are there any issues that are important to you in your present role that have not been explored in the questions above? (Optional) (Comment - <i>I completed the survey with some breaks in between and may thus be biased by some of the later points, there is an inherent bias in the survey that reflects much of my own thinking but may sell FGID short. It may perpetuate our focus on phenomenology and drop the attempt to understand physiology, perhaps with the exception of central sensitization/hypervigilance. Several areas that are generically approached (if at all) by clinicians were left out (e.g., fiber), emerging areas like probiotics or antibiotics to influence colonic (or small bowel) flora, and a central component of gut function (food intake, absorption) have not been touched at all</i>).	

Appendix B: Delphi round comment sheets

• Round 1

5. A biopsychosocial understanding of FGID	
a. FGID may represent an accumulation of biological, psychological, social and environmental contributing factors	<ul style="list-style-type: none"> Domestic abuse It may
b. FGID is a single item diagnosis with no relevance to psychological disorders	<ul style="list-style-type: none"> For some individuals, this may be the case It may be, therefore I am open to that possibility
c. FGIDs can be associated with long-term comorbidity such as depression and anxiety	
d. Comorbid psychiatric conditions increase the likelihood of health seeking behaviour in FGID patients	
e. Psychiatric disorders have minimal impact on FGID	<ul style="list-style-type: none"> Again, some with what satisfies criteria for an FGID may have no psychosocial features. Context is also important: community vs clinic vs tertiary referral centre
f. The presence of additional disease can complicate or interfere with treatment.	<ul style="list-style-type: none"> Don't really understand the question, no mentioning of FGID, do you mean "...with treatment of FGID"? This statement is not clear to me. An additional disease is the same as comorbidity. This one is probably not correctly worded, needs to be rephrased We mean comorbidities that do not exclude FGIDs
g. Awareness of possible comorbidities may modify the assessment, diagnosis and subsequent management of FGID	
h. Other. (optional)	<ul style="list-style-type: none"> Doctors should be trained in the spirit of the BPS model However, far from ALL with FGID have psychosocial comorbidity Pain has been considered to be the primary symptom here. This may not be correct. Asian patients in particular may not have pain always but may have altered stool form, frequency and bloating (Gwee KA, et. al. J Gastroenterol Hepatol 2009; 24:1601-7; Gwee KA et. al. J Gastroenterol Hepatol 2010; 25:1189-1205) Strong family history of chronic pain disorders and history early child abuse The section was a bit redundant. FGID often represent somatoform manifestations of psychiatric problems. However, they fit into a wide spectrum of problems integrating biological, psychosocial and environmental components. This spectrum also includes 'biologically' defined illness (e.g., Crohn's disease), where we learn yet one more time that biological markers of disease do not sufficiently explain clinical manifestations

<ul style="list-style-type: none"> • These comments can apply to any chronic non-fatal illness and not just FGID
Contributors to Functional Gastrointestinal Disorders
6. Physical Origin
a. Previous truncal surgery <ul style="list-style-type: none"> • BUT by definition, this is not FGID • Define more clearly • Surgery is more often a consequence of medical intervention in response to symptoms • This is always important, FGID or not
b. Previous visceral conditions <ul style="list-style-type: none"> • BUT by definition, this is not FGID • Intestinal inflammation/post-infections • No idea what you mean • Relevant for a subgroup (e.g., post-infectious IBS), but not major cause. • This is somewhat non-specific. What is the meaning of previous visceral condition? For example, one can have pneumonia long ago. It may not be important. On the other hand, acute gastroenteritis in past may be important • What visceral conditions?
c. Ongoing visceral conditions <ul style="list-style-type: none"> • BUT by definition, this is not FGID • Including IBD and coeliac disease. "Diverticular disease and diverticulosis" whatever they are, also should be considered. • No idea what you mean • Ongoing disease and experience of/concerns related to disease can trigger relevant disease mechanisms (e.g., hypervigilance, sensitization); IBD may serve as example; inflammation, perhaps increase in intestinal permeability; can secondarily influence affect (nicely shown in animal experiments by Steve Collins and in human data on Crohn's) • Such as Inflammatory Bowel disease
d. Previous physical trauma <ul style="list-style-type: none"> • Association not in my area • Especially during childhood • This is also non-specific. For example, trivial physical trauma long ago may not be important
e. Ongoing myofascial dysfunction <ul style="list-style-type: none"> • This is likely NOT causal in a sense of FM=>FGID, but rather both being linked to underlying central mechanisms
f. Brain injury / condition (effect on CNS - emotion, cognition, personality) <ul style="list-style-type: none"> • Not injury, but affect, stress tolerance
g. Other. (optional) <ul style="list-style-type: none"> • Migraines, severe accidents

<ul style="list-style-type: none"> • Previous acute gastroenteritis may result in a condition called post-infectious IBS (Ghoshal UC et al. J Gastroenterol Hepatol 2010; 25:244-51) • Previous gastrointestinal infection • Previous gastrointestinal infection • Red flags such as anaemia, sedimentation rate, etc. that suggest an organic aetiology
7. Psychological Origin
a. Early life events (eg. maternal separation) <ul style="list-style-type: none"> • Overlap with the psychiatric confounders
b. Effects of life stress resulting in autonomic dysregulation and susceptibility to changes in GI physiology
c. Abuse (physical, psychological, sexual)
d. Psychiatric disorders (effect on FGID experience and behaviour)
e. Personality trait (effect on FGID experience and behaviour)
f. Dietary habit (regularity, nutritional value and side effects) <ul style="list-style-type: none"> • Agree but this is very poorly understood in IBS; yes, we can enquire but we are not skilled at assessing the data and acting upon it • This is a HUGE topic and goes way beyond the conventional "it's all about fiber"
g. Other. (optional) <ul style="list-style-type: none"> • Frequency of lactose malabsorption and celiac disease in the population, milk and milk product intake, wheat vs. rice intake, dietary fiber intake, fruit intake (for fructose intolerance), chilly and spice intake • Significant life events, (death in family, geographical move, divorce or separation) • There is no proof that these disorders are psychosocial IN ORIGIN
8. Gender Differences
a. Perception of pain <ul style="list-style-type: none"> • Higher perception in women • Quite complex as the gender specific difference depend on modality and testing approach
b. Sex hormone effect on GI sensitivity, function and motility <ul style="list-style-type: none"> • It has not been proven
c. Sex hormone effect on nociceptive processing <ul style="list-style-type: none"> • It has not been proven
d. Central nervous system processing of visceral stimuli

<ul style="list-style-type: none"> Here I refer to cognitive appraisal, not nociception
e. Cultural values and beliefs <ul style="list-style-type: none"> It is in the process of research so I don't think at this time we have an answer but I would expect so Very important. For example, in some population, females consult physicians more often. In contrast, in other population, female visit doctors less often. In some population, it is taught from the childhood that one must pass one to two stools daily for good health, in other populations, three stools per week is considered normal
f. Socioeconomic status <ul style="list-style-type: none"> Again, I am not sure we have enough info on this one
g. Increased prevalence of FGID in the female population <ul style="list-style-type: none"> In western populations The female preponderance of IBS is not uniformly seen all over the world. In some population, frequency of IBS among male population is as high as in female population (Gwee KA, et. al. J Gastroenterol Hepatol 2009; 24:1601-7)
h. Other. (optional) <ul style="list-style-type: none"> Differences in response to the different pharmacological treatments that have been developed
9. Impact of Symptoms on Daily Life
a. Reduced health-related quality of life <ul style="list-style-type: none"> Again, depends on the context
b. Increased healthcare seeking behaviour
c. Help from significant others <ul style="list-style-type: none"> This is not clear
d. Increased symptom vigilance
e. Effect of current medication <ul style="list-style-type: none"> Not sure what you mean here While pharmacotherapy is common, the cost-benefit ratio is less impressive
f. Altered daily toilet habits <ul style="list-style-type: none"> Consider daily bowel habit of the local population before subtyping IBS into constipation and diarrhoea types based on stool frequency criteria as daily bowel habit differ from population to population. For example, In India, people pass one to two stools per day (Ghoshal UC, et. al. Indian J Gastroenterol 2008; 27:22-28) Depend on the type of FD

<ul style="list-style-type: none"> Unclear what you mean, as it could refer to encopresis, patients with faecal incontinence or functional dyspepsia (the answer might be strikingly different)
g. Altered daily occupational functioning
h. Altered daily family functioning
i. Other (optional) <ul style="list-style-type: none"> For children and adolescents, school attendance is a key factor Sleep
10. Genetic Polymorphism
a. Contribution of genetic factors to the mediation of psychological disorders <ul style="list-style-type: none"> Each of these may well be relevant but currently the data is either inadequate or conflicting There will be genetic contribution to factors from altered motility to sensation or even healthcare-seeking behaviour. However, this will be difficult to identify, as the ROME criteria as common standard are atheoretical and do not translate into endophenotypes that can even be examined in some meaningful way; with the limited ability to identify or use phenotypes, linking this information to genotypes will be an exercise in futility (see SERT in IBS or also anxiety/depression). As is true for complex traits, there will not be a single candidate gene; many genes and the environment will interact to lead to a phenotype. The genetic variance does matter (see twin studies) We need more research on this factor
b. Contribution of genetic factors to the mediation of gastrointestinal sensory and motor function <ul style="list-style-type: none"> Don't know Not enough info yet on this one but I expect so
c. Contribution of genetic factors to pain modulatory pathways <ul style="list-style-type: none"> The same as the last one
d. Polymorphism in genes that modulate immune and/or neuro-immune functions may contribute to the onset of symptoms in the presence of other exogenous stressors <ul style="list-style-type: none"> Suggested with some emerging data Though genetic factors are potentially important, data on different population are scanty. Hence, more data on these are needed
e. Other. (optional) <ul style="list-style-type: none"> Contribution in the response to different pharmacological treatments (pharmacogenetics) Epigenetics is likely much more relevant and was ignored in this section. The epigenome is an area of convergence for several fields

<ul style="list-style-type: none"> • None proven in FGID
The Therapeutic Relationship
11. Clinicians working with FGID patients need the following qualities and experience
a. Scientific competence. I evidence-based assessment and management procedures) <ul style="list-style-type: none"> • Always, for all patients and diagnoses • The weakness of evidence-based medicine is its basis on an atheoretical approach in a complex phenotype
b. The number of years working with FGID patients <ul style="list-style-type: none"> • At least 1 year of training focused upon this population
c. The ability to work in a multi-disciplinary team <ul style="list-style-type: none"> • If one had one! • Yep
d. Humanistic concern (empathy on the role of the patient as an individual) <ul style="list-style-type: none"> • Always, for all patients and diagnoses • See Tony Lembo's article in the BMJ
e. Awareness of ethical issues (social, economic, legal and cultural)
f. Awareness that symptoms are "real" <ul style="list-style-type: none"> • Believe in what your patient tells you • VITAL
g. Engagement with the patient and their family in selecting and monitoring a given treatment plan
h. Continuity of consultation <ul style="list-style-type: none"> • Again vital and undervalued • Not always by you in person, but by some on the "team"
i. Other. (optional) <ul style="list-style-type: none"> • The ability to listen as if you believe it is all possible because it is indeed the patient's reality and to pick out the pieces where you can look for cause and effect and/or see psychological as well as physiologic connection and the ability to translate that into an idiom that the patient understands to improve their understanding of how it all fits together • Thorough knowledge of socio-cultural issues of the population. Knowledge that some of the patients with symptom-based

<p>diagnosis of IBS may have more organic diseases such as small intestinal bacterial overgrowth</p> <ul style="list-style-type: none"> • Understanding of pathophysiology in functional bowel disorders. Good interview skills
12. The patient-practitioner relationship
<p>a. Patient-practitioner relationship (patterns of communication)</p> <ul style="list-style-type: none"> • In particular education and reassurance
<p>b. Non-directive interview (inviting the patient to talk about their own experienced problems, concerns and fears etc)</p> <ul style="list-style-type: none"> • It is important to focus on how the patient interprets symptoms, and not only on the symptoms themselves
<p>c. Structured interview using direct questions to elicit information about the patient's presentation</p> <ul style="list-style-type: none"> • Also • B&C are complementary • Both non-directive interview and structured interview are important, as without non-directive interview, lot of psychosocial information might not be obtained. Moreover, over-reliance on criteria for the diagnosis of IBS may lead to its under or over-diagnosis as has been shown in several studies showing that as the criteria was changed from Rome I to Rome II, frequency of IBS in the same population was found different • Even though a non-directive interview is preferable, some kind of structure is needed during the interview • sometimes appropriate after the patient has gotten thru their concerns to hone in
d. Role of family (support, influence and corroboration)
e. Developmental history (significant events in the patient's life)
f. Consideration of family illness history
g. Integration of medical and social history
h. Utilisation of past clinical records
<p>i. Past clinicians</p> <ul style="list-style-type: none"> • I don't understand what you want to ask with this one
j. Patient expectation
<p>k. Other (optional)</p> <ul style="list-style-type: none"> • Availability • History of past enteric infections/infestations
13. Cultural Factors

a. The patient's cultural background and attitudes toward abdominal symptoms (ingrained habits such as stoicism and expectation of sympathy) <ul style="list-style-type: none"> Although I agree, we still need more research on this one For example, in some population, there may not be an appropriate terminology in the local language to describe a symptom. For example, in Chinese language, there is no good terminology for heartburn
b. Interpretation of abdominal symptoms (thresholds and description of character)
c. Expression of reporting symptoms (verbal style and level of emotion)
d. Attitudes toward health practitioners
e. Favoured type of practitioner (medical, complementary, alternative) <ul style="list-style-type: none"> Important for many reasons, at least because many patients do not disclose the use of CAM to their conventional providers
f. Favoured type of clinical approach (e.g., pharmacological, physical, psychological)
g. Socio-economic status (family & job security, lack of medical insurance etc)
h. Educational status <ul style="list-style-type: none"> I agree but again, we still need more info to determine in education status makes a difference in the prevalence, symptom patterns of FGID. It may have an influence on the acceptance of the different types of treatments In some population, consultation behaviour is determined by education
i. Spiritual and religious attitudes <ul style="list-style-type: none"> We don't have any info on this factor and I have not seen a difference in my experience
j. Patient gender
k. Other. (optional) <ul style="list-style-type: none"> Views on homosexuality
Areas for Consideration and Possible Measurement.
14. Abdominal Symptoms
a. Abdominal pain <ul style="list-style-type: none"> Abdominal pain, however, may not be universal in all population (Gwee KA, et. al. J Gastroenterol Hepatol 2009; 24: 1601-7). Quality, nature (constant, intermittent), nocturnal awakening, what the pt is doing, eating, thinking when pain occurs food triggers
b. Altered stool character <ul style="list-style-type: none"> Use Bristol stool chart
c. Functional abdominal bloating

<ul style="list-style-type: none"> • Very important • Very important one, it may be more common than pain • When it occurs, as day goes on, post-prandial
d. Increased flatulence
<ul style="list-style-type: none"> • Odour
e. Functional heartburn
f. Aerophagia
g. Functional nausea / vomiting
<ul style="list-style-type: none"> • What is functional vomiting? Do you mean it's CNS in origin or they are making it up?
h. Functional constipation
<ul style="list-style-type: none"> • Stool frequency alone cannot be used to define constipation and diarrhoea as there is variation in frequency in different population. Use Bristol stool chart • Straining, complete evacuation
i. Functional diarrhoea
<ul style="list-style-type: none"> • I worry about diarrhoea as a sole presentation; often indicates an underlying organic problem
j. Functional faecal incontinence
<ul style="list-style-type: none"> • Again- what is this? • Faecal incontinence may mean more organic disease. Investigate including anorectal manometry
k. Changes in appetite
<ul style="list-style-type: none"> • Anorexia may mean more organic disease. Investigate!
l. Other (optional)
<ul style="list-style-type: none"> • All of them are relevant but not all the same way and all in the same patient; if patients have more than one symptom I typically ask them to give me a sense of priority for their different problems • Dysphagia • Epigastric pain / burning, early satiety, post-prandial fullness • Meal related symptoms common in IBS and other FGIDs • Obstructive defecation • Overlap with dyspepsia, GERD and IBS may be common

- Polyuria, dysuria
- Suggested alt phrasing "FGIDs may have a major..." and/or... just to open up for other possible impact scenarios/combinations

15. Patient Description of Abdominal Symptoms

a. Location

- Can be anywhere

b. Duration

- Variable

c. Progression

d. Severity

- Difficult to assess

e. Character

- Not very helpful
- This expression is not clear to me

f. Frequency

g. Exacerbating factors

h. Relieving factors

i. Associated symptoms (bodily regions)

- Ultimately, we are dealing with an illness that in some significantly impairs quality of life with relatively little in terms of objective correlate; anxiety is a driver; cognitive appraisal of symptoms is related to anxiety; the effect of CBT is largely due to focusing on this point

j. Interference of symptoms on social and household functioning

k. Patient control of symptoms (e.g., medication, physical therapy, distraction techniques, relaxation)

l. Patient's interpretation of abdominal symptoms

m. Other (optional)

- Inter-relationships between symptoms
- Previous experiences
- Qs strongly weighted towards agree side (ceiling effect)... possible with alternative/reversed wording/phrasing
-

16. Emotional Functioning
a. Anxiety (difficulty in controlling worry, muscle tension, altered bowel habit and sleep disturbance)
b. Depression and sadness (subjective report or observed by others)
c. Symptom-related fear-avoidance beliefs (avoidance of situations associated with abdominal symptoms) <ul style="list-style-type: none"> Food avoidance
d. Sense of hopelessness (none or minimal expectation for symptom improvement)
e. Diminished interest or pleasure in activities
f. Anger and irritability (hostility)
g. Emotive coping strategies (The patient's adjustments to their symptoms, i.e., diet and toilet habit)
h. Catastrophic misinterpretation of FGID related pain (fear of movement resulting in hypervigilance)
i. Fatigue - inertia and inability to concentrate
j. Other. (optional) <ul style="list-style-type: none"> Interference with school or work
17. Social Functioning
a. Occupational functioning (change in job status due to physical deconditioning, psychosocial dysfunction and / or any resultant interpersonal conflict)
b. Familial functioning (altered interpersonal relationships due to conflict and lack of cohesion between the patient and significant others)
c. Avoidant behaviour (fear of being away from toilet facilities)
d. Reliance on medication and / or significant others (instead of the patient managing their own symptoms) <ul style="list-style-type: none"> Not usually a major factor These are two very different aspects. I strongly agree with the reliance on medications but I am not so sure about the significant others. Anyway, I would separate this one onto 2 factors
e. Other. (optional) <ul style="list-style-type: none"> Qs weighted towards agree side (ceiling effect)... possible with alternative/reversed wording/phrasing?

18. Physical Functioning
a. Functional impairment (assistance with one or more personal care tasks)
b. Kinesiophobia (excessive, irrational fear of physical movement due to the feeling of vulnerability to painful injury or re-injury) <ul style="list-style-type: none"> I have never seen this in patients with FGID
c. Symptom associated disability (disproportionate restriction of physical daily activities for observable dysfunction)
d. Previous physical ability <ul style="list-style-type: none"> I am not sure I understand what you mean with this one
e. The patient's past attitudes towards physical activity
f. The patient's present attitudes towards physical activity
g. Old Age (eg. decreased visceral sensitivity, impaired communication) <ul style="list-style-type: none"> Should be considered as an alarm feature and rule out organic diseases
h. Other. (optional)
Multiaxial Assessment and Formulation (Relevance to FGID)
19. Pros
a. It expands from single-item diagnosis to several axes that provide additional "domains" of information of high clinical value <ul style="list-style-type: none"> One issue is age of patient, Patient reporting is unreliable in children/infants. Will this just apply to patients over a certain age? While I do this integrated into clinical practice, it is obviously not formalized such as in psychiatry; I anticipate at least a potentially better and more comprehensive description of disease in its context; however, I do not know whether we will get closer to the 'endophenotypes' that may provide mechanistic understanding/treatment in some subgroups
b. Reduces reliance on clinical judgement for diagnosis and therefore reduces clinical subjectivity
c. Allows users to systematically approach both physiological and psychological components of FGID <ul style="list-style-type: none"> To the extent that it is desirable for any chronic illness or pain syndrome
d. Conveys large amounts of information related to disorders in the form of clinical shorthand that are otherwise difficult to communicate
e. Promotes structured clinical dialogue based on standardised criteria, compared to self-reporting questionnaires or loosely structured interviews <ul style="list-style-type: none"> A fixed criterion has its own problem. Even those criteria are not well validated all over the world (Gwee KA, Ghoshal UC. Saudi J

Gastroenterol)
f. Allows for quantitative rating of a person's mood, cognition and behaviour, which may create a profile of functioning
g. Should encompass not only multiaxial evaluation but personal idiography that reflect their individual strengths and weaknesses
h. Can often validate the patient's own experience by informing them that others have similar experiences
i. Other. (optional)
20. Cons
a. Is only applicable to psychiatric diagnosis and therefore not appropriate for FGID
b. Is a time consuming exercise and is of little value to the management of FGID <ul style="list-style-type: none"> • 2 different questions: it is time consuming, but we do not know yet about its value • In light of the current standard of practice, this one assertion would be right, but we need to change the perception of this type of approach • It is time consuming. But it is needed • This will be a problem in some clinical settings
c. The term "diagnosis" implies a distinct illness that is therefore not relevant in many cases of FGID <ul style="list-style-type: none"> • Depends... e.g. "pathological anatomical diagnosis" vs "functional diagnosis"... I have seen both being used in the same contexts • In fact, FGID is a symptoms-based diagnosis without gold standard and many syndromes may mimic these symptom-based diagnosis • Psychiatry obviously would be close to non-existent with such an attitude; however, by focusing on syndromic definitions only and perhaps expanding them, we stay in the atheoretical domain that does not look for cause/mechanism
d. Multiaxial diagnostic criteria often lack clear distinctions between normal and abnormal & therefore do not avoid diagnostic consideration of ordinary problems of daily living <ul style="list-style-type: none"> • Need studies to evaluate • That statement is true for any phenomenon that is not truly dichotomous
e. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability <ul style="list-style-type: none"> • It is a matter of standardization, awareness and knowledge of this type of approach
f. Tends to promote "all or nothing" diagnoses when considering an individual's problem (i.e., how many symptoms from a list are required before action is taken) <ul style="list-style-type: none"> • Not at all, as it is multiaxial and on the other hand in disorders such as FGID in which we don't have a biological marker for diagnosis, this is the only type of diagnostic approach • Not sure

g. Other. (optional)
21. Future Research: Identifying Risk Factors and the Replication of Findings (brain-gut-axis)
a. Incorporation of well-replicated neuroscientific data providing bases for diagnosis <ul style="list-style-type: none"> • More studies are needed • Too vague • We have obviously learned quite a lot with functional neuroimaging; however, this approach is based on the assumption that FGID is largely disorders of sensation/central processing; it leaves little room for the explanations of the 'past' (i.e., disturbed motility is the cause) or some of the emerging views related to unique subgroups (i.e., subclinical inflammation in post-infectious disease)
b. Incorporating genetic information in relation to psychological & visceral conditions <ul style="list-style-type: none"> • And symptoms • Hope for the future with no clear direction at present • In even fairly well defined illnesses (e.g., Crohn's disease), GWAS have demonstrated relatively little in my eyes; while we now know about several predisposing genes, genetic mechanisms generally account for a small fraction of the variance only. We may eventually learn about convergent pathways and thus be able to infer/target mechanisms. The problem in FGID is different from IBD or comparable illnesses, as phenotypes are less distinct, often overlap with other GI or non-GI disorder and/or shift over time; thus, genetic studies can only work if we have true phenotypic definition of more homogeneous and hopefully stable subgroups; the recent studies using populations exposed to a waterborne illness show some promise in this context (in this unique case, the only caveat is that the genetics may uncover risk factors for more severe bacterial GI infections as these correlate with higher risk of PI-IBS) • More studies are needed
c. Incorporating pharmacogenomic research relating to therapies targeting of specific diseases <ul style="list-style-type: none"> • Clearly promising, especially if we had medications that are more promising • More studies are needed • Therapies don't exist; pathophysiology is largely unknown
d. Other. (optional) <ul style="list-style-type: none"> • Epigenetics • Immunological and microbiome-gut-brain interactions
22. Multiaxial assessment systems: future development
a. Multiaxial diagnostic formulation will benefit the management of FGID <ul style="list-style-type: none"> • As I mentions before, it will give a more integrative and multidimensional information on the patients disorder

<ul style="list-style-type: none"> • May. Time will tell the truth • Probably good in tricky cases when validated • To be proven
b. Multiaxial diagnostic formulation may benefit the management of FGID depending on results from further clinical research and consensus <ul style="list-style-type: none"> • Everything “may”!
c. Multiaxial diagnostic formulation will only benefit the management of FGID when further clinical "gold standard" diagnostic testing becomes available <ul style="list-style-type: none"> • Before 'gold standard testing' we need mechanistic understanding, assuming that you refer to biomarker when you mention test; the Rome criteria (or comparable approaches) are not tests • Currently, there is no gold standard for the diagnosis of FGID • FGID's are undoubtedly heterogeneous; some may "fall the wayside" as new discoveries are and new "organic" entities are described • On the contrary, its usefulness is based on the absence of a gold standard of diagnosis • There will never be a clinical gold standard
d. Multiaxial diagnostic formulation will never benefit the management of FGID
e. Other (optional) <ul style="list-style-type: none"> • May help reimbursement! But that entire aside. I think GI docs tend to simplify functional disorders but these pts often are very complex with overlapping GI disorders and a psychosocial background that many practitioners either don't care to get into or don't appreciate • Will at the very least contribute to further progression of research hypothesis and ideas for clinical management of this important group of patients
23. What recommendations would you make for future development in any of the areas covered by this survey? (Optional)
<ul style="list-style-type: none"> • Coexistence of other chronic pain syndromes (e.g., chronic pelvic pain, migraine headaches, chronic fatigue syndrome) • Development of diagnostic formulation taking spectrum of FGID all over the world as the spectrum may be different in different regions, patients perception, physician perceptions, needs, socio-economic and cultural background are quite different in different parts of the world. These diagnostic formulation needs to be validated in different population • Excellent research project - will benefit patients - it want to include something on connective tissue disorders (eg EDS) and previous gastroenteritis • I personally missed a definition of FGID in the beginning of the questionnaires. Perhaps this is an intentional strategy. However, it may provide useful to inform the respondents, especially if there are some who are not up to date with the literature and/or

research in this area

- It is ok
- None
- One area, which is ignored in FGID, is the exclusion of other known clinical conditions producing GI symptoms such as connective tissue disorders and neurological and endocrinological disorders. GI symptoms may sometimes be the presenting features of these conditions and are often missed. In the multiaxial assessment, a specific mechanism must be developed to exclude these
- The survey appears to have a priori declared FGID as part of the spectrum of psychiatric illnesses, and introduced a bias. The perpetuations of terms such as functional, psychosocial, etc do not do either the patients or the field a service. While much of the effort to broaden the approach to these patients is laudable, I would argue that they apply equally to many other chronic illnesses- the only difference being the physician's subjective bias as to what is "organic" and what is not. This also leads to absurd nomenclature like "functional vomiting"- try to explain that to a patient who is throwing up!
- Well-designed, prospective studies with clear-cut endpoints

24. Are there any issues that are important to you in your present role that have not been explored in the questions above?

- Epigenetics I put strongly agree for other simply to allow me to complete the survey
- I completed the survey with some breaks in between and may thus be biased by some of the later points, there is an inherent bias in the survey that reflects much of my own thinking but may sell FGID short. It may perpetuate our focus on phenomenology and drop the attempt to understand physiology, perhaps with the exception of central sensitization/hypervigilance. Several areas that are generically approached (if at all) by clinicians were left out (e.g., fiber), emerging areas like probiotics or antibiotics to influence colonic (or small bowel) flora, and a central component of gut function (food intake, absorption) have not been touched at all
- It is ok
- None
- Not much. However, the questionnaire is quite long and some questions are not clear
- Perhaps further targeting the fact that many FGID patients use CAM therapies to manage their problems... what do different provider think about this? Relevance for management, set up of multiprofessional teams, patient recommendations, clinical strategies, etc.?

- **Round 2**

2. A biopsychosocial understanding of FGID	
a. Far from all FGID, patients have psychosocial comorbidity.	<ul style="list-style-type: none"> • English should be clearer. Instead of writing "Far from all" write "not all FGID patients" • In my eyes, the biopsychosocial model of disease is a form of truism. It tells me that (any) illness manifests in an individual with a unique genetic background that interacts with an equally unique biography of exposures (environmental, cultural...). The relative importance of different domains will vary (e.g., common cold vs. fibromyalgia). However, even my reaction to a runny nose fits into the BPS model • this item is poorly written
b. FGID belongs to a biopsychosocial and environmental spectrum that includes "biologically defined" illness such as Crohn's disease	<ul style="list-style-type: none"> • FGID is a diagnosis by symptom-based criteria. Hence, it include true functional diseases as well as more organic diseases such as small intestinal bacterial overgrowth, celiac disease, faecal evacuation disorders, lactose malabsorption etc. • In Light of recent research that IBS may be an inflammatory condition, this may be truth, but more studies are needed in this regard • In my eyes, the biopsychosocial model of disease is a form of truism. It tells me that (any) illness manifests in an individual with a unique genetic background that interacts with an equally unique biography of exposures (environmental, cultural...). The relative importance of different domains will vary (e.g., common cold vs. fibromyalgia). However, even my reaction to a runny nose fits into the BPS model • This may be somewhat vague, e.g. the difference between "biologically defined" vs pathology?
c. FGID have strong associations with a family history of chronic pain disorders	<ul style="list-style-type: none"> • perhaps in the severest cases • We certainly have data on FH & FGID and epidemiologic data on FGID & chronic pain, which again clusters. The strength of other chronic pain disorders & FGID as determined by family history is not established in my eyes, but probably exists mediated by somatisation
d. FGIDs have a strong association with a history of migraine headaches	<ul style="list-style-type: none"> • Strong is overstatement

Contributors to Functional Gastrointestinal Disorders
3. .Physical Origin
<p>a. Previous truncal surgery (e.g. cholecystectomy, bowel surgery)</p> <ul style="list-style-type: none"> • By definition, this is not FGID • Certainly strongly associated but many patients have surgery because of FGID symptoms • Especially those with endometriosis • I think that at FGID are risk factors for surgery, especially as the threshold for operations has been lowered with laparoscopic interventions. Look at 'biliary dyskinesia' as a FGID that even the Rome committee still sees as an indication for surgery. A blank statement cannot separate chicken and egg • May be important in some patients, but not as a general main contributing factor • This may vary in different population. For example, studies from Taiwan documented that. This may be related to the fact that many surgeons and gynaecologists may not be aware of FGID and hence, undertake surgery with a mistaken diagnosis • Though the nature of this association is unclear
<p>b. Previous physical trauma (e.g., occupational and road traffic accidents resulting in truncal injury)</p> <ul style="list-style-type: none"> • By definition, this is not FGID • May be important in some patients, but not as a general main contributing factor • Not well documented but frequently encountered
<p>c. Brain injury / condition (i.e., effect on CNS - emotion, cognition, personality and possible brain-gut response)</p> <ul style="list-style-type: none"> • By definition this is not FGID • I think you may see this more in a tertiary setting • It is 'dangerous' to put emotion into a section on brain injury or condition. While it follows the "chemical imbalance" rhetoric of psychopharmacology, the vast majority of patients with affective spectrum disorders do NOT have a "brain disease" as defined by distinct structural and physiological changes. While CNS injury affects GI function (dysphagia to constipation or incontinence), this impact is distinct of the processes that are the main burden of FGID in clinic • May be important in some patients, but not as a general main contributing factor • Very little data

<p>d. Ongoing connective tissue disorders such as Marfan's and Ehlers-Danlos syndromes (<i>possible resultant abnormalities such as external and hiatus herniae, intestinal diverticula and rectal prolapse</i>)</p> <ul style="list-style-type: none"> • By definition, this is not FGID • May be important in some patients, but not as a general main contributing factor • There is, as yet, very little published data here
<p>e. Previous gastrointestinal infection/infestation (<i>e.g., bacterial, viral, protozoal</i>)</p> <ul style="list-style-type: none"> • AT LEAST IN A SUBGROUP-POST INFECTIOUS IBS • It is important to realize that there is an overlap between PI malabsorption and PI-IBS. Refer to our reviews in J Gastroenterol Hepatol (Ghoshal UC, Park HJ, Gwee KA. Bugs and Irritable Bowel Syndrome: The good, the bad and the ugly. J Gastroenterol Hepatol 2010; 25: 244-51; Ghoshal UC, Ranjan P. 2011 Apr;26 Suppl 3:94-101)
<p>f. Aberrant enteric microbiota (<i>i.e., due to previous infection, dietary change or drugs</i>)</p> <ul style="list-style-type: none"> • Data emerging here but as yet, unclear what the precise changes are and what their primacy may be. • Likely but unproven • new area of research so topical and exciting but very limited data
<p>7. Psychological Origin</p>
<p>a. Dietary habit (<i>socially related regularity, nutritional value and side effects</i>)</p> <ul style="list-style-type: none"> • If so then it is likely more on the macronutrient aspects such as the proportion of different kinds of foods than social related regularity • IN A WAY, THIS IS TRUTH, IN THE SENSE THAT PATIENTS ALWAYS REFER THEIR SYMPTOMS IN RELATION TO DIET. BUT NOT SPECIFICALLY TO DIETARY HABIT AS IT HAS BEEN RELATED IN ALL THE CULTURES AND COUNTRIES • This is very important. For example, if a patient takes too much milk and milk product in an area of world where lactose intolerance is common, one must consider this to contribute to symptoms. On the other hand, in areas of world where diet is rice based with low frequency of celiac disease, exclusion of celiac disease may not be important
<p>b. Significant life events (<i>family bereavement, family/partner separation</i>)</p> <ul style="list-style-type: none"> • Significant life events should include "psychosocial trauma"
<p>c. There is no proof that FGID are psychological in origin</p> <ul style="list-style-type: none"> • But there is a strong association especially in moderate and severe disease • Define "proof" more clearly, e.g. do you mean scientific evidence?

- Goes back to my prior comments about the biopsychosocial model and the atheoretical concept of FGID. As long as we have broad categories defined by nosology rather than ENDO-phenotypes, the apparent phenotype (e.g., IBS-D) may be due to a prior infection (post-infectious IBS), a somatoform manifestation of an affective spectrum disorder, lactose intolerance, celiac disease, microscopic colitis
- Terminology is all important here; I would contend that FGID are not CAUSED by psychopathology but symptoms may be precipitated or exacerbated by psychological factors and presentation significantly influenced by co-morbid anxiety/depression

8. Gender Differences

a. Sex hormone effect on GI sensitivity and motility (*i.e., alterations in GI transit/colonic permeability during follicular and luteal phases of the menstrual cycle*)

- Can have a significant influence on menstrual-cycle related fluctuations in symptom severity and nature; however, data on actual effects in man on motility are conflicting
- Evidence in humans as opposed to animals is wanting
- NOT BEEN PROVEN
- While the statement passes face validity, every 9defined as >90%) attempt at proving the link with a prospective study have failed

b. Sex hormone effect on nociceptive processing (*i.e., modulation of visceral pain via oestrogen receptors expressed in the dorsal root ganglion*)

- as above
- While the statement passes face validity, every 9defined as >90%) attempt at proving the link with a prospective study have
- This puts too much weight on nociception and inferences from reductionist studies. Nociception focuses on high intensity stimuli. FGID manifest with symptoms in response to low/no intensity stimulation, which is at best partially explained by peripheral sensitization. Considering the context of FGID disease, processing of with gender-specific differences in appraisal etc. may matter more. We need to remember differences in the lifetime prevalence of anxiety and depression
- Yes, but not particular to the example given of oestrogen receptors on DRG

c. Socioeconomic status (*trait anxiety in relation to family, job security and lack of medical insurance*)

- Data conflicting
- trait anxiety is not socioeconomic

d. There is insufficient evidence to make claims on statements a and b

e. There are gender differences in response to pharmacological treatments

- Data are conflicting but likely there are minor differences

- Do you mean pharmacological treatments of FGIDs?
- Here, there is insufficient evidence only anecdotal

9. Impact of Symptoms on Daily Life

a. Help from significant others (*aid in everyday activity from family or friends*)

- Don't understand question
- I am not sure about the direction. Social support is important life and certainly in chronic illness
- I think this question is a bit unclear in relation to: "Which areas of contribution do you think are important when considering the evaluation and diagnosis of FGID?" I mean help from significant others is of course important in the overall assessment of the patient and how s/he is able to manage everyday life, but it is not as important in the "diagnosis" of FGID per se. Do you see what I mean?
- Not well documented
- This statement is vague- are you implying that help is impacted by symptoms?

b. Reduced school attendance in children and adolescents

10. Genetic Polymorphism

a. Contribution of genetic factors to the mediation of psychological disorders (*e.g., reduced function polymorphisms in the serotonin reuptake pump in conditions such as anxiety and depression*)

- All these statements may be relevant, but exact mechanisms of action are likely still to be defined, hence the uncertainty and that all first round statements failed to gain consensus
- Data inconclusive
- See my prior comments. Genes matter. However, we have not made true inroads despite a few case series here or there suggesting a genetic basis that is typically not confirmed by subsequent studies
- The examples are the hot buttons with divergent data; I suspect if other examples are used, there may be more consensus
- There is good data to support this independent of IBS not in relation to IBS
- Though there is insufficient evidence. More studies are needed

b. Contribution of genetic factors to the mediation of gastrointestinal sensory and motor function (*e.g., polymorphisms of enteric serotonin transporter genes and alpha-2 adrenoceptors*)

- All these statements may be relevant, but exact mechanisms of action are likely still to be defined, hence the uncertainty and that all first round statements failed to gain consensus. All these statements may be relevant, but exact mechanisms of action are likely

<p>still to be defined, hence the uncertainty and that all first round statements failed to gain consensus</p> <ul style="list-style-type: none"> • Jury is out for most of these, perhaps, with the exception of the small group of patients with the Na channel disorder • Not enough info yet
<p>c. Contribution of genetic factors to pain modulatory pathways (e.g., polymorphisms of serotonin receptors in the dorsal root ganglion)</p> <ul style="list-style-type: none"> • All these statements may be relevant, but exact mechanisms of action are likely still to be defined, hence the uncertainty and that all first round statements failed to gain consensus • Jury is out for most of these, perhaps, with the exception of the small group of patients with the Na channel disorder. • Likely but data not there
<p>d. Polymorphisms in genes that modulate immune and/or neuro-immune functions (i.e., possible contribution to the onset of symptoms in the presence of other exogenous stressors)</p> <ul style="list-style-type: none"> • All these statements may be relevant, but exact mechanisms of action are likely still to be defined, hence the uncertainty and that all first round statements failed to gain consensus • Best evidence here is in PI-IBS • Likely but data not there • Though more studies are needed • Very likely although not proven
<p>e. Epigenetics (heritable changes in phenotype appearance or gene expression caused by mechanisms other than changes in the underlying DNA sequence) is an area of importance in several areas of FGID expertise</p> <ul style="list-style-type: none"> • All these statements may be relevant, but exact mechanisms of action are likely still to be defined, hence the uncertainty and that all first round statements failed to gain consensus • As prev mentioned even more likely to be true but data limited • Likely but data not there • more studies are needed • Same as with the BPS model, we need to avoid stating the obvious. Gene-environment interactions (with culture being part of the environment) determine the resulting phenotype (=living being with or without disease). The question is not whether such a relationship exists, but whether we can identify (epi-) genetic markers to truly correspond with a substantially increased risk for distinct forms of FGID and/or explain more than a marginal fraction of the phenotype. With the pure nosology as underlying

<p>classification system, chances seem slim</p> <ul style="list-style-type: none"> • Though not studied at all in IBS, the heritability gap that exists in IBS could well be explained by epigenetics
The Therapeutic Relationship
11. Clinicians working with FGID patients need the following qualities and experience
<p>a. The number of years working with FGID patients (<i>i.e.</i>, does the number of clinical working years necessarily mean continued professional and reflective development?)</p> <ul style="list-style-type: none"> • All the years of practice in the world will not compensate for an inappropriate attitude to FGID • More experience can be helpful; it can also lead to burnout
<p>b. Healthcare professionals working with FGID patients should be trained in the spirit of the biopsychosocial model</p> <ul style="list-style-type: none"> • I use the BPS here as a surrogate for a comprehensive approach in understanding and treating illness • Part of an holistic approach to these patients • The model is still unclear and certainly not standardized •
<p>c. Clinicians working with FGID patients must have a thorough understanding of the pathophysiology of FGID</p> <ul style="list-style-type: none"> • It is important with understanding of the pathophysiology of the gastrointestinal tract, but then again in relation to FGID, there are reasons it is called "functional" rather than "pathological" • Part of an holistic approach to these patients
12. The patient-practitioner relationship
<p>a. Structured interview using direct questions to elicit information about the patient's presentation</p> <ul style="list-style-type: none"> • Both non-directive interview and structured interview are important, as without non-directive interview, lot of psychosocial information might not be obtained. Moreover, over-reliance on criteria for the diagnosis of IBS may lead to its under or over-diagnosis as has been shown in several studies showing that as the criteria was changed from Rome I to Rome II, frequency of IBS in the same population was found different • Highly debatable • Important for research • Not validated in different cultural / clinical settings
<p>b. Past clinicians (<i>e.g.</i>, medical specialty, complimentary or alternative therapies)</p>

<ul style="list-style-type: none"> • For most of us practicing in tertiary referral centres, extensive use of healthcare resources is the norm. • I DONT UNDERSTAND EXACTLY WHAT YOU ARE ASKING WITH THIS QUESTION, WHETHER IT IS IMPORTANT FOR THE PATIENT, OR FOR THE PHYSICIAN, OR IS IT'S COMMONLY USED BY EITHER ONE? • Just a note... complementary vs "complimentary"
c. The availability of clinician (<i>regular work hours at a given clinic</i>) <ul style="list-style-type: none"> • AGAIN. I AM NOT SURE WHAT YOU MEAN HERE • I am not sure. I do not think that there is any evidence for this • I think that technology will 'outrun' us. Availability will eventually be defined by contact options including electronic interactions
13. Cultural Factors
a. The patient's socio-economic status (<i>family & job security, lack of medical insurance etc</i>)
b. The patient's educational status
c. The patient's spiritual and religious attitudes (<i>interpretation of symptoms & attitudes towards medical treatment</i>)
d. The clinician's outlook on homosexuality and other gender issues <ul style="list-style-type: none"> • DON'T UNDERSTAND EXACTLY WHAT YOU MEAN HERE • I am not sure • Should not have any importance for professional health care providers
e. The patient's understanding of FGID terminology (<i>e.g., the Chinese language has no good terminology for "heartburn"</i>). <ul style="list-style-type: none"> • Clinicians should have clinical competence. This includes the ability to listen to descriptors rather than rely on a set of given terminology • SPANISH HAS NO WORD FOR BLOATING
<ul style="list-style-type: none"> • Cultural dietary factors (<i>e.g., wheat versus rice intake / milk, fibre, fruit, spices intake in relation to intolerance frequency</i>) • May be important to be aware of such factors to build a good relationship and develop a feasible treatment plan for the patient etc • SAME COMMENT AS IN 4A, BUT THIS ISSUE MAY BE MORE SUITED IN FOR THE CULTURAL FACTORS AREA
Areas for Consideration and Possible Measurement
14. Abdominal Symptoms
a. Increased flatulence

<ul style="list-style-type: none"> • All aspects of GI function may be important • FLATULENCE IS IMPORTANT BUT WHETHER INCREASE FLATULENCE IS, I AM NOT SURE. HOWEVER, IT IS VERY IMPORTANT TO INCLUDE ABDOMINAL BLOATING AND PAIN • This section is not helpful, as all these symptoms potentially matter. However, the relative importance depends also on the prevalence of defined disorders/abnormal behaviours. For example, aerophagia is less common than reflux disease. Yet for patients with this problem, it is the primary if not only concern. Obviously, problems with micturition are not defining FGID, but may still matter (e.g., IC) • While not a component of the Rome process, this can be a very disabling symptom for patients
b. Aerophagia <ul style="list-style-type: none"> • All aspects of GI function may be important • Important, not that uncommon and poorly understood
c. Changes in appetite <ul style="list-style-type: none"> • All aspects of GI function may be important
d. Dysphagia <ul style="list-style-type: none"> • All aspects of GI function may be important • Dysphagia is rarely functional in nature. I think, this symptom require a thorough evaluation motility studies rather than including in FGID
e. Epigastric pain <ul style="list-style-type: none"> • All aspects of GI function may be important • Overlap of symptoms are common
f. Postprandial fullness <ul style="list-style-type: none"> • All aspects of GI function may be important
g. Obstructive defecation <ul style="list-style-type: none"> • All aspects of GI function may be important
h. Flatulence odour <ul style="list-style-type: none"> • All aspects of GI function may be important
i. Polyuria

<ul style="list-style-type: none"> • All aspects of GI function may be important • Assume absence of pathology means absence of diabetes • Suggests phenotype.
j. Dysuria <ul style="list-style-type: none"> • All aspects of GI function may be important • Many, many, many others are missing
15. Patient Description of Abdominal Symptoms
a. Patient's previous experiences of similar symptoms
b. Patient's view of inter-relationships between more than one symptom (<i>e.g., bloating and constipation or abdominal pain and referred musculoskeletal pain</i>)
17. Emotional Functioning
18. Social Functioning
a. Reliance on medication (<i>Relief of physical and psychological symptoms</i>)
b. Reliance on significant others (<i>instead of the patient managing their own symptoms</i>) <ul style="list-style-type: none"> • I do not understand • The statement is not clear
19. Physical Functioning
a. Kinesiophobia (<i>excessive, irrational fear of physical movement due to the feeling of vulnerability to recurrence of gastrointestinal symptoms</i>)
b. Previous physical ability (<i>compared to the present</i>)
c. The patient's past attitudes towards physical activity
d. The patient's present attitudes towards physical activity
e. Old Age (<i>eg., decreased visceral sensitivity, impaired communication</i>) <ul style="list-style-type: none"> • Should be considered as an alarm feature and rule out organic diseases
Multiaxial Assessment and Formulation (Relevance to FGID)

20. Pros
a. Reduces reliance on clinical judgement for diagnosis and therefore reduces clinical subjectivity
b. Should encompass not only multiaxial evaluation but personal idiography that reflect their individual strengths, weaknesses and individual goals <ul style="list-style-type: none"> no idea what this means
21. Cons
a. Multiaxial assessment is a time consuming exercise <ul style="list-style-type: none"> I do not know the exercise It is important. In fact, in absence of such system, many patients may not be adequately evaluated
b. Multiaxial assessment is of little value to the management of FGID
c. The term "diagnosis" implies a distinct illness and is therefore not relevant in many cases of FGID <ul style="list-style-type: none"> Functional vs pathological diagnosis The term is indeed relevant in FGID. As do psychiatrist, we currently use the syndromic classification to 'diagnose' and then treat disorders. I agree with the implicit statement that we should try to get beyond a mere symptom-based system and use true endophenotypes. Obviously, we are not quite there yet
d. Multiaxial diagnostic criteria lack clear distinctions between normal and abnormal and therefore do not avoid diagnostic consideration of ordinary problems of daily living <ul style="list-style-type: none"> Important point that needs to be considered
e. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability
f. Tends to promote "all or nothing" diagnoses when considering an individual's problem (<i>i.e.</i>, how many symptoms from a list are required before action is taken)
22. Future Research: Identifying Risk Factors and the Replication of Findings (brain-gut-axis)
a. Incorporating genetic information in relation to psychological & visceral conditions (<i>e.g.</i>, polymorphism in genes that encode opioidergic or serotonergic receptors)

<ul style="list-style-type: none"> • Again, perhaps use different examples • All research questions are important, some in terms of being explorative and to generate hypothesis and some to be deductive and testing hypothesis • But more studies are needed • Important research area, but unlikely that it will be incorporated in clinical practice within the near future
<p>b. Incorporating pharmacogenomic research relating to therapies targeting of specific diseases (<i>i.e., distinct FGID may have different underlying genetic influences, pathological mechanisms and therefore, personalised drug strengths and combinations?</i>)</p> <ul style="list-style-type: none"> • All research questions are important, some in terms of being explorative and to generate hypothesis and some to be deductive and testing hypothesis • But more studies are needed • Important research area, but unlikely that it will be incorporated in clinical practice within the near future
<p>c. Incorporating data on immunological and microbiome-gut-brain interactions (<i>potential for specific modulation of enteric microbiota as a strategy for modulating co-morbid aspects of FGID</i>)</p> <ul style="list-style-type: none"> • All research questions are important, some in terms of being explorative and to generate hypothesis and some to be deductive and testing hypothesis
<p>d. Incorporating epigenetics when considering changes in gene expression caused by mechanisms other than alterations in DNA sequencing</p> <ul style="list-style-type: none"> • All research questions are important, some in terms of being explorative and to generate hypothesis and some to be deductive and testing hypothesis • But more studies are needed • Important research area, but unlikely that it will be incorporated in clinical practice within the near future
<p>23. Multiaxial assessment systems: future development</p>
<p>24. Are there any issues that are important to you in your present role that have not been explored in the questions above?</p>
<ul style="list-style-type: none"> • No. • Perhaps the collaboration of different health care providers, or how patients themselves utilises different health care providers, in the management of FGID should be explored in more depth. For example the potential (pros and cons of) integration of physicians, physios, psychologists and CAM-providers (such as osteopaths)

- **Round 3**

2. A biopsychosocial understanding of FGID

a. Significant numbers of FGID patients do not present with psychosocial comorbidity

- Any chronic condition will be associated with psychosocial comorbidity and I don't think this comorbidity is really very significantly raised in FGID compared to for instance IBD but as far as the statement is concerned, I would have to disagree.
- depends on what is exactly meant by "significant"
- It is necessary to more clearly separate those who have psychological problems because of their FGID that does not respond to therapy nor is recognized by many doctors, from those in whom psychosocial conditions may play a pathogenetic role
- The statement is sufficiently vague. Especially for specialty groups and tertiary centres, the statement is correct

b. FGID belongs to a biopsychosocial and environmental spectrum that includes "biologically defined" illness such as Crohn's disease

- I do not think that it is helpful to blur boundaries between illnesses that are defined by distinct phenotypes and a merely syndromic definition of illness (apples and oranges). I had previously mentioned that the BPS is not a testable hypothesis but a truism that is relevant for us to keep in mind and not exclusively important for FGID. Lastly, Rome criteria lack specificity, unless we exclude 'organic' disease. In this context, we need to remember that 'remission' in IBD is operationally defined by activity indices or degree of inflammation. These phenomena are not dichotomous. Thus, an IBD patient does not bounce back and forth between diseases there or gone, but degrees of severity, all of which occurs in the context of learned illness behaviour
- I don't think there is enough evidence yet for this statement
- IBD is a bad example as increasing evidence suggests that there may be discrete entities within IBD, as there also may be in IBS
- The question then is if a "functional" disorder can/should be considered to belong to a "pathological/biological" spectrum... hmmm not sure about that
- The sentence is true but may be misinterpreted since even cancer can present with dyspepsia and IBS-like symptoms - wording must be very clear in this respect

c. FGID have strong associations with a family history of chronic pain disorders

- I think this may be possible to determine once we can do genetic testing in this area. I don't think we have the data here to support this
- Lack of adequate experimental data
- NOT SURE ABOUT SIMILARITY OF ENDOGENOUS PAIN MECHANISMS
- Really, need to consider two questions here, which make it a bit tricky to give you a clear answer, i.e. "FGIDs share associations with a family history of chronic pain disorders..." (Y/N), if "Y" is this due to similar endogenous pain mechanisms between FGIDs and chronic pain disorders
- Yes, it's the susceptibility to develop chronic pain after injury, inflammation or stress that is relevant here rather than the tissue involved

d. FGID have a strong association with a history of migraine headaches

- Challenging to compare what you suggest might be a "biopsychosocial spectrum disorder" with a more narrow and specific headache disorder... Also, what kind of association? E.g., might a tentative negative association also be an association...? Alt wording "positive association"?
- Lack of adequate experimental data
- Not well documented
- They do share an association with migraine but I don't think there is enough evidence for this statement yet

Contributors to Functional Gastrointestinal Disorders

3. .Physical Origin

a. Previous truncal surgery (e.g. effects of local injury, repair mechanisms & stress response)

- By definition, FGIDs are not secondary to previous surgery
- Genetic testing of pain receptors may give us more information
- Hmm... Do you mean the PHYSICAL "stress response" on the HPA axis?
- The implied causal role/model is wrong. FGID is associated with healthcare seeking. Contact with the healthcare system puts you at risk for interventions. They do not get 'sick' because of surgery, but mostly get surgery because they have symptoms we as MDs interpret as a problem that may respond to surgery (see prior comment to the question)

<p>b. Previous physical trauma (e.g., local injury, repair mechanisms & stress response)</p> <ul style="list-style-type: none"> • All injury/pathology might be relevant to probe... i.e. as for any medical assessment of patients • Do not see this as a relevant mechanism for the vast majority of patients
<p>c. Brain injury / condition (i.e., effect of cognition, personality concerning top-down possible brain-gut response)</p> <ul style="list-style-type: none"> • All injury/pathology might be relevant to probe... i.e. as for any medical assessment of patients • I think this question is not phrased correctly. Brain injury is not a cognitive or personality top-down brain-gut response, but i have agree on the basis of the info in brackets • Not sure that I would classify CNS-injury or disease-related GI symptoms as a FGID • What is "brain condition"? See my prior comment, (I do not see this as a relevant mechanism for the vast majority of patients)
<p>d. Ongoing connective tissue disorders such as Marfan's and Ehlers-Danlos syndromes (high incidences of patients with joint hypermobility- referred to tertiary neurogastroenterological care with unexplained GI symptoms)</p> <ul style="list-style-type: none"> • All injury/pathology might be relevant to probe... i.e. as for any medical assessment of patients • Still awaiting data here • the stem says 'may' which I agree with • There is no doubt that systemic diseases alter gut function. However, we are now making the picture more blurry. A patient with systemic sclerosis may meet criteria for IBS with bloating, altered bowel patterns etc. However, there is an endophenotype, which defines the illness • This should be considered organic and not functional
<p>4. Psychological Origin</p>
<p>a. There is no proof that FGID are psychological in origin</p> <ul style="list-style-type: none"> • Again there are some instances that are and some that are not • Again, they are associated or FGIDs can have psychological comorbidities but there are no evidences to conclude that FGIDs are psychological in origin • Do not want to imply that all FGID are psychological in origin, but there is evidence supporting the role of psychosocial determinants on FGID origin, but not extensive to all disorders and subtypes • Hmm... Do you mean 'scientific peer reviewed evidence'? ;-) • I think it is an exacerbator not a cause.

- Lack of adequate experimental data
- Not only psychological but - in part – psychological
- Psychological factors are important in its development
- Psychological factors are one of the vulnerability factors
- Some may be psychological in origin
- The difficulty is in thinking about the origin of FGIDs rather than associations
- There is evidence that it is important in some. However, we deal with a wide array of problems. I doubt you would put achalasia in here
- Who knows yet about the etiology?

5. Gender Differences

a. Sex hormone effect on GI sensitivity and motility (*Women with FGIDs experience heightened visceral symptoms due to late luteal and menses*)

- My comment on lack of published evidence still stands
- This is a clinical observation of mine but evidence is not great

b. Sex hormone effect on nociceptive processing (*i.e., Oestrogen receptors in the dorsal root ganglion modulate purinergic and opiate receptor signalling*)

- Some evidence for this in literature but this is not yet mature for a definite perspective
- While shown in animal experiments, the relevance on human pain/disease is less well established

c. Socioeconomic status (*trait anxiety in relation to family, job security and lack of medical insurance*)

- Anxiety and depression are not affecting both sexes similarly
- I changed my prev response
- I should agree in general, but also note that some of these statements (lack of medical insurance) do not apply to European countries where health care IS FREE, i.e. SPAIN. Therefore, to avoid ambiguity consider removing that particular statement
- Lack of adequate experimental data
- Not sure, what the question means
- Stress related

<ul style="list-style-type: none"> • The evidence from around the world is conflicting • Unclear what is exactly meant here: effect of socioeconomic status per se or effect of anxiety related to socioeconomic status
d. There is insufficient evidence to make claims on statements a and b <ul style="list-style-type: none"> • A and b are probably correct but lack of evidence is the problem • Lack of adequate experimental data
e. There are gender differences in response to pharmacological treatments <ul style="list-style-type: none"> • Lack of adequate experimental data • Not enough evidence to support this statement. Usually enough men are not studied • There are gender differences but there is no evidence to prove that it is in relation to menstrual or menopausal women
6. Impact of Symptoms on Daily Life
a. Help from significant others <i>(e.g., fear avoidance relating to movement)</i> <ul style="list-style-type: none"> • Lack of adequate experimental data • No evidence to support this • Statement is unclear to me. if you mean that getting help contributes to perpetuate symptoms, I disagree; If you mean the opposite, I agree • Truism. I may similarly say that social support improves coping
7. Genetic Polymorphism
a. Contribution of genetic factors to the mediation of psychological disorders <i>(e.g., Modification of behaviour, mood etc due to effect of altered serotonin transport & adenosine A2a receptor mechanisms)</i> <ul style="list-style-type: none"> • Lack of adequate experimental data • Not enough evidences for this statement • Recent literature is in favour of this • See prior comments; genetic studies have been inconclusive; even in well-defined diseases (e.g., IBD) genetic markers explain less than 10% of the variance
b. Contribution of genetic factors to the mediation of gastrointestinal sensory and motor function <i>(e.g., Changes in colonic motility due to variations within cannabinoid metabolism and serotonin transport).</i> <ul style="list-style-type: none"> • Lack of adequate experimental data

<ul style="list-style-type: none"> • Not enough evidences for this statement • Probably true as a general rule • Some emerging, but not impressive data (see comment above)
<p>c. Contribution of genetic factors to pain modulatory pathways (<i>e.g., genetic variants of genes controlling G-Protein synthesis described in IBS and low catecholamine-O-transferase activity associated with pain sensitivity</i>).</p> <ul style="list-style-type: none"> • As above • Lack of adequate experimental data • No evidence that it is relevant in FGID as of yet • Not enough evidences for this statement
<p>d. Polymorphisms in genes that modulate immune and/or neuro-immune functions (<i>i.e., gene polymorphism in TNAα & IL6 due to inflammatory response during and after gastroenteritis in PI-IBS</i>).</p> <ul style="list-style-type: none"> • 'May' is the key word here • Evidence sketchy • Lack of adequate experimental data • Likely but data not available
<p>e. Epigenetics (<i>heritable changes in phenotype appearance or gene expression caused by mechanisms other than changes in the underlying DNA sequence – eg microbiota-mediated changes in HPA axis</i>) is an area of importance in several areas of FGID expertise.</p> <ul style="list-style-type: none"> • as above • Lack of adequate experimental data • Likely to be true but I think unproven • Little or no data but plausible • Missing evidence
The Therapeutic Relationship
8. Clinicians working with FGID patients need the following qualities and experience
<p>a. The number of years working with FGID patients (<i>i.e., does the number of clinical working years necessarily mean continued professional and reflective development?</i>)</p> <ul style="list-style-type: none"> • As long as the clinicians are updated on the newest knowledge

- Depends on one's ability to listen and to change and adapt
- Does the number of clinical working years necessarily mean continued professional and reflective development? - Though this question is important but practically speaking the number of years is an important quantitative measurement of experience
- Hopefully
- That is true for most disorders (experience matters in professions), so truism
- This could go both ways. Good if you keep your perspective. Not so good if you have burnout
- What matters is interest not (only) experience
- You need to get into the mind-set of the pts and have be able to practice evidence based and also experience based medicine to help some of these pts.

9. The patient-practitioner relationship

a. Structured interview using direct questions to elicit information about the patient's presentation

- Eg with experience comes "flexibility" and less need for a structured interview
- For research yes but for patient management no
- Generally not validated in different cultural / linguistic settings
- I think after the 'open question' phase direct questions are useful in clarifying what the pt means

b. The availability of clinician (*regular work hours at a given clinic & management continuation*).

- But can be a pain!
- Depends upon the patient, not for all, but important for small number.
- Discharge them before resolving their problems and they will just reappear somewhere else.
- I would have gone for important but am prepared to agree with consensus of v important.
- Not only FGID.

10. Cultural Factors

a. The patient's socio-economic status (*family & job security, lack of medical insurance etc*)

- Data not there
- I am not aware of evidences in this matter
- See remark above (I don't understand what is exactly meant here)
- Stress related

<p>b. The patient's educational status (lower educational levels interpret pain as a “signal of harm”)</p> <ul style="list-style-type: none"> • Stress related • Note that frequently we face the opposite, that is, highly educated people-not related to medicine- may misinterpret clinical manifestations just because they have more access to information. Thus, educational status is important, but level of education may affect differently
<p>c. The patient's spiritual and religious attitudes (<i>interpretation of symptoms & attitudes towards medical treatment</i>)</p> <ul style="list-style-type: none"> • Few data • Stress related and coping mechanisms related
<p>d. The clinician's outlook on homosexuality and other gender issues</p> <ul style="list-style-type: none"> • I don't understand this one; I don't think clinicians look at this • I don't really understand the question! • I would stick with this being of little or no importance • Important if they are negative only • No idea
<p>e. The patient's understanding of FGID terminology (<i>e.g., the Chinese language has no good terminology for "heartburn"</i>)</p> <ul style="list-style-type: none"> • In Spanish, there is not a word for Bloating • Very under-appreciated. In Spanish, only one word for bloating and distension
<p>Areas for Consideration and Possible Measurement</p>
<p>11. Abdominal Symptoms</p>
<p>a. Aerophagia</p> <ul style="list-style-type: none"> • Can be relevant to bloating • I don't think that aerophagia is important for functional heartburn patients, it is important as an independent disorder, but very uncommon • Is it different from belching? • More common than imagined and poorly understood • This is of particular importance in Paediatric patients

b. Changes in appetite <ul style="list-style-type: none"> • Anorexia is a red flag • Appetite changes are important in any event • Changed (gone with consensus) • Quite common and often missed
c. Flatulence odour <ul style="list-style-type: none"> • Can be suggestive of bacterial overgrowth but is subjective; ask the spouse • Lack of adequate experimental data
d. Polyuria <ul style="list-style-type: none"> • Frequency of micturition rather than polyuria (which means more urine in volume than usual) • Important as it may be a comorbid manifestation as any other comorbidity • Not so common in my experience.
e. Dysuria <ul style="list-style-type: none"> • Changed (gone with consensus) • Important as it may be a comorbid manifestation as any other comorbidity • Yes - common in pts in tertiary care that i see
12. Patient Description of Abdominal Symptoms
13. Emotional Functioning
13. Social Functioning
a. Reliance on medication (<i>Changes in personality and social withdrawal due to increased medication</i>) <ul style="list-style-type: none"> • Unclear what is exactly meant here • We often make people sick by over prescribing
b. Old Age (<i>under-recognised due to associations of younger age groups associated with FGID</i>) <ul style="list-style-type: none"> • FGID in the elderly under-appreciated and little researched • It is true that older patients have less frequently FGIDs, but this is an important group to consider as organic disorders are more possible as underlying causes of symptoms and they need dosage adjustments of the different treatments as well • Long standing symptoms often interpreted as consequence of old age if proper hx is not taken

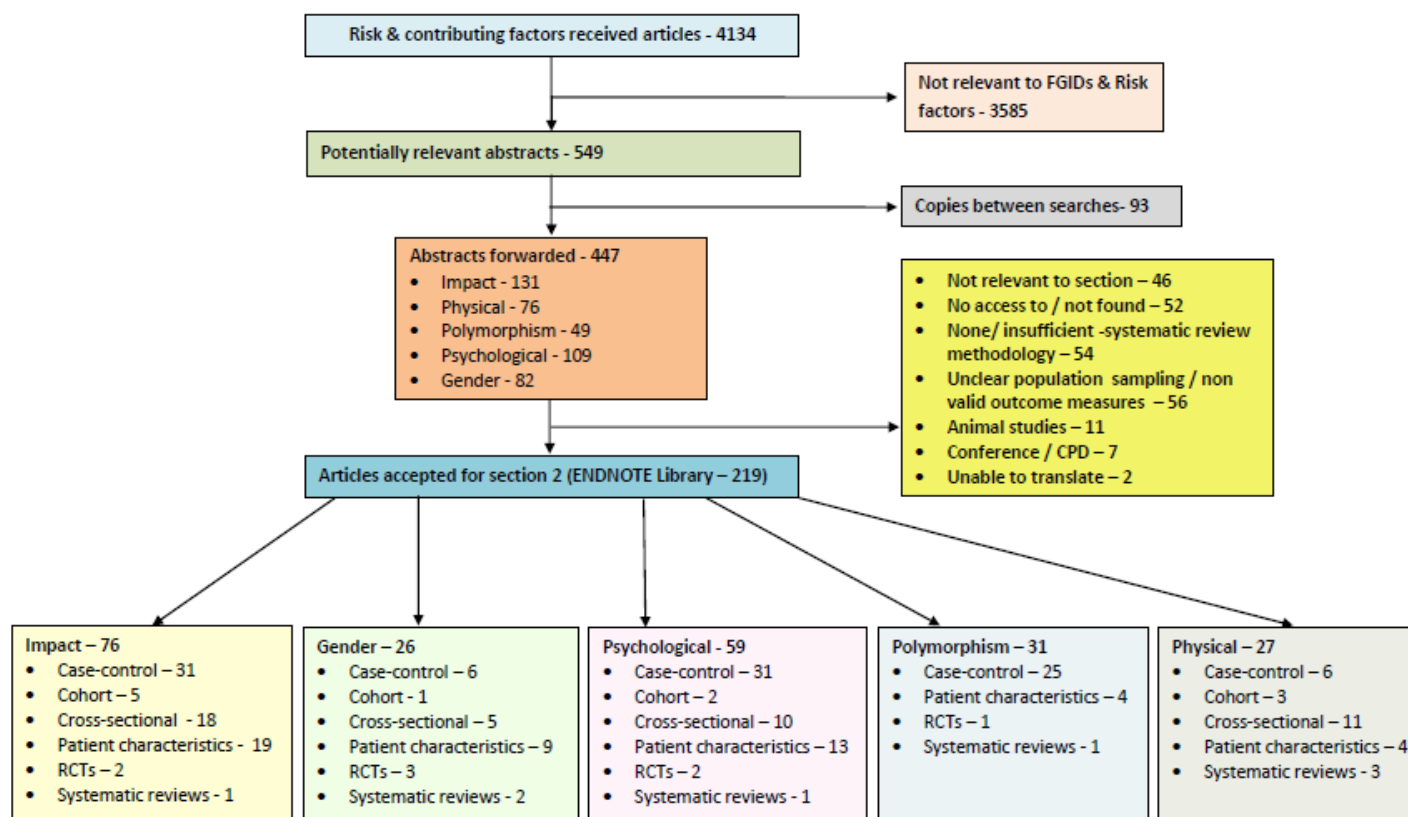
<ul style="list-style-type: none"> • No idea • Unclear what is exactly meant here.
14. Physical Functioning
<p>a. Kinesiophobia – <i>an excessive, irrational fear of physical movement due to the feeling of vulnerability to recurrence of gastrointestinal symptoms (flatulence, belching in relation to quick, unexpected movement)</i></p> <ul style="list-style-type: none"> • I have never seen this nor have read about it • lack of physical activity has important implications on digestive functions • Uncommonly seen is more severe end of spectrum especially those who develop allodynia on abdominal wall after recurrent surgery
<p>b. Previous physical ability compared to the present (<i>were patients physically active before onset of FGID, & do symptoms prevent them from being as active?</i>)</p> <ul style="list-style-type: none"> • Changed (gone with consensus) • Lack of physical activity has important implications on digestive functions • Useful to know this
<p>c. The patient's past attitudes towards physical activity</p> <ul style="list-style-type: none"> • I am not quite sure I understand this statement. I think that aside from previous attitudes towards physical activity, this may change because of the presence of FGIDs • Not sure about this but prepared to go with consensus • lack of physical activity has important implications on digestive functions
<p>d. The patient's present attitudes towards physical activity</p> <ul style="list-style-type: none"> • For example, some patients with diarrhoea have fear of running or exercising because they may get diarrhoea or pass gas in a gym • lack of physical activity has important implications on digestive functions
Multiaxial Assessment and Formulation (Relevance to FGID)
15. Pros
<p>a. Reduces reliance on clinical judgement for diagnosis and therefore reduces clinical subjectivity</p> <ul style="list-style-type: none"> • Hopefully not • No idea

<ul style="list-style-type: none"> • Reliance on clinical judgment is quite crucial • We have found that answering questionnaires is better to classify IBS subtypes, for example. However, it does not replace the clinical judgement • Will have to see how it performs
b. Should encompass not only multiaxial evaluation but personal idiography that reflect their individual strengths, weaknesses and individual goals <ul style="list-style-type: none"> • Will have to see how it performs • I still don't understand this question • No idea
16. Cons
a. Multiaxial assessment is a time consuming exercise <ul style="list-style-type: none"> • Might be; eg with experience comes "flexibility" and less need for a structured interview
b. Multiaxial assessment is of little value to the management of FGID. <i>(designed for psychological disorders and are therefore of little value to FGIDs)</i> <ul style="list-style-type: none"> • Psych AND soma are interconnected
c. The term "diagnosis" implies a distinct illness and is therefore not relevant in many cases of FGID <ul style="list-style-type: none"> • Definitely an issue • FGID are "distinct" diseases provided doctors know what they are talking about • Having a diagnosis is important for the patient as well as for the caregiver to direct a therapy • True if disease is defined by mechanisms rather than nosology only
d. Multiaxial diagnostic criteria lack clear distinctions between normal and abnormal and therefore do not avoid diagnostic consideration of ordinary problems of daily living <ul style="list-style-type: none"> • Clear distinctions are more difficult in the absence of pathology.
e. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability <ul style="list-style-type: none"> • Changed (gone with consensus) • Often

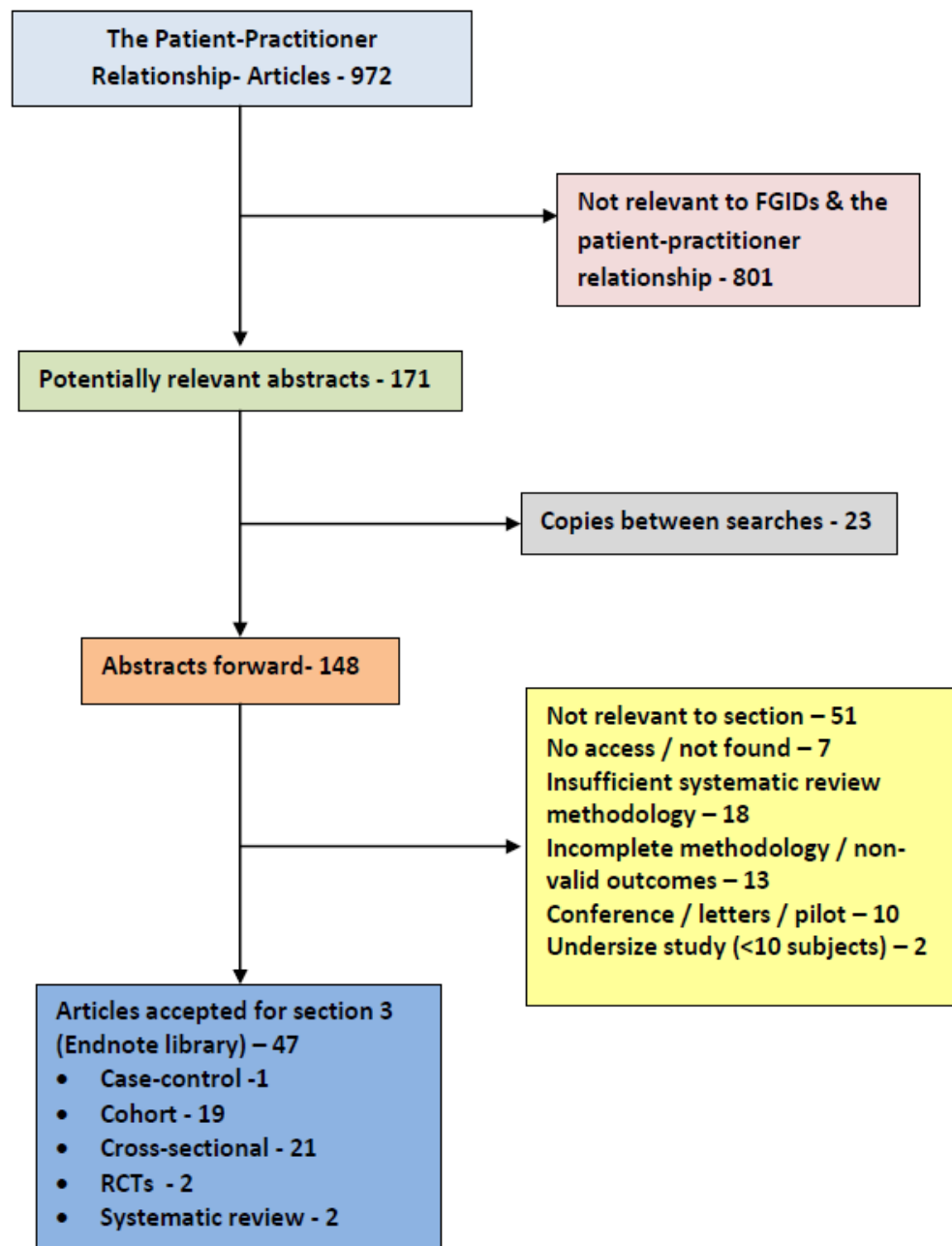
<p>f. Tends to promote "all or nothing" diagnoses when considering an individual's problem (<i>i.e.</i>, how many symptoms from a list are required before action is taken)</p> <ul style="list-style-type: none"> • In some cases yes, hence the need for additional and complementing clinical judgement • Language and cultural factors will also impact here • Yes, this may a possibility as it is when using Rome questionnaires to diagnose patients. If they don't satisfy the criteria, they will not be diagnosed with a FGID. In the clinic interview however, one may think that a patient has that FGID
<p>18. Future Research: Identifying Risk Factors and the Replication of Findings (brain-gut-axis)</p>
<ul style="list-style-type: none"> • Incorporating epigenetics when considering changes in gene expression caused by mechanisms other than alterations in DNA sequencing <ul style="list-style-type: none"> • I think microbiota brain gut signalling is an important topic for the future • If you believe in doing genetics at all
<p>19. Multiaxial assessment systems: future development</p>
<p>a. Multiaxial diagnostic formulation will only benefit the management of FGID when further clinical "gold standard" diagnostic testing becomes available</p> <ul style="list-style-type: none"> • Could complement these approaches • I think that multiaxial assessment will be helpful irrespective of whether a gold standard test is available and it is unlikely that such a test will become available in the foreseeable future • No idea
<p>20. Are there any issues that are important to you in your present role that have not been explored in the questions above?</p>
<ul style="list-style-type: none"> • Good questions Phil, although many were a tad difficult to provide a straight answer for... (Hence no consensus yet)... anyway, I hope the answers and comments may help you in your analysis and interpretation. Good luck! • None

Appendix C: Search protocol flow charts for Delphi survey sections

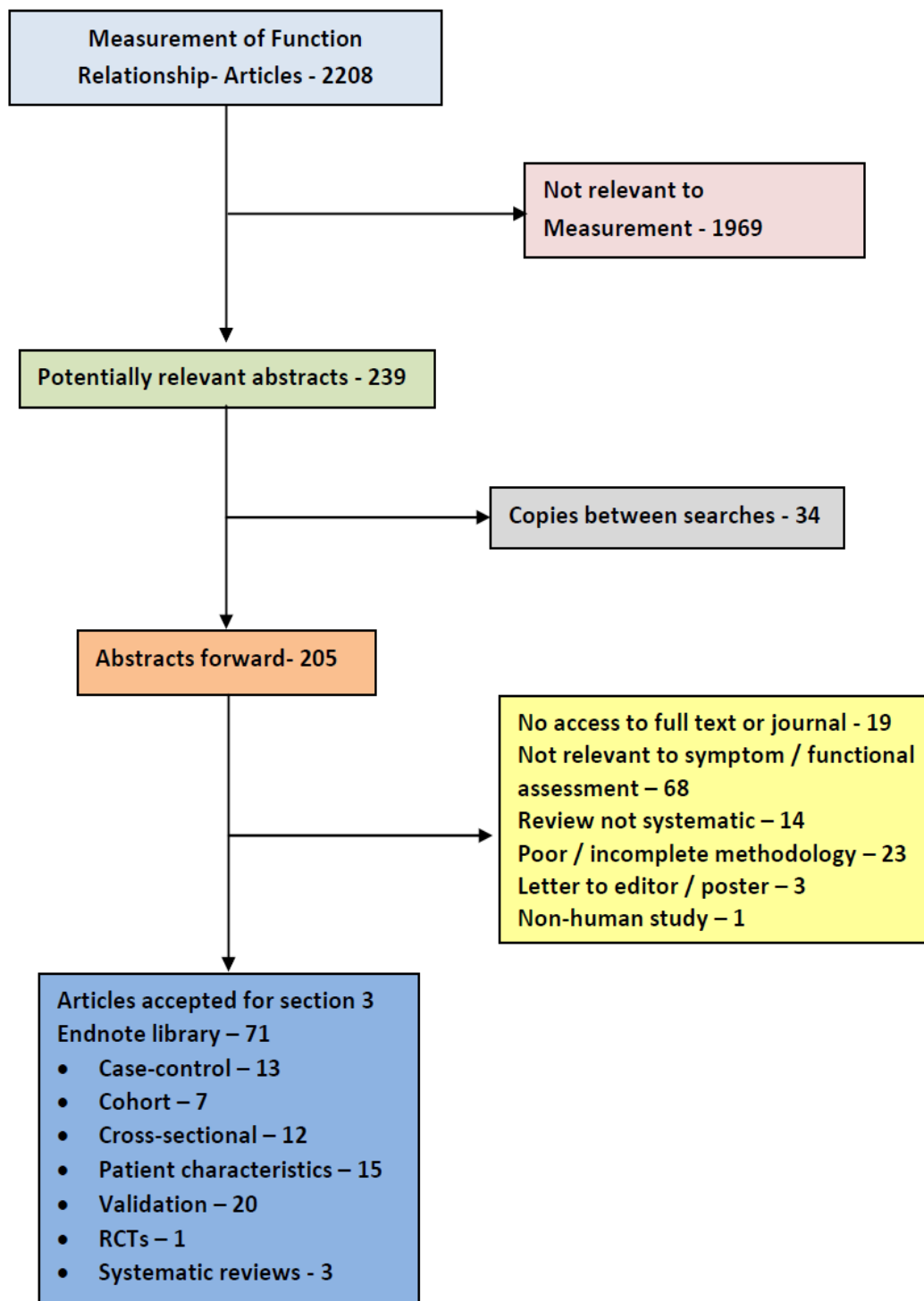
A Flow chart describing search protocols for articles examining risk factors associated with FGIDs



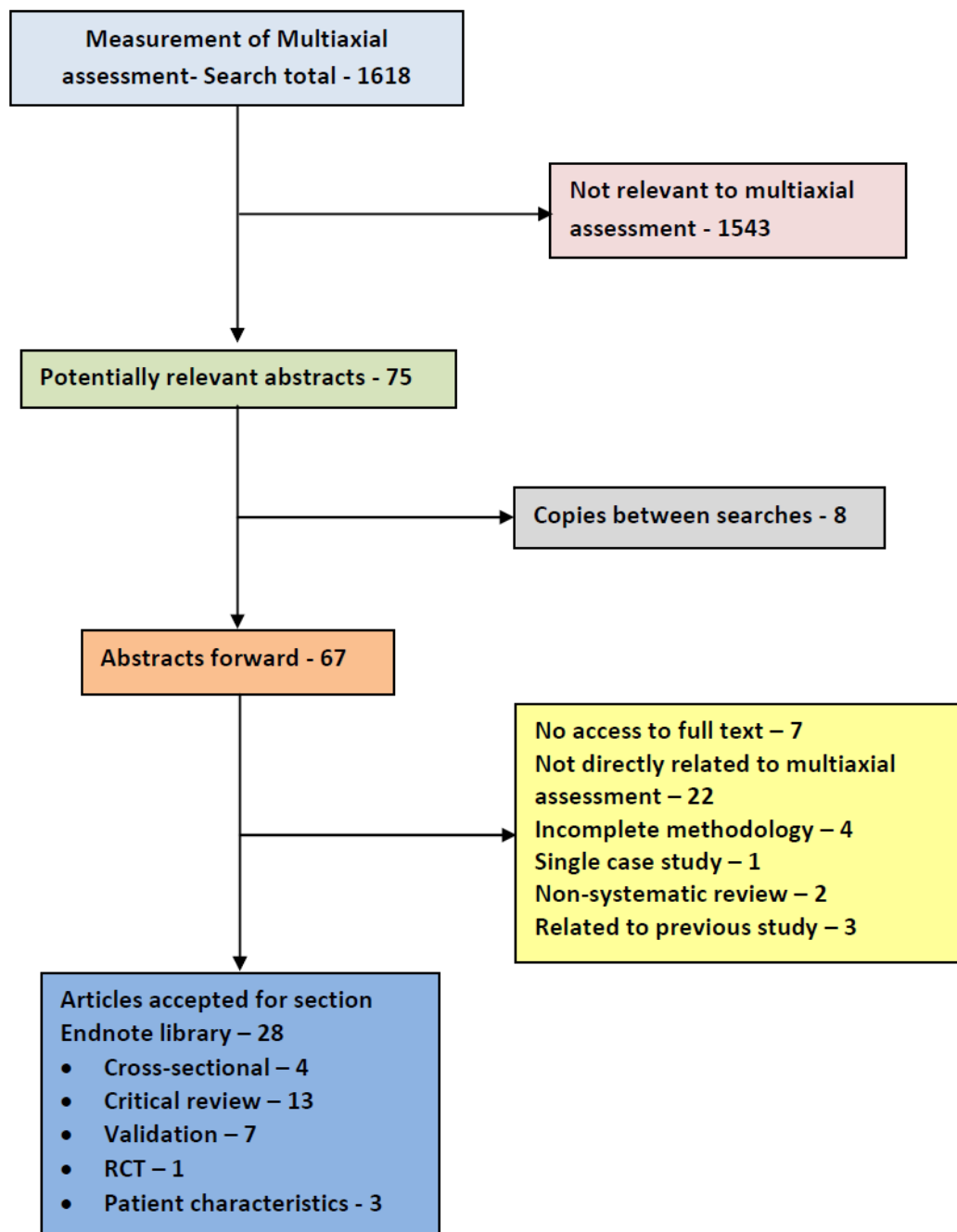
A flow chart describing the search protocols for articles examining the patient-practitioner relationship and the presentation of FGIDs



A flow chart describing search protocols for area of consideration and possible measurement in patients presenting with FGIDs



A flow chart describing search protocols for articles investigating the use and validity of multiaxial assessment criteria



Appendix D: Tables of Articles used in the Systematic Review

Risk factors

- Gender
- Case control studies

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Cain et al 2009	Prospective	Community adverts screened by telephone	Menstruating (n=89) Post-menopausal (n=660)	USA	Healthy matched	Gender-related differences in GI and somatic symptoms are apparent in IBS subjects. Stronger in postmenopausal women.
2	Franke et al 2010	Retrospective	IBS Outpatients at university hospital	IBS (n=149)	France	Healthy matched male	Contribution of SERT activity is no uniform and is possibly gender specific. Results suggest that assessment of SERT function may help assess patients on drugs affecting 5-HT system
3	Houghton et al 2009	Retrospective	Outpatients from local GPs and adverts	IBS (n=73)	UK	Healthy matched male	Male and female subjects have raised concentrations of 5-HT. 5-HT concentrations normalise at menses in IBS women
4	Houghton et al 2000	Retrospective	Outpatients department	IBS-Male (n=50)	UK	Healthy matched	Results support need for further exploration of role of male sex hormones in IBS pathophysiology
5	Miller et al 2004	Retrospective	Outpatients from secondary care clinic	IBS male (n=70)	UK	Healthy matched	Men with IBS exhibit lass male characteristics, but it remains to be determined if this is cause or effect
6	Saito et al 2010	Retrospective	Adult outpatients	IBS (n=477) First degree relatives (n=1492)	USA	Healthy matched	IBS aggregates strongly in families that varies somewhat by relationship to proband

• **Patient Characteristics**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Anbardan et al 2012	Retrospective	Single centre study	144	Iran	NA	Gender – important in IBS although comparisons were statistically insignificant. Rome III suggests higher prevalence of bowel movement and looser stools in males and more nausea reported in females
2	Berman et al 2000	Retrospective	IBS diagnosed patients recruited from UCLA GI clinic	30	USA	NA	Greater insula activation with visceral stimulation in male as compared to female IBS patients
3	Keogh et al 2007	Retrospective	Healthy adult staff & students at a University	50	UK	NA	Results are inconsistent with predictions of pain thresholds being related to digit ratio in women but not men
4	Kumar et al 2010	Retrospective	1 st to 3 rd year medical students	40	India	NA	Pain studies should only be carried out during a particular phase of menstrual cycle in order to increase reproducibility
5	Nisenblat et al 2010	Retrospective	Patients recruited from assisted reproduction unit	31	Israel	NA	Although pain perception at different oestrogen levels remained constant, enhancement of pain at supra-physiological levels suggest possible role of sex hormones
6	Schmulson et al 2010	Retrospective	Database of subjects completing Rome II questionnaire	1021	Mexico	NA	Gender differences exist with both IBS and FD being more common in women, especially bloating and constipation
7	Sullivan et al 2000	Retrospective	subjects enrolled on a psychology course at university	80	Canada	NA	Women report more intense pain and display pain behaviour for longer duration than men. Women also scored higher on catastrophic thinking than men
8	Taub et al 1995	Retrospective	Psychology students routinely completing questionnaires	1344	USA	NA	The 3 core Manning symptoms have equal applicability to both gender and African-Americans as well as Caucasians
9	Tousignant-Laflamme et al 2009	Retrospective	Local advertisement & convenience sample of first 32 subjects	32	Canada	NA	Women have greater dorsal nucleus inhibitory centre activation during ovulatory phase. Higher pain perception reported during menstrual cycle

- **Cross-sectional studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Chang JY et al 2009	Retrospective	Random sampling stratified by age and gender	653	USA	NA	Chronic diarrhoea in the absence of IBS is common. Self-reported food sensitivity, male gender and lower level of education = risk factors
2	Lackner et al 2005	Retrospective	IBS subjects referred to clinics involved in NIH multisite study	281	USA	NA	Dysfunctional attitude scores elevated in IBS subjects. 11% variance in pain affect when controlling gender, age and psychopathology
3	Landau et al 2008	Retrospective	Recruited teenage military service personnel	466855	Israel	NA	GI diseases are not uncommon. Associations found between BMI with IBS and GERD , while female gender was associated with gallbladder disease
4	Lee OY et al 2001	Retrospective	Populations from UCLA FGID centre and adverts for clinical trials program	714	USA	NA	Female subjects report higher levels of a variety of non-intestinal sensory symptoms despite similar levels of IBS severity, abdominal pain & psych symptoms
5	Usai et al 2010	Retrospective	Village and town populations stratified by gender and age	980	Italy	NA	Higher prevalence of IBS in females. In towns, but no gender differences in rural population

- **Randomised controlled studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Kakeda et al 2011	Prospective	Recruited from billboard adverts on university campus	40	Japan	NA	Analgesic effect of sweet stimulus on pain thresholds is influenced by gender differences in human adults
2	Lampe et al 1993	Prospective	Healthy university community	80	USA	NA	Significant gender differences in bowel function and bile secretion observed when eating the same high fibre diet
3	Zheng et al 2008	Prospective	Postmenopausal subjects recruited from US clinics between 1993 - 1998	27347	Sweden	NA	Oestrogen treatment alone, but not with progestin may cause GERD in postmenopausal women. Increased weight = increased risk of developing GERD

- **Systematic reviews**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Popescu et al 2010	Retrospective	Medline, EMBASE, Biosis, Web of Science, PsychINFO, Cochrane	17 studies included	USA	NA	Gender differences in DNIC effect depend on both experimental methodology and modes of measurement of the effect
2	Racine et al 2012	Retrospective	Medline, EMBASE, CINHALL, BIOSIS, PsychINFO	172 studies included	Canada	NA	10 years of research have not been successful in producing clear consistent patterns of sex differences in human pain sensitivity

- **Cohort studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Lee et al 2007	Prospective	Patients recruited from digestive diseases clinic	372	South Korea	NA	GI symptoms are more frequent in women especially during menstrual phase

- **Impact of daily life**

- **Case-control studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Akerhurst et al 2002	Prospective	Multi primary care site recruitment	IBS (n=161)	UK	Healthy matched	IBS affects patients through reduced HRQoL, more time off work and greater health care utilisation
2	Balboa et al 2006	Retrospective	Multi primary care site recruitment	IBS (n=517)	Spain	Healthy matched	Upper GI tract symptoms present in IBS patients impair HRQoL, psychological status, especially IBS-C and IBS-M than IBS-S
3	Bray et al 2006	Retrospective	New patients at hospital GI clinic	IBS (n=32)	UK	Non-IBS matched	Symptom interpretation does not differ between IBS and non-IBS patients. Tendency to attribute physical disorders does not explain why IBS patients seek care
4	De Vries et al 2007	Retrospective	Tertiary GI clinic and advert recruitment	GERD (n=263)	Holland	Healthy matched	GERD patients with IBS symptoms have much lower HRQoL and higher prevalence of health care seeking

5	Dibonaventura et al 2011	Retrospective	Data from 2007 National health and wellness survey	IBS (n=789)	USA	Healthy matched	IBS-C – poorer HRQoL, greater work productivity loss, activity impairment and greater health care use
6	Eugenio et al 2012	Retrospective	Adverts and database follow up info on RCT for self-management program for female IBS	Sexual dysfunction (n=89)	USA	IBS- no sexual dysfunction	IBS severity did not differ between groups, Self-management can reduce the effect of IBS on sexual dysfunction
7	Faresjo et al 2006	Retrospective	Random selection of medical databases in Sweden and Greece	IBS Sweden (n=90) IBS Greece (n=30)	Sweden	Health matched	Living in different cultures could perceive disease differently. Greek female IBS patients affected more seriously than Swedish counterparts
8	Frank 2003	Retrospective	Multiple chronic disease recruitment from community and clinics and published US population norms	IBS (n=140) FD (n=126) GERD (n=516) Asthma (n=375) Migraine (n=303)	USA	<ul style="list-style-type: none"> Published data on national norms Untreated community 	IBS – associated with impairment of HRQoL relative to US norms and all other chronic illness tested. Greater HRQoL appears greater in IBS with panic disorder
9	Guthrie et al 2003	Retrospective	IBS patients not responding to treatment at 7 UK hospitals	107	UK	Healthy matched	Marked differences across 3 rectal distension threshold groups were marked by the level of psychological comorbidity and child abuse
10	Halder et al 2004	Retrospective	Population random sample	112	USA	Those reporting no abdominal pain and < 2 other GI symptoms	HRQoL in community is impaired in IBS subjects and FD. However, much of this association can be explained by psychological factors
11	Houghton et al 1996	Retrospective	Refractory IBS patients treated with hypnotherapy	25	UK	IBS patients on hypnotherapy waiting list	Hypnotherapy profoundly improves symptoms of IBS and HRQoL

12	Huerta et al 2002	Retrospective	Consecutive patients to GI clinic	46	Mexico	Healthy matched	Under basal conditions there are differences in anxiety, weakness, attention and arousal between IBS and controls. The Spanish stress symptom rating questionnaire is a good instrument to evaluate stress during physiological paradigms
13	Jones et al 2007	Retrospective	IBS patients recruited from GI clinic in university hospital	84	USA	Matched IBD patients	Both groups had significantly poorer HRQoL especially internet responders versus clinic-based patients
14	Jones et al 2006	Retrospective	Consecutive patients from GI clinic	IBS (n=74) IBD (n=48)	USA	Health matched	Compared to controls, IBS & IBD have increased levels of psychological stress and poorer HRQoL & greater reliance on coping strategies
15	Kanazawa et al 2004	Retrospective	Young adults seen for annual health check	417	Japan	Healthy matched	Parental history of bowel problems and acute gastroenteritis are significant risk factors for IBS development in Japan

16	Koloski et al 2005	Retrospective	Subjects whom participated in random population surveys	IBS / FD	Australia	Healthy matched	The belief that the presence of serious pathology characterises community subjects with IBS and FD, but not health care seeking
17	Koloski et al 2003	Prospective	Random sample from previous study database	IBS (n=361)	Australia	Health matched	Psychological stress, not related to GI symptoms over a 1-year period. It is however, linked to persistent GI symptoms and frequent health care seeking behaviour
18	Lackner et al 2006	Prospective	Randomly sampled from previous NIH trial for psychological treatment for IBS	IBS (n=6)	USA	Healthy matched	A regimen of cognitive therapy is associated with changes in neural activity deemed aberrant in IBS patients when testing visceral distension and brain scan devices
19	Lee et al 2010	Retrospective	Recruitment of FGID patients requiring upper and lower GI endoscopy	IBS / FD (n=279)	South Korea	Matched patients (no FGID symptoms)	Depressive mood was significantly related to FD and FD / IBS overlap, but not IBS alone based upon Rome III criteria
20	Levy et al 2004	Retrospective	Families from data base of large health care organisation	IBS mothers (n=208) and children (n=296)	USA	Healthy matched	Children with GI complaints whose mothers have IBS are not explained by mother's biased perceptions Children of mothers with IBS have more non-GI as well as GI symptoms, disability days and clinic visits
21	Morken et al 2009	Retrospective	Consecutive IBS patients with post metronidazole for G lamblia	27	Norway	Healthy matched	Patients with post giardiasis IBS suffer very little somatic comorbidity suggesting the aetiology if this form of post infectious IBS is predominantly biological
22	Naliboff et al 2012	Retrospective	IBS & IBD subjects seen at 2 university GI centres	IBS (n=564)	USA	Matched IBD patients	Psychological stress is less dependent on GI symptom severity in IBS compared to IBD even though impact HRQoL is similar
23	Pace et al 2003	Retrospective	Consecutive IBS & IBD patients at tertiary clinic	IBS (n=80)	Italy	Matched IBD patients	IBS patients show HRQoL, psychological distress and recent stressful events that are comparable to matched IBD patients

24	Park et al 2009	Retrospective	Consecutive patients from 12 research institutes	IBS (n=664)	South Korea	Matched general population	IBS-related symptoms – great effect on HRQoL in Korean patients. Results show significant prevalence and social impact
25	Piche et al 2010	Prospective	IBS and IBD patients from 5 university hospitals	IBS (n=40)	France	Matched Crohn's patients	IBS-like symptoms are elevated in quiescent Crohn's disease. This is probably associated with fatigue and depression disorders.
26	Posserud et al 2009	Retrospective	Consecutive patients referred to outpatient clinic	IBS (n=36)	Sweden	Matched organic GI disease	Compared to organic GI disease, IBS patients appear to be hypervigilant for information regarding GI sensations
27	Ringstrom et al	Retrospective	Recruitment of subjects via adverts and primary care	IBS (n=218)	Sweden	Matched non-consulting IBS subjects	GI symptom severity alone cannot explain IBS illness behaviour. HRQoL and psychological symptoms are important factors for health care seeking behaviour
28	Seres et al 2008	Retrospective	Recruitment of IBS & IBD patients from tertiary GI clinic	IBS (n=88)	Hungary	Matched IBD patients	Results suggest that there are differences between IBS and IBD in the role of physical and psychological factors in HRQoL and emphasises the importance of cognitive processes in IBS

29	Tkalcic et al 2010	Retrospective	Consecutive outpatients from GI clinic	IBS (n=56)	Croatia	Matched IBD patients	Patients with IBS are more prone to the effect of psychosocial variables on GI symptoms compared to IBD patients. IBS patients experience higher levels of anxiety and neuroticism.
30	Wang et al 2010	Retrospective	Random sample of long terms patients from a previous survey	IBS + pelvic floor disorders (n=204)	USA	Non-IBS with pelvic floor disorders	Women with IBS are more likely to report symptoms of pelvic organ prolapse and sexual dysfunction and poorer HRQoL
31	Whitehead et al 1996	Retrospective	Undergraduate student recruitment	IBS non consulter (n=41) IBS consulter (n=91)	USA	Healthy matched	IBS patients showed greater reductions in HRQoL than non-consulters who in turn showed greater impairment than controls. SF 36 may be a useful outcome measure

• Patient Characteristics

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Abdulmajeed et al 2011	Retrospective	Patients attending family health centre	117	Egypt	NA	IBS has significant impact on work, lifestyles and social well-being
2	Brun-Stang et al 2007	Retrospective	Sample of 1 st 4 IBS patients per physician over 3-month period	504	France	NA	IBS involves large medical costs to France and that IBS patients experience poor HRQoL than general population
3	Bryant et al 2011	Retrospective	Consecutive IBS patients at hospital outpatient clinic	162	Australia	NA	FGID symptoms in IBD patients with greater psych comorbidity and poorer HRQoL
4	De Gucht et al 2003	Retrospective	Nursing staff at university hospital	207	Holland	NA	Personality trait of neuroticism – significant predictor of somatisation whereas alexithymia predicted longer, more severe somatisation. All predicted psychological stress
5	Farndale et al 2011	Retrospective	Patient info from 2 research databases	18	UK	NA	Findings confirm both extensive impact of IBS on daily living, well-being and self-identity

6	Farrokhyar et al 2006	Retrospective	Consecutive In-patients with IBD	149	Canada	NA	Many IBD patients present with FGID symptoms which are associated with impaired HRQoL and increased health-care seeking
7	Frankhuisen et al 2009	Retrospective	Consecutive achalasia patients at outpatient clinic	131	Holland	NA	FD & IBS symptoms in patients treated for achalasia are common and have negative impact on HRQoL
8	Lackner et al 2010	Retrospective	Consecutive IBS patients recruited from GPS	105	USA	NA	There is a link between perceived adequacy of social support and global severity of IBS symptoms (especially pain)
9	Lee et al 2008	Retrospective	Subjects recruited from RCT for FGID self-care	578	UK	NA	Psychological & physical factors are significantly associated with impaired HRQoL in IBS patients in primary care
10	Longstreth et al 2005	Retrospective	Patients recruited from neuroenteric disease program	155	USA	NA	The IBS-impact scale is a short, user-friendly instrument with excellent psychometric properties
11	Nojkov et al 2010	Retrospective	Nursing staff at university hospital	552	USA	NA	Rotating shift work is associated with the development of IBS that is also independently associated with sleep quality
12	Reilly et al 2004	Retrospective	Employed patients with IBS from 5 US clinics	135	USA	NA	Discriminative validity of the work productivity and impairment questionnaire in IBS was established
13	Si et al 2004	Retrospective	Consecutive IBS patients from 10 hospitals	662	China	NA	IBS can cause generalised body discomfort and psychological problems that affect HRQoL. SF-36 may be a useful instrument for Chinese IBS patients
14	Simren et al 2001	Retrospective	Direct referrals from GI clinics and adverts	343	Sweden	NA	HRQoL is affected by gender but not IBS subgroup. Fatigue is a common symptoms in IBS that correlates with well-being

15	Spiegel et al 2004	Retrospective	Consecutive patients at university referral centre	770	USA	NA	HRQoL in IBS patients is primarily related to extraintestinal symptoms rather than GI
16	Spiegel et al 2005	Retrospective	Consecutive IBS patients with no previous colon surgery or recent colonoscopy	458	USA	NA	There is no independent association between negative colonoscopy and reassurance of improver HRQoL in IBS patients < 50 years
17	Tang et al 2012	Retrospective	Consecutive 1 st time outpatients at Nanjing hospital	452	China	NA	There are significant gender differences in symptoms, psychological ratings and HRQoL in IBS. Somatic symptoms, anxiety and depression all contribute to negative impact of IBS
18	Ten Berg et al 2006	Retrospective	Patients from pharmacies attending for mebeverine	375	Holland	NA	The burden of illness of IBS in Netherlands is substantial in IBS patients treated with mebeverine have low HRQoL
19	Trevidi et al 2011	Retrospective	Military personnel deployed in middle east	121	USA	NA	High prevalence of FGID symptoms in military personnel returning from deployment., Including reduced physical and mental HRQoL

- **Cross-sectional studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Aro et al 2011	Retrospective	Random sample from 2 Swedish populations	2122	Sweden	NA	FD impacts all main domains describing physical, mental and social aspects of HRQoL in general population. FD/IBS overlap impacts domain related to body pain
2	Deann et al 2005	Retrospective	Random sample of bank employees	1776	USA	NA	Reduced work productivity and diminished HRQoL found in this study may have substantial financial impact on employers
3	Faresjo et al 2006	Retrospective	Random sample from register data from 3 randomly sampled primary care centres	723	Sweden	NA	IBS patients appear to be heavier users of primary care than of those who attend. The majority are managed by the GP
4	Ford et al 2008	Prospective	Random sample of subject previously selected for H Pylori program	3873	UK	NA	Poor QoL at baseline – strong predictor of new onset IBS, but not IBS related
5	Hililila et al 2007	Retrospective	Random sample Finnish population	5000	Finland	NA	IBS consulters and non-consulters demonstrate high rates of comorbidity, also seeking health care for abdominal complaints rather than psychiatric comorbidity
6	Hu et al 2002	Retrospective	Random computer telephone number selection	1649	China	NA	IBS & FD associated with anxiety, depression and significant social morbidity, health care use and days off work
7	Jafri et al 2007	Retrospective	Random population attending primary care clinic	1048	Pakistan	NA	IBS is seen in both urban and suburban communities. Health care seeking behaviour is common in males with abdominal pain

8	Jeong et al 2008	retrospective	Random sample of urban and farming communities	217	South Korea	NA	HRQoL was significantly impaired in subjects with GERD, un-investigated dyspepsia and IBS
9	Kaii et al 2010	Retrospective	Workers attending routine health check	2680	Japan	NA	Overlap of GERD, FD and IBS were common and poorer HRQoL in Japanese general population
10	Katsinelos et al 2009	Retrospective	Patients attending 3 university hospitals seeking medical consultation	2397	Greece	NA	Prevalence of IBS in Northern Greece is relatively high, mainly affecting female participants living in urban areas
11	Koloski et al 2000	Retrospective	Random sample of population electoral role	2910	Australia	NA	FGIDs impair HRQoL, particularly in those that consult for health care
12	Masud et al 2001	Retrospective	House to house survey	2542	Bangladesh	NA	IBS is a problem for rural people in Bangladesh with prevalence almost identical to other countries. Only a minority seek health care
13	Minocha et al 2006	Retrospective	Convenience sample of adults in 9 locations including colleges, medical centres and churches	1000	USA	NA	Although IBS occurs less frequently in older adults, no statistical difference between IBS in older versus younger people. IBS affects HRQoL in all ages, social functioning was better in older compared to younger IBS subjects
14	Moghimidehkordi et al 2011	retrospective	Random sample of Tehran post codes, house to house questionnaires, non-responders replaced	18180	Iran	NA	IBS is moderately high in Iran and imposes heavy financial burden on Iranian national health system due to high prevalence and impact on HRQoL
15	O'Keefe et al 1995	Retrospective	Random sample of Olmstead county in > 65 years	704	USA	NA	FGIDs appear to interfere with daily living and HRQoL in elderly persons
16	Olafsdottir et al 2011	Prospective	Random sample of national register	799	Iceland	NA	Heartburn is common as a chronic condition. Subjects with BMI below or above the average are more likely to experience heartburn which had impact on daily activities, sleep and HRQoL
17	Talley et al 1997	Retrospective	Random sample of suburban residents from electoral role	730	Australia	NA	Psychological factors do not seem to explain health care seeking among IBS subjects

18	Williams et al 2006	Retrospective	Sampled from previous US population IBS survey (random dialling)	1697	USA	NA	Health care seeking behaviour among IBS patients was determined by presence of comorbidities and extent of IBS affecting HRQoL
----	---------------------	---------------	--	------	-----	----	--

• Cohort studies

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Cain et al 2006	Retrospective & prospective	Community adverts & local health groups	242	USA	NA	Abdominal pain – most disruptive of IBS and symptom that has an independent & significant impact in those with diarrhoea
2	Drossman et al 2000	Prospective	Female IBS patients from university CBT and patient education program	156	USA	NA	IBS-QOL is responsive to treatment in a referral based clinical population of FGID patients
3	Nojkov et al 2008	Prospective	Consecutive patients undergoing oesophageal / gastro endoscopy	101	USA	NA	IBS and psychological stress impact GERD symptoms and HRQoL before and after proton pump inhibitor treatment
4	<u>Pare</u> et al 2006	Prospective	IBS Sample with > 1 symptoms recruited from 5 Canadian provinces including primary care, gastroenterologists, etc)	1555	Canada	NA	Evaluation of baseline data indicates that patients with IBS symptoms have a history of multiple treatments and have significant symptoms, burden impaired HRQoL, reduced work and low health status
5	Resendiz-Figueroa et al 2008	Prospective	Consecutive Tertiary care patients at national institution	74	Mexico	NA	Anxiety and depression negatively impact the number of days, frequency of symptoms and HRQoL in patients with moderate to severe IBS

• Randomised controlled studies

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Guglielmetti et al 2011	Prospective	Subject recruitment from multi centres and adverts	122	Italy	NA	Probiotic MMBb75 alleviates global IBS and improves HRQoL

2	Choi et al 2011	Prospective	IBS patients recruited at University hospital	67	South Korea	NA	S. boulardi improves IBS-QOL better than placebo but not superior for individual symptoms in IBS-D patients
----------	-----------------	-------------	---	----	-------------	----	---

• **Systematic reviews**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	El-Serag 2002	Retrospective	Medline, EMBASE & Cochrane (English and non-English – 1980-2001)	17 articles	USA	NA	There is reasonable evidence for a decrease in HRQoL in patients with moderate to severe IBS. However, data is conflicting regarding the impact of IBS in daily life in population-based studies of non-consulters

- **Physical risk factors**

- **Case-control studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Mearin et al 2005	Prospective	Local residents affected by salmonella outbreak	677	Spain	Healthy matched	Salmonella gastroenteritis is a significant risk factor for both IBS (1 in 10) and FD (1 in 7) at 1-year follow up.
2	Muhammed et al 2010	Retrospective	Referred patients with rectal evacuatory dysfunction	100	UK	Healthy matched	The greater prevalence of joint hypermobility demonstrated significantly higher frequencies in morphological abnormalities than those without
3	Myasoedova et al 2011	Retrospective	Subject data from epidemiological project on rheumatoid (RA) patients	493	USA	Healthy matched	Several upper and lower GI disorders were significantly more prevalent in RA versus non-RA subjects especially aged < 60. Physical impairment was also associated with GI disorders
4	Parry et al 2003	Prospective	Identified gastroenteritis asked to take part maximum 2 weeks post stool sample	500	UK	Healthy matched	Symptoms consistent with IBS-D are more frequent compared to controls even after careful exclusion of those with pre-existing FGIDs. Frequency is similar at 3 and 6 months
5	Ruigomez et al 2007	Prospective	Information taken from primary care research database	5894 with gastroenteritis	Spain	Healthy matched	IBS risk in individuals after gastroenteritis was 2 fold greater than the general population. Pre-existing psychological and GI comorbidities independently increase the risk of developing IBS
6	Sperber et al 2008	Prospective	Scheduled group to undergo gynaecological surgery	132	Israel	Outpatients not undergoing surgery	Only psychosocial variables predicted pain development. Among women undergoing non-pain related surgery, development of IBS was no greater than controls

- **Patient characteristics**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Burgmann et al 2006	Retrospective	Manitoba criteria patients contacted in telephone survey	396	Canada	NA	Older patients and those with likely and possible pre-existing IBS are more likely to experience longer symptom duration before IBD diagnosis
2	Neal et al 1997	Retrospective	Postal questionnaire to patients with previously confirmed gastroenteritis	386	UK	NA	Persistent bowel symptoms commonly occur after bacterial gastroenteritis and is responsible for morbidity and health care costs
3	Petrone et al 2011	Retrospective	Colorectal surgery databases	77	USA	NA	Physicians should consider small intestinal bacterial overgrowth in differential diagnosis of patients with normal anatomical findings and lower GI complaints
4	White et al 2010	Retrospective	Consecutive female veterans	337	USA	NA	Women veterans report high frequency of physical and sexual trauma. A lifetime history of as broad range of traumas is independently associated with an increased risk of IBS

- **Cross-sectional studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Aquas et al 2011	Retrospective	1 st degree relatives of IBD patients using health care database	360	Spain	NA	IBS prevalence in 1 st degree relatives of IBD patients is significantly greater which may suggest genetic and psychological involvement
2	Gulewitsch et al 2011	Retrospective	Students from multi universities	2196	Germany	NA	IBS is common in German university students with associated with reduced HRQoL. Student reporting childhood abdominal pains are at risk of IBS
3	Jung et al 2007	Retrospective	Random community population sampling	2273	USA	NA	IBS and GERD overlap is common and does not occur by chance

4	Kubo et al 2011	Retrospective	Workers consulting for annual health check	6334	Japan	NA	IBS was common and associated with young age, female gender, low BMI and allergic diseases in Japanese adults
5	Lee et al 2009	Retrospective	Multi-stage systematic sampling of community subjects	1443	South Korea	NA	Overlaps between GERD and IBS are common in the general population and are predominantly in individuals with anxiety
6	Locke et al 2000	Retrospective	Random community population sampling	643	USA	NA	IBS is significantly associated with analgesic use. This is confounded by somatic pain complaints. IBS symptoms are also associated with many food allergies which need further investigation
7	Marshall et al 2006	Retrospective	Subjects drawn from epidemiological study on long term outcomes of water contamination	4315	Canada	NA	PI-IBS is common after gastroenteritis from water contamination and is often IBS-D
8	Son et al 2009	Retrospective	Recruitment from 5 high schools	405	South Korea	NA	IBS prevalence in female adolescent students is 25.7% and has significant associations with anxiety and depression
9	Talley et al 1995	Retrospective	Stratified random sample of local community	3022	USA	NA	The onset of IBS may not be limited to early adulthood and that IBS subgroups based on bowel pattern may not identify clinically distinct entities
10	Tuteja et al 2008	Retrospective	Employees at veterans affairs health care systems	723	USA	NA	Bloating is common in healthy adults and is often but not predictive of FGIDs. Smoking and high dose aspirin are associated, but physical activity is not
11	Zarate et al 2009	Retrospective	Consecutive patients attending neurogastroenterology clinic	146	UK	NA	This preliminary study found high incidences of joint hypermobility in patients referred for tertiary GI care. Symptoms and functional tests suggest dysmotility

- **Cohort studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Favreau et al 2012	Prospective	Consecutive post haemorrhoidal surgery	369	France	NA	Persistent functional pain remains the long-term factor associated with dissatisfaction post-surgery
2	Neal et al 2002	Prospective	Disease record system. Residents with food poisoning contact at 6 years	436	UK	NA	PI-IBS differs from non-infective IBS by having more diarrhoeal features Less than half of both PI-IBS and non-infective IBS recover in 6 years. A history of anxiety and depression is enough to warrant treatment
3	Spence et al 2007	Prospective	Consecutive sampling of confirmed campylobacter patients	620	NZ	NA	Patients with high stress and anxiety levels are more prone to develop IBS post gastroenteritis. Additional risk factors include a tendency to interpret illness in a pessimistic fashion

- **Systematic reviews and meta-analysis**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Halvorson et al 2006	Retrospective	Medline, EMBASE, Cochrane	81 articles	USA	NA	Evidence for PI-IBS as a sequela to gastroenteritis with a 7-fold increase in the odds of developing IBS following gastroenteritis. Primary intervention may reduce likelihood of PI-IBS
2	Suares et al 2011	Retrospective	Medline, EMBASE, EMBASE classic	3278 articles	UK	NA	Pooled evidence of chronic constipation showed higher rates in women, older individuals, lower socioeconomic status and the presence of IBS
3	Spiller et al 2007	Retrospective	EMBASE, Medline, world of science, Cochrane and personnel databases		UK	NA	Better ways of identifying which patients with IBS will respond to specific treatment are urgently needed

- **Psychological risk factors**

- **Case-control studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Beesley et al 2010	Retrospective	Consecutive IBS patients at outpatients GI clinic	151	UK	Matched Crohn's disease	Higher levels of anger characterise IBS patients compared to organic bowel disease groups, but does not explain link between IBS and childhood abuse
2	Blomhoff et al 2000	Retrospective	Female subjects recruited from primary care data base and local adverts	IBS (n=40)	Norway	Healthy matched	IBS patients appear to have been hyperactive to auditory stimuli compared to controls. Later elements (P300, N400) of stimulus processing were influenced by emotion and personality
4	Butt et al 2012	Retrospective	IBS patients visiting GI outpatient clinic	82	Pakistan	Neurological and medical patients	Findings show that common mental disorders are more common and strongly associated with IBS compared to chronic disease
5	Chang L et al 2009	Retrospective	Recruitment from local adverts and FGID clinic	IBS (n=25)	USA	Healthy matched	While dysregulation in stress-responsive systems such as the HPA axis and mucosal immune function are demonstrated in IBS, they do not show a primary role in modulating symptoms
6	Delvaux et al 1997	Retrospective	IBS patients recruited from 8 university hospitals and other diseases from appropriate clinics	IBS (n=196)	France	<ul style="list-style-type: none"> • Organic GI, • ophthalmic patients • healthy matched 	There is a high prevalence of sexual abuse among IBS patients consulting in GI clinics. Some of these patients would benefit from appropriate therapy, namely psychotherapy.

7	Dinan et al 2006	Retrospective	Subject recruitment from hospital outpatient GI clinic	IBS (n=76)	Ireland	Healthy matched	IBS is characterised by an over-activation of the HPA axis and pro-inflammatory cytokine increase
8	Eslick et al 2011	Retrospective	Subjects recruited from 2 previous population-based studies	Unexplained chest pain (n=27)	Australia	Healthy matched	A history of childhood emotional / verbal abuse is a risk factor for unexplained chest pain. The association may be moderated by psychological distress, specifically depression
9	Heitkemper et al 2011	Retrospective	Community adverts between 2001-5	IBS (n=40)	USA	NA	Women with IBS who self-report childhood abuse, neglect are more likely to report disturbed sleep, somatic symptoms and psychological distress
10	Hobbis et al 2002	Retrospective	Chronic constipation (CIC) and IBS subjects recruited from GI tertiary care clinic	IBS (n=50) CIC (n=53)	UK	<ul style="list-style-type: none"> • Crohn's disease • Healthy matched 	Results challenge current assumptions that past abuse experiences may be significant in later presentations of FGIDs, but suggest that previous abuse may be related to a general level of psychopathology
11	Koloski et al 2006	Retrospective	Subjects whom previous participated in a random population survey	385	USA	No abdominal symptoms for > 1 year	Abuse occurring as an adult was significantly associated with IBS and / or FD but not an important factor when psychological factors were controlled for these disorders
12	Kurland et al 2006	Retrospective	Recruitment from 2 private rheumatology clinics	Fibromyalgia (n=105)	USA	Non-fibromyalgia patients	The prevalence of IBS and depressive symptoms was higher in fibromyalgia patients compared to control population

13	Lee et al 2010	Retrospective	Patients visiting hospital GI clinic	124	South Korea	Healthy matched	Depressive mood was significantly related to FD and IBS overlap but not to IBS based on Rome III. IBS / FD overlap have poorer HRQoL than FD or IBS alone
14	Porcelli et al 1998	Retrospective	Consecutive patients referred for FGIDs to tertiary care clinic	129	Italy	Gallstone disease	Findings show associations between FGID symptoms and eating disorders and show that these symptoms may persist after the recovery from the eating disorder
15	Reilly et al 1999	Retrospective	Consecutive patients from 2 outpatient clinics	IBS (n=40) Non-epileptic attack disorder (n=40)	UK	<ul style="list-style-type: none"> • Crohn's disease • epilepsy 	Presentation of functional neurological and abdominal symptoms are characterised by a history of abuse. Compared to organic disease groups
16	Ringel et al 2003	Retrospective	Recruitment from local adverts and FGID centre	IBS (n=6)	USA	Healthy matched	Findings replicate previous alterations in brain responses to rectal distension in patients with IBS compared to healthy controls
17	Ringel et al 2004	Retrospective	Female IBS patients enrolled on a multicentre clinical trial	IBS + abuse history (n=74)	USA	IBS – no abuse history	Severe sexual and physical abuse is associated with higher urge and pain thresholds for rectal distension
18	Ringel et al 2008	Retrospective	All subjects recruited from campus, hospital general and GI clinics	IBS (n=10)	USA	No IBS	Pain ratings during rectal distension are associated with activation of dorsal cingulate regions implicated in homeostatic afferent processing which prior abuse enhances its activation
19	Ross et al	Retrospective	Subject recruitment from tertiary GI clinics and newspaper adverts	IBS (n=29)	USA	<ul style="list-style-type: none"> • IBD • Other GI disorders 	The subjects in the 3 groups did not differ in dissociative experience scale or SCL-90. However, subjects with IBS reported much higher rates of childhood sexual abuse and psychosomatic symptoms

20	Rubenstein et al 2007	Retrospective	Subjects recruited from tertiary GI clinic and medical procedure unit	Proton pump non-responders (n=9)	USA	Proton pump responders	Heartburn subjects have heightened oesophageal sensation suggesting that oesophageal hypersensitivity may persist despite therapy due to its association with psychiatric disease
21	Salmon et al 2003	Retrospective	Subject recruitment from Tertiary GI clinic at teaching hospital	IBS (n=64)	UK	IBD patients	Findings consistent with model in which childhood abuse is linked to IBS because it causes a tendency to dissociates which then causes general increase in physical symptoms
22	Saps et al 2011	Retrospective	Data from children diagnosed with pyloric stenosis during infancy	100	USA	Healthy matched	Findings suggest that infant pyloric stenosis and associated perioperative care represent risk factors in development of chronic abdominal pain at long-term follow up
23	Tang et al 1998	Retrospective	IBS Subjects referred by gastroenterologists selected from university	60	Canada	Healthy matched	Findings suggest the need for more comprehensive understanding of IBS and eating disorders. No significant correlations
24	Taymur et al 2007	Retrospective	IBS Subjects consulting university hospital	34	Turkey	Health matched	Psychiatric disorders can have a role in occurrence or exacerbation of IBS
25	Tousignant-Laflamme et al 2006	Retrospective	IBS subjects recruited from tertiary GI outpatient clinics	14	France	Healthy matched	IBS subjects demonstrate different autonomic nervous system response to pain
26	Vidlock et al 2009	Retrospective	IBS subject recruitment from GI clinic and local adverts	45	USA	Health matched	HPA axis hyper-responsiveness to visceral stressors is related to early adverse life events than the presence of IBS. The HPA axis reactivity has a moderating effect on IBS symptoms
27	Walker et al 1990	Retrospective	Subject recruitment from	IBS (n=28)	USA	IBD patients	Significantly more IBS subjects had

			tertiary GI clinic, family medical centre and private GI clinic				lifetime diagnoses of major depression, somatisation, generalised anxiety disorder and phobic disorder
28	Whitehead et al 1997	Retrospective	Newspaper adverts requesting IBS and health controls	IBS abuse (n=17), IBS non-abuse (n=15)	USA	Healthy matched With abuse and non-abuse history	A history of sexual abuse does not contribute significantly to rectal pain threshold sensitivity
29	Woodman et al	Retrospective	IBS subject recruitment from university GI centre	20	USA	Previous laparoscopic cholecystectomy	IBS subjects had significantly more lifetime psychiatric illness , than controls. There was no significant difference between IBS subjects and relatives
30	Yilmaz et al 2011	Retrospective	Subjects referred to university research clinic	IBS + asthma (n=101)	Turkey	Asthma without IBS	Psychiatric disorders were more common with IBS + asthma. Asthma + IBS patients had significantly less forced expiration volumes compared to non-IBS asthma patients
31	Zapata et al 2005	Retrospective	Inpatient services recruitment	282	Peru	Non-Rome criteria patients	Depressive symptoms and lower GI symptoms have a high frequency in hospitalised patients. However there was no relationship between IBS and hospitalisation

- **Patient characteristics**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Abraham et al 2011	Retrospective	Consecutive female patients admitted for eating disorders	160	Australia	NA	The presence of IBS (but not other FGIDs) in eating disorder patients are strongly related including psychological feelings. The poorer the QoL eating disorder, the poorer the IBS HRQoL
2	Baccini et al 2003	Retrospective	Consecutive patients to GI clinic with over 1 year of GI symptoms	226	Italy	NA	A history of physical and/or sexual abuse has a high prevalence in patients with chronic FGIDS irrespective of organic diagnosis. Abuse history has no relevant role in the pathogenesis of either organic or functional GI disorders but can affect clinical expression
3	Blanchard et al 2004	Retrospective	Patients seeking non-drug treatment for IBS symptoms	196	USA	NA	Patients with abuse history have significant increase in criteria for DSM axis I (especially substance abuse disorders, dysthymia and GAD
4	Boyd et al 2005	Retrospective	Consecutive female inpatients attending specialised eating disorder unit	101	Australia	NA	In patients with eating disorders, especially psychological traits predict FGID type and the presence of multiple coexisting FGIDs
5	Drossman et al 1990	Retrospective	Consecutive female patients at university GI clinic	206	USA	NA	A history of sexual or physical abuse is a frequent yet hidden experience in women seen in referral based GI practice and is common in those with FGIDs. A history of abuse is associated with greater symptom .reporting and lifetime surgeries
6	Farnam et al 2008	Retrospective	Continuous IBS patients attending university outpatient clinic	166	Iran	NA	Evidence suggests differences between IBS patients and general population that shows personality dimension according to symptoms-based subtypes of IBS

7	Geeraerts et al 2009	Retrospective	Consecutive patients with newly diagnosed FD	233	Belgium	NA	Abuse history was not associated with differences in gastric emptying. Abuse history is associated with alterations in gastric sensorimotor functioning in FD, particularly sexual abuse
8	Hansel et al 2010	Retrospective	Data collected from consecutive adult patients undergoing glucose and hydrogen breath tests for GI symptoms	230	USA	NA	Type D personality was associated with decreased perceived HRQoL and reported more severe GI symptoms
9	Heitkemper et al 2001	Retrospective	2 samples recruited from local community adverts plus telephone screen	125	USA	NA	Childhood abuse history is elevated among women with IBS, however, those with an abuse history do not appear to be different from female IBS patients with no history of abuse
10	Perkins et al 2005	Retrospective	Subjects recruited from eating disorder unit	234	UK	NA	Preliminary findings suggest that eating disorders may increase the risk of developing IBS
11	Riedl et al 2009	Retrospective	IBS Outpatients consulting at university medical centre for FGIDs	161	Germany	NA	Subjective theories of illness can have significant implications for IBS symptom severity as well as for physical and mental QoL
12	Scarinci et al 1994	Retrospective	Consecutive GERD, NCCP and IBS patients recruited for ongoing pain perception study	5079	USA	NA	Data suggest that relationships between abuse, disability, multiple pain syndromes and health care seeking behaviour are mediated by abnormal pain perception, psychiatric disorders, disruption of physical function and environmental stressors.
13	Talley et al 1995	Retrospective	Consecutive patients at large clinic	IBS (n=997)	USA	NA	Outpatients who report abuse are more likely to have IBS type symptoms

- **Cross-sectional studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Hililila et al 2008	Retrospective	Random sample of Finnish national registry	5000	Finland	NA	Depressive symptoms are prevalent in the general population which are associated with high rates of GI symptoms leading to increased use of health care services and work absenteeism
2	Klooker et al 2009	Retrospective	Dutch famine birth cohort (1943-7)	816	Holland	NA	Findings suggest that exposure to severe wartime conditions in early life is associated with increased risk of developing IBS attributed to stressful environment, under-nutrition and increased prevalence of infectious diseases
3	Lackner et al 2004	Retrospective	Subject referral to academic, behavioural, medical unit	244	USA	NA	Relationships between depression and pain is not direct or linear, but works partly through the patient's beliefs regarding their pain and pain catastrophising in specific
4	Lee et al 2009	Retrospective	Random sample of domestic households via telephone survey	2005	China	NA	IBS and general anxiety disorder was common and added to impairment in the community. The strong association between psychiatric comorbidity and IBS was observed in referral centres was not a consequence of help seeking behaviour
5	Lembo et al	Retrospective	Randomly selected twin pairs from twin registry	3083	USA	NA	Extraintestinal symptoms are independently associated with both IBS and psychiatric disorders. Evidence suggests a genetic basis to these symptoms , but association with IBS and psychiatric disorders is not primarily explained by genetic influences

6	Minocha et al 2010	Retrospective	Prospective collection sample of adults in 9 locations	990 (570 African American) (320 Caucasian)	USA	NA	Substantial similarities as well as differences in IBS patients of 2 races support the concept that while there is an important role for a biological component for IBS, it may not be an exclusive determinant
7	Perona et al 2005	Retrospective	Sample from women's' attention services	70	Spain	NA	Most women suffering domestic violence have FD and/or IBS plus elevated psychological distress
8	Thjissen et al 2010	Retrospective	Subjects recruited from secondary and tertiary outpatient clinics	268	Holland	NA	Dysfunctional cognitions independently influence physical and mental QOL and symptom severity
9	Talley et al 1998	Retrospective	Random sample from electoral role	730	Australia	NA	There is an association between abuse and IBS in the community but may be explained by other psychological factors
10	Ålander et al 2008	Retrospective	Random sample of National Swedish population registry	FGID (n=141)	Sweden	Symptom-free controls	Previous abuse is common in women with FGID and must be considered by the physician

- **Cohort studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Creed et al 2005	Prospective	Subjects recruited from 7 GI clinics	257	UK	NA	In severe IBS, the association between self-reported sexual abuse and impaired function is mediated by general tendency to report numerous bodily symptoms. A reported history of abuse is associated with marked improvement following psychological treatment
2	Han et al 2009	Prospective	Clinical referral and local adverts in local population	61	USA	NA	History of abuse did not appear to have any significant clinical correlation at baseline and did not predict treatment response

- **Randomised controlled studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Creed et al 2005	Prospective	Subject recruitment secondary and tertiary care clinics	257	UK	NA	Findings show that the number of psychological disorders associate with dose-response fashion. Improved depression was associated with increased role functioning. Depressive and panic disorders contribute to poor outcomes in sever IBS
2	Marks et al 2008	Prospective	Subjects recruited through clinical referrals and newspaper adverts	76	USA	NA	History of depressive and/or anxiety disorders was not associated with response of IBS symptoms to paroxetine.

- **Systematic reviews**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Tak et al 2011	retrospective	EMBASE, Medline, PsychINFO	85 articles	Holland	NA	No evidence was found to consider all 3 functional somatic disorders as hypocortisolemic as significant reductions in basal cortisol levels compared to healthy controls was only found in chronic fatigue syndrome, females with fibromyalgia, but not IBS

- **Genetic polymorphisms**

- **Case-control studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Andersen et al 2006	Retrospective	IBS subjects on Large clinic database	233	USA	Healthy matched	In contrast to reported association with FD, GNbeta3-C825T polymorphism is not associated significantly with lower FGID, with different IBS or FAP phenotypes
2	Camilleri et al 2010	Retrospective	population from prior publication which was sampled from local community	466	USA	Healthy matched	Expression of several neuropeptides is induced upon NPS-NPSR1 signalling. NPSR1 variants are associated with colorectal transit in FGID
3	Camilleri et al 2007	Retrospective	Sample recruited from local population (150 mile radius)	482	USA	Healthy matched	The association of genetic variation in metabolism of endocannabinoids due to FAAH polymorphism with symptom phenotype in IBS-D and IBS-M and with faster colonic transit in IBS-D supports hypothesis that cannabinoid mechanisms may play a role in controlling colonic motility

4	Camilleri et al 2009	Retrospective	IBS patients sampled from local community	466	USA	Healthy matched	No significant association of mitochondrial DNA genotypes tested and stomach volumes, small bowel or colonic transit and rectal compliance. Therefore, the association may lie with satiation, gastric emptying and possibly pain
5	DeVries et al 2009	Retrospective	Prospective sampling from hospital outpatient clinics	GERD (n=363)	Holland	Healthy matched	GERD is associated with GNB3C825T due to increased signal transduction upon GPCR activation associated with the 825T allele
6	Jun et al 2011	Retrospective	Sample from 3 case-control study DNA databases (female)	IBS (n=190)	USA	Healthy matched (female)	No significant associations between tryptophan hydroxylase polymorphisms with IBS
7	Karling et al 2011	Retrospective	Subjects consecutively referred from primary care to GI outpatient clinic	IBS (n=70)	Sweden	General population data from multiple outcome study	There is an association between Val/val genotype of the val158met COMT gene and IBS as well as to specific IBS related bowel patterns
8	Kilpatrick et al 2011	Retrospective	Subjects recruited from digestive disease clinic and adverts	IBS (n=26)	USA	Healthy matched	C/C genotype in HTR3A compared to T-carrier status is associated with increased anxiety and amygdala responsiveness during emotional and non-emotional tasks. This polymorphism is also associated with IBS symptom severity
9	Kohen et al 2009	Retrospective	Community adverts for subjects also part of a CBT intervention study	IBS (n=188)	USA	Health matched	Carriers of rare G allele rs 25531 increase odds threefold of IBS than controls. Findings suggest that further investigation of the possible role of SERT in the aetiology of IBS is warranted
10	Lee et al 2004	Retrospective	IBS patients referred to GI outpatient clinic	33	South Korea	Health matched	There was no relationship between SERT gene polymorphism and IBS. However S/S genotype was most prominent in subjects and controls

11	Li et al 2007	Retrospective	Consecutive IBS patients referred to GI clinic	87	China	Healthy matched	This study suggests that individuals with L/L SERT genotype are vulnerable to development of IBS-C. These patients also respond poorly to routine dose of tegaserod
12	Markoutsaki et al 2011	Retrospective	Consecutive patients to GI outpatients clinic	124	Greece	Healthy matched	The study suggests that the carriers of A allele of the 1438 (G/A) polymorphism of 5-HT2A receptor gene have a high risk of IBS
13	Markoutsaki et al 2001	Retrospective	As above	124	Greece	Healthy matched	The results suggest that SERT and GNB3 gene polymorphisms might be associated with IBS in Greeks
14	Niesler et al 2010	Retrospective	Subjects recruited from GI outpatient clinic	IBS-D (n=97) IBS-C (n=99)	UK	Healthy matched	Male IBS-D patients have reduced frequency of 5-HTTLPR (s/s) genotype which contradicts three earlier studies of similar size
15	Park et al 2006	Retrospective	Patients visiting for routine check up with no prior history of GI disease	IBS (n=190)	South Korea	Healthy matched	Significant association observed between SERT polymorphisms and IBS especially IBS-D
16	Park SY et al 2010	Retrospective	IBS outpatients having routine check up	80	South Korea	Healthy matched	CCK receptor polymorphisms (receptor intron 1779T>C) were associated with IBS-C and IBS-A
17	Pata et al 2004	Retrospective	Consecutive outpatients from university clinics	IBS (n=55)	Turkey	Healthy matched	This study suggests that patients with homozygote C allele of the 102T/C or homozygote A allele of the 1438G/A polymorphisms in the 5-HT2A receptor gene have a high risk of IBS
18	Saito et al 2009	Retrospective	IBS subjects seen at GI outpatients and data gained from human genetic cell repository	49	USA	Healthy matched	The G298S-SCN5A miss-sense mutation causes marked reduction of whole cell sodium current and loss of function of Na ^v 1.5 suggesting that SCN5A as a candidate gene for IBS pathophysiology

19	Sikander et al 2010	Retrospective	Subjects recruited from GI outpatients clinic	IBS (n=151)	India	Health matched	Significant associations between alpha 1291C>G polymorphism and IBS-D
20	Van der Veek et al 2005	Retrospective	Subjects recruited through GI outpatients clinic	IBS (n=111)	Holland	Healthy matched	Support for emerging hypothesis that genetically determined immune activity plays a role in pathophysiology of IBS with increases in TNF-alpha and reduction of IL-10 producers
21	Vazquez-Roque et al 2011	Retrospective	Subject whom had previously provided DNA samples for previous studies	IBS (n=94)	USA	Healthy matched	IBS-D patients who were positive for HLA-DQ8 or for both HLA-DQ2 and DQ8 have faster small bowel transit
22	Wang et al 2012	Retrospective	Subjects recruited from GI outpatients clinic	IBS (n=254)	China	Healthy matched	Polymorphisms in the promoter region of the SERT gene can influence the expression of SERT mRNA and levels of SERT protein in the colonic mucosa, thereby playing a key role in motility related IBS symptoms
23	Yeo et al 2004	Retrospective	Female patients whom provided leukocyte DNA samples for a previous phase III clinical trial	IBS (n=194)	USA	Healthy matched	Significant associations between IBS-D and SERT-P deletion genotype
24	Zang et al 2009	Retrospective	Patients referred to university hospital clinic	IBS (n=73)	China	Healthy matched	The TRPV1 gene polymorphism is associated with IBS development through the presence of C/C and C alleles
25	Zucchelli et al 2011	Retrospective	2 independent cohorts from Sweden and USA from secondary and tertiary care clinics	IBS (n=861)	Sweden	Healthy matched	NFSF15 is a susceptibility gene for IBS and IBS-C. As TL1A, the protein encoded by TNFSF15 contributes to the modulation of inflammatory responses, the results support the role of the immune activation in IBS

- **Patient characteristics**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Jarrett et al 2007	Retrospective	Participants of an ongoing intervention study recruited from community adverts	IBS (N=128)	USA	NA	Results support a biopsychosocial model of IBS in which SERT genotype modifies the risk for depression episodes
2	Kim et al 2004	Retrospective	Subjects selected from administration database of IBS patients in local community	IBS-C (n=90) IBS-D (n=128) IBS-M (n=38)	USA	NA	Functionally distinct alpha (2A) and (2C) adrenoreceptor and 5-HT transporter polymorphisms are associated with constipation and high somatic symptoms in lower GI disorders. Strength of genetic contribution is still unclear
3	Muiakovic et al 2011	Retrospective	Patients involved on multicentre RCT with dyspepsia without alarm symptoms	664	Holland	NA	The HTR3ac-42T allele is associated with severe dyspeptic symptoms, the stronger association among patients carrying the 5-HTTLPR-L allele
4	Saito et al 2007	Retrospective	IBS Outpatients seen in hospital GI clinic	103	USA	NA	This study suggests that the 5-HTTLPR polymorphism may be associated with IBS-M but not IBS overall. No association was observed for GNBeta3 C825T polymorphism with IBS

- **Randomised controlled studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Cremonini et al 2005	Retrospective	Direct mailing to patients from motility clinic	IBS (n=40)	USA	NA	Dexloiglumide (CCK receptor antagonist) accelerates gastric emptying and delays proximal, but not overall colonic transit

- **Systematic reviews**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Van Kerkhoven et al 2007	Retrospective	Medline, PubMed and world of science	8 eligible articles	Holland	NA	Genetic polymorphism in the gene encoding for activity of the SERT protein is not associated with IBS

- **The patient-practitioner relationship**

- a. **Case-control studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Heitkemper et al 2002	Retrospective	Random sample of IBS patients from earlier study and non-IBS from outpatient centres	1014	USA	<ul style="list-style-type: none"> • Non- IBS patients • Medical specialities 	Findings suggest differences between women with and without IBS in symptom perception as well as differences between physician and the patient

- **Patient characteristics**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Abioye et al 2010	Retrospective	Consecutive literate patients at university health centre	300	Nigeria	NA	A fifth of study subjects were dissatisfied with their clinician-patient relationship. Lack of interpersonal skills and application of medical knowledge were cited
2	Barakzai et al 2007	Retrospective	Random selection of IBS patient charts from 3 medical centres	139	USA	NA	Findings suggest that the clinician have a clearer understanding of IBS, appropriate treatment and the impact of culture on perception and description of symptoms
3	Casiday et al 2009	Retrospective	GP sample from UK, Holland	UK (n=30) Holland (n=15)	UK	NA	GPs diagnostic procedures for IBS are at odds with patient expectations and current guidelines
4	Hamberg et al 2004	Retrospective	Intern physicians	289	Sweden	NA	Findings suggest that gender bias is involved in medical management of IBS with men and women showing differing patterns of gender bias

- Cross-sectional studies

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Casiday et al 2009	Retrospective	Recruited from primary care network in UK and Holland	51	UK	NA	Clinicians should be aware of extensive impact of IBS on daily life activities. Clearly explaining guidelines for IBS diagnosis and the range of treatment options may help patient make sense of processes
2	Chassany et al 2006	Prospective	GPs recruited at Medical observation centre, spontaneous consulting patients for 3 different complaints	GPs (n=325) Patients (n=694)	France	NA	Patient and clinician related outcomes differed in addition; their relationship was not consistent across diseases. PROs are , therefore, essential to take account of all aspects of diseases
3	Dickman et al 2011	Retrospective	Hospital and community-based gastroenterologists and nurses nationwide	300	USA	NA	Physicians and nurses have different attitudes to perceptions of IBS and IBS concerning duration, treatment efficacy, personal control and illness coherence
4	Forrest et al 2002	Retrospective	Random digit dialling to households in 60 health care markets	18049	USA	NA	Managed health plans that loosen restriction on provider choice relax gate-keeping arrangements. Patients are likely to experience higher satisfaction with GP. Lack of health insurance impedes development of patient-practitioner relationships
5	Franke et al 2009	Retrospective	GPs in urban practices around large city	260	Germany	NA	Rome criteria and IBS pathogenetic Models are largely unknown among GPs. It is likely a small minority are referred to gastroenterologists
6	Halpert et al 2010	Retrospective	IBS patient sampling from multiple primary care centre, online adverts and newsletters	1242	USA	NA	Many patients think that IBS is caused by dietary habit developing into cancer, colitis causing malnutrition or worsening with age
7	Halpert et al 2006	Retrospective	IBS patient sampling from multiple primary care centre, online adverts and newsletters	1242	USA	NA	IBS patients expect more education than they receive. Patient learning preferences can be highly valuable in development of educational interventions

8	Harris et al 2008	Retrospective	IBS patients randomly selected from registers of 8 GP surgeries	8646	UK	NA	Most patients were willing to accept various forms of treatment. However, this population wanted focus to be directed at allaying fears and misconceptions about the IBS
9	Lapid et al 2009	Retrospective	Post graduate psychiatric residents from 7 training facilities	249	USA	NA	Findings suggest a perceived need for more education on many topics pertaining to boundaries and relationships. Those encountering ethical dilemmas wanted more education
10	Letson et al 1996	Retrospective	Opportunity sample of ward nurses from 18 hospitals	253	UK	NA	This study shows that majority of nurses have negative attitudes to IBS sufferers
11	Lionis et al 2009	Retrospective	Patient records from 15 primary care trusts in both remote and community hospitals	32117	Greece	NA	Gastroenteritis is a frequent health problem and GPs fail to adequately diagnose FD and IBS. Further training is needed
12	Lu et al 2009	Retrospective	Female IBS subjects recruited from university GI clinics and bulletin boards	12	Taiwan	NA	The bodily experiences and practices of women with IBS are developed from personal , cultural views of body and gender norms
13	Mercer et al 2012	Retrospective	Patients recruited from large university clinic	136	USA	NA	IBS and IBD patients look to clinicians as trustworthy advisors regarding the use of probiotics as an alternative supplements.
14	Sajjadi et al 2009	Retrospective	Drs sampled from 9 medical specialties	69	Iran	NA	Patient's satisfaction with different medical specialties is different from physicians' common-sense assumptions. Patients were more satisfied with gynaecologists rather than psychiatrists.
15	Schattner et al 2004	Retrospective	Patients recruited from a 600 bed teaching hospital	445	Israel	NA	Patients studied want their physician to be highly professional and expert clinicians to show humaneness, support, and most of all, respect their autonomy.
16	Shi et al 2003	Retrospective	Random sample from 60 random health care markets	18409	USA	NA	Enhancing primary care performance may reduce barriers to care experienced by vulnerable populations, thereby improving patient relationships with GPs

17	Smith et al 2009	Retrospective	Random telephone sampling for study evaluating lower and higher education groups	73		Australia	NA	Higher educated conceptualise their involvement in decision making in diverse way, sharing responsibility and acting as a resource in health decisions. Lower educated conceived as consenting to an option recommended by a Dr who makes the ultimate decision and describe the role of friends and relatives to help in such decisions
18	Sood et al 2011	Retrospective	Sampling of general and GI paediatricians affiliated to 4 hospitals where authors were working	<ul style="list-style-type: none"> GI Paeds (n=1351) Gen Paeds (n=500) 		USA	NA	Findings show a wide variety in evaluation and management of children with pain-related FGIDs. Rome III criteria have not reached the desired impact on practice behaviour
19	Spiegel et al 2010	Retrospective	Random sample of 4 medical specialities	950		USA	NA	Most community providers believe that IBS is diagnosed by exclusion and associated with increased resource use. Experts comply more closely with guidelines to diagnose IBS
20	Wolliscroft et al 1994	Retrospective	All 1 st year internal medicine residents and patients in pulmonary services	<ul style="list-style-type: none"> Patients (n=625) Residents (n=70) 		USA	NA	Patients, attending physicians, supervisors and nurses all view humanistic attributes differently. Nurses and supervisor ratings are more reproducible, but nurses perceptions correlate closely with the patient

• Cohort studies

	Author / date	Type of data collection	Population studied	Sample size		Location of study	control group	Findings
1	Bellini et al 2006	Prospective	Random sample of 28 GPs in large city	IBS patients (n=60)		Italy	NA	Findings suggest that the gap between specialists and GPS in the management of IBS should be examined to develop effective strategies for primary care
2	Besharat et al 2007	Prospective	Consecutive referral to hospital eating disorder unit	136		UK	NA	Videotapes assessed by 2 independent reviewers scored good inter-rater reliability in the Patient Response Style Scale, Self-Disclosure and Emotional Engagement self-report surveys

3	Cash et al 2011	Prospective	Online and flyers for educational initiatives to clinicians at a medical school	275	USA	NA	Learners achieved mastery in topics to IBS-C regardless of baseline knowledge or speciality. These data included continued medical education activities employing confidence-based learning to address knowledge gaps
4	Conboy et al 2010	Prospective	Subjects drawn from previous RCT testing impacts of patient-practitioner relationships on IBS	289	USA	NA	Opportunity for patient to discuss illness improved outcomes among patients. Supportive patient-practitioner relationships may overcome provider expectations for subjects
5	Coutts et al 2000	Prospective	Student performance data during 6-week family medicine rotation	428	USA	NA	Correlation between humanism and other performance measures is quite low. Humanistic students perform better than less humanistic peers, but current scoring methods do identify humanism
6	Dumitrescu et al 2006	Prospective	GPs from 2 counties invited for a course on IBS	100	Romania	NA	GPs appeared well-trained in recognising IBS. 66% estimated the prevalence of IBS as between 1 and 10%
7	Håkinsson et al 2011	Prospective	IBS patients on educational program waiting list	56	Sweden	NA	Coping patterns changed and symptoms were significantly less among participants on the education program
8	Keefer et al 2008	Prospective	Consecutive new patients at GI outpatient clinic	new patients (n=159) gastroenterologists (n=38)	USA	NA	Anxiety and depression were prevalent in this study yet gastroenterologists did not accurately detect these conditions. They were however twice as likely to diagnose anxiety patients as having an FGID
9	Lacy et al 2007	Prospective	All adult IBS patients screened over 3-year period	261	USA	NA	IBS patients believe that anxiety, diet and depression cause IBS, contrary to clinicians' views
10	Longstreth et al 2003	Prospective	GPs in major city area medical care plan	20	USA	NA	GPs had attitudes about IBS patients and lacked knowledge that could interfere with patient care. A single class improved short-term knowledge but had little effect on attitudes on IBS
11	Parkin et al 2006	Prospective	Patients and health	209 appointments	UK	NA	Professional centered training can be effective in

			professionals at hospital diabetes centre				improving patient perception of the consultation and increased patient – professional agreement on recall of clinical decisions
12	Raine et al 2004	Prospective	GPs and mental health patients identified from medical database	177 patients	UK	NA	Guidelines cannot be based on data alone. Clinical judgement is unavoidable and may be elicited and aggregated with nominal group technique in a structured way
13	Ringstrom et al 2009	Prospective	IBS patients referred to secondary care unit	86	Sweden	NA	IBS patients appear to have knowledge about their disorder, but do lack useful knowledge as they are suffering from a disabling, yet medically harmless disease
14	Roberts et al 2011	Prospective	Random sample of 4 th -year medical students from 23 Osteopathic med schools		USA		Findings are consistent with interpretation that humanistic clinical skills such as professionalism, interpersonal relationships and patient-practitioner communication
15	Saito et al 2004	Prospective	IBS patients referred to GI specialist clinic for multidisciplinary education	403	USA	NA	Class attendance was associated with improvement in health promotion, lifestyle scores but not with pain ratings, HRQoL or health care use
16	Saunders et al 1999	Prospective	Subject recruitment from primary care with low back pain	<ul style="list-style-type: none"> 5 year follow up (n=1213) Pain register baseline (n=1040) Randomised trial (n=255) 	USA	NA	Self-care intervention education modified attitudes to beliefs predicting health care. Group intervention was associated with changes in attitudes to Dr services
17	Shapiro et al 2002	Prospective	Medical members from 4 different sites, socioeconomic and cultural backgrounds All patients were below federal poverty line	Faculty (n=18) Residents (n=21) Patients (n=14)	USA	NA	All stakeholders recognise the importance of doctor-patient communication. Concern lay with various stakeholders to engage in person blame models
18	Terry et al 2007	Prospective	Resident cohort at family	10	USA	NA	The use of standardised patients is a feasible and

			practice program				useful option for evaluating family medicine resident decision making
19	Van den Heuvel-Janssen et al 2006	Prospective	IBS and FD Subject recruited from registration network of family practices	291	Holland	NA	Once non-specific abdominal complaints have become labelled, as chronic by physician, little improvement can be expected

• **Randomised controlled studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Aki et al 2012	Prospective	Subjects recruited from residents of 4 internal medicine, 2 family practice programs	371	USA / Canada	NA	No wording approach was clearly superior in conveying the strength of health care recommendations
2	Kelley et al 2009	Prospective	Subjects sampled from health care providers	289	USA	NA	Personality and gender influenced the placebo response to patient-practitioner relationship , but only in the warm, empathetic augmented group. Practitioners differed markedly in effectiveness, despite standardised interactions

- **Systematic reviews**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Dhaliwal et al 2004	Retrospective	PubMed, Cochrane database, PsychINFO, Cinahl, EMBASE, Web of Science and reference list searches	Articles (n=4)	Canada	NA	Evidence suggests that some IBS patients in primary care experience dissatisfaction and negative attitudes to GP interactions

- **Domains of measurement**

- b. **Case-control studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Basler et al 2008	Retrospective	Chronic low back patients from orthopaedic and neurological departments	Low back pain (n=103)	Germany	Healthy matched	Findings suggest that fear-avoidance belief measurement is also valid in the elderly
2	Bengtsson et al 2011	Retrospective	Consecutive IBS, dysmotility disorder and Sjögren's syndrome at hospital outpatients clinic	IBS (n=39) Dysmotility (n=21) Sjögren's (n=26)	Sweden	Healthy matched	The VAS-IBS could be used to assess the level of GI symptoms but does not differentiated between IBS and other dysmotility orders
3	Cho et al 2011	Retrospective	IBS Subjects recruited from hospital GI clinic	124	South Korea	Healthy matched	Anxiety and depression were frequently found in Korean IBS patient's that lead to symptom severity and poor HRQoL. The above should be an important part of IBS evaluation
4	Crane et al 2004	Retrospective	Convenience sample of patients from outpatients and local community	IBS (n=25)	UK	IBD patients	IBS are more likely than IBD patients to revert to illness related social learning in childhood, parental reinforcement of illness, and passive coping behaviour

5	Lagier et al 1999	Retrospective	Subjects recruited through local adverts	Healthy aged (n=12)	France	Healthy young controls	Rectal sensory thresholds triggered by distension are increased in aged healthy subjects while compliance and tone are not different
6	Piche et al 2007	Retrospective	Consecutive IBS patients from hospital GI clinic	51	France	Healthy matched	Significantly more IBS patients express fatigue than controls. Fatigue was associated with blood leptin levels independent from age, sex and body mass
7	Poynard et al 1992	Retrospective	Consecutive IBS, FD and organic disease patients from outpatient clinics and private GI specialists	IBS (n=120) FD (n=46) Organic disease (n=94)	France	Healthy matched	Manning criteria is reliable but not highly sensitive for a French population. While it was specific compared to health controls, it was not with organic disease.
8	Rey et al 2009	Retrospective	IBS patients recruited from primary and secondary care clinics	73	Spain	Healthy matched	IBS subjects do not show lower rational intelligence than controls. However, experiential intelligence is associated with IBS
9	Seres et al 2008	Retrospective	IBS patients recruited from 3 tertiary GI clinics	88	Hungary	Ulcerative colitis patients	IBS patients have higher levels of psychological distress, pain severity and maladaptive coping strategies than ulcerative colitis patients.
10	Talley et al 1990	Retrospective	All subjects recruited from patients reporting for routine exam for GI complaints	IBS (n=82) FD (n=33) Organic disease (n=101)	USA	Healthy subjects	Manning criteria can discriminate IBS from organic disease and FD. IBS is still difficult to diagnose in patients with atypical colonic symptoms
11	Walker et al 1998	Retrospective	Adolescents recruited complaining of recurrent abdominal pain at a single medical centre	76	USA	Subjects reporting with acute minor illness or injury from which they recovered	Female patients with recurrent abdominal pain may have the increase risk of IBS during adolescence or young adulthood, more functional disability, increased health care utilisation and lower academic al and social competence

12	Welch et al 1985	Retrospective	IBS outpatients at hospital clinic and IBS non-reporters at blood donor centre	26	UK	Healthy subjects	Outpatients and non-reporters were psychologically similar and showed more somatic stress than controls. The number of women referred to outpatient clinic may reflect sociological factors rather than symptom severity
13	Wong et al 2010	Retrospective	IBS patients presenting more than once to primary care clinic over 3 month period	1100	USA	Healthy matched	Patients identified by Rome III for chronic constipation and IBS-C are not distinct groups

• Patient characteristics

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Addolorato et al 2008	Retrospective	Consecutive patients at GI clinic	1641	Italy	NA	Most patients who seek medical consultation for gastrointestinal complaints show associated affective disorder. Management should involve psychiatric and gastroenterological expertise
2	Baber et al 2008	Retrospective	New patients referred to paediatric GI clinic	368	USA	NA	Changes in Rome criteria make Rome III more inclusive, allowing classification of 67% of paediatric patients with unexplained chronic abdominal pain
3	Chang et al 2001	Retrospective	Consecutive patients at FGID specialist centre	542	USA	NA	Bloating and visible abdominal distension may arise from two distinct physiological processes. The sensation of bloating may be related to enhanced visceral afferent sensitivity while abdominal distension may be related to visceromotor reflex action
4	Engel et al 1996	Retrospective	Convenience sample of IBS and IBD patients consulting to a medical centre over a 7 month period	103	USA	NA	Dissociation among this sample is a long term coping pattern that is associated primarily with past sexual trauma, chronic emotional distress, alcoholism and physical disability

5	Farnam et al 2007	Retrospective	Subjects with IBS diagnosis at University outpatients clinic	150	Iran	NA	Differences were observed between IBS patients and general population as well as between IBS subtypes in terms of personality factors. IBS subtype patients may benefit from psychological intervention
6	Hammer et al 2004	Retrospective	All patients consulting at hospital GI clinic over 4 years	568	Australia	NA	Symptoms in over 50s patients and blood on toilet paper is an alarm symptom. IBS diagnosis associated with female gender, pain over 6 times in the previous year and looser bowel motion
7	Keefer et al 2005	Retrospective	Subjects were part of an RCT for cognitive behavioural therapy trial for IBS	46	USA	NA	Anxiety and worry are independent predictors for symptom severity especially in IBS-C compared to IBS-D
8	Kovacs et al 2010	Retrospective	Consecutive new patients at tertiary care clinic	IBS (n=70) UC (n=55) Non-erosive reflux disease (n=80) Erosive reflux disease (n=74)	Hungary	NA	IBS is associated with significant psychological distress than any other group. The IBS group also has significantly higher levels of depression and somatisation
9	Lackner et al 2004	Retrospective	IBS Patients previously enrolled on a psychological study	179	USA	NA	Patients report high levels of catastrophising, which causes increased interpersonal problems. Both the above remained after removal of general symptom distress
10	Lembo et al 1999	Retrospective	Consecutive new IBS patients to tertiary care centre	443	USA	NA	Abdominal pain reported in only one-third of IBS patients. Pain predominance correlates with development of rectal hypersensitivity
11	Mearin et al 2004	Retrospective	FGID Subjects from epidemiological study originally from random population sample	211	Spain	NA	Many subjects meeting Rome I criteria for IBS do not meet Rome II criteria. One-quarter of subjects have insufficient symptom duration of frequency , and almost half are now considered as having other FGIDs
12	Pals son et al 2012	Retrospective	IBS subjects recruited at	392	USA	NA	Stool consistency varies greatly within

			FGID centre and IBS website				individuals. IBS-D – loose watery stools with bowel movement-related pain versus IBS-C – increased pain unrelated to bowel movement, bloating and life interference. Daily symptom diary – more sensitive and reliable than questionnaire
13	Rey et al 2008	Retrospective	IBS consulters and non-consulters recruited in primary and secondary care or approached at offices and leisure centres	280	Spain	NA	HRQoL is impaired in all subgroups of IBS sufferers with degree depending mostly on symptom severity and psychological factors
14	Schmulson et al 1999	Retrospective	Consecutive Rome positive patients recruited from clinical referrals and adverts	625	USA	NA	IBS-C patients show greater prevalence of wide-ranging symptoms referred to the upper and lower abdomen, musculoskeletal and constitutional function
15	Yao et al 2008	Retrospective	Army subjects with IBS	478	China	NA	Rome III criteria can improve the early diagnostic rate of IBS. IBS-D is the most common accounting for 2/3 cases followed by IBS-unsubtyped, IBS-C, IBS-M.

• **Cross-sectional studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Barakzai et al 2007	Retrospective	Medical records from 3 large medical centres of non-English speaking Mexican population	139	USA	NA	63% of Mexican-American patients reported symptoms that met any of the nine Rome II criteria. No significant difference between English or Spanish concerning their complaint
2	Bharucha et al 2000	Retrospective	Local community random sample	2800	USA	NA	Patients with painful constipation resemble IBS-C than painless constipation. Consideration should be given to separating painful from painless constipation in Rome criteria
3	Dansie et al 2012	Retrospective	Mid-Atlantic twin	4590	USA	NA	Findings support that clinically based

registry mailed survey							samples of chronic fatigue like illness is frequently comorbid with chronic widespread pain and IBS and / or depression disorder
4	Ghoshal et al 2008	Retrospective	22 medical centres across India based upon IBS publications	<ul style="list-style-type: none"> Complainants (n=2805) Non-Complainants (n=4520) 	India	NA	Most IBS patients are middle-aged men that complain of incomplete evacuation and mucous in stools. Abdominal pain is not universal and stool frequency is similar regardless of IBS-D or C
5	Hahn et al 1997	Retrospective	Records from 3 IBS groups	<ul style="list-style-type: none"> Self-reported IBS (n=6,412,895) Rome Criteria (n=5,410,829) Manning criteria (n=14,388,312) 	USA	NA	Findings suggest that there is a large undiagnosed population with numerous symptoms consistent with IBS
6	Hellstrom et al	Retrospective	6 urban hospital GI clinics in 5 countries	158	Sweden, Belgium, Germany, Denmark, USA	NA	The frequency of severe pain attacks in IBS patients was 1.4 per week and affected daily activities. However, most pain attacks were untreated in IBS patients
7	Lackner et al 2004	Retrospective	Consecutive IBS patients referred to behavioural clinic	244	USA	NA	IBS patients with greater depression report greater pain severity due to more catastrophic thinking. This relationship is not linear but works through patient's beliefs
8	Pilotto et al 2008	Retrospective	Random sample of local GP lists for 60+ seeking medical help for diarrhoea in a 2 week period	5387	Italy	NA	Diarrhoea is common in elderly outpatients and its prevalence increases with age as does the severity of disability and the number of drugs
9	Sugaya et al 2011	Retrospective	Questionnaire distribution to University students during class hours	1343	Japan	NA	Severe anxiety sensitivity in individuals with IBS is related to symptoms-related cognition. Altered cognition increases anxiety
10	Talley et al 1992	Retrospective	Random sample of local	328	USA	NA	Complaints of FGIDs are common in the

			community				elderly, but symptoms are a poor predictor of presentation for medical care
11	Talley et al 2000	Retrospective	Random samples of communities in 4 countries	USA (n=2200) Australia (n=1135) Germany (n=500) Sweden (n=1517)	See previous column	NA	Findings suggest that GI symptoms and groupings of individuals are similar across Western cultures and are consistent with Rome criteria for upper and lower FGIDs
12	Walter et al 2011	Retrospective	Random sample of every 7 th person in 2 urban districts	1705	Sweden	NA	Current criteria for IBS that rely on recall of relationships between abdominal pain and bowel disturbances may overcall this association when measured prospectively

• Cohort studies

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Adeniji et al 2004	Prospective	10-13 year follow up on IBS subjects who took part in a colon function study	196	USA	NA	92% considered that their symptoms had not improved; nearly half had repeated structural evaluation of the colon with no new diagnoses. Clinicians are advised to use clinical criteria for specific and durable diagnosis of IBS
2	Dorn et al 2009	Prospective	Female IBS subjects enrolling for multicentre trial	248	USA	NA	Rome II and III IBS subtypes are in high agreement and behave similarly over time
3	Lembo et al 1996	Prospective	Rome criteria IBS	20	USA	NA	Findings suggest that patients with short symptom duration and fewer psychological symptoms have a better prognosis than patients with a long history of IBS and associated psychological distress
4	Lewandowska et	Retrospective	Medical records of patients	581	Poland	NA	Knowledge of the Rome III criteria is insufficient

	al 2008		over 7 years at Hospital GI clinic				among family Drs with low agreement levels between preliminary diagnoses of IBS. Neoplasms and IBD diagnoses found in 60 patients with preliminary diagnosis of IBS
5	Rutter et al 2007	Prospective	Invite letter via GP surgeries to IBS subjects	42	UK	NA	Illness coping behaviours and psychological outcome (HRQoL, anxiety and depression) were stable over a 12-month period. Coping played no part in predicting outcomes
6	Vanner et al 1999	Prospective	Consecutive IBS patients to GI clinic (retrospective) and subjects meeting Rome criteria and n red flags	Retro group (n=98) Prospective group (n=95)	Canada	NA	Retrospective test – Rome criteria with no red flags – positive prediction value of 100%. No patients needed diagnostic revision after 2 year follow up Rome criteria with a lack of red flags has a very high predictive value for IBS diagnosis
7	Walker et al 2004	Prospective	Parents of new paediatric patients referred to GI clinic	114	USA	NA	Findings show that children with symptoms of recurrent abdominal pain are consistent with symptom criteria for several other FGIDs defined by Rome criteria.

- **Randomised controlled studies**

	Author / date	Type of data collection	Population / databases studied	Sample size	Location of study	control group	Findings
1	Hyphantis et al 2009	Prospective	Subjects recruited from 7 GI clinics	257	UK	NA	Results indicate that improvement in interpersonal problems in IBS patients appear to be primarily associated with reduced psychological distress. Improved health status following psychotherapy suggests that specific help with interpersonal problems

- **Systematic reviews**

	Author / date	Type of data collection	Population / databases studied	Sample size	Location of study	control group	Findings
1	Andresen et al 2011	Retrospective	Medline, Premedline, PsychINFO, Cambase, Cochrane Central Register	243	Germany	NA	Variable symptom constellation caused by multiple levels of GI regulation. This German guidelines on IBS translates up to date scientific knowledge as represented in current publications
2	Ford et al 2008	Retrospective	EMBASE, Medline	2355 patients in 10 articles	Canada	NA	Individual symptoms have limited accuracy for diagnosing IBS patients referred with lower GI tract symptoms. The accuracy of Manning criteria and Kruis scoring were only modest. Despite strong advocacy for Rome criteria, only Rome I has been validated

3	Lundberg et al 2011	Retrospective	Cinahl, Embase, PubMed, PsychINFO, Web of Science	37 articles	Sweden	NA	Weak construct validity implies that no measure can currently identify who is fearful when assessing Fear-Avoidance Beliefs Questionnaire, Fear Avoidance of Pain Scale, Fear of Pain Questionnaire, Pain and Anxiety Scale and Tampa Scale for Kinesiophobia
----------	---------------------	---------------	---	-------------	--------	----	---

• **Validation studies**

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Chachamovich et al 2007	Prospective	Elderly patients recruited from elderly health care communities	424	Brazil	NA	THE WHOQOL-BRIEF instrument show suitable psychometric performance in a sample of Brazilian older adults measuring HRQoL
2	Digesu et al 2010	Prospective	Females attending tertiary care at gynaecological, urological and colorectal outpatients	201	UK	NA	The Rome III criteria questionnaire is reliable and reproducible but does not appear to be a valid instrument for diagnosing constipation
3	George et al 2009	Prospective	Low back pain patients seeking treatment at university hospital	Reliability cohort (n=60) Validity cohort (n=108)	USA	NA	The fear of daily activities questionnaire is a potentially viable measure for specific activities in physical therapy settings. The measure is not appropriate as a screening tool
4	Gholamrzaei et al 2011	Prospective	IBS subjects referred to GI research clinics	141	Iran	NA	Results indicate that the Persian version of the IBS-QOL is reliable with sufficient psychometric requirements to assess QoL in Iranian IBS patients

5	Grotle et al 2012	Prospective	Consecutive female patients reporting post-partum pelvic pain	Main survey (n=87) Retest (n=42)	Norway	NA	Pelvic Girdle questionnaire show satisfactory discriminant validity and is recommended for symptom evaluation with pelvic pain
6	Haghighyegh et al 2012	Prospective	IBS subjects recruited from medical centres	126	Iran	NA	The Persian version of the IBS-QOL 34 is valid and reliable for use in research and clinical trials
7	Labus et al 2004	Prospective	External and internal physicians (IP, EP), patients and non-patients recruited from local adverts	IP (n=9) EP (n=3) Patients (n=11) Participants (n=100)	USA	NA	The Visceral Sensitivity Index is a reliable and valid measure for GI symptom-specific anxiety that may be useful in clinical assessment and treatment outcome studies
8	Labus et al 2007	Prospective	2 samples of students enrolled on psychology courses. 1 sample from local community	Sample 1 (n=290) Sample 2 (n=215) Sample 3 (n=82)	USA	NA	Overall the visceral Sensitivity Index demonstrated excellent psychometric properties providing further support for its use with IBS presentation
9	Lee et al 2012	Prospective	Malaysian Rome III translation	Clarity assessment volunteers (n=10) IBS volunteers (n=31)	Malaysia	NA	The translated Malay language IBS Rome III, red flags and psychological Alarm symptom questionnaires are valid and reliable
10	Longstreth et al 2005	Prospective	IBS outpatients from FGID research centres	155	USA	NA	The IBS-Impact Scale is a short user-friendly instrument with excellent psychometric properties that has potential use in clinical trials
11	Molinder et al 2009	Prospective	Random samples of subjects on ongoing GI study and patients seeking GP advice for any complaint	1123	Sweden	NA	The accuracy of the Swedish version of Rome II criteria is doubtful for clinical practice and research.
12	O'Keefe et al 1992	Prospective	Random sample of local residents, independent and consecutive outpatient samples	653	USA	NA	The Elderly Bowel Symptom Questionnaire is a reliable and valid measure of GI symptoms in older persons

13	Poitras et al 2002	Prospective	Consecutive, invited and advert response FGID outpatients	207	Canada	NA	The GI Index and QOL index were developed to detect FGIDs and follow its evolution in response to treatment. This offered significant contribution to management of FGID patients undergoing psychotherapy
14	Reisswitz et al 2010	Prospective	FD patients responding to adverts	109	Brazil	Healthy controls	The Rome III questionnaire for FD is ready to be used in clinical research and has been successfully validated in Portuguese
15	Rey et al 2010	Prospective	Random sample of local community	900	USA		The revised Talley Bowel Disease Questionnaire is reliable with excellent reproducibility and validity
16	Roalfe et al	Prospective	Random sample of IBS patients in county community	379	UK		The Birmingham IBS Questionnaire shows suitability to self-completion and acceptable for patients. The questionnaire has good reliability and external validity
17	Schmulson et al 2007	Prospective	Female IBS patients in Mexico and matched patients in USA	63 per group	USA / Mexico		Further validation of the Mexican version of the IBS-QoL shows Mexican female patients have lower HRQoL than those North Carolina
18	Sperber et al 2011	Prospective	Subjects recruited from FMS and GI clinics over 1 year	FMS (n=30) IBS(n=27)	Israel	Healthy matched	The new symptom-based diagnostic criteria for FMS can be used in clinical studies in which physical examination is unfeasible. Gastroenterologists can now investigate comorbid FMS in IBS patients

19	Talley et al 1995	Prospective	Volunteers attending endoscopy, private GI outpatients, hospital workers and	325	Australia	random sample of local population	The Bowel Symptom Questionnaire provided reliable and valid data in GI symptoms
20	Talley et al 1990	Prospective	Consecutive GI outpatients	361	USA	Healthy controls	Symptoms can be used to diagnose IBS but Manning criteria are not highly sensitive

• Multiaxial assessment criteria

c. Cross-sectional studies

	Author / date	Type of data collection	Population studied	Sample size	Location of study	control group	Findings
1	Mezzich et al	Retrospective	Members of the World Psychiatric Association (WPA)	175	USA	NA	The most frequent recommendation offered for the advancement of diagnostic symptoms include the improvement of patient evaluation procedures and the greater use of multiaxial diagnosis and its empirical validation
2	Salloum et al 2011	Retrospective	Literature reviews and consensus from WPA meetings over a 2 year period outlining the basis for person-centered integrative diagnosis	Not specified	USA	NA	Each domain will be evaluated with standardised categories and dimensions as well as narratives. Specific attention is paid to the interactive evaluation process via clinicians, patients, family, and other carers
3	Weiner et al 2001	Retrospective	Patients over 65 attending university interdisciplinary pain clinic	108	USA	NA	Results indicate that older patients are heterogeneous in their response to persistent pain.

- **Critical reviews**

	Author / date	Type of data collection	Topics studied	Sample size	Location of study	control group	Findings
1	Berganza et al 2005	Retrospective	Reviewed the concept of illness and the consideration for biopsychosocial framework	NA	USA / Guatemala	NA	Flexible diagnostic assessment and a systematic approach may be helpful in advancing the appraisal of mental disorders when classification systems are developed
2	Craddock et al 2007	Retrospective	Discusses issues on ways forward with both DSM-V and ICD-11	NA	UK	NA	<ul style="list-style-type: none"> • Valid diagnostic classification is crucial for clinical research and practice • Data, not opinion must inform classification • Current classifications are inhibiting clinical research progress •
3	Healy 2011	Retrospective	Review the case for a biopsychosocial framework in correcting the limitations of the DSM model in psychiatry	NA	Australia	NA	It is hoped that iterations of the DSM model will take greater account for the need for a biopsychosocial framework facilitating integrated , and holistic patients centered care
4	Jablensky 2005	Retrospective	Discuss option for future revisions of classifications versus dimension, prototypes of psychiatric diagnosis	NA	Australia	NA	A means of identifying natural clinical groups by a combination of discriminant function analysis and mixture analysis has been demonstrated two distinct patterns of somatisation. Prototype models have been used to assign relative to .endophenotypes classes for schizophrenia
5	López-Ibor 2003	Retrospective	To argue that too many cultural adaptations are unacceptable to the science and profession of psychiatry	NA	Spain	NA	Efforts should go toward development of new classifications based on pathogenesis, genetic and neuroimaging that have no cultural relevant component. Diagnosis should be international

6	McGuire 1992	Retrospective	To discuss multidimensional assessment of pain in a clinical setting	NA	USA	NA	A systematic process with appropriate tools and documentation procedures enhance the clinician's ability to evaluate pain carefully via objective and subjective data
7	Mezzich et al 2009	Retrospective	Guideline formulation for a practical approach to prepare a cultural formulation as a component of cultural competent care.	NA	USA	NA	Future work should examine the actual process of cultural formulation and the impact of its systematic use on clinical outcomes
8	Mezzich et al 2010	Retrospective	Review the conceptual bases for "person-centred" integrative diagnosis as a component to person-centred psychiatry	NA	USA	NA	Person-centred integrative diagnosis is aimed at appraising the overall health through pluralistic descriptions and evaluative partnerships
9	Mezzich et al 1995	Retrospective	Reviews the historical development of multiaxial diagnosis and contributions to comprehensive diagnosis	NA	USA	NA	It may be appropriate to consider diagnosis as part of a process accompanying a therapeutic endeavour from which specific interventions results
10	Narrow et al 2011 (DSM task force)	Retrospective	To present information on planned changes to the DSM-V	NA	USA	NA	<p>Implications for health care provision</p> <ul style="list-style-type: none"> • Patient reported assessment will aid more comprehensive and systemised characterisations at baseline <p>Implications for health policy</p> <ul style="list-style-type: none"> • DSM-V proposals for patient reported measures can serve as a method for gauging treatment outcome <p>Implications for further research</p> <ul style="list-style-type: none"> • DSM-V will continue research that will serve as the basis for further measurement refinement

11	Roberts et al 2005	Retrospective	Explore ways genetics and genomics, in conjunction with neurosciences and other biological disciplines can help shape diagnostic classification	NA	USA	NA	Robustness analysis helps us to appreciate the complexity of mental illness and its facilitates the development of adequate psychiatric nosology
12	Schneider et al 1995	Retrospective	Discuss advantages and disadvantages of descriptive multiaxial diagnostic models	NA	Germany	NA	It is important to have varying diagnostic dimensions with their very specific methods of inquiry
13	Spitzer et al 1999	Retrospective	Examines whether clinical significance criteria achieves its purpose and consider the broader impact of diagnostic validity	NA	USA	NA	In the process of revising DSM-IV, the generic use of clinical significance should be reconsidered. It may be also advantageous to raise symptom thresholds that excludes normal reactions to psychological stress
14	Starcevic 2006	Retrospective	Debates the future of somatoform disorders by critically reviewing relevant terminology and validity for separate classifications	NA	Australia	NA	Characterised mainly by medically unexplained physical symptoms that are real. It is a matter of ongoing debate as to how somatoform disorders are to be named and classified. This debate between psychiatrists and non-psychiatrists is crucial

- **Validation studies**

	Author / date	Type of data collection	Population / Test studied	Sample size	Location of study	control group	Findings
1	Klapow et al 1993	Prospective	Consecutive male patients complaining of daily low back pain	96	USA	NA	There were 3 identifiable groups; <ul style="list-style-type: none"> • Chronic pain syndrome with high pain levels, impairment and depression. • Positive adoption with high pain levels, low impairment and depression • Good pain control, low pain levels, impairment and depression. Multiple outcomes did not change across time
2	Lobo et al 1993	Prospective	Outpatients referred to psychosomatic liaison services	48	Spain	NA	This new interview allows use of different research diagnostic criteria including DSM-III, ICD-10 and Goldberg's criteria for somatic symptoms. The interview fulfils face and content validity
3	Turk et al 1990	Retrospective	Consecutive chronic low back pain, headache and temporomandibular disorders (TMD) <ul style="list-style-type: none"> • Multiaxial Assessment of Pain taxonomy 	<ul style="list-style-type: none"> • Chronic low back pain (n=200) • TMD (n=200) • Headache (n=100) 	USA	NA	Although patients were classified into 3 groups using Multiaxial Assessment of Pain taxonomy, the profiles based on inter-relationships among assessment scales were remarkably similar. Data suggests that psych and behavioural responses assist with chronic pain management
4	Turk et al 1987	Prospective	Subjects attending university pain centre and clinicians from different specialities <ul style="list-style-type: none"> • IASP taxonomy system 	Patients (N=115) Clinicians (n=4)	USA	NA	Results indicate that the IASP taxonomy system is on the right track and will eventually evolve into a scientifically sound diagnostic system

- **Randomised controlled studies**

	Author / date	Type of data collection	Population / Test studied	Sample size	Location of study	control group	Findings
1	Vollenbroek-Hutton et al 2004	Prospective	Outpatients attending back rehabilitation program	163	Holland	NA	The overall effect of multidisciplinary treatment is disappointing; however, multi-axial assessment before admission may be of value in clinical practice resulting in more effective treatments for chronic pain

- **Patient characteristics**

	Author / date	Type of data collection	Population / outcome measure studied	Sample size	Location of study	control group	Findings
1	Böker et al 2000	Retrospective	Psychiatric and orthopaedic patients attending hospital clinics	<ul style="list-style-type: none"> • Psychiatric (n=127) • Orthopaedic (n=34) 	Switzerland	NA	Idiographic results help to differentiate the spectrum of affective disorders. The underlying importance of interpersonal dimension of depression may be used as the basis for therapeutic appraisal
2	Talo et al 1994	Retrospective	Sub-sample of chronic low back pain patients in Finnish Invalid Foundation (4 rehabilitation centres)	173	Finland	NA	Disability and handicap measures of function increase effective predictors than impairment measures. Predictive power of psychological impairments, disabilities and handicaps varied with different patient groups

Appendix E: Response rates

Response and non-response rates (1 – responders, 2 – non-responders)

Bootstrap Specifications

Sampling Method	Simple
Number of Samples	1000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
response * location	90	100.0%	0	.0%	90	100.0%
response * title	90 ^a	100.0%	0	.0%	90	100.0%
response * gender	90 ^a	100.0%	0	.0%	90	100.0%
response * Specialism	90 ^a	100.0%	0	.0%	90	100.0%

Response (location)

Crosstab

			location			Total
			N America	Europe	Asia and Australia	
response	1	Count	12	22	2	36
		% within response	33.3%	61.1%	5.6%	100.0%
		% within location	24.0%	68.8%	25.0%	40.0%
	2	Count	38	10	6	54
		% within response	70.4%	18.5%	11.1%	100.0%
		% within location	76.0%	31.3%	75.0%	60.0%
Total	Count		50	32	8	90
	% within response		55.6%	35.6%	8.9%	100.0%
	% within location		100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	17.104 ^a	2	.000
Likelihood Ratio	17.287	2	.000
Linear-by-Linear Association	4.962	1	.026
N of Valid Cases	90		

Response (academic rank)

Crosstab

			Academic rank			Total
			Prof	Ass Prof	Research Fellow	
response	1	Count	26	2	8	36
		% within response	72.2%	5.6%	22.2%	100.0%
		% within title	40.0%	28.6%	44.4%	40.0%
	2	Count	39	5	10	54
		% within response	72.2%	9.3%	18.5%	100.0%
		% within title	60.0%	71.4%	55.6%	60.0%
Total		Count	65	7	18	90
		% within response	72.2%	7.8%	20.0%	100.0%
		% within title	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.529 ^a	2	.768
Likelihood Ratio	.544	2	.762
Linear-by-Linear Association	.045	1	.832
N of Valid Cases	90		

Response (gender)

Crosstab

			gender		Total
			Male	Female	
response	1	Count	29	7	36
		% within response	80.6%	19.4%	100.0%
		% within gender	43.3%	30.4%	40.0%
	2	Count	38	16	54
		% within response	70.4%	29.6%	100.0%
		% within gender	56.7%	69.6%	60.0%
Total	Count	67	23	90	
	% within response	74.4%	25.6%	100.0%	
	% within gender	100.0%	100.0%	100.0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.178 ^a	1	.278	.330	.202
Continuity Correction	.703	1	.402		
Likelihood Ratio	1.206	1	.272		
Fisher's Exact Test					
Linear-by-Linear Association	1.165	1	.280		
N of Valid Cases	90				

Response (Specialist field)

Crosstab

			Specialist field						
			Others	Physiology	Gastro	General surgery	Psychology	Paediatrics	Total
response	1	Count	11	2	18	3	3	3	36
		% within response	30.6%	5.6%	50.0%	8.3%	100.0%	8.3%	100.0%
		% within Specialism	61.1%	25.0%	36.7%	50.0%	40.0%	50.0%	40.0%
	2	Count	7	6	31	3	54	3	54
		% within response	13.0%	11.1%	57.4%	5.6%	100.0%	5.6%	100.0%
		% within Specialism	38.9%	75.0%	63.3%	50.0%	60.0%	50.0%	60.0%
Total	Count		18	8	49	5	6	90	6
	% within response		20.0%	8.9%	54.4%	5.6%	6.7%	6.7%	100.0%
	% within Specialism		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.227 ^a	5	.204
Likelihood Ratio	8.602	5	.126
Linear-by-Linear Association	1.637	1	.201
N of Valid Cases	90		

Late versus early response rates

Bootstrap

Bootstrap Specifications

Sampling Method	Simple
Number of Samples	1000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

Early / Late (Academic rank)

Crosstab

			title			Total
			Prof	Assoc Prof	Research Fellow	
Early Late	1	Count	18	1	5	24
		% within Early Late	75.0%	4.2%	20.8%	100.0%
		% within title	69.2%	50.0%	62.5%	66.7%
	2	Count	8	1	3	12
		% within Early Late	66.7%	8.3%	25.0%	100.0%
		% within title	30.8%	50.0%	37.5%	33.3%
Total		Count	26	2	8	36
		% within Early Late	72.2%	5.6%	22.2%	100.0%
		% within title	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.389 ^a	2	.823
Likelihood Ratio	.375	2	.829
Linear-by-Linear Association	.175	1	.676
N of Valid Cases	36		

Early / Late (location)

Crosstab

			location			Total
			North America	Europe	Asia and Australia	
Early Late	1	Count	7	16	1	24
		% within Early Late	29.2%	66.7%	4.2%	100.0%
		% within location	58.3%	72.7%	50.0%	66.7%
	2	Count	5	6	1	12
		% within Early Late	41.7%	50.0%	8.3%	100.0%
		% within location	41.7%	27.3%	50.0%	33.3%
Total		Count	12	22	2	36
		% within Early Late	33.3%	61.1%	5.6%	100.0%
		% within location	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.989 ^a	2	.610
Likelihood Ratio	.974	2	.615
Linear-by-Linear Association	.173	1	.677
N of Valid Cases	36		

Early / Late (Specialism)

Crosstab

			Specialism					
			Others	Physiologists	Gastroenterol	General surgery	Paediatrics	Total
Early Late	1	Count	7	1	13	1	2	24
		% within Early Late	29.2%	4.2%	54.2%	4.2%	8.3%	100.0%
		% within Specialism	63.6%	50.0%	72.2%	50.0%	66.7%	66.7%
	2	Count	4	1	5	1	1	12
		% within Early Late	33.3%	8.3%	41.7%	8.3%	8.3%	100.0%
		% within Specialism	36.4%	50.0%	27.8%	50.0%	33.3%	33.3%
Total	Count	11	2	18	2	3	36	
	% within Early Late	30.6%	5.6%	50.0%	5.6%	8.3%	100.0%	
	% within Specialism	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
	NB Psychologists had no respondents							

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.795 ^a	4	.939
Likelihood Ratio	.774	4	.942
Linear-by-Linear Association	.028	1	.868
N of Valid Cases	36		

Early / Late (gender)

Crosstab

			gender		Total
			Male	Female	
Early Late	1	Count	22	2	24
		% within Early Late	91.7%	8.3%	100.0%
		% within gender	75.9%	28.6%	66.7%
	2	Count	7	5	12
		% within Early Late	58.3%	41.7%	100.0%
		% within gender	24.1%	71.4%	33.3%
Total	Count		29	7	36
	% within Early Late		80.6%	19.4%	100.0%
	% within gender		100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.675 ^a	1	.017	.029	.029
Continuity Correction	3.746	1	.053		
Likelihood Ratio	5.399	1	.020		
Fisher's Exact Test					
Linear-by-Linear Association	5.517	1	.019		
N of Valid Cases	36				

Appendix F: Participant correspondence

Research outline letter (paper and electronic)

Dear

As part of a Ph.D. thesis in the School of Clinical Sciences and Community Health at the University of Edinburgh, we are developing a biopsychosocial assessment system for functional gastrointestinal disorders (FGID). We are asking experts to give us the benefit of their knowledge in order to gain consensus on ways to maximise the coverage of all significant information when evaluating FGID patients.

The central feature is the development of a comprehensive assessment that incorporates several informative domains of potential clinical value. This multiaxial model will include both standardised and personal idiographic assessment that recognises the patient's lifelong experiences and ensuing influences on their personality, predisposition to ill health and ways of approaching adversity. Additionally, it is hoped that future integration of well-replicated neuroscientific, genetic and pharmacogenomic data may also provide additional bases for diagnoses and therapies targeting specific visceral diseases.

Before designing the multiaxial assessment and classification system, the project team needs to address a number of important issues

- Which contributing factors should be considered when evaluating FGID?
- Which epidemiological factors should be considered when evaluating FGID patients?
- Which domains of function should be measured when evaluating the symptom of FGID?
- Which areas of communication and clinical competence are important when establishing a therapeutic relationship with FGID patients?
- How would such a multiaxial system advance scientific, humanistic and ethical practice concerning the presentation of FGID?
- How would such a multiaxial system be structured?

Consensus will be gained by using the Delphi technique, a structured group communication method for soliciting expert opinion about complex problems through a series of questionnaires and controlled feedback. In total, the various rounds of questionnaire will take approximately 80 minutes of your time over a period of about nine months. To help you follow your responses in relation to collective opinion, simple statistical summaries will be added to ongoing rounds. The Delphi process will begin in February 2011 and completed the following year.

If you have any questions, please contact us at the above address or telephone number.
Yours sincerely

Invitations (electronic)

Invitation to round one

Dear

Please find below the link for the Delphi consensus study concerning the development of a multi-axial assessment system for functional gastrointestinal disorders.

As we would like to send you your results in relation to all other participants, you will be asked for your name at the beginning of the survey. The level of consensus has been set at 75% and will be available as part of a summary sheet accompanied by that round's questionnaire.

This online survey allows participants to enter and leave the questionnaire at any time, however, please follow the "finish later" instructions carefully. This facility serves all experts whom are unable to give a full 20 minutes due to busy work schedules.

Your expertise and participation is vital to the validation and reliability of this study in order that we can maximise the coverage of all significant areas of information relevant to describing the patient's disorders, dysfunction and disabilities.

Many thanks for completing this round of the Delphi. This study will be open for one month. You will receive the second round after the data has been analysed, and the second round questionnaire developed in May 2011.

<https://www.survey.ed.ac.uk/round1>

Many thanks for your contribution

Yours sincerely

Invitation to round two

Dear

Thank you for returning the first round Delphi questionnaire. You will receive a link for the 2nd round of the Delphi study concerning the development of multiaxial assessment criteria on the May 25th. Also attached to the following email will be three PDF files containing:

1. Personal scores and comments
2. Data analysis summary
3. General group comment and opinion

Many areas of knowledge surrounding FGID assessment gained consensus in round one. Additionally, there were several comments and suggestions submitted by participating experts. As you will observe, this data has been thematically analysed and subsequently inserted as additional statements in the appropriate sections of the second round survey.

This second round survey will take approximately 15 minutes to complete.

LINK - <https://www.survey.ed.ac.uk/round2>

Again, I would like to say thank you for your continued participation in this study.

Yours sincerely

Invitation to round three

Dear

Ref: The Development of a Multiaxial Assessment for Functional Gastrointestinal Disorders: A Worldwide Delphi Expert Consensus Study.

Thank you for your time in collaborating with this study and willingness to participate in the third and final Delphi round. This round contains those statements concerning the potential content and development of a multiaxial assessment tool for FGIDs that have not yet reached consensus.

Over the course of this Delphi study, validation has been strengthened by inserting examples of current high-level research pertaining to each statement where possible between each round. It is hoped that such statement validation may help to further increase cross-disciplinary awareness that is vital to the collaboration of the many fields of expertise both in assessment and management settings.

On September 1st, as for Round 2, you will also receive your individual response sheet, a general group comment sheet and a Round 2 data summary sheet.

Yours sincerely

Reminders (electronic)

Online Pilot study

Dear

This is a reminder to please take a moment to fill out this pilot questionnaire regarding multi-axial assessment for functional gastrointestinal disorders.

The goal is to test the format and timing of the questionnaire before its launch in February. Your participation is essential to further enhance the impact of this study over the coming year. Thank you for your time in this pilot study. As your questionnaire has not yet been received, I am including your username password and link to the questionnaire.

Link - <https://www.survey.ed.ac.uk/pilot>

Your username.....

Your password.....

Could you please contact me if you are unable to participate at the above email address? This survey closes on the 9th of December 2010. Additionally, a third telephone reminder will be sent five days before the closing date.

Yours sincerely

First reminder (all rounds)

Dear.....

The round 1 survey has now been open for two weeks. So far, the first round has over...% of the surveys returned, again with a great deal of interesting comment and opinion.

It is now becoming clear which areas of FGID multi-axial assessment are favoured and those where there is much indecision, especially those areas where further research is needed to allow evaluation within the clinical setting.

Please find below, the link to the setting.

Link - <https://www.survey.ed.ac.uk/round1>

Again, if you do not wish to continue with the study, please let me know. If your time is at a premium, we can arrange for you to complete the survey into August as your participation is far more important than specific time scales.

Yours sincerely

Second reminder (all rounds)

Dear

We are now entering the final week of the first round of the Delphi study seeking expert opinion on the future development of a biopsychosocial assessment system for functional gastrointestinal disorders (FGID). It would be truly appreciated if you can make time to complete this survey as the next round is constructed entirely on yours and other experts responses from this round.

To date, it appears that consensus will be gained on many areas significant to the biopsychosocial assessment of FGID, therefore, the next round will be shorter and take up less of your time. Additionally, there have been many interesting and highly useful comments made by expert participants, all of which will be anonymously inserted into the data sheets that will accompany the next round.

Confidentially, all data collected in this survey is held anonymously and securely.

I look forward to your participation

Link - <https://www.survey.ed.ac.uk/round1>

Yours sincerely

Final reminder (all rounds)

Dear

This is the final reminder for the University of Edinburgh's Delphi study that is gaining worldwide expert opinion on the development of multiaxial assessment criteria for functional gastrointestinal disorders. The study closes on the 31 of March, so if you can find the time, please access the questionnaire on the link below to complete this survey.

If you have any queries, please do not hesitate to email me at the above address or call me on (telephone number)

Your time is very much appreciated

Link - <https://www.survey.ed.ac.uk/round1>

Yours sincerely

Thanks for participating letter (electronic)

Dear

Many thanks for taking part in this study, your help and feedback have been invaluable to this study. I will inform you of the results and any relevant publications in the New Year. On a more personal note, your work has been of great assistance to the development of this study.

Yours sincerely

Appendix G: No requirement for ethical approval

If the study involves NHS staff only then there is no requirement (policy-wise or legally) for NHS ethical review unless the study involves any of the following:

People who lack the capacity to give informed consent to take part in the research
processing of confidential patient information without consent where this would otherwise breach confidentiality material consisting of or including human cells, which has been taken from the living or the deceased. Legally required, if it involves analysis of DNA in material from the living and consent for research not in place (UK-wide)
patients who are cared for in private and voluntary sector nursing homes (in England, Wales and Northern Ireland) and/or residents of residential care homes (in Northern Ireland only)
exposure to ionising radiation
medical devices that are not CE-marked or CE-marked devices that have been modified or are being used for a new purpose
investigational medicinal products
practising midwives conducting a clinical trial
protected information from the Human Fertilisation and Embryology Authority register

So, in short there is no requirement for NHS ethical review, and this information has now filtered down to journals etc., so there won't be a problem with publishing.

Alex Bailey
Scientific Officer
South East Scotland Research Ethics Service
Waverley Gate
Edinburgh
EH1 3EG
Phone: 0131 465 5679 (3567)

Appendix H: Rome Foundation

Letter of permission for use of research publication

- An international Delphi study to assess the need for multiaxial criteria in diagnosis and management of functional gastrointestinal disorders”
- **To be used as support for the new Rome IV multi-axial criteria**

Date: Sat, 15 Jun 2013 17:39:22 +0000 [15/06/2013 18:39:22 BST]

From:

[Drossman, Douglas Arnold <douglas_drossman@med.unc.edu>](mailto:douglas_drossman@med.unc.edu)

To:

[Philip Austin <P.D.Austin-2@sms.ed.ac.uk>](mailto:P.D.Austin-2@sms.ed.ac.uk)

Subject: RE: Delphi study article

Very nice. With your permission, I'd like to forward to Rome Board since we have developed multi-axial criteria and this can be supportive.

Doug

Appendix I: Poster Presentations and peer review article

British Pain Society ASM, Edinburgh, June 2011

Biopsychosocial Assessment Criteria for Chronic Unexplained Visceral Pain: A Pilot Consensus Study.

¹School of Clinical Sciences and Community Health, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, UK

P.D. Austin ¹, S.E. Henderson ¹, I Power ¹



INTRODUCTION

Chronic, unexplained visceral pain (CUVP) is especially common in the general population (1). Pain and other visceral symptoms, such as bloating and diarrhoea which make up the spectrum of functional gastrointestinal disorders (FGID), occur due to a range of mechanisms according to the organ and their afferent pathway. Advances in our understanding of neuro / psychobiological mechanisms, together with ever-increasing epidemiological and gender based information of FGID, may allow for the development a multiaxial assessment model that goes beyond 'disease only' analysis and moves toward idiographic diagnosis relating to the uniqueness of the patient.

Before development can commence, expert consensus worldwide must be gained on issues such as contributing factors considered when evaluating FGID; the therapeutic relationship; areas of communication and clinical competence; domains of function considered for measurement and the pros and cons of multiaxial assessment relevant to FGID. For this study, the Delphi technique was selected as it has proved to be a robust form of structured group communication within clinical and social health sciences (2). Due to the vast area of knowledge surrounding FGID, opinions and ideas from differing areas of expertise will hopefully emerge to form objective statements that embody relevant aspects the knowledge spectrum that apply purposely to the study.

References

1. Halder SL, Locke II GR, Schleck CD, Zinsmeister AR, Melton III LJ, Talley NJ. Natural History of Functional Gastrointestinal Disorders: A 12-year Longitudinal Population-Based Study. *Gastroenterology*. 2007;133(3):799-807.e1.
2. Skulmoski GJ, Hartman PT, Krahn J. The Delphi Method for Graduate Research. *Journal of Information Technology Education*. 2007;6:1-21.

AIMS

- Gain insight into any specific factors which are highlighted as important when considering the assessment of FGID.
- To test the validity and reliability of the Round 1 Delphi survey.
- Gain feedback on the format of 132 statements of the Round 1 Delphi survey.

METHODS

The Bristol Online Survey was used as a means of developing and distributing the Delphi questionnaire. Invitations were sent to 8 experts in the field of pain management and integrative medicine at Stockholm's Karolinska Institute and Sydney's Pain Management Research Institute.

The feedback and comments concerning question validity and reliability were assessed, and the 5 item Likert scale coding (5-strongly agree to 1-strongly disagree) was descriptively analysed.

RESULTS

Of the 8 invitations, 6 accepted, and of the 6, 3 completed and gave feedback on the questionnaire. Three failed to complete the questionnaire. All completing participants reported that the question format was easy to follow, while giving valuable feedback on possible additional statements and more "disagree" statements to reduce the occurrence of pattern answering. The issues of the non-finishers were appropriately rectified. Descriptive analysis revealed highest median scores (4+) to the importance of the following factors when considering assessment of FGID:

- Psychological and impact of daily living as the most likely contributors to FGID
- The qualities of a practitioner and the patient/practitioner relationship
- Areas of consideration: abdominal symptoms, the patient's description of symptoms coupled with emotional and social functioning
- The pros of multiaxial assessment concerning CUVP and other unexplained visceral symptoms

CONCLUSION

This pilot study highlights the potential of Delphi technique as a chance to reach consensus concerning the biopsychosocial diagnosis of FGID, and therefore adapted to other research and clinical settings. As all feedback concerning the survey format was very positive coupled with the consistent levels of agreement bodes well for a successful Delphi study amongst more than 80 experts in the field of FGID.

Scottish Pain Research Community ASM, Dundee, March 2012

The Development of Multiaxial Assessment Criteria for Chronic Unexplained Visceral Pain: A Worldwide Delphi Expert Consensus Study (Preliminary Findings)

P.D. Austin¹, S.E. Henderson¹, I. Power¹, M. Jirwe², T. Ålander²

¹School of Clinical Sciences and Community Health, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, UK

²Department of Neurobiology, Care Sciences and Society, Division of Nursing, Karolinska Institutet, Stockholm, Sweden

Email: P.D.Austin-2@sms.ed.ac.uk



INTRODUCTION

Functional gastrointestinal disorders (FGIDs) are common in the general population¹. Pain and other visceral symptoms, such as bloating and diarrhoea which make up the spectrum of FGIDs, occur due to a range of mechanisms according to the organ and their afferent pathway. Advances in our understanding of neuro/psychobiological mechanisms, together with ever-increasing epidemiological and gender based information of FGID, may allow for the development of a multiaxial assessment model that goes beyond 'disease only' analysis and moves toward idiographic diagnosis relating to the uniqueness of the patient.

Before development can commence, expert consensus worldwide must be gained on issues such as contributing factors considered when evaluating FGID; the therapeutic relationship; areas of communication and clinical competence; domains of function considered for measurement and the pros and cons of multiaxial assessment relevant to FGID. For this study, the Delphi technique was selected as it has proved to be a robust form of structured group communication within clinical and social health sciences². Due to the vast area of knowledge surrounding FGID, opinions and ideas from differing areas of expertise will hopefully emerge to form objective statements that embody relevant aspects of the knowledge spectrum that apply purposefully to the study.

REFERENCES

1. AUSTIN PD & WINDGOSTEN R (2010) Biopsychosocial assessment criteria for Functional Chronic Visceral Pain: A global review of concept and practice. *Pain Medicine*, 11(12), 1828-1837.
2. HALDORSEN, LOCH II SR, SCHUCK CD, ZIMMERMAN AR, MCELROY LI & TALLEY NJ (2007) Natural History of Functional Gastrointestinal Disorders: A 12-year Longitudinal Population-Based Study. *Gastroenterology*, 133, 799-807.e1.
3. HARRISON P & KIRKPATRICK (2011) Enhancing rigor in the Delphi technique research. *Technological Forecasting and Social Change*, 78, 1569-1576.

AIMS

Gain expert judgement and opinion on both standardised and idiographic domains of information concerning the following:

- contributing and risk factors
- the patient-practitioner relationship
- domains of measurement
- pros and cons of multiaxial evaluation
- future integration of neuroscientific, genetic and pharmacogenomic data for diagnoses of specific gastrointestinal diseases.

METHODOLOGY

The modified Delphi technique was used to gain consensus amongst experts worldwide. The inclusion criteria required that experts worked within the many fields of FGIDs, were currently active with the Rome Foundation, The International Foundation for Functional Gastrointestinal Disorders and/or working as principal investigators with FGIDs.

The first round of this study was developed using the results of a comprehensive review that allowed options for additional comment with each item and section. This was employed to reduce content bias, unambiguous broad statements and reduce allocated time to busy experts. The Bristol Online Survey was used as a means of developing and distributing the Delphi survey rounds. Invitations were sent to 90 experts, 68 of whom were eligible to take part.

The 5-item Likert scale coding (5-strongly agree to 1-strongly disagree) was descriptively analysed over 3 survey rounds. The consensus level was set at 75%.

RESULTS

Response Rates	Items Gaining Consensus	Main Areas of Agreement (>75%)	Main Areas of Non-Agreement (< 86%)
Survey Round 1 <ul style="list-style-type: none"> • Number of respondents: 36 • Expected number of respondents: 68 • Response rate: 52.9% • Launch date: 25 February 2011 • Close date: 31 March 2011 	Total: 132 items <ul style="list-style-type: none"> • Items gaining agreement consensus: 93 • Items gaining disagreement consensus: 4 • Items forwarded to Round 2: 35 • Items added from expert feedback analysis: 34 	<ul style="list-style-type: none"> • the psychological impact of daily living as the most likely contributor to FGIDs • the qualities of a practitioner and the patient-practitioner relationship • areas of measurement: abdominal symptoms, the patient description of symptoms in association with emotional and social functioning • the pros of multiaxial assessment concerning FGIDs 	<ul style="list-style-type: none"> • role of gender • genetic polymorphism • measurement of physical function • cons of multiaxial assessment • future research (risk factor identification/replication of findings)
Survey Round 2 <ul style="list-style-type: none"> • Number of respondents: 31 • Expected number of respondents: 35 • Response rate: 85.1% • Launch date: 25 May 2011 • Close date: 28 July 2011 	Total: 68 items <ul style="list-style-type: none"> • Items gaining agreement consensus: 19 • Items gaining disagreement consensus: 0 	<ul style="list-style-type: none"> • the roles of previous GI infection and aberrant enteric microbiota (additional items suggested by experts) • future research (risk factor identification and replication of findings) 	<ul style="list-style-type: none"> • role of gender • genetic polymorphism • cultural factors (additional items suggested by experts) • measurement of physical function • cons of multiaxial assessment
Survey Round 3 <ul style="list-style-type: none"> • Number of respondents: 31 • Expected number of respondents: 31 • Response rate: 100.0% • Launch date: 31 August 2011 • Close date: 11 October 2011 	Total: 60 items <ul style="list-style-type: none"> • Items gaining agreement consensus: 9 • Items gaining disagreement consensus: 0 	<ul style="list-style-type: none"> • associations with a family history of chronic pain disorders • physical functioning: the patient's previous physical ability and present attitudes toward physical activity 	<ul style="list-style-type: none"> • role of gender • genetic polymorphism • cultural factors (additional items suggested by experts) • measurement of physical function • cons of multiaxial assessment

In total, 121 of the 167 items gained consensus

CONCLUSION

This study not only highlights the use of Delphi technique as a chance to gain opinion and consensus concerning the biopsychosocial evaluation of FGIDs, but also in drawing attention to areas of knowledge where diagnostic indecision persists. Expert feedback suggests this indecision is due largely to the lack of credible human research in fields such as genetic polymorphism, gender and physical measurement.

The Need for Multiaxial Assessment and the Importance of Psychosocial Factors in Functional Gastrointestinal Disorders: An Analysis of Worldwide Expert Opinion

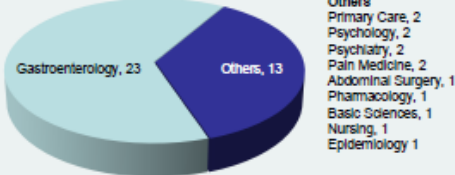
P.D. Austin¹, S.E. Henderson¹, I. Power¹, M. Jirwe², T. Ålander²

¹School of Clinical Sciences and Community Health, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, UK.

²Department of Neurobiology, Care Sciences and Society, Division of Nursing, Karolinska Institutet, Stockholm, Sweden.

Email: P.D.Austin-2@sms.ed.ac.uk



PURPOSE	RESULTS	RESULTS																
<ul style="list-style-type: none">To survey expert opinion worldwide.To derive consensus on relevant dimensions of management.Take into consideration risk factors, areas for evaluation, and therapeutic aspects.	<ul style="list-style-type: none">33 additional items were generated from participant feedback using thematic analysis.31 participants (86%) responded to Rounds 2 and 3 with a further 28 items gaining consensus and a total of 167 (74%) items gained consensus. <div><p>Others Primary Care, 2 Psychology, 2 Psychiatry, 2 Pain Medicine, 2 Abdominal Surgery, 1 Pharmacology, 1 Basic Sciences, 1 Nursing, 1 Epidemiology 1</p></div> <ul style="list-style-type: none">Comparison of gastroenterologists and other specialists found disagreement for only 13 items (10%) in round 1 and for 3 items (6%) in round 3.	<table><tr><th>Future development of multiaxial assessment criteria.</th><th>Agree %</th></tr><tr><td>Multiaxial assessment criteria will benefit the management of FGIDs.</td><td>75</td></tr><tr><td>Multiaxial assessment criteria may benefit the management of FGIDs (depending on results from further clinical research).</td><td>83</td></tr><tr><td>Multiaxial assessment criteria will only be valuable when further "gold standard" diagnostic testing is available.</td><td>23</td></tr><tr><td>Multiaxial assessment criteria will never benefit the management of FGIDs.</td><td>8</td></tr></table>	Future development of multiaxial assessment criteria.	Agree %	Multiaxial assessment criteria will benefit the management of FGIDs.	75	Multiaxial assessment criteria may benefit the management of FGIDs (depending on results from further clinical research).	83	Multiaxial assessment criteria will only be valuable when further "gold standard" diagnostic testing is available.	23	Multiaxial assessment criteria will never benefit the management of FGIDs.	8						
Future development of multiaxial assessment criteria.	Agree %																	
Multiaxial assessment criteria will benefit the management of FGIDs.	75																	
Multiaxial assessment criteria may benefit the management of FGIDs (depending on results from further clinical research).	83																	
Multiaxial assessment criteria will only be valuable when further "gold standard" diagnostic testing is available.	23																	
Multiaxial assessment criteria will never benefit the management of FGIDs.	8																	
METHODS		CONCLUSION																
<ul style="list-style-type: none">Delphi study; a 5 point Likert-scale with 132 items, generated from review work on the development of multiaxial assessment criteria for functional gastrointestinal disorders (FGIDs), sent to a purposive sample of 90 experts from the Rome Foundation and the International Foundations for FGIDs.The survey was divided into 4 sections:<ol style="list-style-type: none">Physiological and psychosocial risk factorsThe therapeutic relationshipPhysiological and psychosocial domains of measurementPros and cons of multiaxial assessmentAn online survey tool was used to develop, distribute and analyse the Delphi rounds.Item consensus levels, defined as at least 75%.	<table><tr><th>Strong agreement concerning multiaxial assessment. (% = strongly agree plus agree Likert points)</th><th>Agree %</th></tr><tr><td>A systematic approach for both physiological and psychosocial components of FGIDs.</td><td>91</td></tr><tr><td>Expansion from a single-item diagnosis to axes that provide additional "domains" of information of clinical value.</td><td>89</td></tr><tr><td>Its application in conjunction with laboratory and self-reporting findings.</td><td>83</td></tr></table> <table><tr><th>Divided opinion on the importance of physiological risk factors.</th><th></th></tr><tr><td>Genetic association.</td><td>71</td></tr><tr><td>Previous truncal surgery.</td><td>53</td></tr><tr><td>Previous physical trauma.</td><td>53</td></tr></table>	Strong agreement concerning multiaxial assessment. (% = strongly agree plus agree Likert points)	Agree %	A systematic approach for both physiological and psychosocial components of FGIDs.	91	Expansion from a single-item diagnosis to axes that provide additional "domains" of information of clinical value.	89	Its application in conjunction with laboratory and self-reporting findings.	83	Divided opinion on the importance of physiological risk factors.		Genetic association.	71	Previous truncal surgery.	53	Previous physical trauma.	53	<ul style="list-style-type: none">This study supports the use of Delphi technique for gaining consensusNon-agreement was observed between specialists on those domains relevant to the evaluation of FGID.Expert consensus shows a current need for multiaxial assessment criteria (regardless of gold-standard testing).Judgment was split on physiological risk facts and genetic association, citing paucity of reliable human data.Further research is now warranted to address these knowledge gaps.
Strong agreement concerning multiaxial assessment. (% = strongly agree plus agree Likert points)	Agree %																	
A systematic approach for both physiological and psychosocial components of FGIDs.	91																	
Expansion from a single-item diagnosis to axes that provide additional "domains" of information of clinical value.	89																	
Its application in conjunction with laboratory and self-reporting findings.	83																	
Divided opinion on the importance of physiological risk factors.																		
Genetic association.	71																	
Previous truncal surgery.	53																	
Previous physical trauma.	53																	
RESULTS																		
<ul style="list-style-type: none">From 13 countries, representing 10 areas of expertise, 36 out of 68 eligible participants (52%) responded to Round 1.96 items gained consensus.																		



An international Delphi study to assess the need for multiaxial criteria in diagnosis and management of functional gastrointestinal disorders



Philip Austin ^{a,*}, Sarah Henderson ^a, Ian Power ^a, Maria Jirwe ^b, Ture Ålander ^b

^a Department of Anaesthesia and Pain Medicine, University of Edinburgh, Edinburgh, United Kingdom

^b Department of Neurobiology, Care Sciences and Society, Division of Nursing, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:

Received 22 March 2013

Received in revised form 13 May 2013

Accepted 14 May 2013

Keywords:

Functional gastrointestinal disorders

Multiaxial assessment criteria

Consensus

Delphi study

Comorbidity

Psychosomatic

ABSTRACT

Purpose: To assess agreement, consensus, and disagreement between experts in different domains in the evaluation of functional gastrointestinal disorders (FGIDs).

Methods: We conducted a modified Delphi study using 90 purposively sampled experts using an online survey tool to develop, distribute and analyse the Delphi rounds. Experts were sent 132 items generated from a literature review examining present and proposed future FGID evaluation. The survey items examined contributory factors and were divided into four sections examining risk and contributing factors, the therapeutic relationship, domains of measurement and the pros and cons of multiaxial assessment. The consensus level was set at 75%.

Key results: 36 of 68 eligible participants (52%) replied to round one and 96 items gained consensus. Using expert feedback, we used thematic analysis to generate 33 additional items for round two. 31 of 36 participants (86%) replied to rounds two and three. In round two, 19 items gained consensus, and in round three, nine items gained consensus. Agreement was high concerning systematic approaches for both physiological and psychosocial components of FGIDs (91%) using laboratory and self-reporting findings (83%). Opinion was divided regarding physical risk factors such as previous surgery (53%) and genetic association (71%). Overall, 124 of the 167 items gained consensus.

Conclusion and inferences: We have identified expert consensus and disagreement on domains of information relevant to the evaluation of FGIDs. Experts agreed there is an immediate need for multi-axial assessment. Physiological and genetic risk factors are not fully accepted and require further study.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Functional gastrointestinal disorders (FGIDs) are a highly common group of persistent and recurring disorders that can affect any part of the gastrointestinal tract [1]. Conditions such as irritable bowel syndrome and functional dyspepsia are chronic relapsing disorders with spectrum of GI symptoms, most notably abdominal pain, bloating, constipation, diarrhoea and non-cardiac chest pain [2]. Visceral hypersensitivity and altered motility are also considered as key pathophysiological mechanisms that can originate in both peripheral or central nervous systems and caused by cognitive and affective factors [3,4]. In clinical practice, history taking includes searching for symptoms proposed by FGID diagnostic algorithms as well as alarm symptoms and features of organic disease such as GI bleeding, and unexplained weight loss. Currently, the Rome III self-reporting questionnaire to diagnose one or more FGIDs in adults incorporates alarm symptoms and red flags [5].

FGIDs are also associated with comorbid conditions such as personality disorders and psychological distress that affect health related quality of life (HRQoL) [6]. Evidence shows that FGIDs are a group of multifactorial conditions that occur in families, supporting genetic factors, while stress, especially in early life has also been shown to predispose an individual to developing FGIDs [7,8]. Recent evidence also indicates that FGIDs may be the adverse outcome of an acute episode of infectious gastroenteritis [9].

Present FGID diagnostic methods

There are currently no known biomarkers or gold-standard tests for the diagnosis of FGIDs. The Rome Foundation offers symptom-based criteria for 28 adult and 17 paediatric FGID diagnoses developed through expert consensus using 14 investigative committees representing 18 countries worldwide, which at present is considered the accepted FGID diagnostic resource [10,11]. The Rome Foundation aims to develop further scientific understanding of FGIDs and on issues such as the contributory role of aberrant intestinal microbiota [12], reliability of psychometric evaluation [13] and gender.

* Corresponding author at: Karlavägen 75a, 11449 Stockholm, Sweden. Tel.: +46 707484684.

E-mail address: P.D.Austin-2@sms.ed.ac.uk (P. Austin).

Multidimensional assessment

Multiaxial assessment criteria (MAC) are grounded in the biopsychosocial model that considers multiple factors of a disorder [14,15]. The first purpose of MAC is to systematically approach and communicate the fundamental components of a disorder where axes developed in various diagnostic systems reflect a disorder as a whole [16]. Secondly, MAC create an overview of the patient's condition which also accounts for its impact on the functioning of the patient [16]. However, while the biopsychosocial model of health attempts to conceptualise how these contributing factors interact in FGIDs and other functional somatic syndromes (FSSs), we remain caught in a circular argument that struggles to go beyond mind–body dualism. This presents a challenge for both patient and practitioner that is reflected in the number of clinician visits, diagnostic tests and secondary economic losses due to work absenteeism [17]. Hence, MAC aim to integrate the contribution of multiple factors, both peripheral and central, current and past that can contribute to symptom severity and impact on HRQoL that is at present used in psychiatry and psychosomatic medicine including chronic pain.

There are no published studies on the formal development or implementation of MAC for FGIDs. Given the multifactorial nature of FGIDs, separate axes of information may be developed to describe clinical disorders including FGID symptoms, psychological disorders, adaptive social, physical and occupational functioning and levels of quality of life. Furthermore, given that, FGIDs are common in patients with a history of childhood adversity, chronological relationships between the onset of both physical and psychological symptoms may also be of some value helping solve the much-maligned chicken and egg question. Inclusion in Rome criteria should make MAC and treatment more widely applied outside the field of experts working in clinical research. MAC are relevant to FGIDs as they occur commonly in the population with no biological markers for diagnostics that are presently based on expert consensus.

This study examined a broad range of subject matter relevant to FGID assessment. Therefore, due to the large amount of data available, this article will primarily present and discuss MAC for FGIDs accompanied by brief overviews of other domains of information shown in Table 1. The aims of this study were to assess levels of expert agreement, and disagreement on what evidence is currently available and which areas require further research concerning risk and contributing factors and areas of possible measurement in the evaluation of FGIDs using MAC.

Table 1
Outline of the round one Delphi survey for future development of MAC for FGIDs

Survey section	Subsection
FGID related comorbidity	
Risk and contributing factors	Physical Psychosocial Impact on daily life Gender differences Genetic polymorphism
The therapeutic relationship	Clinician qualities and experience The patient–clinician relationship Cultural factors
Areas of possible measurement	Abdominal symptoms (in the absence of pathology) Patient description of abdominal symptoms Emotional function Social function Physical function
Multiaxial assessment criteria	Advantages Disadvantages
Areas of future research (brain–gut axis)	Identification of risk factors Replication of findings

Methods

Modified Delphi approach

The Delphi approach refers to a structured, iterative process of collecting and summarising opinions from experts with the goal of forecasting the prediction of future events and group consensus building on a specific issue based on the opinion of experts [18–20]. Delphi technique can also help address problems of clinical practice when there is incomplete knowledge on a subject as is the case with this study [21,22]. This modified Delphi method surveyed a panel of experts who provided both quantitative and qualitative feedback about an evolving set of statements during several rounds of data collection. We employed three rounds of data collection with the aim of gaining over 75% consensus among participants for each item and suggestions for items to be considered in round two. Participants were asked to re-rate and further comment on items in light of their own previous rating and overall panel response on items that had less than 75%. Delphi approach has the advantage of anonymity, thus decreasing the chance of some participants modifying their response based upon the opinion of highly influential experts [19].

Participants

We purposively sampled experts from working teams and advisory boards of The Rome Foundation and The International Foundation of Functional Gastrointestinal Disorders who are currently involved in the advancement of FGID research and clinical management. Participants must have published in the last 5 years and be consistently employed in their field of expertise. While panel sizes for Delphi studies vary depending on the aims of the project, we invited 90 experts to participate. In round one, 53% ($n = 36$) of eligible participants ($n = 68$) returned completed surveys. In round two, 86% ($n = 31$) of round one participants returned completed surveys and 100% ($n = 31$) in round three yielding a 77% response rate across survey rounds. However, the overall response rate for those participating in the entire study was only 45%. There were no differences between eligible and ineligible participants. Ineligible participants responded to the invitation explaining their ineligibility or those whose email addresses were not recognised. Participation from Asia was reduced to one expert in India. Japanese participants showed willing but were unable to take part due to the 2011 tsunami, email addresses in China were returned and one expert from Singapore refused the survey due to lack of time and lack of “bandwidth”.

Survey item development

Guided by validated MAC and FGID frameworks such as the Diagnostic Statistical Manual–IV, International Classification of Disease–10, The International Association for the Study of Pain Taxonomy of Chronic Pain, The Rome Foundation and related literature, we reviewed pathophysiology of the brain–gut axis, multiaxial assessment, diagnosis and classification. We also reviewed FGID related comorbidity, novel neuroimaging techniques, endophenotypic approaches to FGIDs and the patient–practitioner relationship from both the patient and clinician's perspective. We converted review information into 132 items for round one to ensure content validity and an item set that sufficiently represented the clinical domains of FGIDs (Table 1) [18]. All items were previewed by full-time faculty at the University of Edinburgh and Karolinska Institutet in Stockholm. Furthermore, one paper and one online pilot survey were conducted whereby some items were reworded, removed or others items included both improving the reliability and testing of the survey measurement methods.

Procedures

All contact between the researchers and participants was conducted by email. Due to financial constraints, we were unable to organise face-to-face meetings with experts before or after the study. All correspondence was sent individually to maintain privacy. Two weeks before the study, potential participants received an email introducing the purpose of the study and nature of their participation, including the goal of building consensus with other experts by completing up to three surveys. When the survey began, each potential participant received a link to the online round one survey, after which up to three emails were sent to encourage participation. Similar information was provided to participants for rounds two and three.

Data analysis

Quantitative analysis

Non-parametric tests measuring central tendency and levels of dispersion were calculated for each item. We calculated agreement percentages by totalling participant “strongly agree” and “agree” Likert scores [23,24]. We also tested for internal consistency of opinion among participants using Cronbach's α as homogeneity in ranking also shows reliability with consensus among panelists [25,26]. We used Mann Whitney U test to compare survey answers over the three Delphi rounds between gastroenterologists versus other specialities, while Pearson's Chi square test was used to compare responders versus non-responders on geographic location, speciality, academic title and gender. Two sided exact significance and bootstrapping tests with a confidence interval of 95% were used as sample and inter-sample group sizes were small [27]. Statistical Package for the SPSS Version 19 was used to calculate all quantitative data.

Qualitative analysis

We analysed the comments from round one to categorise the qualitative data into specific themes. Comments were grouped from each theme into categories relating to each set of items. We identified key themes which were added to the round two survey as statements to be scored and commented on by the experts [28,29]. Qualitative and quantitative data summary sheets were sent to all participants at the end of each round.

Results

Participant characteristics

Self-reported professional background and experience among participants who completed all three round ($n = 36/31$) confirmed that the sample met the goal of reaching clinical and research experts with extensive knowledge of FGIDs (Appendix A). Participants were mostly heads of research departments, while all participants had authored or co-authored peer-review publications that were empirical investigations related to FGIDs. When we analysed the characteristics of responders versus non-responders, only job location was significant ($P = 0.0001$). There were no overall significant differences between academic title ($P = 0.768$), gender ($P = 0.278$) and speciality ($P = 0.204$). However, while psychiatrists and other experts working in psychosomatic medicine responded to the invitation, psychologists ($n = 6$) were the only profession whom none responded either to the survey or replied to any correspondence. This was possibly due to the survey being developed predominantly for medically trained experts.

We observed non-significant differences in agreement between gastroenterologists and other specialities showing differences in only 10% of items in round one, 7% in round two and 6% items in round three. Cronbach's α in round one was 0.956 showing excellent internal consistency in homogeneity of participant opinion. However, while Cronbach's α decreased through rounds two (0.853) and three (0.749), good levels of internal consistency were maintained. The reduced score shows inconsistency of participant agreement on fewer items in latter rounds. This was probably due to the wide variety of job speciality resulting in non-homogeneity of opinion regarding specific areas of FGIDs. Additionally, consensus gained on round two and three items was likely due to the inclusion of further detail to each item between rounds or experts verbally stating their following of consensus.

Delphi process findings

Overview

In round one, 96 out of 132 items gained consensus. 33 additional items were generated for round two using qualitative feedback from round one. 19 out of 69 items gained consensus, and in round three, nine items out of 50 gained consensus. 124 items from 167 gained consensus (74%) over three survey rounds.

Multiaxial assessment criteria for FGIDs

Over three Delphi rounds, experts agreed on the current need for MAC and FGIDs with 75% of participants agreeing that MAC should be developed regardless of gold-standard diagnostic testing (Table 2). Experts agreed with 78% of items exploring the advantages of MAC (Table 3) especially regarding the benefit of a systematic approach to physiological and psychological components of FGIDs (92%). Experts disagreed on the benefit of reductions in reliance on clinical judgement (55% in round three), commenting that while MAC may improve classification of FGID subtypes; it does not replace the importance of clinical judgement when evaluating FGID patients. However, many participants believe that at present MAC are the only diagnostic approach available for FGID where no biological markers exist, and should be considered now. Similar comment continued on MAC promoting “all or nothing” diagnoses in individuals with FGIDs (32% in round three), with experts stating that this issue presently occurs when using Rome questionnaires in that if patients don't meet the criteria they will not be diagnosed with an FGID. However, concerning the term “diagnosis” as implying a distinct illness with no relevance to FGIDs (19% in round three), several experts suggested a need to move beyond symptom-based systems toward the description of multidimensional endophenotypes that are quantifiable factors in gene-to-sensorimotor/behaviour pathways. Idiographic evaluation reflecting the patient's individuality caused confusion with three participants commenting that they did not understand the concept or context of the item. However, over the survey rounds agreement increased from 61% to 71%. Finally, most agreement concerning disadvantages of MAC related to time required to implement MAC and its relevance to FGIDs (65% in round three). Experts, who did comment, suggested that while MAC is time consuming, it is required, and that perception within FGID practice must change to accommodate such an approach. Further comments on this item suggest that without such a system, many patients would not be adequately evaluated.

Contributing and risk factors to FGIDs

Consensus was gained on psychosocial risk factors, the impact of symptoms on daily life and contribution of previous gastrointestinal infections to the onset and maintenance of FGIDs. Experts disagreed on physical trauma, previous surgery (both 53% in round three) and connective tissue disorders such as joint hypermobility syndrome (47% in round three) as being significant contributing factors for FGIDs. Some experts viewed surgery as an intervention in response to abdominal symptoms, or health-care seeking behaviour, whereas others commented that laparoscopic procedures have reduced post-surgical FGID risk factors. However, in contrast to physically identified causes of FGIDs, experts agreed that myofascial disorders such as fibromyalgia (94%) do contribute to the onset of FGIDs with comments suggesting these conditions are linked, having similar underlying central mechanisms.

Experts agreed with higher FGID prevalence in the female population (92%), however, opinion differed over valid explanations for these relationships. Expert agreed that increased perception of symptoms might be related to central nervous system processing of visceral stimuli (83%). However, disagreement continued over likely pathophysiological mechanisms of sex hormones on GI sensitivity and motility. Only one item gained consensus regarding contribution of genetic factors to the mediation of GI function and psychological disorders. Experts commented that while they agreed in principle with the gender and genetic-based items, lack of human data and the need to translate present diagnostic criteria into endophenotypic analysis needed to be more rigorously scrutinised.

The therapeutic relationship

82% of items relating to the therapeutic relationship, qualities and experience of the clinician and the patient's cultural background gained consensus in round one.

Table 2

Overall agreement on the future development of MAC for FGIDs over three Delphi rounds

Future development of multiaxial assessment criteria	Median Likert score	Agree%
Multiaxial assessment criteria will benefit the management of FGIDs.	4	75
Multiaxial assessment criteria may benefit the management of FGIDs (depending on results from further clinical research).	4	83
Multiaxial assessment criteria will only be valuable when further “gold standard” diagnostic testing is available.	3	28
Multiaxial assessment criteria will never benefit the management of FGIDs.	2	8

Table 3

Item content, Likert scale ratings and agreement ratings for the advantages and disadvantages of multiaxial assessment for FGID assessment over three rounds (R1, 2, 3)

Multiaxial assessment and formulation	Median	Range	Agreement %
(Relevance to FGIDs)	R1 R2 R3	R1 R2 R3	R1 R2 R3
<i>Advantages</i>			
a. It expands from single-item diagnosis to several axes that provide additional “domains” of information of high clinical value.	4	3	89
b. Reduces reliance on clinical judgement for diagnosis and therefore reduce clinical subjectivity.	3.5 4	4 3	50 55
c. Allows users to systematically approach both physiological and psychological components of FGIDs.	4	2	92
d. It can be applied in conjunction with laboratory testing (e.g., functional MRI), well validated psychological tests and self-reporting criteria in FGID patients.	4	4	83
e. Conveys large amounts of information related to disorders in the form of clinical shorthand that are otherwise difficult to communicate	4	3	83
f. Promotes structured clinical dialogue based on standardised criteria, compared to self-reporting questionnaires or loosely structured interviews.	4	3	78
g. Allows for quantitative rating of a person's mood, cognition and behaviour, which may create a profile of functioning.	4	4	81
h. Should encompass not only multiaxial evaluation but also personal idiography that reflect their individual strengths and weaknesses.	4 4	3 2	61 61
i. Can often validate the patient's own experience by informing them that others have similar experiences	4	3	78
<i>Disadvantage</i>			
a. Is only applicable to psychiatric diagnosis and therefore not appropriate for FGIDs.	2	4	25 (consensus disagreement)
b. Multiaxial assessment is a time consuming exercise and is of little value to the management of FGIDs – split into 2 questions for round 2	2 4	3 3	17 65
c. Multiaxial assessment is of little value to the management of FGIDs (round 2)	2	3	07
d. The term “diagnosis” implies a distinct illness that is therefore not relevant in many cases of FGIDs.	2	3	03
e. Multiaxial diagnostic criteria often lack clear distinctions between normal and abnormal & therefore do not avoid diagnostic consideration of ordinary problems of daily living.	2 3	3 3	14 19
f. Multiaxial assessment and diagnostic systems often sacrifice descriptive diagnostic validity for increased inter-practitioner reliability.	2 3	3 3	16 11
g. Tends to promote “all or nothing” diagnoses when considering an individual's problem (i.e., how many symptoms from a list are required before action is taken).	3 3	3 3	16 36

Participants disagreed on how clinicians share information with patients and with the effects of socioeconomic environment, educational status and spiritual attitudes on health. Experts commented that epidemiological data is lacking concerning these factors in FGID patients.

Areas for possible measurement

Consensus was gained on 75% of items relating to abdominal symptoms, the patient's description of symptoms (100%), emotional and social functioning (both 100%). However, experts agreed with only 57% of items relating to measurement of physical function. While consensus was achieved on functional impairment, experts disagreed on the need for measurement of physical disability such as kinesophobia in relation to fear of physical movement due to exacerbation of GI symptoms such as flatulence and incontinence. Experts also disagreed on the consideration of old age in FGID assessment. Some experts commented that this group is seen as important only in ruling out organic disease as FGIDs are associated with younger populations.

Future research

Experts agreed on future incorporation of well-replicated neuroscientific (80%), genetic and pharmacogenomic information (87%) that both provide stronger bases for diagnosis and therapies targeting specific FGIDs. Epigenetics as a heritable form of disorder, due especially to the effect of aberrant microbiota and its potential effect on brain–gut signalling also gained consensus (84%).

Many participants commented that while they disagreed with many items, they only did so due to a lack of reliable human data. Therefore, Table 4 shows a summary of items where consensus was not achieved, but where experts believe that further research may benefit MAC for FGIDs. Fig. 1 below summarises the mean percentages of agreement for each survey section over three Delphi rounds.

Discussion

Results from this study show that MAC is currently required for FGID evaluation in the absence of gold standard testing. This was demonstrated with agreement on the consideration of psychological comorbidity, the impact of psychosocial risk factors and psychosocial and emotional function measurement as important to MAC for FGIDs. However, this study also shows expert disagreement on evaluating gender, genetic and physical contributors citing a lack of valid and reliable human studies. Most experts agreed however, that there is a current requirement for further human research in areas of physical, gender and genetic risk factors.

Obvious areas of future research lie with differentiating between individual FGIDs and between FGIDs and related FSS sub-phenotypes. This will help us understand the relevance of the pathophysiology and endophenotypes to single-syndromic and multi-syndromic FSSs [30] which include the FGID spectrum. Currently, the somatic pain research field has made progress-investigating changes in brain structure when investigating emotional modulation and cognition using novel neuro-imaging techniques such as voxel based morphometry and arterial spin labelling [31,32]. Recently, a study showed associations between cognitive factors and anxiety in individuals with IBS [33]. Concerning physical risk factors, there is encouraging research concerning the prediction and prevention of post surgical and injury pain, especially concerning the use of psychological predictors for chronic post surgical pain [34,35], however, scant evidence based research has been carried out concerning the contribution of surgery or non-abusive injury to FGIDs. If psychological factors are pivotal to post trauma or surgical pain, investigations into centrally mediated pathophysiological mechanisms and assessment of pre, peri and post operative or post-traumatic patient characteristics should be considered sooner rather than later.

Our findings also show that experts recognise the importance of the patient–practitioner relationship. However, several factors such as socio-economic and educational status were not considered important to FGID patient evaluation. We also found similar findings concerning the patient's spiritual and religious attitudes to health care. Lower levels of economic and educational statuses have been shown to have significant effects both on the aetiology and the course of FGIDs and other FSSs [36–39]. Issues such as job security, lack of medical insurance, the number of household rooms, religious attitudes toward health and beliefs that symptoms are a “signal of harm” unrelated to the emotional experience are all important areas of consideration when evaluation these patients.

Methodological rigour

We chose the online Delphi technique as it enabled consultation with geographically dispersed multiprofessional FGID experts more easily than other consensus methods where data can be collected and analysed immediately. We obtained quantitative and qualitative data on key themes shown in the results and discussed above. However, Delphi methods are hard to apply to healthcare research. This is due to the transfer of measurements across quantitative (Likert scales) and qualitative (comment and opinion) models. Continuous modifications and the lack of guidelines make it difficult to test for

Table 4

Levels of agreement for statements where consensus was not achieved in subject areas

Statements	Round 1	Round 2	Round 3
Physical risk and contributing factors			
• Previous truncal surgery (effects of repair and stress mechanisms on GI function)	53%	62%	53%
• Previous physical truncal injury (effects of injury and stress mechanisms on GI function)	67%	42%	53%
• Ongoing connective tissue disorders (secondary abnormalities such as hiatus hernia and rectal prolapse)	Inc in R2	38%	47%
Gender differences			
• Effect of sex hormones on GI sensitivity and motility (e.g., female FGID patients experience more intense symptoms such as bloating and abdominal pain)	62%	74%	71%
• Effect of sex hormones on nociceptive processing (e.g., influence of oestrogens on nociceptive signalling pathways)	64%	74%	65%
• Effects of sex hormones in response to pharmacological treatment (e.g., hormone receptor modulation as a therapeutic option in menstruating women)	Inc in R2	70%	71%
Genetic polymorphism			
• Contribution of genetic factors to mediation of psychological disorders in FGID patients (effects of altered behaviour on GI motility)	61%	71%	74%
• Genetic factors modulate visceral pain pathways (e.g., genes controlling G-protein synthesis)	58%	68%	71%
• Genetic factors affecting neuro-immune function in FGID patients (e.g., polymorphism of TNF α during inflammatory response)	64%	69%	68%
• Epigenetics; the heritable change in gene function without changes in DNA sequencing in FGID patients (e.g., microbiota-mediated changes in regulation of GI immune system)	Inc in R2	55%	65%
Consideration in therapeutic relationship			
• The patient's socioeconomic status (e.g., job security, medical insurance circumstances)	67%	61%	58%
• The patient's education status (e.g., lower educational levels open beliefs that symptoms are a "signal of harm" unrelated to the emotional experience)	58%	58%	48%
• The patient's religious and spiritual attitudes (e.g., interpretation of symptoms and attitudes toward treatment)	39%	45%	42%
Consideration for physical function measurement			
• Old age and FGIDs (symptoms are under-recognised due to associations with younger age groups)	33%	39%	52%
• Kinesiophobia – fear of physical activity due to recurrence of GI symptoms (e.g., flatulence, belching etc with quick unexpected movement)	36%	29%	36%
Multiaxial assessment			
• Multiaxial assessment reduces reliance on clinical judgement and therefore reduced clinical subjectivity	50%	55%	55%
• Must encompass standardised multiaxial assessment criteria and idiographic evaluation of the patient as an individual	61%	61%	71%

rigour. However, by twice piloting and refining the survey, our first Delphi survey gained excellent inter-rater reliability scores using Cronbach's α . Cronbach scores declined due to inconsistency of participant agreement on fewer items; however, good reliability was maintained through the three Delphi rounds. Reliability was also improved by ongoing iteration and increased expert feedback over the course of the study, while content and face validity as improved by developing a closed ended first round survey using empirically based data. Although the lack of accountability for views expressed by participants were not verified, purposive sampling of high ranking experts from both academic and research fields in combination with low drop-out rates after round one helped to establish internal validity. "Trustworthiness" is now considered more appropriate than reliability and

validity to measure the effectiveness of Delphi technique. The four strategies for establishing trustworthiness are shown below and therefore, as recommended, we also met trustworthiness criteria for this Delphi study [40,41] (Appendix B).

Study limitations

Our methods were tailored to the resources available. Face-to-face and online in-depth interviews were considered. These formats are flexible and are an excellent means of obtaining ideas by probing interviewees about their opinion [42]. However, the time required to complete interviews on such a large number of items was considered too long for busy experts. Additionally, the costs of travelling worldwide

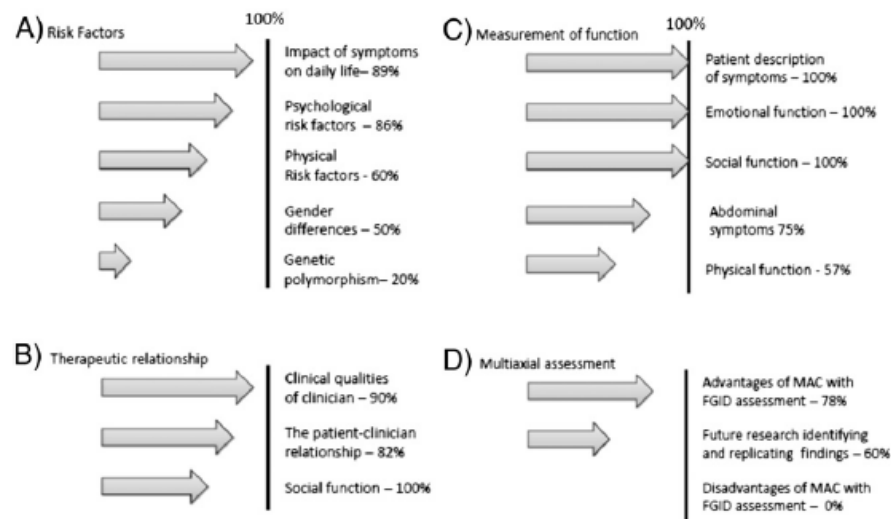


Fig. 1. Mean percentages of consensus over 3 survey rounds for items relating to A) risk factors, B) the therapeutic relationship, C) domains of measurement and D) Multiaxial assessment criteria (MAC) concerning FGIDs.

to meet each expert or unreliable internet bandwidth to complete full interviews made these options impracticable.

Research on FGIDs and MAC is extensive as was the variety of expert sampled. This had a negative effect on response rates and item bias. Firstly, professions such as physiologists and nurses felt that the parts of the survey were not in their area of expertise. Secondly, due to the extent of information in this study, there was an uneven distribution both in numbers of experts from each area of expertise and positive-scoring items in round one. We recommend that future studies of this kind focus on smaller areas of knowledge, sampling experts from one area of expertise. The terms “FGIDs” and “multiaxial assessment” are especially broad and as such, probably hindered expert contribution due to the large number of items and the time required to complete the survey.

Potential FGID assessment criteria framework

Features of the World Psychiatric Association's International Guidelines for Diagnostic Assessment (IGDA) formulation (Appendix C) are well-suited to the nature of FGIDs and with some adaptation, would be an appropriate model of assessment [43]. This approach considers a wide range of areas that can be tailored toward the individuality of the patient. One may further take a “stepwise approach” as recommended by Gatchel R who on biopsychosocial assessments noted that evaluation could have greater impact when the order of steps are coordinated to meet the needs of a specific patient [44], in this case, depending on the importance of each axis.

Diagnostic evaluation is an essential feature of clinical care. It involves the gathering of information to describe and understand the patient's clinical condition and to manage effective care [45]. Concerning this study and from current literature that, care of FGID patients should be conducted from the beginning with a clear therapeutic goal, engaging both the patient and their family using a comprehensive range of standardised, reliable and valid multiaxial assessment tools that reflect the many domains of the FGID experience. This of course requires the clinician to be both scientifically competent and humane in their approach to FGIDs. This can only happen if current data is available concerning both identifiable risk factors and replicated findings. As a final thought, symptom-only based labels do not describe underlying pathophysiological processes and lead to nomenclature like “functional vomiting syndrome”. How do we explain these types of diagnoses to patients who are vomiting and have come to the end of their care strategy?

Conclusion

This Delphi technique not only shows expert consensus, but also differences of opinion on domains of information relevant to the evaluation of FGIDs. Experts agree that there is a current requirement to develop MAC now, regardless of gold standard testing and to the importance of psychosocial risk factors and measurement. We recommend that further human research addressing physical, gender and genetic risk factors be shared by research groups from authorities concerning FGIDs who may then develop a template for MAC that provides further dimensions to present diagnostic criteria.

Conflicts of interest

There are no financial and personal relationships with other people or organisations that could potentially and inappropriately influence (bias) their work and conclusions. Examples of potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and research grants or other funding.

Acknowledgments

We wish to thank Dr Gordon Drummond, Prof Douglas Drossman and Dr Tobias Sundberg for their support and invaluable advice on the design, development and publication of this study.

Appendix A. Delphi participants completing in all three rounds

Variable		Round 1 N (%)	Rounds 2 & 3
Academic title	Professor	26 (72)	24 (77)
	Associate Professor	2 (6)	1 (3)
Primary occupational setting	Senior research fellow	8 (22)	6 (19)
	Gastroenterology	19 (53)	15 (48)
	Psychiatry	3 (8)	2 (7)
	Paediatrics	3 (8)	3 (10)
	Primary care	2 (6)	2 (7)
	General surgery	2 (6)	2 (7)
	Pain medicine	2 (6)	2 (7)
	Physiology	2 (6)	2 (7)
	Integrative medicine	1 (3)	1 (3)
	Nursing	1 (3)	1 (3)
Consecutive years experience with FGIDs	Pharmacology	1 (3)	1 (3)
	Median: 20 years		
Job location	Range: 4–30 years (26)		
	Europe	22 (61)	20 (64)
	North America	12 (33)	10 (32)
	Asia & Australia	2 (14)	1 (3)

Appendix B. Criteria for trustworthiness

Credibility	<ul style="list-style-type: none"> • The use of an audit trail throughout the study • Ongoing iteration • Increased expert feedback in later rounds • Low participant drop-out rate after round one
Dependability	<ul style="list-style-type: none"> • The range and representative expert sample (wide demographic and range of expertise)
Conformability	<ul style="list-style-type: none"> • Detailed description of data collection and analysis
Transferability	<ul style="list-style-type: none"> • Verification of applicability of Delphi findings <ul style="list-style-type: none"> ○ Worldwide use of information (range of expert sample) ○ Potential use of information in the development of multiaxial assessment criteria for FGIDs ○ Highlight gaps in knowledge that warrant further research

Appendix C. IGDA. 7: Standardised multi-axial diagnostic formulation. Source: World Psychiatric Association

Axis	Factors
I	Clinical disorders (mental and general medical conditions)
II	Disabilities (in personal care, occupational functioning, functioning with family, and broader social functioning)
III	Contextual factors (interpersonal and other psychosocial and environmental problems)
IV	Quality of life (primarily reflecting the patients self-perceptions)

References

- [1] International Foundation for Functional Gastrointestinal Disorders. Reporter's guide to functional gastrointestinal disorders; 2009 [Milwaukee].
- [2] Halder SLS, Locke GR. Epidemiology and socioeconomic impact of visceral and abdominal pain syndromes. In: Pasricha PJWW, Gebhart GF, editors. Chronic abdominal and visceral pain: theory and practice. New York, London: Inroma Healthcare; 2007.
- [3] Barbara G, Cremon C, De Giorgio R, Dotthel G, Zecchi L, Bellacosa L, et al. Mechanisms underlying visceral hypersensitivity in irritable bowel syndrome. *Curr Gastroenterol Rep* Aug 2011;13:308–15.
- [4] Cervero F. Visceral versus somatic pain: similarities and differences. *Dig Dis* 2009;27:3–10.

- [5] Rome Foundation. Rome III diagnostic questionnaire for the adult functional GI disorders (including alarm questions) and scoring algorithm. Raleigh: Rome Foundation Inc.; 2006.
- [6] Fortea J, Prior M. Irritable bowel syndrome with constipation: a European-focused systematic literature review of disease burden. *J Med Econ* 2013;16:329–41.
- [7] Mikocka-Walus A, Turnbull D, Moulding N, Wilson I, Andrews JM, Holtmann G. Psychological comorbidity and complexity of gastrointestinal symptoms in clinically diagnosed irritable bowel syndrome patients. *J Gastroenterol Hepatol* Jul 2008;23:1137–43.
- [8] Camilleri M, Carlson P, Zinsmeister AR, McKinzie S, Busciglio I, Burton D. Mitochondrial DNA and gastrointestinal motor and sensory functions in health and functional gastrointestinal disorders. *Am J Physiol Gastrointest Liver Physiol* Mar 2009;296:G510–6.
- [9] Barbara G, Cremon C, Pallotti F, De Giorgio R, Stanghellini V, Corinaldesi R. Postinfectious irritable bowel syndrome. *J Pediatr Gastroenterol Nutr* Apr 2009;48:S95–7.
- [10] Chang L, editor. From Rome to Los Angeles – the Rome III criteria for the functional GI disorders. Digestive diseases week, Medscape gastroenterology; 2006.
- [11] The Rome Foundation. Rome III disorders and criteria. Mclean VA: The Rome Foundation; 2006 [cited 2010 Oct 19th].
- [12] Sperber AD, Drossman DA, Quigley EM, Simren M, Barbara G, et al. The global perspective on irritable bowel syndrome: a Rome Foundation-World Gastroenterology Organisation symposium. *Am J Gastroenterol* 2012;107:1602–9.
- [13] Spiegel B, Camilleri M, Bolus R, Andresen V, Chey WD, Fehnel S, et al. Psychometric evaluation of patient-reported outcomes in irritable bowel syndrome randomized controlled trials: a Rome Foundation report. *Gastroenterology* Dec 2009;137:1944–53 [e1–3].
- [14] Mezzich JE, Salloum IM. Towards innovative international classification and diagnostic systems: ICD-11 and person-centered integrative diagnosis. *Acta Psychiatr Scand* Jul 2007;116:1–5.
- [15] Oken D. Multiaxial diagnosis and the psychosomatic model of disease. *Psychosom Med* March 1 2000;62:171–5.
- [16] Banzato CEM, Jorge MR, Kastrup M. Multiaxial schemas for psychiatric diagnosis. In: Salloum IM, Mezzich JE, editors. *Psychiatric diagnosis: challenges and prospects*. Wiley; 2009. p. 281–2.
- [17] López-Colombo A, Morgan D, Bravo-González D, Montiel-Jarquín A, Méndez-Martínez S, Schmulson M. The epidemiology of functional gastrointestinal disorders in Mexico: a population-based study. *Gastroenterol Res Pract* 2012;2012:8.
- [18] Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. *Technol Forecast Soc Change* 2011;78:1695–704. <http://dx.doi.org/10.1016/j.techfore.2011.04.005>.
- [19] Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007;12:1–8.
- [20] Keeney S, Hasson F, McKenna H. The Delphi technique. Delphi technique in nursing and health research. Oxford: Wiley Blackwell; 2011 10–2.
- [21] Linstone H, Turoff M. The Delphi method: techniques and applications. Reading, MA: Addison-Wesley Publishing Company; 2002.
- [22] Keeney S, Hasson F, McKenna H. Conducting the research using the Delphi technique. The Delphi technique in nursing and health research. Oxford: Wiley Blackwell; 2011.
- [23] McIlpatrick SJ, Keeney S. Identifying cancer nursing research priorities using the Delphi technique. *J Adv Nurs* 2003;42:629–36.
- [24] Jirwe M, Gemish K, Keeney S, Emani A. Identifying the core components of cultural competence findings from a Delphi study. *J Clin Nurs* 2008;18:2622–34.
- [25] Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol* 2003;56:1150–6.
- [26] Bland JM, Altman DG. Cronbach's alpha. *BMJ* Feb 22 1997;314:572.
- [27] Hinneburg A, Mannila H, Kaislaniemi S, Nevalainen T, Raunio Brunberg H. How to handle small samples: bootstrap and Bayesian methods in the analysis of linguistic change. *Lit Linguist Comput* June 1 2007;22:137–50.
- [28] Aronson J. A pragmatic view of thematic analysis. The qualitative report [serial on the internet]. 2(1); 1994. [Available from: <http://www.nova.edu/ssss/QR/Backissues/QR2-1/aronson.html>].
- [29] Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;3:77–101.
- [30] White PD. Functional somatic syndromes may be either “polysyndromic” or “monosyndromic”. *J Psychosom Res* 2013;74:2–3.
- [31] Roy M, Piche M, Chen JI, Peretz I, Rainville P. Cerebral and spinal modulation of pain by emotions. *Proc Natl Acad Sci U S A* Dec 8 2009;106:20900–5.
- [32] Van Oudenhove L. Understanding gut–brain interactions in gastrointestinal pain by neuroimaging: lessons from somatic pain studies. *Neurogastroenterol Motil* 2011;23:292–302.
- [33] Sugaya N, Nomura S, Shimada H. Relationship between cognitive factors and anxiety in individuals with irritable bowel syndrome. *Int J Behav Med* Sep 2012;19:308–15.
- [34] Powell R, Johnston M, Smith WC, King PM, Chambers WA, Krukowski Z, et al. Psychological risk factors for chronic post-surgical pain after inguinal hernia repair surgery: a prospective cohort study. *Eur J Pain* Apr 2012;16:600–10.
- [35] Hinrichs-Rocker A, Schulz K, Järvinen I, Iefering R, Simanski C, Neugebauer EAM. Psychosocial predictors and correlates for chronic post-surgical pain (CPSP) – a systematic review. *Eur J Pain* 2009;13:719–30.
- [36] Soares NC, Ford AC. Prevalence of, and risk factors for, chronic idiopathic constipation in the community: systematic review and meta-analysis. *Am J Gastroenterol* Sep 2011;106:1582–91 [quiz 1, 92].
- [37] Bytzer P, Howell S, Leemon M, Young LJ, Jones MP, Talley NJ. Low socioeconomic class is a risk factor for upper and lower gastrointestinal symptoms: a population based study in 15,000 Australian adults. *Gut* Jul 2001;49:66–72.
- [38] Johannes CB, Le TK, Zhou X, Johnston JA, Dworkin RH. The prevalence of chronic pain in United States adults: results of an internet-based survey. *J Pain* Nov 2010;11:1230–9.
- [39] Gundel H, Ladwig KH, Wolowski A, Fischer A, Grubl A, Marten-Mittag B, et al. Psychic and somatic findings in jaw or facial pain of unclear origin. Comparison of patients with severe and mild symptoms. *Schmerz* Aug 2002;16:285–93.
- [40] Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. *Technol Forecast Soc* 2011;78:1695–704. <http://dx.doi.org/10.1016/j.techfore.2011.04.005>.
- [41] Day J, Bobeva M. A generic toolkit for the successful management of Delphi studies. *Electron J Bus Res Methods* 2005;3:103–16.
- [42] Gordon TJ. The Delphi method. *Future Res Methodol* 1994;3:29.
- [43] World Psychiatric Association. IGDA. 7: standardised multi-axial diagnostic formulation. *Br J Psychiatry* May 1 2003;182:s52–4.
- [44] Gatchel RJ. A biopsychosocial overview of pretreatment screening of patients with pain. *Clin J Pain* Sep 2001;17:192–9.
- [45] Mezzich JE, The WPA. International guidelines for diagnostic assessment. *World Psychiatry* 2002;1:36–9.