This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, D Clin Psychol) 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.
GOOGLE SEARCH AND THE MEDIATION OF DIGITAL HEALTH INFORMATION

A CASE STUDY ON UNPROVEN STEM CELL TREATMENTS

Anna Couturier
Doctoral Thesis
Science and Technology Studies
The University of Edinburgh
2022
For Mum and Dad,
Clea and Tristan.
Google Search occupies a unique space within broader discussions of direct-to-consumer marketing of stem cell treatments in digital spaces. For patients, researchers, regulators, and the wider public, the search platform influences the who, what, where, and why of stem cell treatment information online. Ubiquitous and opaque, Google Search mediates which users are presented what types of content when these stakeholders engage in online searches around health information. The platform also sways the activities of content producers and the characteristics of the content they produce. For those seeking and studying information on digital health, this platform influence raises difficult questions around risk, authority, intervention, and oversight.

This thesis addresses a critical gap in digital methodologies used in mapping and characterising that influence as part of wider debates around algorithmic accountability within STS and digital health scholarship. By adopting a novel methodological approach to Blackbox auditing and data collection, I provide a unique evidentiary base for the analysis of ads, organic results, and the platform mechanisms of influence on queries related to stem cell treatments. I explore the question: how does Google Search mediate information that people access online about ‘proven’ and ‘unproven’ stem cell treatments?

Here I show that, in spite of a general ban on advertisements of stem cell treatments, users continue to be presented with content promoting unproven treatments. The types, frequency, and commercial intent of results related to stem cell treatments shifted across user groups including geography and, more troublingly, those impacted by Parkinson’s Disease and Multiple Sclerosis. Additionally, I find evidence that the technological structure of Google Search itself enables primary and secondary commercial activities around the mediation and dissemination of health information online. It suggests that Google Search’s algorithmically-mediated rendering of search results – including both commercial and non-commercial activities - has critical implications for the present and future of digital health studies.
Google Search plays a vital role in how we gather information on our own health – particularly when we face life-altering diseases and conditions, like Parkinson’s Disease and Multiple Sclerosis, and search for potential treatments. At the same time, Google Search presents difficult challenges to study as it is not possible to see directly what decisions the platform makes when it selects what information will be included, excluded, or promoted in results.

This research demonstrates that information presented to users around stem cell treatments includes advertisements for unproven treatments (despite a ban on these advertisements by Google) as well as significant portion of commercial results. Additionally, my research indicates that Google Search is significantly influenced by for-profit activities that treat users as a product for content creators (like private clinics and Charities/NGOs). My research further explores the implications of these findings on how we think about the health information we find through technology platforms like Google Search. Finally, my research offers new ways for social scientists to study internet platforms like Google Search using data from human and virtual study participants.
The first and foremost - thank you to my advisors, James Mittra and Fadhila Mazanderani. Thank you for your efforts, your commiseration, your reassurances, your unflappability, and your invaluable perspectives. The guidance you both offered so enthusiastically has carried me and my research through a gauntlet of technical setbacks, funding obstacles, ever-changing Google policies, and *an actual worldwide pandemic*. I am ever grateful.

This research would not have been technically possible without my collaboration with a brilliant team of researchers at the Algorithm Accountability Lab of the University of Kaiserslautern. A heartfelt thank you to Tobias Krafft (for your fellowship and insight), Roman Krafft, and group leader Prof. Dr. Katharina Zweig. In particular, I owe this project to the industry, patience, and problem-solving of Martin Reber.

I am sincerely grateful for the above-and-beyond encouragement and support from Amanda Waite and Prof. Clare Blackburn of EuroGCT/EuroStemCell at the University of Edinburgh. I am immensely proud of the challenges we have taken on together since 2016. This research would not have happened without you both. I am also indebted to the current and former public engagement team based at the Centre for Regenerative Medicine for all that they taught me over the years: Dr. Jan Barfoot, Dr. Cathy Southworth, Dr. Robin Morton, Ellie Rogers, Dr. Hsin-Yu Kuo, and Jennifer Lorigan. Thank you.

I would like to extend my appreciation for the staff and faulty in Science, Technology, and Innovation Studies at the University of Edinburgh for their assistance at each stage of my studies. I would be remiss to forget all the logistical help in organising my in-person workshops provided by Margaret Acton and Munro.

I would also like to thank the Faculty and Administration of the Graduate School of Social and Political Science for their financial assistance throughout my studies as Graduate Scholar. This project has appreciatively received additional funding and material support from the department of Science, Technology and Innovation Studies
and the Centre for Regenerative Medicine at the University of Edinburgh as well as from the European Commission through its H2020 project, EuroStemCell and EuroGCT.

This project is deeply indebted to the participation, intuition, and feedback from the members of Parkinson’s UK (Alison Williams, David Adams, David Melton), Edinburgh Parkinson’s Research Initiative (Martin Taylor), Young Onset Parkinson’s Disease (YOPD) (Gaynor Edwards), the Anne Rowling Clinic (Alison Irving, Pamela MacDonald, Dawn Lyle, and Judith Newton) and the many other patient advocacy groups and individuals who participated. Whether through becoming a data donor or through the conversations we had, thank you for sharing your knowledge, experience, and advice.

Special thanks to Prof. Megan Munsie at the University of Melbourne and the late Prof. Andrew Webster of York University. I greatly appreciate your early inspiration and the time you spent brainstorming with me as I navigated this academic space that you both have so clearly impacted for the better.

My parents, Richard Couturier and Maureen Smith, and my sister, Kate, who said so reassuringly at that dinner in Amsterdam, “Yes. Of course. Do it.” This research grew from the first-hand knowledge you have shared with me of how we navigate health in a modern world together – all the good and the challenging. I can only hope to pay forward the love and support you have given me. Thank you to Isobel, Robert, Ben, Betty, Geof, Fred, Sy, and wee Dex. I will never stop finding new ways to appreciate getting to call you all “family”.

Clea and Tristan: If you are reading this, you’ve probably picked up this big dusty book while on the hunt for embarrassing old yearbooks or weights for a Shooter’s Sandwich. I know it must be rather disappointing to just see hundreds of pages of boring text. Just know that whatever you do (even if you ever decide that you want to write one of these!), I will be right there cheering you on. Thank you both for being the beautiful humans that you are.

And finally – Faye. For absolutely everything.
TABLE OF FIGURES

FIGURE 1: Snapshot of Google Search for the query “stem cell” featuring ‘Knowledge Graph’ returns, Captured by author 12.4.2022. ................................................................. 94
FIGURE 2: Snapshot of Google Search page “Removing Content from Google”, audited by author 12.4.2022. .................................................................................................................. 96
FIGURE 3: Screenshot of www.eurostemcell.org/datadonation, Provided by author 2021. .... 133
FIGURE 4: Unified Modelling Language (UML) activity diagram of the EuroGCT Data Donation registration process, courtesy of Martin Reber, University of Kaiserslautern 2019. .... 134
FIGURE 5: EuroStemCell Data Donation plugin as it appeared in the Mozilla Firefox store, Captured by author 29.8.2019. ........................................................................................................... 135
FIGURE 6: Screenshot of the EuroStemCell Data Donation survey displayed to data donors and required in order to install the plugin. Captured by author 29.08.2019. ................................. 136
FIGURE 7: Sketch of the plugin-server-communication of the EuroStemCell Data Donation, courtesy of Martin Reber, University of Kaiserslautern 2019. .............................................................. 137
FIGURE 8: Example of raw Google Search Result page source code for the query “stem cell treatment”, captured by author 22.1.2022. ........................................................................................................... 138
FIGURE 9: Raw Markup of Google Ad on Search Result Page for Query: “stem cell treatments” : Cells4Life, Captured by Author, 1.11.2021. ......................................................................................................... 139
FIGURE 10: Raw Markup of Google Organic Result on search result page for query “stem cell treatments”: NHS.uk, captured by author, 1.11.2021. .................................................................................. 140
FIGURE 11: “Thank You” and example results page provided to donors after scheduled data scrape, provided by author 2019. ........................................................................................................... 143
FIGURE 12: Data Donations. Figure courtesy of Reber, Krafft, Krafft, Zweig, and Couturier 2020. .......................................................................................................................... 156
FIGURE 15: Screenshot from Parkinsons.org.uk. Captured by author 12.2.2021. ........... 163
FIGURE 16: Screenshot from Stemaidinstitut.eu. Captured by author 2.6.2021. ............... 164
FIGURE 17: Screenshot from sct4me.com. Captured by author 2.6.2021. ......................... 165
FIGURE 22: Screenshot from Sciworthy.com. Captured by author 01.7.21. ................. 169
FIGURE 23: Example image of “prescription treatment websites” unreadable ads within the dataset. By author 20.7.2021. ......................................................................................................... 170
FIGURE 25: Example image of ‘Other’ unreadable ads within the dataset. By author 20.7.2021. .......... 172
FIGURE 26: Screenshot from private clinic advertising an unproven stem cell treatment, EmCell.com. Captured by author, 15.12.2020. ......................................................................................................... 173
FIGURE 27: Screenshot from private clinic advertising, SwissMedica. Captured by author, 5.3.2021. ................................................................. 174
FIGURE 28: Screenshot from the Google Health Policy Blog, Captured by Author 02.09.2019 ................................................................. 180
FIGURE 29: Depicting Absolute Rate Of All ads collected .................................................. 183
FIGURE 30 Absolute Rate Of All ads collected by Condition .................................................. 184
FIGURE 31 Ad Returns from US and UK Donors ................................................................. 184
FIGURE 32: Total Ad Returns by Query .............................................................................. 185
FIGURE 33: Top 25 Advertising Sources .............................................................................. 186
FIGURE 34: Ad Return Category percentage Share ............................................................... 187
FIGURE 35: Screenshot from clean research browser showing Google Search query for “Parkinson’s stem cell treatment” with advertisements from two clinics promoting unproven treatments. Captured by author 8.7.2021 ................................................................. 188
FIGURE 36: Ad return categories ......................................................................................... 189
FIGURE 37: Ad returns of Charities/NGOs vs Clinics over time ............................................ 190
FIGURE 38: VPS vs. Human Data Donor ad returns over time in UK Parkinson’s Study .............. 191
FIGURE 39: Fraction of ads in total donations per participant, Figure courtesy of Reber, Krafft, Krafft, Zweig, and Couturier 2020 ................................................................. 192
FIGURE 40: Returns of ads including explicit reference to unproven treatments vs all other categories over time ............................................................................................. 195
FIGURE 41: Target Audience over time .............................................................................. 196
FIGURE 42: Non-Commercial vs. Commercial ad returns over time ..................................... 197
FIGURE 43: Screenshot from Leiden University, Captured by author 22.6.2022 .................................................. 200
FIGURE 44: Screenshot taken from private clinic www.cellmedicine.com, Captured by author, 15.12.2020 ........................................................................................................ 219
FIGURE 45: Snapshot of Google Search for the query “Parkinson’s stem cell treatment” highlighting ads (green), organic (orange), and ‘Knowledge Graph’ (purple) returns, audited by author 14.1.2021 .................................................. 238
FIGURE 46: Snapshot of Google Search for the query “stem cells” highlighting ads (green) and organic (orange) returns, audited by author 1.10.2019 .................................................. 239
FIGURE 47: Snapshot of Google Search for the query “stem cells multiple sclerosis” highlighting ads (green), organic (orange) and ‘Knowledge Graph’ (purple) returns, audited by author 4.16.2018 .................................................. 239
FIGURE 48: Snapshot of Google Search for the query “stem cell cure” highlighting organic (orange) and ‘Knowledge Graph’ (purple) returns, audited by author 1.10.2019 .................................................. 240
FIGURE 49: Rich Text Snippet (purple) for EuroStemCell and SwissMedia, captured by author 20.2.2020 ........................................................................................................ 244
FIGURE 50: Screenshot of https://www.mayoclinic.org/diseases-conditions/parkinsons-disease, Captured by Author 22.2.22 .................................................. 250
Figure 51: Screenshot of Source Code for HTTPS://WWW.MAYOCLINIC.ORG/DISEASES-
            CONDITIONS/PARKINSONS-DISEASE, Captured by Author 22.2.22. ............................................250
Figure 52: Top 100 Webpages for Multiple Sclerosis and Parkinson’s Disease across dataset
            with inclusion or exclusion of Schema.org markup. .................................................................252
Figure 53: Screenshot of WWW.STEMCELLMEXICO.ORG, Captured by Author 29.12.2021. .................252
Figure 54: Screenshot of Microdata Markup included in WWW.STEMCELLMEXICO.ORG, Captured by
            Author 29.12.2021 ....................................................................................................................253
Figure 55: Screenshot of Yoast.com Homepage, Captured by Author 22.2.2022 ..................................257
Figure 56: Screen Capture of Query “Parkinson’s Stem Cell Treatment” on Google Search,
            Captured by Author 13.1.2020 ....................................................................................................259
Figure 57: Popularity of search query “Stem Cell Treatment” 2004-2022, scraped from Google
            Trends, retrieved by Author 01.2022 ..........................................................................................264
Figure 58: Worldwide popularity of search query “Stem Cell Treatment Cost” 2004-2022,
            scraped from Google Trends, retrieved by Author 01.2022 ......................................................265
TABLES OF TABLES

TABLE 1: BREAKDOWN OF DONOR TYPES  ................................................................. 154
TABLE 2: BREAKDOWN OF DONOR RESULTS BY COUNTRY  .............................................. 155
TABLE 3: BREAKDOWN OF CONDITION ASSOCIATION OF THOSE IDENTIFIED AS IMPACTED BY DISEASE ....... 155
TABLE 4: BREAKDOWN OF AD RETURNS BY P VALUE AND SIGNIFICANCE COMPUTED FROM SIMPLE LINEAR REGRESSION ................................................................. 189
TABLE 5: BREAKDOWN OF CLINICS BY TREATMENTS OFFERED, AD COUNT, AND SHARE OF RETURNS. ........... 194
TABLE 6: VPS ORGANIC RETURNS FOR STUDY 6 - PARKINSON’S DISEASE IN THE UNITED KINGDOM BY ‘TYPE’ .................................................................................................. 211
TABLE 7: VPS SEARCH RETURN AVERAGE POSITION BY TYPE FOR PARKINSON’S DISEASE IN THE UNITED KINGDOM ...................................................................................... 211
TABLE 8: TOP 20 VPS RESULT SOURCES FOR PARKINSON’S DISEASE IN THE UK BY FREQUENCY .......................................................... 215
TABLE 9: VPS ORGANIC RESULTS FREQUENCY AND AVERAGE POSITION FOR STUDY 9 - PARKINSON’S DISEASE IN THE UNITED STATES BY ‘TYPE’ .................................................................................................. 216
TABLE 10: TOP 20 VPS RESULT SOURCES FOR PARKINSON’S DISEASE QUERIES IN THE US BY FREQUENCY AND POSITION .................................................................................................. 218
TABLE 11: PERCENTAGE SHARE OF ORGANIC VPS RESULTS BY ‘TYPE’ – UK AND US ........................................... 220
TABLE 12: PERCENTAGE SHARE OF ORGANIC VPS RESULTS BY ‘TYPE’ - UK VPS VS UK HUMAN DATA DONORS ............................................................................................................. 221
TABLE 13: TOP 20 HUMAN DATA DONOR RESULT SOURCES FOR PARKINSON’S DISEASE QUERIES IN THE UK BY FREQUENCY AND POSITION .................................................................................................. 222
TABLE 15: SEARCH RETURN SHARE BY QUERY FOR PARKINSON’S STUDY - UNITED KINGDOM ........................ 225
TABLE 16: SEARCH RETURN SHARE BY TYPE OF “CAN STEM CELL’S HELP ME?” QUERIES VS. ALL QUERIES IN THE UNITED KINGDOM .................................................................................................. 225
TABLE 17: SEARCH RETURN SHARE BY TYPE FOR “CAN STEM CELLS HELP ME?” QUERIES VS. ALL QUERIES IN THE UNITED STATES .................................................................................................. 226
TABLE 18: SEARCH RETURN SHARE BY TYPE FOR ALL QUERIES VS “CURE” QUERIES IN THE UNITED KINGDOM (PARKINSON’S DISEASE) .................................................................................................. 227
TABLE 19: SEARCH RETURN SHARE BY TYPE FOR ALL QUERIES VS “CURE” QUERIES IN THE UNITED STATES (PARKINSON’S DISEASE) .................................................................................................. 227
TABLE 20: SEARCH RETURN SHARE BY TYPE FOR “COST” QUERIES VS. ALL QUERIES IN THE UNITED KINGDOM .......................................................................................................... 228
TABLE 21: SEARCH RETURN SHARE BY TYPE FOR “COST” QUERIES IN THE UNITED STATES ............................................. 229
TABLE 22: CLINIC CONTENT DISPLAYED IN UNITED KINGDOM BY QUERY .................................................................................. 230
TABLE 23: CLINIC CONTENT DISPLAYED IN UNITED STATES BY QUERY .................................................................................. 230
It’s three a.m. You’ve been recently diagnosed with a life-altering condition. You pull out your phone or laptop and begin to Google. What do you find?

Since 2016, as the Digital Project Manager for the European Union-funded EuroStemCell project, I have used these words in meetings, talks and workshops to prompt discussion. The decade long EuroStemCell¹ project is a network of over 400 stem cell researchers throughout the EU focused on bringing stem cell scientists and publics together through science communication, educational resource creation, and training. Its centrepiece is a website of factsheets and educational resources on stem cells visited by one million people each year.

Jokingly, my colleagues referred to this turn of phrase as *Anna’s 3 A.M. Test*; a yardstick against which we could measure real impact, value, and motivation to undertake public engagement work in digital spaces. However, throughout these discussions an uninterrogated spectre loomed, Google Search² itself. While we would throw the conversation back and forth with researchers on what *proven* or *unproven* treatment meant, the platform of Google Search was assumed to be infrastructure; *de facto* public commons by virtue of its overwhelming share of global internet usage. When we would listen to patients recounting their first days of diagnosis and their personal search for information around their health trajectories, Google Search was interacted with as a neutral set piece. As regulators, teachers, patient advocates, healthcare providers and members of the public expressed their frustrations around *which* treatments were available *when* and *how* and *why* and *where*, they placed Google Search in the role of a blank automaton, lifting offerings

---

¹ [https://www.eurostemcell.org](https://www.eurostemcell.org)

² Throughout my research, I refer to my subject as “Google Search” when referring to the platform of Google Search and the section of Google, the company, tasked with the policy, development, and publishing of the Google Search platform. When I use the term “Google” I am speaking about the larger corporate entity that includes not only Google Search but other corporate departments, technologies and platforms, including YouTube.
of health content when its lever was pulled.

As one of my responsibilities included the development and administration of the EuroStemCell website, I was privy to the technical space that this digital resource occupied within digital health discovery platforms like Google Search. Crucially, 80% of the web traffic to the site came directly from organic³ Google Searches; ‘What are stem cells?’ ‘Can stem cells treat my disease?’ ‘Stem cell treatment cure’. This indicated that – deliberately or not – the digital space developed through our work was part of some flow of information Google facilitated.

I also observed and administered hundreds of patient enquiries about stem cell treatments each year. Reading through these direct enquiries made it clear to me how much is at stake for people affected by serious health conditions and their desire to gain access to stem cell therapies. While contributing to a research paper with colleagues from the University of Regina and Stem Cells Australia, I was struck by the timbre of language used by enquirers. One individual impacted by a serious condition wrote, “Don’t use lab rats – use me instead.” (Zarzeczny et al., 2018).

My research has developed from and been informed by this experience within a niche space of biomedical research and public engagement. My work forced me to engage with both the activity of public-facing information creation and the technical aspects of web development reliant on discovery through algorithmically mediated platforms like Google Search. Through this undertaking, it became obvious that there needed to be a way to bring Google Search into discussions about the implications of digital technologies for health in a grounded, evidence rich way. This thesis was sparked by those conversations – with stem cell scientists, public engagement practitioners, computer scientists, and patients.

In my research, I explore the impact of the Google Search platform on the mediation of information for patients and other members of the public engaging in search within the online healthcare context. Although I explore this concept later in this chapter, I

³ Within digital spaces, organic search refer to traffic drawn to a website through discovery on search engines or other secondary search platforms. This is compared to web traffic created through a user’s direct entering of a URL into their browser.
define mediation on Google Search as the algorithmically-facilitated inclusion and exclusion of specific results based on platform-specific considerations. Specifically, my research is an interdisciplinary, in-depth case study of the impact of Google Search on the mediation of information as a foundation for what information people find about stem cell treatments. My overarching research question is: how does Google Search (i.e. its governing algorithm and content-mediating mechanisms) mediate the different sources of information that people access online about stem cell therapies? And, crucially, why does this mediation matter?

I explore the complexities of studying Google Search as an algorithmically-mediated platform and epistemological machine (Manvovich 2013). Here and throughout my research, the term algorithmically-mediated platform refers to digital technology – in this case, a website – in which users engage with a function that is determined by an unseen algorithmic technology – in this case, search indexing and display. In this way, I am defining Google Search as an epistemological machine as it renders, through the indexing and display of content information, different types of knowledges. These knowledges are rendered through Google Search’s governing schema and mechanisms for display to the user as a search result page – or as explored later on in this chapter, Shove et. al’s conceptualisation of the moment “when practices are enacted” (2012).

The moment a user enters their health-related query into the search box and presses return, this rendering of knowledge is captured in snapshot. In this moment, every byte of information and every actor, every appeal and offer, every need and desire, is collapsed into a single space. This collapsed moment - analysable through the dissection of the single return page - is a potential litmus test for Google’s algorithmic ideology and its influence on the flow of stem cell treatment information through Google Search.

My research is fundamentally meant for an audience interested in the exploration of a specific digital ethnography of health information online. It starts from a singular focus on the material subject of stem cell treatments on Google Search and engages with broad sources of theoretical and philosophical tools of conceptualisation from STS literature when and where it is useful to gain a deeper understanding of what I
am able to observe. In doing so, my research provides an intellectual contribution through my application of novel methods for analysing large-scale opaque digital platforms. My research provides a unique application of STS and digital health sociological analytical tools as a novel and potentially wider reaching reflection on health, algorithmically-mediated platforms, and our lives online. More specifically interesting to those concerned with emerging gene and cell therapies, my research addresses a critical area of risk now facing the public by providing evidence-based reflection on the information eco-system in the high-risk area of commercially promoted stem cell treatments.

1.1 CONTEXT AND APPLICATION

Although narratively couched in promises and expectations, stem cell research has currently yielded few proven, effective, and safe treatments. More specifically, stem cell treatments derived from reprogrammed adult stem cells – ‘induced pluripotent stem cells’ (iPS) – have received a significant focus within medical research, with their promise of personalized therapies derived from autologous – a patient’s own – cells. However, as of 2022, iPS cells and their capabilities remain far from common clinical application.

Despite this, the past decade has seen a boom in direct-to-consumer marketing of autologous stem cell treatments for major diseases and conditions in digital spaces. Championed by private clinics, stem cell treatments are presented as established, if artificially restricted, medical interventions available for a price (Peterson, Munsie, and Tanner 2017). Moreover, these clinics actively engage in one of the primary revenue generating activities for the Google Search platform, Search Engine Marketing (SEM) on Google Ads. A 2017 study conducted by Alan Petersen, Megan Munsie, Claire Tanner, Casmir MacGregor and Jane Brophy found that private clinics promoting unproven stem cell treatments self-reported spending over 30,000 USD per month on Google Ads (Petersen et al., 2017).

Alongside seeking to understand general stem cell treatment information on Google Search, I have placed particular focus on information around stem cell treatment for
Parkinson’s Disease and Multiple Sclerosis\(^4\). I have chosen these specific diseases due to their impact on patient’s lives, the clinical research stage stem cell treatments are currently at in addressing them, and the established links to patient communities for these two conditions made possible through the EuroStemCell network. Additionally, Parkinson’s Disease and Multiple Sclerosis are uniquely targeted by private stem cell clinics due to their incurability and lack of effective treatment options.

It is important to note here that my project focuses on unproven stem cell treatments as interventions for major conditions or diseases, rather than those for cosmetic purposes. I have made this distinction for two reasons. First, I would like to position this project in relation to discussions within the medical research community on stem cell research aimed at treating conditions and diseases that require medical intervention for patient survival rather than elective intervention. This position allows me to ground the discussion of proof and evidence of certain stem cell treatments within an established body of medical literature. Secondly, cosmetic treatments, such as those marketed as stem cell-derived treatments, open different questions of risk, construction of patienthood and regulatory oversight that I do not want to explore in this thesis\(^5\).

Finally, my research captures a significant event in Google’s search platform policy related to digital health and stem cell treatments. In September 2019, only weeks before my data collection was set to launch, Google announced that it would ban advertisements related to all unproven or experimental gene and cell therapy treatments, including stem cell treatments on its platform by October 1\(^{st}\), 2019. This gave me a unique opportunity to capture this moment of policy-driven intervention to exclude advertisement content on the platform. This decision – enacted during my research design and data collection phase.

\(^4\) My initial data collection and research design included collection of results for queries around Diabetes (Type I and Type II). Difficulties around participant recruitment as well as errors in the scheduling of machine-assisted data collection led me to the decision to drop this category from my analysis. Although it is mentioned in a few points in my research, I made the conscious choice to not include analysis of diabetes-related results due to these challenges.

\(^5\) I have, however, explored the subject of cosmetic stem cell treatments specific to the UK in parallel collaboration with Dr. Sonja Erikanian and Dr. Sarah Chan through the paper ‘Marketing Experimental Stem Cell Therapies in the UK: Biomedical Lifestyle Products and the Promise of Regenerative Medicine in the Digital Era’ (Erikanian, Couturier, and Chan 2019).
research period – allowed me to track Google’s intervention represented by the presence or absence of ad returns and illustrate the intervention’s efficacy. It also created an urgency to my project’s development and deployment. This novel experience of real-time research adaptation is documented in Chapter 5 – Ad Results.

1.2 RESEARCH QUESTIONS

My research project is motivated by the need to understand how both information about and access to stem cell treatments – as complex, contested, emerging medical interventions – are proliferated through Google Search. I seek to address the central research question: how does Google Search mediate information that people access online about ‘proven’ and ‘unproven’ stem cell treatments?

To address this overarching question, I have identified several sub-questions that will be answered through the project. They are:

How has ‘unproven stem cell treatment’ been defined and by whom? What practices of power are performed through these definitions?

How can Google Search – as a proprietary, algorithmically-driven search engine – be studied empirically? What methodologies can be employed and why is this decision important?

How is the mediation of information on stem cell treatments shaped by structural factors within the platform of Google Search? What impact did Google’s policy change have on paid search advertisements relating to stem cell treatment queries? Does the information found differ depending on the user’s or query’s characteristics? When? And, hopefully, why?

\[6\] e.g., Are users with specific health conditions, geographies or query structures more often targeted by direct-to-consumer advertisements of unproven treatments?
Beyond an investigation into the state of unproven stem cell treatments in digital search platforms, this project addresses a critically under investigated methodological strategy for STS scholars and medical sociologists dealing with controversial subjects in health online. I sought to address these challenges by adapting a digital ethnographic approach in which the practice of interrogation is deeply rooted within technology itself. As Pink et al. write in their book, *Digital Ethnography: Principle and Practices*, a key component to conducting remote ethnographic research is to engage directly with the tools, structures, and modes of interaction native to the digital landscape (2015). I could not simply attach established modes of interrogation – like interviews or surveys – to address the questions I wanted to ask. Instead, I needed to find ways to observe and then produce findings rooted in the language of the technology I chose to study. In the case of Google Search, however, this mandate of research practice creates a unique obstacle for observation.

There are two core challenges I identified in studying empirically decentralised result generation platform like Google Search. The first is that what I see when I engage in Google Search is different from what you see, what your neighbour sees, what an individual in another country sees. The second is that Google Search is a proprietary technology in which observation is limited to *input* and *output* – a ‘Blackbox’\(^7\). This limits the ability of researchers to interrogate the internal decision-making within an algorithmically-mediated platform.

While much social science attention has been directed at digital technologies and health, methodologically this has usually been based on analyses of researcher-specific online content and/or interviews with different users (as seen in Safiya Umoja Noble’s self-auditing in *Algorithms of Oppression* [2018] and Deborah Lupton and Sarah Maslen’s focus group work on women and search [2019]). However, these approaches to the study of digital search technologies have several limitations

---

\(^7\) This term and its place within digital studies is explored in greater depth in Chapter 4.
that I explore in Chapter 4, including a lack of scale and anecdotal evidentiary support.

I recognised early on the development of my methodological approach to Google Search needed to address the platform’s inaccessibility and questions of scale of evaluation of the differences of results displayed to users. I, therefore, consider my Methodology chapter (4) to be a finding in and of itself. It answers the question; How can Google Search – as a proprietary, algorithmically-driven search engine – be studied empirically? What methodologies can be employed and why is this decision important?

To address this, I pursued the development and deployment of a novel methodological approach based on ‘data donation’. Although this will be explored in more detail in Chapter 4, data donation involves the collection of results on a digital platform by dispersed ‘data donors’ – real individuals or automated. These data donors participate in research by downloading a browser extension that facilitates the regular snap-shotting of results from an algorithmically-mediated website over a given time period. This results in the significant database of results that can be interrogated on a large scale and create a rich evidence base on which to map Google’s influence.

In collaboration with colleagues at the Algorithm Accountability Lab, Martin Reber8, Tobias Krafft9, and Roman Krafft10, we modified the data donation plugin developed by the non-profit research and advocacy organisation Algorithm Watch (https://algorithmwatch.org/) to collect results for queries around stem cell treatments. With their guidance, I was able to utilise their approach to Blackbox analysis and interrogation currently used within algorithmic accountability studies. This included the development of the anonymised browser-based plugin, creation of Virtual Personal Server (VPS) data donors, and publishing and recruitment of participants over the study period. I was able to fund the technical development and deployment of this data collection plugin through the generous support of

---

8 Masters Student – University of Kaiserslautern.
9 Doctoral Candidate – University of Kaiserslautern.
10 Masters Student – University of Kaiserslautern.
EuroStemCell, the Centre for Regenerative Medicine, and the School of Social and Political Science at the University of Edinburgh.

Deploying this data collection plugin, I undertook a data donation campaign through which I collected results for specific queries over a period of 6 months in 2020. The campaign was promoted through EuroStemCell’s network and website and social media as well as direct inclusion of patient advocates in Edinburgh. These results were provided by data donors – both real individuals and virtual donors. The results included both ad and organic returns comprising the first page of Google Search results for queries around stem cell treatments.

The result was a database of over 200,000 Google results from over 100 anonymous data donors, including self-identified patient and carers impacted by Parkinson's Disease or Multiple Sclerosis. I then analysed this dataset along specific questions related to ads (Chapter 5), organic results (Chapter 6), and the overall technical structure of a single Google Search results page (Chapter 7) across the study time period. This analysis is then contextualised and deepened through the application of STS and digital health discourse within my conclusion.

1.4 KEY CONCEPTS: SEARCH AND MEDIATION IN DIGITAL HEALTH

As mentioned earlier, my research is primarily focused on contributing to digital methodological research and health studies around the topic of stem cell treatments online. In my engagement with STS literature, I made the conscious effort to find theoretical conceptualisations that served to illuminate my subject, rather than adhering to a strict philosophical model. The directionality is from subject to insightful theoretical lenses, rather than theory to subject. While this prioritises materially-grounded digital health and methodology literature, I found value in engaging, where appropriate, with deeper philosophical conceptualisations of technology and health found in STS literature.

Before situating my work within this wider literature, I would like to briefly examine two core concepts within digital health scholarship that informed the development of my research questions. They are the importance of search within digital health and
the role of search engines as mediators. In the next sections, I will outline my approaches to these crucial aspects of digital health as essential groundwork for the work that follows.

1.4.1 IMPORTANCE OF SEARCH WITHIN DIGITAL HEALTH

To address the role of Google Search in the dissemination of information on stem cell treatments specifically, I found it critical to first examine the role of search more broadly within social science research on digital health. Specifically, the space search inhabits within existing digital health scholarship, where it is addressed, and where it is left out. In answering this, I can situate my research and its contributions within the wider multidisciplinary field of digital health.

Throughout my thesis, I adopt Deborah Lupton’s definition of digital health as a useful reference. She is not the only recent scholar to offer a helpful definition of the wide ranging and somewhat unwieldy realm of digital health. Here I am particularly noting Alan Petersen’s often stem cell treatment-specific application in his book Digital Health and Technological Promise: A Sociological Inquiry (2019). Petersen’s extensive work on stem cell tourism and promissory health interventions might have made his definition an easier fit for my work. However, there are a few key considerations that I found useful when applied to my particular focus on Google Search. Lupton outlines digital health as:

…a wide range of technologies directed at delivering healthcare, providing information to lay people and helping them share their experiences of health and illness, training and educating healthcare professionals, helping people with chronic illnesses to engage in self-care and encouraging others to engage in activities to promote their health and well-being and avoid illness. (Lupton, 2018)

This definition is particularly helpful for my purposes due to its inclusion of lay people through terms like “share”, “self-care”, “encouraging others” and “promote”. This definition alludes to the inherently social activity enabled or disabled by the “wide range of technologies”. In my application, the activity that provides these limits and
opportunities is the act of engaging in search through the Google Search platform accessed at www.google.com. This activity of search is undertaken by publics; patients, lay people, users. Interestingly, the role of the technology is to provide the means, possibilities, and limits of these activities.

Additionally, Lupton’s inclusion of the motivating factor of “avoiding illness” speaks to the unique inclusion of risk assessment as well as risk avoidance and acceptance in the motivations of those studied and studying digital health. This builds upon both Lupton and Petersen’s earlier 1996 work in The New Public Health: Health and Self in the Age of Risk which presciently warned health studies researchers of the future pitfalls wrought from underestimating the impact of digital technologies on healthcare in natively developed digital terms. This earlier work argued for a move away from binary, individualised blame for risk-adopting behaviour as it has material impacts like limiting healthcare access and discrimination within treatment. I would push this further with the privilege of hindsight to argue that the widespread proliferation of free-to-use and easily accessible search platforms like Google, which provide information in a practical sense, complicates the question of risk, responsibility, and often obscured power in the healthcare decision making of the public. This theme is explored throughout the following chapters.

Although I make a conscious effort to focus my research away from lay – or indeed patient – voices, I found this focus in other sociology of digital health useful. It allowed me to conceptualise digital health as being fundamentally (perhaps as guiding principle) about people even when focused on technologies. This may mean a focus on decision-making approaches at Google, the activities of commercial and non-commercial content creators, and invariably end users. This definition influences the way I conceptualise patients and users in Chapter 3.

In my initial approach to my subject, I identified existing literature around the act of search and health information online. While many STS and health sociology scholars addressed issues of digital health information, the analytical lenses often focused on behaviour and impacts on the user’s side, rather than the platform itself. Much of the work on sociology of digital health concerns the act of health information seeking by publics (e.g. Michael Hardey’s ‘Doctor in the House’ 1999), Deborah Lupton’s work
on self-care and monitoring, 2013), what publics do to inform their health decision making (e.g. Btihaj Ajana, Joaquim Braga and Simone Guidi’s 2021 work on the quantification of bodies through smart phone usage) and what they think of the resources they gather (e.g. Tereza Hendla and Bianca Jansky’s work on women’s empowerment through digital health technologies, 2020). These approaches may cross conditions and diseases or be developed specifically to address the concerns of those impacted by specific conditions, diseases or patient community (e.g. Steven Epstein’s 1996 work on HIV/AIDS activism). Importantly, early digital health work (particularly that informed by the medical and healthcare perspectives) involved the analysis of online content. These analyses were largely focused on static content on websites with little emphasis on the understanding of people’s actual practices and responses to this content. Moreover, interrogation of the technology is informed largely by the user behaviour it provokes (engagement, education, peer-support), rather than the platform activities in rendering the opportunities for patient engagement.

I am shifting the focus to the platform of Google Search as a digital health technology. In placing Google Search at the fore of my research, I explore Deborah Lupton’s definition of “digital health technologies” as “sociocultural products located within pre-established circuits of discourse and meaning” (Lupton 2014). Accordingly, I centre Google Search – in its capacity as a digital health technology – and treat it as an actor itself replete with motivations, ideology, and practices to be interrogated. This includes positioning Google Search as an active actor within the constellation of digital health around stem cell treatments. I am guided especially by Astrid Mager’s work and the notion of “algorithmic ideology” – or the socially-founded logic enacted and expressed through Google search as an algorithmically-driven platform (Mager 2009). Mager suggests that Google Search “performs, renews and fosters the capitalist ideology” through its very code due to its reliance on extractive commercial operations. Informed by traditional Marxist critique of labour-transformed-to-product, Mager suggests that the influence of this fundamental aspect of Google Search as an algorithmically-mediated platform is essential to integrate into any subsequent analysis (Mager 2014). In this way, Lupton and Mager’s work allows me to address Google Search as a technology influenced by its own socio-economic position and agency. Additionally, through my work, I
interrogate Mager’s positioning of Google Search vis-à-vis capitalist ideology through analysis of the material mechanisms I am able to observe through large scale analysis of returns.

To explore the importance and role of a widely used and pervasive technology like Google Search within digital health, I would like to take a moment to explore how my approach to Google Search interacts with recent novel approaches in addressing tensions within STS and digital health sociology. In my reading, I found a unique approach explored in the special issue of Flis Henwood and Benjamin Marent, two social scientists with a background in both STS and medical sociology. Navigating the parallel and often overlapping fields of digital health and STS, Henwood and Marent describe the longstanding tension in STS and sociological work on technology more generally:

… around the long-standing challenges of avoiding technological determinism while maintaining a clear focus on the materiality and agency of technologies and recognising enduring sets of relations that emerge in new digital health practices while avoiding social determinism. (2019)

Applied to the question of Google Search and health, this tension creates a challenge. Google Search as a platform is sprawling, dynamic, and opaque. While novel methodological means may grant the ability to collect empirically-based and indicative evidence, parsing clear boundaries between the agency of Google – *the platform* – and Google – *the company* – is challenging. For my purposes, it is important to note these differences as, in many ways explored in Chapter 7 and 8, they represent different models of agency. One model – *Google* the company run by humans – exerts and is exerted upon within the socioeconomic, regulatory, and policy differently than its dynamic yet software-mediated counterpart, Google Search. To equate them would be to lose the nuance between human endeavours and the technology deployed in service of those endeavours.

To address this tension, Henwood and Marent point to digital health work focused on four general themes; promissory digital health, (re)configuring knowledge, (re)configuring connectivity, and (re)configuring control.
First, promissory digital health highlights contradictory virtues within discourses that configure digital health. Second, (re)configuring knowledge outlines ambivalences of navigating new information environments and handling quantified data. Third, (re)configuring connectivity explores the relationships that evolve through digital networks. Fourth, (re)configuring control explores how new forms of power are inscribed and handled within algorithmic decision-making in health. (Henwood and Marent 2019)

I found these lenses informative when applied to Google Search throughout my research. They emphasise the importance of meeting lively technologies where and how they are in the moment of interaction with digital health actors. They also recognise that the influence of technology is not linear, but dynamic and sometimes unexpected. These themes provide a critical lens for my findings related to advertisements and organic search results in Chapter 5 and 6 as well as the structural documentation of Google influence in Chapter 7. By situating Google Search within and against these themes, I was able to keep sight of both the material impact of Google Search as a technology while acknowledging the power, relationships, and environmental contexts in which it operates.

1.4.2 GOOGLE AS MEDIATOR

The rise and dominance of Google Search as a dominant platform for informational search means that, within the Anglosphere of digital health, most, if not all, roads lead through Google. While I will discuss the practical integration of Google Search into widespread use in all facets of online life in Chapter 3, first it is helpful to situate the platform as what Michel Callon refers to as an “obligatory passage point” (1986) and characterise what that means for my observations of its rendering of results.

There are many ways of for conceptualising Google. In my initial research, several applicable terms for Google with specific STS significance arose, including “mediator”, “intermediary”, “constructor” (von Glasersfeld 1995) and “co-constructor” (Kukla 2002). Each of these terms carry implications for the roles, opportunities, engagements, and limitations applied to Google within the act of Search.
Analytically, I initially engaged with the wider history of conceptualisation beginning with positioning Google Search within Bruno Latour’s discussions on mediation within *Reassembling the Social* (2005). Rather than a neutral vehicle for reproducing relationships between actors, Google Search platform becomes “society made durable” (Latour 2005), a mediated and active platform that creates opportunities for specific interactions and information display. From this perspective, Google Search does not act as an intermediary that simply replicates edges of connection between digital actors and nodes of information. Rather, by its design, it fills the role of a mediator – promoting and withholding particular nodes of information to the user (Gillespie 2018).

Drawing on the early work of Mackenzie and Wajcman on social shaping theory, the importance of recognising the economic, social and political power that shape the agency and positionality of Google Search becomes clear. Google Search is “affected at a fundamental level by the social context in which it develops” (1999). This is further expanded upon by Astrid Mager who suggests that the platform includes its own ‘algorithmic ideology’ that informs its representation of results – and for my purposes, the mediation of knowledge – to users (2012). I will return to Mager’s approach to algorithmic ideology in Chapter 3.

Within Deborah Lupton’s definition of digital health, Google Search falls within the facilitating role as one of the “wide range of technologies”. It “delivers”, “provides”, and “helps”. This points to a salient consideration in developing my research questions; which term should be used for the role and activities of Google Search when search is engaged? And how does this definition shape my conceptualisation of Google Search through related digital health literature?

Applying Callon’s “obligatory passage point” (1986) through which information is both institutionalised (through record) and standardised is quite clear; if content is to be displayed as a Google Search result it - by technical necessity - must meet a set standardised of algorithmic machine readability and, in so doing, becomes institutionalised within the Google platform. However, what secondary intervention is applied by Google? Does it simply pass along information from one source to a
recipient (intermediary)? Does it construct a collection of results with itself understood as a knowledge creator (constructor)? Does it co-create an index of results with a user-operator (co-constructor)? Or does it curate based on internal and functional demands and considerations (mediator)?

At the start of my research, I did not have a firm grasp on which conceptualisation of Google Search was most accurate. However, through the process of engaging with the platform and finding relevance in my findings, I became increasingly convinced that Google Search most clearly enacted its influence through mediation. I would like to outline my rationale briefly here for adopting mediation as a key theme and concept throughout my research.

Since the early 2000s, STS and digital health researchers have tracked the politics of search engines. This work situates search engines like Google Search as both active producers of knowledge and innately political. Most notable is Laura Granka’s reflective work on the impact of search engines on democratic institutions and civil society between 2000-2010. Granka’s work found the often-utopian view of web search technologies adopted in the early 2000s gave way to complicated realities of platform consolidation, increasing opacity in operations, and calls for the development of algorithmic accountability (Granka 2010). Scholarship also includes critical work on to challenge the assumed neutrality of search engines and the introduction of bias through algorithmically-mediated processes (Introna and Nissenbaum 2000, Diaz, 2008).

Of these, Abbe Mowshowitz and Akira Kawaguchi’s 2002 research ‘Assessing bias in search engines’ is unique in its approach to measuring bias through interrogation of search returns and queries. While determining their definition algorithmically-mediated bias, Mowshowitz and Kawaguchi note the significance of technology companies positioning bias as a platform feature, rather than negative characteristic. This distinction of bias-as-feature-not-bug is borne out through my research. Mowshowitz and Kawaguchi write:

Bias may also be viewed as a tool in the marketing arena. Here it is not necessarily something to be expunged from an information system, but rather
it may be seen as a system feature to be manipulated so as to obtain the maximum exposure for a given product or service. Marketers want to ensure that their company's product name appears close to the top of the list of items retrieved by a search engine in response to a query about products of that type. (Mowshowitz and Kawaguchi 2002)

In *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* (2018), Tarleton Gillespie – a sociologist who has critically studied content mediation in both academic and commercial research contexts – makes a strong argument for platform as mediator. He defines mediation as the platform actively determining inclusion and exclusion of content returned and available to users based on policy, technical structure, and functionality (e.g. commercial). Drawing on his technical and developmental understanding of aggregating social platforms like Google Search, Gillespie argues that mediation is fundamental to the form and function of social platforms. I found that this definition of mediation successfully underlined the importance of active mediation performed by the platform itself as well as the inherent role vis-à-vis user. Mediation is decision-making made algorithmically viable with display decisions made on behalf of both user and the internal interests of the platform.

To return to the definition of intermediary, Niels Kerssen’s describes search engines as intermediaries:

> Different from traditional reference work, online search intermediaries transformed user needs into a search and, finally, relevant information, through a systematic and planned approach that was documented in written form in a so-called search strategy. (Kerssens 2017)

I argue that this definition does not go far enough to acknowledge the active intervention and agency expressed through the mechanisms within search platforms. As will be seen in Chapters 5, 6, and 7, Google Search’s practice of returning results cannot be separated from the mechanisms of influence that allow the platform to function. This functionality and relation to structural behaviour is not addressed by the terms intermediary (transient), constructor (non-responsive to user or content
provider input), or co-constructor (limited role of secondary contributor like user). I explore these differences more deeply in Chapter 2 in my exploration of stem cells in health.

At the same time, Google Search’s ubiquity as a point of entry for users to discover information, positions it as an “obligatory passage point” (Callon 1986) that fundamentally renders relationships between actors within the digital hierarchy of search and information. To better situate this conceptualisation of Google Search as mediator, I first sought to position my research in relation to established STS and digital health literature.

1.5 THEORETICAL GROUNDING

This project sits at the intersection of four areas of literature: Science and Technology Studies (STS), Digital Methods, algorithmic accountability studies, and sociological discussions around digital health. Due to the unique opportunity presented by the case of stem cell treatments in Google Search, I made the conscious decision to privilege digital methods and health studies as the primary contribution I sought to make. As a result, my reading of STS literature is broad and wide-reaching; I use them to offer theoretical contextualisation of my material and methods-focused findings. Future work might reverse this emphasis and grant more space to specific STS analytical frameworks. However, for my purposes and audience in this research project as a contribution to digital methods, I found it necessary to apply a hierarchy of literature and perspective.

Current sociological research on stem cell treatments tend to focus on patient perspectives (Scott 2007, Peterson, Seear and Munsie 2014, Langstrup 2011) and the issues surrounding access, including stem cell tourism (Peterson, Munsie, Tanner, MacGregor and Brophy 2017) and regulation (Harmon 2008, Mahalatchimy 2017). Currently there is no research specifically focused on the impact of digital search platforms on unproven stem cell treatments. My project has evolved practically from the discussions of the direct-to-consumer marketing flagged as problematic by the stem cell research community. However, it is the insight enabled by Science and Technology Studies analytical frameworks and the use of novel
digital methods that will allow this project to contribute to the sociology of digital health.

My decision to focus on Google Search as the main digital platform is informed by several areas of concern within STS and social sciences more broadly. The first and perhaps most important practical reason is the sheer ubiquity of Google Search. Simply put, it is used across purposes, settings, and user groups, across healthcare domains and topics beyond that of any other platform, as will be explored in Chapter 3. Simply put, everyone uses it.

In exploring Google Search’s ubiquity, I confront what it means to be a user group engaged both with the technology of search and as a subject of biomedical research and discovery in public consciousness. To mark boundaries of these individuals as reckonable groups not only defined by their interaction with the search platform, I borrow Mike Michael’s approach to public understanding of science. I am informed by his consideration of *Publics-in-General* and *Publics-in-Particular*. He defines publics in general as “undifferentiated whole that is distinguished from science that is itself characterized globally in terms of some key dimension” and Publics-In-Particular as “those publics that have an identifiable stake in particular scientific or technological issues or controversies.” Although I am not uniquely focused on the user experience of Google Search, I am interested in the way user groups are rendered through their interactions with the platform. This is determined largely by users’ “identifiable stake” – in my case, their relationship to a condition or disease. In my discussions of the conceptualisation of users (Chapter 3) and patient groups (Chapters 5, 6, and 7), I have a strong focus on Publics-in-Particular and their interactions with the Google Search platform.

Despite its central importance, there is limited research on Google Search within digital health (Mager 2014, 2015, 2018), and what there is focuses on patient (or other user) search behaviour or the analysis of content found online (explored in Chapter 2). This may be due in part to the methodological challenges inherent in tackling Google as a subject. This project attempts to address these issues with a multi-disciplinary approach informed first and foremost by digital methods and STS. It should be noted that this may raise conflicting issues. As Bryman (2004) notes,
different disciplines inherently stem from distinctive epistemological groundings. These challenges and my proposed approach to address them will be discussed later in my Methodology Chapter (Chapter 4).

Wrestling with Google’s relationship to scientific claims to authority is critical to mapping how the development of Google’s practice – its act of mediating information through rendering results – when a user searches. This is because – as we will see in Chapter 3 – appeals to the perceived neutrality and authority of scientific knowledge and approval processes are used by Google Search itself in its approach to policy implementation. To address this in Chapter 3, I am guided by the work of Harry Collins and Robert Evans’ (2002) approach to expertise concerning dynamic and user-gathered knowledge enabled by technology.

Moreover, the STS focus on deconstruction and critique (Latour 2004) of scientific structures leaves us navigating difficult terrain when analysing potentially harmful ‘bleeding edge’ treatments from varying perspectives. In focusing on unproven stem cell treatments, this project discusses the delineation of proof for treatments available to the public. This touches on questions of claims to authority within medicine and questions of public access to treatments outside of traditional medical structures. Although these is a wealth of literature on ‘alternative medicines’ (Ross 2012, Attwell, Ward, Meyer, Rokkas and Leask 2018, Brosnan, Vuolanto, and Brodin 2018), this project situates unproven stem cell treatments as its own peculiar subject – enabled by the configuration of for-profit clinics operating within specific regulatory and economic spaces.

For my purposes, this project explores internal discussions within the stem cell research community around mesenchymal stromal cells as a ‘jumping off point’ for the delineation between “proven” and “unproven”. As will be discussed later, I am using the specific example of mesenchymal stromal cell-derived treatments to delineate unproven stem cell treatments. These treatments are derived from specific cell lines including adipose (or fat) cells which are controversial within the medical research community (Sipp and Turner 2015) and yet are the most oft-cited cell lines used in direct-to-consumer marketing of stem cell treatments (Turner 2016).
The decision to focus on unproven stem cell treatments situates this research within broader questions of risk (Peterson, Munsie and Tanner 2017) as well as STS conversations on ‘promissory expectation’ within medical research and treatment (Abraham and Davis 2013). Here risk – or the potential real harm of undergoing unproven treatments themselves – is brought into conversation with narratives of promissory expectation of cures prevalent within discussions of stem cell research. Finally, there is a strong focus on economic aspects that contributes to the choice of unproven stem cell treatments as a subject. The findings from Peterson, Munsie and Tanner (2017) that for-profit medical clinics promoting unproven stem cell treatments funnel significant capital into search engine marketing each month underlines the importance of studying the connection between unproven stem cell treatments and Google Search.

Finally, it is helpful to enter these findings with an understanding of my deployment of types of knowledge as adopted within this analysis of search engine marketing. Unlike Carl Mitcham’s (1994) technology theory of knowledge that relies on larger theoretical characterisations of rules, maxims, etc., here I am deploying types of knowledge as a form of categorisation. This is an important decision; here I am using STS approaches not as a jumping off-point for deeper theoretical analysis, but as an informant for the types of material considerations I grapple with in my pursuit of a novel digital method. In this socio-economic analysis, each search result represents a type of knowledge to which I ask the question: cui bono? Who benefits from its display? What is being asked of the user to whom it is displayed? Who has developed it and for what purpose? In this way with each search return I am looking at types of knowledge distinguished by their motivations and resulting practices. To reach the next practical steps of analysis, I must first explore the central concepts of performativity and practice.

1.6 PERFORMATIVITY AND PRACTICE IN DIGITAL HEALTH

Lupton’s definition of digital health offers opportunities for exploring performativity as applied to search platforms in terms of the actual practice and implications of technologies. This includes the performative nature of search as it both reiterates and contains its query subject, the collaborative impact of search, and the
characteristics of power exhibited during the act of search. Having established that Google functions as a mediator, how – given the challenges of accessing its inner workings as a proprietary platform – could I seek to observe its performativity in a meaningful way? At first overwhelmed, I was aided by Annemarie Mol’s advice to adopt praxiography – or the practice of pursuing practice-driven research (2002). Her 2010 work encouraging the embracing of the ‘multiplicity of reality’ was a valuable suggestion that freed me from the necessity of direct observation of the platform’s inner workings as a one-to-one relationship; an irreplicable interaction only between one unique user and the platform. Instead, I was encouraged to interrogate those practices that could be seen.

To this end, I found it helpful to draw from Andreas Reckwitz’s ‘material turn’ of social theory within STS in which technology functions as active agent and participant within the constellation of interaction and decision-making (2002). To apply this agency to Google Search – or rather enable my analysis of the platform as a moving, interacting agent – I conceptualised the way search platform might perform and how that performance might be both similar and different from the ways of human-centred agents. This required interrogation of the terms performativity and practice as the means for observing these interactions.

I was drawn to Elizabeth Shove, Mika Pantzar, Matt Watson’s The Dynamics of Social Practice: Everyday Life and how it Changes (2012) as a particularly helpful resource at the intersection of sociology and STS. It notes that practice – characterised through the key elements of meanings, materials, and competences – is understood as performance. Individuals – and in my case, Google Search – is a “carrier” of these practices through their enaction of practice. Here again the distinction between Google – the company – and Google Search – the platform – becomes relevant in addressing Google as a unique technology. Although Google’s human actors have the capacity to act as agents of practice, my research suggests that Google Search – through its algorithmically-mediated and policy-guided meanings, materials, and competencies – also engages in performance.

Within digital health several researchers have integrated this approach into their analyses of questions of public health. Helpfully, these analyses touch on different
constellations of actors and influences. For example, whereas Tim Harries and Ruth Rettie’s *Walking as a Social Practice: dispersed walking and the organisation of everyday practices* analysed the uptake of individualised health improving practice within the context of public health strategies (2018), others included non-human centred influences. These include marketing and advertisement on the expansion of vaping as a practice amongst teenager groups (Keane et al. 2017), and the commercial promotion of personal blood monitoring technology as part of individual health practice (Williams et al., 2018). The former examples provided particularly useful insight into integrating commercial concerns (or what I will adapt as *platform concerns*) into digital health analysis of practice. However, these three prominent examples still adopted a particular focus on individuals, rather than the technologies through which individuals accessed or expressed their healthcare decision-making practices.

Fundamentally, this research is a case study in relational networks (or ‘assemblages’) and the importance of performance as an end point that can be analysed (Latour 2005, Mol 2002, Knorr-Cetina 1997). To that end, I needed to establish where and in what moments I could observe practice taking place. To address this, I was aided by Shove et al.’s dynamic approach to analysing social practice that states: “Practices consist of elements that are integrated *when practices are enacted*” [emphasis added] (Shove et al. 2012).

From this grounding, I adopted a critical conceptual decision that appears throughout my research. Google’s Search’s *performance* is the moment of decision making between the action (*user queries*) and reaction (*search platform returns*) from which I can create a snapshot of the assemblages coalescing around stem cell treatments in Google Search.

I would like to expand briefly upon this. Here again Reckwitz’s de-centralisation of human actors (2002) and Mol’s encouragement to focus on everyday practices (2010) is particularly useful in navigating the unique performance undertaken by Google as a search platform. Just as Callon’s “obligatory passage point” was helpful in determining the character of Google as mediator, Reckwitz’s decentralisation
opens up an opportunity to find moments of agency undertaken by the technology of Google Search.

Indeed, while the human end-user/searcher within this study acts as instigator of the interaction, the complexity lies in the way in which that action then prompts a cascade of decisions of the character of the searcher. It then returns results in a moment of enaction on not only the behalf of the searcher, but of other actors exerting influence within the platform itself (e.g. advertisers and content creators). While the user “submits” the search query seeking information, it is the content creators and ultimately the platform itself – through its algorithmic logic – that makes the decision on what material to return in the form of advertisements and organic results. It is the details of the characterisation of the Human-qua-End User by both the ad creator and the search platform that gives my analysis meaning.

I am guided by Jeanette Pols' work Care at a Distance (2012). Informed by a material-semiotic approach, Pols takes a particular interest in applying domestication theory (pg. 18, Pols 2012) to human-technology interactions, which helps to demonstrate collaborative assemblage building. In particular, she focuses on the moments of unleashing and taming.

The unleashing of technology on humans and humans on technology offers opportunities for both parties to push the boundaries of what response actions can be elicited. For example, a user might learn to turn to Google search before a doctor’s appointment, establishing new habits of healthcare decision making. Alternatively, a user might find alternate use for a calendar function within their email to track symptoms across time, a creative use outside the technologies’ initial purpose.

On the other hand, the taming of humans by technology and vice versa presents moments of restriction and behaviour modification. The ur-example of this modification of humans by technology being Foucault’s panopticon in Discipline and Punishment (1977). Within digital technology, this is often seen in commercial interventions. For example, a user might be trained to sit through advertisements long enough to reach their desired content on a video sharing site like YouTube.com.
Humans can also shape their technology through selective use. A software function that is unused by users will quickly be abandoned in future iterations as users restrict their engagement to programs and functionalities they know and understand. Some example of this *taming* within digital health includes Jeannette Pols and Dick Willems exploration of telecare technology adoptions (2011) and Alexander Peine, Vivette van Cooten, and Louis Neven exploration of e-bike technology for elderly patients (2016).

Relevant to Google Search’s ubiquity and integration into digital health practices, domestication of the user fundamentally includes a gradual move towards everyday use, practice, and identity (Silverstone, Hirsch, and Morley 1992). The widespread adoption of Google Search in online behaviour will be explored in Chapter 3. However, a telling example of this domestication of both Google Search and users within digital health is writ large in the sobriquet “Dr. Google” (e.g. “Dr. Google Will See You Now”, Wall Street Journal 2017) and the widespread adoption of Google Search within user’s personal health decision making (as explored in Chapter 2).

In this study, I have focused on four main actor groups; the searcher (real or virtual), the ad content creators, the search platform governed by its own algorithmic rationality and processes, and the algorithmic overseers (Google, the company). *Unleashing* and *taming* in this context indicates the presence of competing motivations and practices. There is the searcher who desires information on their condition or disease. The private clinic that wants to recruit new patients or the charity looking for donations. There is the search platform whose goal is to transform a search query into a click-through on the first results page. Finally, the Google employee who wishes to balance revenue generation with compliance with regional regulations. Crucially, I argue that Pols’ work also issues a stark and prescient warning – although health-oriented technologies may aim to support or enable different forms of material care, it is not inevitable that they will.

From these crucial theoretical considerations of Google Search’s importance, role as mediator, and practices, I can now situate my research within a larger digital health context.
1.7 IMPACT

This project is first and foremost aimed at informing and challenging through the adaptation of a unique digital method through an adapted digital ethnographic study of stem cell treatment information online. It aims to push forward the integration of novel and impactful digital methods into digital sociology and health studies to bridge gaps between social science researchers and both patient and medical research communities. In this project, I contribute to larger discussions around digital search epistemology within health-centred digital sociology literature – how we situate, observe, and analyse Google Search as mediator of knowledge in digital spaces. In doing so, I provide new insights based on underexplored and natively digital research methods. While this project aims to fill a critical gap using the subject of unproven stem cell treatments, it also contributes to the broader question of direct-to-consumer marketing of healthcare and the role of digital platforms in health knowledge mediation and dissemination. The project’s focus on these treatments serves as a ‘canary in the coal mine’ that will allow my research to contribute to ongoing social scientific discussions of healthcare in digital spaces.

Throughout this thesis, I present a material application of wider STS and digital health literature in several key areas. Firstly, I articulate empirically the network interplay between economics, particular platforms, and ideas around risk and health. I find that these assemblages are so intertwined to create questions around agency that have not yet been addressed within digital health literature. Additionally, I contribute to larger conversations around the challenges that result from the outsized role of platforms in health information and public health. Finally, my thesis raises questions of complexity and consequences of platform mediation on major actors within digital health, including patients and Charities/Non-Governmental Organisations.

Studying Google Search empirically is difficult. This project aims to tackle this difficulty by providing an original methodological approach based on digital methods first deployed outside the area of STS and health studies. It also provides insight into the challenges and pitfalls of technically complex methodological approaches to data gathering and analysis. This project is not fundamentally a formal statistical or
‘internet measurement’ investigation, but rather an organised attempt to explore digital ethnography with rich data on dynamic technology (Murthy, 2008). I strive for rich data to support my findings as a catalyst of ignition for future conversations around emerging therapies and the digital spaces in which they are communicated.

The health information we, as patients and citizens, gather at 3 a.m. matters. My thesis is an exploration of what hidden currents of how different interests, forms of power, and types of knowledge interact and manifest in seemingly neutral search result pages. I attempt to bring Google Search into the conversation about stem cell treatments and digital health due to its implications for those seeking treatment, those sharing civically-minded knowledge, and for all of us who have integrated digital search platforms into our healthcare decision-making pathways.

1.8 CHAPTER OUTLINES

This thesis begins with two contextual chapters on stem cell treatments and Google Search respectively. Instead of a separate literature review, I have organically integrated my examination of relevant literature in Chapters 2 and 3 as I present the context of stem cells and Google Search in digital health. These are followed by my methodology chapter, three results chapters, and a conclusion. Each of the chapters pursues my central research question by providing context, evidence, analysis and consideration for future research.

Chapter 2 addresses my secondary research question: How has ‘unproven stem cell treatment’ been defined and by whom? What practices of power are performed through these definitions? It introduces stem cell treatments as emerging and contested treatments through analysis of relevant literature. It sets the stage for stem cells as a search subject while firmly grounded within medical discourse of authority and risk. It provides definitional support for stem cell treatments as an observable subject within the subsequent evidence collection and analysis. It contextualises stem cells within digital health literature and previous work undertaken on the promotion of stem cell treatments in digital and non-digital spaces. It explores what “unproven” means in the context of actors engaged in producing content that is included in Google Search results.
Chapter 3 provides an overview of Google Search in terms of its rise to ubiquity and opacity both generally and within digital health through a review of relevant literature. It contributes a historical and technical grounding that gives context to the Google Search platform as an observable subject addressed by my initial research question. In this chapter, I also situate the end user within existing literature on users and patienthood.

Chapter 4 explores the methodological groundwork and novel work I undertook to collect a large evidence base. As a finding chapter in and of itself, it addresses directly my secondary research question: How can Google Search – as a proprietary, algorithmically-driven search engine – be studied empirically? What methodologies can be employed and why is this decision important? My methodology chapter addresses the question of how Google Search can be studied given its expansive and dynamic nature. In this chapter, I provide an overview of algorithmic governance and digital methodology literature. I also describe in-depth the undertaking of developing stand-alone data collection software and study design specifically deployed for my research.

Chapter 5 addresses directly my secondary research question: What impact did Google’s policy change have on paid search advertisements relating to stem cell treatment queries? It explores the advertising returns captured in my data donation campaign. In this chapter, I explore advertisement returns captured in my data collection and glean insight into who, how, and when direct-to-consumer advertisements around stem cell treatments are deployed.

Chapter 6 expands this analysis of Google Search results through an interrogation of organic – that is, non-advertising – results returned through my data donation campaign. This chapter explores the who, how, and when of organic search results returned when users query around stem cell treatments. This chapter provides evidentiary support for the types of information content contributing to user knowledge on stem cell treatments. This contributes to my overarching questioning of Google’s influence.
Chapter 7 takes a structural approach to answering my secondary research question: *How is the mediation of information on stem cell treatments shaped by structural factors within the platform of Google Search? Does the information found differ depending on the user’s or query’s characteristics? When? And, hopefully, why?* It analyses the full results page and additional mark-up that was returned and informed by the data donation campaign. This chapter takes into account the full results page as a moment of observable practice by the Google Search platform. It explores the mechanisms – including additional mark-up and indexing – that shape information presented.

Finally, chapter 8 is comprised of my conclusion in which I return to the findings of my previous chapters and discuss both the implications of those findings on existing digital health research and future implications for STS and digital health research. It allows me to return to my primary research question: *how does Google Search mediate information that people access online about ‘proven’ and ‘unproven’ stem cell treatments?*
For patients seeking solutions to complex medical conditions and diseases, stem cell treatments have become the modern Balm of Gilead – the oft-touted universal cure from within, rather than without, available for the right price. The reality of research on stem cells and possible derived treatments is much more complex as the area of research occupies both contested and cutting-edge spaces within regenerative medicine.

In this chapter, I introduce the history and state of stem cell treatments as novel and emerging medical interventions. Stem cell treatment is an umbrella term comprising any treatments derived from the use and manipulation of stem cells both within a patient’s body and without. While some stem cell treatments have been in clinical use for decades (as described in this chapter), there are other stem cell treatments in the context of my study that are newer, riskier, and – perhaps – unsupported by evidence. I use the term “unproven” here to indicate these are emerging and contested stem cell treatments, rather than established clinical treatments. In examining these definitions, I speak to my secondary research question: How has ‘unproven stem cell treatment’ been defined and by whom? What practices of power are performed through these definitions?

To address this, I contextualise stem cell treatments as an area of medical research to reach a common definition for unproven stem cell treatment. I come to a shared definition to answer the question “what are stem cell treatments?” and explore why stem cells have captivated the scientific and public eye. I then expand on the state of debate within the stem cell research community as to the definition of “proven” and “unproven” treatments with a particular focus on mesenchymal stromal cell-derived treatments as a bellwether of contestation. A discussion of this authority across research and public spheres will be followed by an analysis of the rise of private clinics and the marketing of treatments directly to patients or consumers. Finally, this
chapter begins to explore the compression of “stem cell treatments” as seen in Google Search results as a dynamic and shifting object.

In this chapter, I will focus on the science and technical details of stem cell treatment, as this provides important context for my study. One might question why attending to the science of stem cell treatments is necessary for a social science dissertation focused on the impact of the subject (‘stem cell treatments’) in a non-biological context (‘Google Search’). My inclusion of the science behind the field of stem cell treatments is a conscious choice informed by my own experiences within the field of public engagement and education in stem cell research. As someone with a non-scientific background, my foray into public engagement around stem cells in both training researchers and developing educational resources for different publics was initially hampered by my own limited understanding of stem cell properties and clinical possibilities. However, the more I immersed myself in the fundamental biology and debates within the basic\textsuperscript{11} stem cell research community, the easier it became for me to identify and respond to misinformation around treatments. As we will see later in my discussion of unproven treatment claims, absence of specificity in the biological reality of stem cells and their derived treatments tills fertile ground for misinformation\textsuperscript{12}, over-claims, and evidentiary manipulation.

By exploring in some detail the underlying science of stem cells, I ensure that my analysis is based on the current best understanding of what stem cells are, what they can do, and the debates within the research community on their possible applications and limitations. This enables me to assign plausibility; a fundamental property of my definition of proven and unproven treatments that will be discussed in all steps of this dissertation. In this chapter, I recreate the crossing I made in my own relationship to understanding stem cells and their derived treatments. From outsider to insider, this chapter serves as an intellectual anchor for my research. The discussions and biomedical context presented within this chapter forms the basis

\textsuperscript{11} Fundamental is the separation of “basic” or foundational medical research and “translational” medical research which seeks to bring foundational theories into practice and clinical settings.

\textsuperscript{12} I define “misinformation” in the area of stem cell treatments as incorrect or misleading information around the safety and efficacy of proposed treatments as well as biological understandings of stem cell biology.
from which I built my data collection and subsequent analysis around the unique mediation of stem cell treatments online.

WHY STEM CELLS AND STEM CELL TREATMENTS?

It should be noted here that this thesis is limited to stem cell treatments exclusively. My reasoning for this is threefold; the ability to limit consideration of treatments within and without a specific area of emerging treatments in regenerative medicine, my own access to competencies around stem cell treatments and the widespread adoption of the term “stem cells” and “stem cell treatment” itself.

There are expanding applications under the larger remit of gene and cell therapies (many of which rely on stem cells as part of their clinical application, but are not strictly “stem cell treatments”). These clinical applications are showing surges in approved treatment protocols. For example, in 2021, over 350 gene and cell therapy products were approved by the European Medicines Agency - the body of the European Union charged with evaluation and supervision of medicine products within the EU (EuroGCT, 2022).

While these gene and cell therapies may offer expanded treatment offers for myriad conditions, they include a wider range of intervention – often only tertiarily reliant on the unique biological activities enabled by stem cells. Limiting my scope to stem cell treatments only is to limit my observations to an established and specific clinical research area. Regulatory oversight of named stem cell treatments – both autologous and allogenic – are addressed by agencies like the MHRA (Medicines and Healthcare products Regulatory Agency) [UK], FDA (Federal Drug Administration) [USA] or international research community as represented by the ISSCR (International Society for Stem Cell Research) and European Medicines Agency.

Secondly, my own relative proximity to stem cell researchers through my work at EuroStemCell and the Centre for Regenerative Medicine at the University of Edinburgh gives me support. Throughout my research, I was able to check in with
researchers through these connections to better understand the scientific evidence provided as support for stem cell treatments presented.

Finally, the popular terms “stem cell treatment” and “stem cells” have been adopted widely outside of the research community in digital health spaces of patients seeking medical intervention. In these spaces, “stem cell treatment” is often presented as a singular intervention – sometimes “stem cell treatment” is misunderstood as an umbrella term or condensed into one “treatment” applied to many conditions and diseases (ISSCR Patient Handbook, 2022). This provides a difficult yet interesting intersection of information mismatch; what stem cell treatments are may not be what they are understood to be in digital public health conversations.

WHY MULTIPLE SCLEROSIS AND PARKINSON’S DISEASE?

Although many marketed stem cell treatments fall under cosmetic regulatory guidelines (Turner 2016), I have limited the scope of unproven stem cell treatments to those aiming to treat major conditions and diseases. I came to this decision because it allowed me to connect my research to other literature on risk and stem cell research as well as high risk topics on Google Search. This focus on major conditions and diseases builds upon the early work of Lee Turner and Amy Zarzeczny’s work on media reporting on private pursuit of stem cell treatments for major conditions and diseases (Zarzeczny et al. 2010) and Alan Petersen, Kate Seear, and Megan Munsie’s work on stem cell tourism and hope (2014).

Additionally, I am informed by my own work alongside Amy Zarzeczny, Jan Barfoot, Claire Tanner, Clare Blackburn, and Megan Munsie on the stark language used by patients requesting additional information on stem cell treatments for major conditions and diseases (Zarzeczny et al. 2018). The collection of enquiries addressed in this research underlined the wide range of conditions patient publics sought for stem cell treatment and the risks these enquirers were willing to take on to address their condition. In this study of over one thousand direct enquiries to the organisations EuroStemCell and Stem Cells Australia, a majority of enquirers were either patients or close relations to patients seeking information on specific conditions and diseases. Over 70% of enquirers were looking for access to stem cell
treatments with only 26% including reference to the availability of clinical trials (Zarzeczny et al. 2018). Moreover, it was fundamentally clear across the responses that stakes were raised for enquirers when focused on conditions and diseases that require medical rather than elective intervention.

Both Multiple Sclerosis and Parkinson’s Disease have established and active patient communities in English-speaking countries. This is reflected in the significant amount of information available online around these conditions (see Chapter 3) and creates a large patient group with which to engage through my research. Later in this chapter, I will discuss the particular disease trajectories and how they relate to proposed stem cell-derived treatments.

2.1. PROVEN SAFE AND EFFECTIVE AND PLAUSIBLE

Before exploring the state of the science behind stem cells, I would like to begin with a central definition that we can build on. Namely to agree on an answer to the question; what is a proven treatment? This is the point from which all others flow. Drawing on guidelines from wording used by the ISSCR (ISSCR Patient Handbook, 2022) and the Federal Drug Administration (FDA 2022), I suggest the following definition:

A proven treatment is a medical intervention that has been proven safe, effective, and scientifically plausible at treating a condition or disease to an established regulatory standard.

This central definition requires us to agree on three conclusions: that a medical intervention is better than nothing if non-intervention or other available interventions do more harm. We must also agree that, theoretically, the treatment should be linked to understood mechanisms involved in the expression of a disease or condition. Finally, we must agree that the status of a treatment can be at least partially

---

13 On a personal note, my father is himself diagnosed with Parkinson’s Disease and my initial interest in the area came from anecdotal explorations of information around his diagnosis online.

14 Of any medical kind.
determined by its status vis-à-vis established regulatory frameworks in which these fundamental characteristics of safety, effectiveness, and plausibility are evaluated.

This definition derives from established norms within biomedical research. For a new treatment to be deemed “proven”, it must pass a number of functional thresholds for use; safety including lack of significant harmful side-effects, effective (as or more effective than currently available treatments) and scientifically plausible in the treatment of a stated condition or disease (ISSCR 2018). These thresholds of proof shift between institutions and national healthcare systems (as well as public and private actors like private clinics), but form a theoretical test for any new treatments moving towards integration within patient protocols (Halpern 2004).

Grounding all analysis of treatments as proven, or more importantly for my purpose its converse, unproven, within this central definition is essential for my analysis in this thesis. These three characteristics - safe, effective, and plausible - become a quick reference for the analysis that follows. Questions like “how safe compared to associated risks within disease progression?” “Effectivity measured by whom and with what evidence?” and “Plausibility, on what basis of current biomedical understanding?” are all avenues to be explored from this fundamental agreement.

I use proven treatments to serve as a litmus test throughout my work in mapping the concept of unproven stem cell treatments, because it creates a basis for the data-driven analysis in subsequent chapters. In contrast to their antonym, unproven treatments become more easily recognisable and definable throughout the subsequent discussions. It should be noted that there is an additional antonym that should be recognised: disproven. I am not addressing disproven treatments as these would be limited to formerly accepted treatments or those that have had active research to disprove their efficacy and safety (an extreme example of which might be lobotomies for mental health disorders or alternative cancer treatments [Vickers 2014]).

Further complicating this is the opacity of treatment particulars in the public-facing information around offered stem cell treatments. This content is specifically targeted to the Public-In-Particular (Michael 2009) grouping of patients seeking information
around their condition or disease. Since promotional material for privately marketed stem cell treatments does not include details of the medical protocols and practices deployed in these medical interventions, I must demarcate the line myself.

I would argue that, within the sphere of stem cell treatments in Google Search, many of the themes covered by a broad STS literature are collapsed into single spaces where narrative, institutions, and market actors exert their influence on patient publics. These include discussions on promissory potential of regenerative medicine (Petersen, Tanner, MacGregor, and Munsie 2018), conversations about commercialisation (Mahalatchimy and Faulkner 2017), accessibility (specific to ageing – Joyce and Loe 2010) and ethics (Chan 2017).

I would like to pause here to address a particularly strong influence on my approach, Professor Andrew Webster from the University of York. I was fortunate enough to work closely with Professor Webster before his passing in 2021 and his perspective on my research approach became a cornerstone of how I approached my conceptualisations of stem cells in the digital moment I studied them. In particular, I found his discussion of Responsible Research and Innovation (RRI) within regenerative medicine a particularly helpful foil to the approach to treatment development guided by ‘innovation at all costs’. While Professor Webster notes the “non-linear, complex and messy process” of treatment development, he also raises the bar for the quality of what we should expect of information on emerging stem cell treatments when they enter the public sphere (Webster 2017). In his work with John Gardner, he expanded on the social aspect of regenerative medicine as part of larger societal discussions of healthcare treatment options (Gardner and Webster 2016). However, the “messiness” that he harkens back to throughout his work reminded me to consider these differences in definitional application as used by different actors and stakeholders within the field.

In their 2016 paper ‘The Social Management of Biomedical Novelty: Facilitating Translation in Regenerative Medicine’, Webster and Gardner draw particular focus to the idea of social scaffolding and socio-technical assemblages that form regenerative medicine through which the process of clinical translation “constitutes a form of collective organising and social change, propelled by the promise” of a
healthier and wealthier future (Gardner and Webster 2016). Gardner and Webster suggest that the novelty (alternately messiness) inherent in these treatments are twofold. They provide novel approaches to how the rendering of biological substances are understood, engineered, and manipulated (Meztler and Webster 2011). They are also novel in that they present unique regulatory and societal challenges to society at large (van Est and Stemerding 2012). This novelty or ‘messiness’ in turn shapes the conversations around stem cell treatments. This conceptualisation guided me throughout my research. Namely, it forced me to keep in mind the question of what happens to the characteristics of stem cell treatments when these “non-linear, complex, and messy” processes are compressed into a single point through the mediation of Google Search.

Compression through the mediation of stem cell treatment information within digital search spaces - as I will discuss in future chapters and throughout this research – raises questions about implied authority and risks that are often implicit or hard to detect. As this chapter demonstrates, the borders of proven and unproven stem cell treatments change depending on the speaker and audience. Much of the detail of the often-antipodal properties ascribed to stem cell treatments is lost in the compression. To regain them and make some sense of the mess Webster described requires me to begin from a shared point of understanding, which I choose to be primarily informed by the biomedical research community’s evidence-based definitions.

2.1 STEM CELLS – THEIR HISTORY AND APPLICATIONS

In this section, I will explore the foundations of stem cell-derived treatments within biomedical research including their history, regulation, and applications for the possible future treatment of condition and diseases like Parkinson’s Disease and Multiple Sclerosis. This section discusses questions of delineation between proven and unproven treatments within the biomedical research and regulatory community.
While the foundational idea of cell therapy may be long standing, regenerative medicine as a distinct field of research was only established in the early 1990s as a reaction to rapid technological change within the medical community (Kaiser 1992). At its core, regenerative medicine is a field of medical research that focuses on the “process of replacing, engineering or regenerating human cells, tissues or organs to restore or establish normal function”. Moreover, as it focuses on “normal function” through the study of abnormal function in the case of major conditions and diseases, the field is considered “translational” or concerned with the movement of research from the lab to clinical applications (Mason and Dunnill, 2008). The field explores a wide swath of applications including tissue engineering, CRISPR/Cas-9, genomics, and the possibility of developing tissue within a lab for transplantation. Crucial to the field is the integration of the body’s own source of cellular construction, stem cells.

For biomedical researchers, stem cells are complex cell types whose essential characteristics and properties must be understood fully in order to analyse the impact of their introduction within public conversations about decision making and treatment. In the simplest terms, stem cells are the cellular building blocks of the human body. From our earliest stages as a smattering of cells within a blastocyst to embryonic and foetal cells to our lived adult bodies, stem cells serve two purposes: they can reproduce themselves and then can specialise. These intrinsic qualities allow for the formation of complex biological systems and organs during human development. It also allows for a turnover of fresh cells within our adult systems like those that support production of skin and blood cells.

There are three main subsets of stem cell origins used within regenerative medicine research; *embryonic stem cells*, *adult stem cells* and *induced pluripotent stem cells*. *Embryonic stem cells* are cells collected from the inner cell mass of an early blastocyst; a pre-implantation embryo. These cells are ‘pluripotent’ or able to be turned into any cell found within the human body (except for earlier cells like those contained in the amniotic sac and the placenta). *Adult stem cells* are found in young
or adult tissue and continue to play a role as they enable the replenishing of cell
types (e.g. blood stem cells within bone marrow and skin stem cells) as well as aid
the turnover of our cellular structures. Adult stem cells are ‘multipotent’ or capable of
specialising into particular cell lines with restrictions. Finally – and most importantly
for our purposes due to their clinical possibilities - induced pluripotent stem cells are
adult stem cells that have been returned to a more naïve pluripotent state by
reprogramming with gene encoding transcription factors\(^\text{15}\) (Nordin, et al., 2011).
These induced pluripotent stem cells (iPS cells) have the capacity to specialise
outside of their original cell lines and open up the potential for autologous – or
obtained from the same source – cell transplants. It must be stressed that the
pathways to cell specialisation are long and complex. Cells differentiate through
many intermediate and “precursor” steps before becoming myriad specialised cells.
Each cell line includes many factors and steps that influence the final phenotypical
expression of a cell within a functional niche. These pathways from early embryonic
stem cells to the specialised cells of our bodies are still not completely mapped.

Early stem cell research was dictated by a simple concept; if cells were faulty or
depleted due to trauma or disease, transplantation of healthy cells from an internal or
external donor site would allow organ systems to recover. This concept is
deceptively simple and - as I will demonstrate in Chapter 5 - is used as an
onboarding conceptual tool within much of the lay language within digital spaces in
which stem cell treatments are discussed. In particular, it is used as a hook for
plausibility within the promotional material produced by private clinics offering stem
cell treatments.

Stem cells, however, are not simply safe agents of positive, linear growth\(^\text{16}\). Cancer
researchers have long theorised that many cancers are in fact the result of faulty
stem cells: reproducing without control and specialising without purpose (Batlle et al.,

\(^{15}\) Transcription factors are proteins that enable or disable RNA’s “reading” of DNA. These factors act
like switches on top of DNA to regulate all levels of the cell cycle – from division to death.

\(^{16}\) For my purposes, I am focusing on the question of the use of stem cells in the treatment of major
acquired conditions and diseases (as opposed to genetic conditions or diseases). Because of this, I
am not delving into the issues of ethics or feasibility of genome editing pre-implantation or the
creation/editing of zygotes (post-implantation eukaryotic cells).
The propensity of stem cells to “misbehave” is a critical complication to proposed or implemented stem cell treatments.

An essential distinction across these treatments is the question of cell origin; donor cells can either be ‘*allogeneic*’ (coming from a different organism) or ‘*autologous*’ (from the target organism itself). The earliest established adult stem cell treatments were derived by applying this approach to blood systems. Bone marrow transplants in which blood cell-creating stem cells found within bone marrow are transplanted from a matching donor allowing for the creation of healthy blood cells within the recipient (Healy 2006). In the case of the treatment of Leukaemia and blood diseases, bone marrow transplants have become a staple of treatment protocols and have heavily relied on the inclusion of donor-matching. Tissue donation and the ‘search for the best match’ for these allogeneic transplants has become a major economic driver within medical charities and non-governmental organisations – navigating the murky waters of altruism, economic exchange and the “right to medical treatments” (Healy 2006).

However, there is a fundamental risk associated with these types of stem cell treatments. Outside of the difficulty in finding and matching patients to donors, allogeneic transplants run the risk of donation rejection – or graft-vs-host (GvHD) reaction (Simonson 1985). In these cases, the foreign donor cells spark an auto-immune reaction within the patient’s body that can cause life-threatening complications. In order to suppress this reaction, recipients of allogeneic transplants are often given immunosuppressant treatments that – while reducing the risk of GvHD – raise the risk of other infections and complications. With these complications in mind, how could regenerative medicine find new avenues of treatments?

In 2006, a major development in the understanding and reprogramming of stem cells opened an entirely new route for developing treatments. Japanese researcher Shinya Yamanaka discovered the first protocol to “reprogram” adult “multipotent” cells into a naïve “pluripotent” state for which he won a Nobel Prize in 2012. This discovery prompted a massive shift within the field of regenerative medicine away from “allogeneic” transplant risk reduction and towards a new frontier of “autologous” treatments (Delgado-Morales 2018). Where once patients would rely on the altruism
of donors, reprogramming of adult stem cells offered the opportunity to use the patient's own body to replace missing or damaged tissue while skirting the ethical considerations of embryonic stem cell transplants or the economic and logistical limitations of donation (Papapetrou and Eirini P 2016). This resulted in research focus shifting towards Induced Pluripotent Stem Cells (iPS cells) as a new source of stem cells. Additionally, interest in iPS cell research grew for treatment development for a range of newly considered diseases and conditions – from Alzheimer's to spinal cord injuries to schizophrenia (Robinton and Daley 2012).

Despite the increase in funding and focus on iPS cells within medical research, there have been relatively few breakthroughs that have translated to patients, as of 2022, because iPS cells raise as many questions as they provide solutions. In particular, the technology for producing naïve cells from adult stem cells often produces unknown knock-on effects like tumorigenicity (or the unwanted growth of benign or malignant tumours). For example, while researchers may be able to bring an adult stem cell back to a more naïve state, they are not able to necessarily direct its further directed specialisation or assess the impact of any signalling factors that may be left over from the original, differentiated cell (Lindner et al. 2010).

STROMAL CELLS

Finally, I would like to take a moment to explore a specific type of stem cells that are vital to understand in order to engage with our later themes; mesenchymal stem (or stromal) cells. MSCs are thought to be multipotent cells found in stromal (connective) tissue which, in vivo, differentiate into osteoblasts (bone), adipocytes (fat) and chondrocytes (cartilage). Unlike other cell lines, MSCs make up cells found throughout the body as connective inner tissue.

The term “mesenchymal stem cells” itself is hotly contested with the field of regenerative medicine with some researchers opting for the alternative term “mesenchymal stromal cells” or simply “stromal cells” (Caplan 2017). This contestation springs from our current lack of understanding of the essential nature of these cell types including their ability to differentiate and self-renew and their in-vivo characteristics (Lindner et al. 2010). There is strong debate within the research
community as to the suitability of these cells for inclusion within human clinical trials due to their volatility (Wang et al. 2012).

Importantly for my analysis, mesenchymal stem cells (whether misnamed or not) form the backbone of treatments advertised by private clinics as they enable allogeneic interventions that fall outside traditional regulatory oversight (Turner 2016). This will be discussed in later sections and serves as an indicator of scientific questionability as I review publicly-targeted marketing material for unproven treatments returned through Google Search. For my purposes, I will refer to these cells as Mesenchymal Stem Cells (MSCs) as this reflects the term used most commonly within public conversation of marketed stem cell treatments despite the potential lack of scientific accuracy.

2.1.2 BIOMEDICALLY PROVEN STEM CELL TREATMENTS

Stem cell-derived treatments have been effectively deployed curatively in a few individual cases for treating major conditions and disease. These have, thus far, been limited to deployment for hematopoietic stem cell transplants, skin, and corneal disease (EuroStemCell, 2022).

Bone marrow or hematopoietic stem cell transplants for Leukaemia and other blood diseases rely on the fundamental characteristics of stem cells for success. This process uses chemotherapy to destroy blood cell-producing stem cells within bone marrow and transplanting healthy (either autologous or allogeneic) stem cells into a patient, thereby “resetting” their immune system to produce healthy cells (Bishop 2009).

Stem cells have also been used in the treatment of large-scale skin conditions through skin grafts since the 1980s. This includes the growing of healthy skin using donor cells from the patient’s own body in the lab for transplantation onto damaged sites. In 2017, biomedical researchers performed the first large-scale skin graft using reprogrammed skin stem cells for the treatment of a child with Epidermolysis Bullosa – a genetic disease caused by missing DNA within skin cells which results in the absence of connectivity between layers of skin (Hirsch et al. 2017). This process
involved the harvesting of epithelial cells from the patient, using CRISPR/Cas-9 to edit the DNA of these cells to include the missing genetic code, growing a large-scale skin graft from these edited stem cells and transplanting this healthy skin onto the patient.

The first stem-cell derived therapy to be proven safe and effective for a major condition or disease outside of those of blood or skin is for the treatment of injured corneas. This therapy – known as Holocar – was approved for use within the European Union in 2014 and involves the growing of corneal grafts using healthy limbal stem cells found between the sclera and cornea within the eye. This therapy has proven effective in the treatment of corneal damage (Knapton 2014). Outside of the treatments listed here, there are currently no other proven safe or effective stem-cell derived treatments for major conditions or diseases (EuroStemCell, 2022).

2.1.1 REGULATION

The process of moving gene and cell therapies from the lab to clinics and eventually to patients faced with challenging conditions and diseases requires robust legal and regulatory frameworks. To assure the safety of patients, the development of stem cell-derived therapies must comply with myriad regulatory obligations. These regulations may be binding legal requirements or only guidelines published for best practice adherence.

Additionally, these regulatory requirements are applied throughout the lifecycle of these treatments – from fundamental research, pre-clinical studies, clinical studies, manufacture, commercialisation, patient/public involvement, pricing/reimbursement, patentability, and post-administration product vigilance. Each iterative and cyclical stage includes specified regulatory best practices, basic regulation standards, and guidelines.

Development-stage specific regulations and guidelines are also varied according to geographic authority and often require adherence to more than one regulatory standard. For example, in the United States, while the Federal Drug Administration
(FDA) provides regulatory standards and guidelines at a federal level, individual states may have their own health and safety requirements that provide augmentation. Regulation of stem cell-derived treatments – across borders and contexts – is therefore complex both scientifically and according to which competent regulatory authorities and agencies maintain oversight.

In recent legislative trends, there is a growing trend towards adopting acceleration-friendly regulation of stem cell treatment translation in countries like the US, UK, Canada, and Australia (Bateman-House, Kimberly, Redman, 2015). Ongoing and incoming legislation in these countries often encourages more rapid translation of gene and cell therapies from the lab to clinics and patients. These are sometimes (in the case of both federal and state legislation in the US) couched in relation to patient choice movements like the “Right to Try” and licensing exemptions (Dresser 2015). They are also often supported by publicly funded acceleration efforts of gene and cell therapy Catapults (Figoli, Rush, Sapsed 2017). These efforts aim to open possibilities for faster movement of therapies through the iterative clinical trial process and address standards of best practice for manufacture and production of emerging therapies (Rao 2011).

The process of regulatory oversight in emerging treatments is largely invisible to patient publics, yet its impact on shared conversations around medical claims to expertise, authority, risk, and trust are significant. This process – with its evolving complexities and challenges – forms the backbone of what proven and unproven mean regarding emerging treatments. Before discussing this important influence, it is first helpful to explore what this process has produced as a benchmark for my target conditions of Multiple Sclerosis and Parkinson’s Disease.

2.1.2 BENCHMARKS FOR MULTIPLE SCLEROSIS AND PARKINSON’S DISEASE

Because my data collection within the thesis relies on the analysis of publicly-available search results, I have placed an emphasis on claims of stem cell treatment
for specific diseases and conditions. While I discuss the decision to focus data collection on the two common neurological diseases Multiple Sclerosis and Parkinson’s Disease in my Methodology (Chapter 4), here I would like to briefly outline the state of the science vis-a-vis stem cell treatments and these conditions. This provides a basis for my delineation of proven/unproven treatments in later coding of stem cell treatment offerings related to these diseases. The bottom line here is clear: although there is significant research into stem cell treatments of each of these diseases, there is currently no proven safe or effective treatment derived from stem cells for Multiple Sclerosis and Parkinson’s Disease.

MULTIPLE SCLEROSIS

Multiple Sclerosis (MS) is the most common, non-traumatic disabling immune-mediated disorder to impact people under the age of 50 (Dobson and Giovannoni 2019). Although the basic pathology of the disease is not yet understood, Multiple Sclerosis’ primary impact is on the myelin sheath insulating nerve and spinal cord resulting in potentially severe disabling as well as lowered life expectancy. The progression of the disease is marked by two stages (often referred to as the “two types” of Multiple Sclerosis) including inflammation-triggered relapsing and remitting disease (Relapsing-Remitting Multiple Sclerosis) followed by neurodegeneration known as Secondary and Primary Progressive Multiple Sclerosis (Leray et al. 2010). For patients, this stage differentiation has a marked impact on quality of life; a period of potentially milder and abating neurological symptoms and an irreversible and ultimately fatal loss of neurological functionality. The length of these stages is unique to each patient’s disease progression. This also creates space for different characteristics with the Publics-In-Particular of Multiple Sclerosis patients; those with high functionality in public discourse around the disease and those for whom the disease progression logistically hinders potential direct engagement.

In recent years, bone marrow transplants have been developed as an alternate approach to treatment for aggressive Multiple Sclerosis by “resetting” a patient’s immune system (Patani and Chandran 2012). By combining chemotherapy with the selection and transplantation of healthy bone marrow cells, clinical researchers are able to stop errant T-cells from destroying the myelin sheath around nerve fibres and
give the body the chance to heal. The use of hematopoietic stem cells to influence immunological function is similar to the treatment protocol deployed in established treatments for blood mediated diseases and conditions like leukaemia. However, this approach has only been proven effective in a subset of all disease development types and patient profiles (aggressively developing disease in its primary progressive stage). Additionally, the protocol results in immunocompromisation in patients (Rahmin et al. 2017).

As of 2022, these clinical trials are still ongoing. Indeed, the UK’s National Institute for Health and Care Excellence (NICE) - the executive public body of clinical guidelines within the NHS England and Wales - still listed the stem cell derived guidelines for “Autologous unmanipulated peripheral blood stem cell therapy for multiple sclerosis” as under review (NICE 2022), rather than approved.

At the time of this project’s data collection, these clinical trials had just published their initial findings all of which - as must be emphasised - are only applicable to a small subset of patients. Moreover, no treatments derived from stem cells were - at the time of data collection and writing of the thesis - integrated into public treatment protocols in the healthcare markets targeted outside of clinical trials (CDC 2020 [US], NICE Guidelines - MS 2022 [UK], Freedman et al. 2020 [CAN], and MS Australia 2022 [AUS]).

For the purpose of my research, the new deployment of stem cell-derived protocols for the treatment of MS creates a grey area around the plausibility of stem cell-derived treatments. Here is an ideal confluence of factors as stem cell treatments for Multiple Sclerosis move from the theoretical ground of biomedical research into general knowledge for patients. Promising yet only specifically applicable results from ongoing clinical trials encourages a general positive association of “stem cell treatments” with effective treatment of Multiple Sclerosis as it is rendered within the digital space of Google Search.
Parkinson’s Disease

Parkinson’s Disease is a common progressively degenerative disease of the brain in which the loss of nerve cells in the substantia nigra results in a reduction in dopamine production and uptake (Davie 2008). Dopamine is crucial in many systematic functions including coordination of movement, sleep, and mental functioning. Additionally, patients often present with clusters of alpha-synuclein proteins known as Lewy Bodies (NHS 2022). The pathology of Parkinson’s Disease is unknown although there appears to be links to both genetic and environmental factors for development. Symptoms are generally treated through dopamine replacement therapy and treatments aimed to alleviate secondary impacts on patients (NHS 2022). For patients with Parkinson’s Disease, disease progression is dependent on individual diagnosis. However, on average, patients receiving the dopamine-replacement therapy (via pharmacological interventions like Levadopa) experienced a 5-year amelioration of symptoms before severe disease progression (Parkinson’s UK 2019). Symptoms include physical symptoms of tremors, slow movement and inflexibility as well as secondary symptoms of depression, anxiety, loss of smell, sleep and memory problems (NHS – Parkinson’s Disease, 2021).

Publics-In-Particular impacted by Parkinson’s Disease skew largely older with most diagnosed after the age of 60. However, there are additional sub-groups impacted by emerging treatments for Parkinson’s Disease. These include Early-Onset Parkinson’s Disease (comprising 5-10% of people diagnosed) in those under the age of 50 as well as the broad umbrella of Parkinsonism indicated by similar motor-neurological conditions to Parkinson’s Disease.

Stem cell research in Parkinson’s Disease has tended to focus on the replacement of dopamine mediating neurons to alleviate the need to supplement dopamine and halt or reverse disease progression (Barker et al. 2017). Like with Multiple Sclerosis, this approach to disease management is rooted in plausible biomedical understanding of the disease. However, no treatments derived from stem cells were - at the time of data collection and writing of this thesis - integrated into public treatment protocols in the healthcare markets targeted outwith clinical trial
applications (CDC 2022 [US], NICE Guidelines - Parkinson’s Disease, 2022 [UK], Grimes et. al 2019 [CAN], Waller et al. 2021 [AUS]).

UNPROVEN TREATMENTS

Based on the exploration of stem cell treatments, their history and their internal debates within the basic research community, my thesis adopts a strict definition of “unproven treatment” for Multiple Sclerosis and Parkinson’s Disease. For the purposes of this research, an “unproven treatment” is any offer of treatment that includes language that implies stem cell derivation of treatment protocol. It should be noted that this is not a definition to be applied to clinical trial settings, only to publicly marketed treatments within digital spaces. This definition is only meant for application to direct offers for or secondary promotion of stem cell treatments for Multiple Sclerosis and Parkinson’s Disease.

2.2 ESTABLISHING EXPERTISE, AUTHORITY AND RISK IN DEFINING STEM CELL TREATMENTS

At the core of the debates on stem cell science and therapies are some fundamental questions; how is authority determined in the definition of “unproven stem cell treatments” and how does it shift across actor perspectives? What happens when these different conceptualisations converge in one digital space?

Here it might be helpful to explore unproven stem cell treatments as potential ‘boundary objects’ “which inhabit several intersecting social worlds and satisfy the informational requirements of each of them” (Star and Griesemer 1989). These boundary objects can exist – maintaining some core essential characteristics – across contexts. Stem Cell Treatments are clear candidates for this consideration. Whether discussed by lay audiences, regulators, scientists, or clinicians, stem cell treatments are treatments that rely on cells that can replicate and specialise. These two characteristics – indeed, the definitional characteristics that grants stem cells their “steminess” – persist.
I concur with Prof. Webster’s understanding that unproven and proven stem cell treatments are ‘messy’ objects (Scoles 2018) that – beyond their practical definitions as concrete protocols for treatment – create movement between entities (Star 2010). These objects, according to Webster, may hold some characteristics across contexts but are heavily influenced by the socio-technical assemblages through which they are observed (Webster 2017). Their ‘messiness’ comes from the way in which actors – presenting different positions, claims to authority and implied meanings – are pushed together into a shared space within digital healthcare discourse.

Webster’s introduction of “bio-objects” as messy boundary objects is helpful here, not least for his application to stem cells in clinical settings. Webster suggests that bio-objects occupy a particular form of boundary object as they move out and away from biomedical contexts into new settings. They are distinct, biological entities that take on new characteristics and forms as they move through socio-technical contexts (again harking back to Gardner and Webster’s exploration of messiness and novelty).

Within discussions of unproven stem cell treatments' there are different expressions of risk and reward that critically change their implication from actor to actor. Each actor wrestle with issues of expertise, authority, and risk in their handling of stem cell treatments. The decisions made in line with these relationships to authority and risk constitute an actor-specific form of mediation. That is to say, there is already mediation of information made by these actors before it is integrated into the secondary filter of Google Search. I would like to focus on this to explore the “unproven stem cell treatments” as messy boundary objects from four perspectives; traditional healthcare research and infrastructure, commercial actors, patient publics, and technology platforms (in my case, Google Search).

2.2.1 EXPERTISE, AUTHORITY, RISK, AND HEALTHCARE RESEARCH AND INFRASTRUCTURE

Within traditional structures of linear therapeutic innovation, there is a distinct emphasis on presumed measures of checks and balances in authority as ideas move from the lab to clinical trials and finally to patient uptake. I should emphasise
here that this may not be true linear innovation – as clearly critiqued throughout established STS literature. Perhaps most foundational here is Woolgar’s 1976 discussion of the conceptualisation of discovery within science history as well as Woolgar and Latour’s dissection of laboratory science practices in 1979. Instead and within health innovation, this movement from basic research and through clinical trials mask the more circular, iterative processes that are central to all forms of innovation (overviewed by Pinch and Leuenberger 2016).

However, the view of clinical translation as linear serves a more representational purpose within lay understanding of medical innovation. The process of clinical trial’s functional existence is cited as support in claims to authority and as a mechanism for the mediation of knowledge (Halpern 2004). I first lay out how these are constructed within the medical research community with a particular focus on risk and benefit. I then explore two recent examples within stem cell research that give insight into the pressure points of these systems.

As Steven Shapin emphasises in his 1994 book *The Way We Trust Now: The Authority of Science and the Character of the Scientist*, credibility forms the foundation of healthy scientific research and trust in those systems is essential for scientific practice (1994). Mediation of knowledge is theoretically grounded in the Scientific Method’s checks and balances including consensus building mechanisms (clearly elucidated in classical approaches like Mertonian Sociology of Science [Hargens 2004]).

In theory, peer-reviewed journals are one mechanism that provide inter-institutional checks contributing to evidentiary consensus. These journals – often private publications that require paid access for review and submission – create an epistemological space to collectively organise the evidence base for stem cell research and development of clinical treatments. First an article is submitted to the journal and it is then reviewed by researchers chosen by the editors. Finally, the article – if proven to have evidential value – is published. These building blocks then become the basis for further findings, information sharing, decision making and consensus building within and without the research community.
Risk - the potential for exposure to unsafe medical interventions and secondary effects - is mediated through evidence for effectiveness and plausibility of intervention. In this way the three foundation principles of safety, effectivity, and plausibility support each other. It is worthwhile to note here that this is a particularly technical definition of risk which I am deploying. However, there is an equally important and subjective social dimension of risk that must be appreciated. This risk experienced by a human subject is particularly relevant in the context of individual patients seeking information on experimental stem cell treatments. This type of risk – which falls beyond the ethical discussions within clinical practice – is inherently personal and (as Lupton argues) social (2006).

Here I would like to briefly raise a corollary concept that may be relevant: studies of knowledge and expertise. Harry Collins and Robert Evans present a helpfully realist and perhaps materialistic approach to expertise that creates opportunities to bring expertise into conversation with the structures of Google Search’s practice. Collins and Evans explore the impact of dynamic and quickly shifting information technologies in the public domain. They suggest that the expansion of internet communication technologies moves public dialogue (in their focus, around political subjects) faster than the process of scientific consensus formation. This, they suggest, has led from a “Problem of Legitimacy” (or the question of claims of authority) to a “Problem of Extension” whereby decision-making rights are extended indefinitely across actors (Collins and Evans 2002).

Applied to the question of health care decision-making vis-à-vis emerging and unproven treatments, this raises questions of decision-making caused by indefinite extension. Collins and Evans might suggest that – while traditional authorities within medical practice and research may maintain their expertise – the question of who gets to decide what is and is not appropriate for public access is expanded to include non-traditional actors. This may include patients themselves. It may also include commercial actors whose motivations for involvement may fall outside medical ethics. Helpfully, Collins and Evans also offer a proposal for addressing this critical

---

17 These include expertise as the institutionally-driven parts of organisations (Jasanoff 2004), discrete networks (Eyal 2013) and stark realist terms integrated power and relative position to subject (Collins and Evans 2002).
challenge of extension. They propose the adoption of new categories of science and expertise – with varying constellations of public involvement – that might create inclusive, yet delineated actor groups for scientific decision-making in the public domain (Collins and Evans 2002).

For my research, these new categories of knowledge and expertise take the form of the types of information sources presented in a search result page. As will be discussed in my Methodology Chapter (4), I took significant steps to identify and categorise sources of information that were included in results related to stem cell treatment searches. If I can extend this metaphor of extension, these sources represent different modalities of knowledge and expertise. In the case of private clinics offering stem cell treatments, claims of knowledge, expertise, and – crucially - authority often mimic or reconfigure claims made by similar sources like research institutions and national healthcare providers. I will keep this conceptualisation of expertise and knowledge throughout my analysis, in particular in my coding of results in Chapters 5, 6, and 7.

RISK ACROSS CONTEXTS

Risk within emerging treatments is discussed on several levels that is helpful to unpack. I will discuss a few key elements of risk and risk management that are particularly salient for future analysis of information presented by Google Search. These areas of focus are: risk within medical research and clinical translation, patient and public involvement in research, and social risk as expressed by researchers and the public.

Discussions on risk within the medical research and clinical practice community offer an interesting insight into the sources of logic when weighing the cost of new treatments. Halpern discusses the idea of the “regulatory morality of risk” or the logic by which risk is evaluated. He suggests that individual medical research professionals cite “indigenous morality” that cannot be disentangled “from convictions about the theories and methods of science” yet support a “lesser-harm reasoning” (Halpern 2004). For example, in interviewing transplantation surgeons, sociologist Renée Fox and Judith Swazy found that surgeons weighed the ‘statistical
morality' based on complication rates and mortality when deciding between the use of a new surgical treatment or established treatments. When asked whether or not they would use novel surgical procedures over established treatment, the surgeons responded, “One weighs the morality of the disease against the morality of the operation” (Fox and Swazy 1978). In other words, clinicians must make a judgement call on the risk to both life and quality of life presented by the disease as well as their intervention through treatment. Likewise, Ducey and Nikoo's investigations into ‘formats of responsibility’ in the context of elective surgeries suggests that risk in opting for treatments with limited bodies of evidence is regulated by a structured weighing of push and pull factors from both patient and clinical practitioner. This is reflected in the risk-benefit calculations used by regulators when approving clinical trials (Ducey and Nikoo 2018).

Within translational work through which products and treatments are brought from basic research towards clinical application, there is added the weighing of risk by funding bodies. Conventional approaches to pharmaceutical development face a unique challenge in approaching personalised treatments like stem cell treatments in which live cells must be manipulated. Large pharmaceutical companies are major funders in translation and were initially reluctant to move towards live biological interventions and tissue engineering due to their complexity (Higham and Webster 2017, Faulkner, Kent, Geesink, and Fitzpatrick 2006). Risk, for these companies, is a question of adaptation to new models of treatment development and grant pharmaceutical companies – as funders of this work – an outsized voice in determining what is and is not an acceptable level of risk in rolling out emerging treatments.

Bioethics play a vital role in mediating information on stem cell treatments vis-a-vis public involvement at all stages of research and clinical development. This includes the well-integrated field of Public and Patient Involvement (PPI) and Public and Patient Engagement (PPE) which produces best practice literature on integrating patient voices at early stages of treatment development through which risk can be discussed. There is a significant body of literature on the impact of regenerative medicine in relation to cosmetics and the societal treatment of ageing (Lafontaine 2015, Choi, Yun, and Kwon 2015).
However, I would argue that risk and benefit are seen as part and parcel of the individual’s functional embeddedness within medical research and clinical practice. That is to say that questions of risk and benefit are only accessible by individuals in limited situations wherein a patient may only choose from pre-filtered options. For example, questions around manipulation of cell lines within gene and cell therapy (Chan 2017) play out often as questions in which the decision-making actors are largely scientists and doctors. Public decision making is integrated on an individual patient level, rather than in the communal, commercial, or societal level.

Social risk – or risk assumed by the individual – is expressed by publics, patients, and researchers themselves. There is a significant work within digital health on patient voices vis-à-vis risk and emerging treatments. Here I am not discussing risk in terms of the marking of patients as “at-risk” (as in Scott, Prior, Wood, and Gray 2005), but rather the willingness of patients to take on risk in their treatment decision-making. This includes Chris Gillespie’s take on patients experiencing risk as “measured vulnerability”. He writes:

As more aspects of everyday life are pathologised, risk becomes a central component in how people manage their health, and their lives. One of the arguments of this research is that as risk becomes a condition to be treated in itself, it has come to be viewed as an illness. And, as such, it has begun to be experienced as an illness. (Gillespie 2011)

To choose to engage as a patient, a recipient, of emerging treatments is quite demonstrably a “measured vulnerability” – even when only searching for information that might place an individual into this position. An individual searching for information around their disease takes onboard the risk of bad information and the potential positive impact of useful information. How, then, does this vulnerability translate to the practice of information gathering on digital platforms like Google Search?

I observed patient’s perspective on risk through a study I undertook with my colleagues Amy Zarzeczny, Claire Tanner, Megan Munsie, Clare Blackburn, and Jan
Barfoot in analysing patient perspectives on risk as seen in direct messages to stem cell research organisations (Zarzeczny, Tanner, Barfoot, Blackburn, Couturier, and Munsie 2018). In this study, we analysed over 1000 individual requests for information and aid submitted directly to the EuroStemCell and Australian Stem Cell Network websites. These anonymised submissions recorded patients seeking more information and access to stem cell treatments for a range of conditions and diseases. When patients were making contact for information around major conditions and diseases, three query themes emerged: Can you treat me? Where should I go? Can stem cell treatments help and would you recommend them? These questions suggest willingness to engage with stem cell treatments despite potential risk, a personal seeking for where offers an available treatment now, and a desire for authority on recommendations specific to individual health realities. While the term “risk” only appeared in 3-4% of enquiries with few patients directly referencing the risk of stem cell treatment or regulatory status, the subject of risk was present throughout responses. This presents a complex relation of individual patients to risk – not simply technical, but social and personal.

This context in which patients are seeking treatments within and without traditional medical spaces is explored in depth by several STS and digital health researchers from the perspective of patients. This includes Fadhila Mazanderani et al.’s exploration of the embodiment of hope in patients seeking contested treatments for Multiple Sclerosis (2018) and Priscilla Song’s investigation of stem cell tourism to China for neurodegenerative diseases (2010). These patient-led perspectives suggest that conversations of risk for patients are both deeply personal and exist outside of the public-wide discussions undertaken within traditional medical research contexts. Moreover, it underlines that the question of risk – what patients are willing to consider as acceptable intended and unintended consequences – may differ from those adopted by the medical community.

From the perspective of biomedical researchers, there exists a pointed focus on researchers, themselves, navigating uncertainty within public conversations of stem cell-derived treatments and raising red flags in emerging treatments (Kewell 2013). Interestingly, a significant number of these commentaries on the social implications of stem cell treatments come from biomedical researchers publishing across fields.
They include dire warnings from researchers about unregulated stem cell industries (Lysaght, Lipworth, Hendl, Kerridge, Lee, Munsie, Waldby, and Stewart 2017) and calls to confront “hype” around emerging stem cell treatments (Caulfield, Sipp, Murry, Daley, Kimmelman 2016).

We can see similar processes of terminology shift within public health discourses played out over time. For example, Dancy-Scott et al. documented the shift in terminology used in representation of HIV/AIDS text mining abstracts from the International AIDS society from 1989 to 2014 (Dancy-Scott et al., 2018).

The term “AIDS epidemic” was dominantly used from 1989 to 1991 and then declined in use. In contrast, use of the term “HIV epidemic” increased through 2014. Beginning in the mid-1990s, the term “treatment experienced” appeared with increasing frequency in the abstracts. Use of terms identifying individuals as “carriers or victims” of HIV rarely appeared after 2008. Use of the terms “HIV positive” and “HIV infected” peaked in the early-1990s and then declined in use. The terms “men who have sex with men” and “MSM” were rarely used until 1994; subsequently, use of these terms increased through 2014. The term “sex worker” steadily increased in frequency throughout conference years, whereas the term “prostitute” decreased over time. (Dancy-Scott et al. 2018)

Dancy-Scott et al. note the movement of terminology that they contribute to two factors; change in medical advances and consensus within the research and clinical community and the public de-stigmatisation of topics around HIV/AIDS. This is a stark documentation of language shifting the borders of disease and those who interact with it as patients, publics, or medical practitioners. Although my data does not yet show a shift of this magnitude in conversations around stem cell treatments, this may be an area of rich data in the future.

Perhaps applicable to stem cell treatments, Maren Klawiter points to the idea of shifting “disease regimes” and associated language as materially impacting patient experience of disease. She suggests that diseases – much like the messy objects of
stem cell treatments – shift in their experiential (or social) meaning over time. Klawiter writes:

To claim that different illness experiences become available at different times is simply to acknowledge that experiences of disease are shaped not only by the individual circumstances of disease sufferers and the particular character of their pathologies, but by culturally, spatially and historically specific regimes of practices. (Klawiter 2004)

The practices performed by Google Search and explored in my research suggest a specific characteristic of these “culturally, spatially, and historically specific regimes”.

For stem cell treatments as emerging medical interventions, shift in discourse creates opportunity for shifts in search results returned according to not only advances supported by scientific consensus within the medical research community, but consensus in discourse within the wider publics. It is important to note here that shift in discourse might also include results that do not reflect consensus. In fact, sudden public excitement around new theories or treatments may result in new terms and language outwith the context of scientific consensus and simply reflect growing interest.

This is uniquely demonstrated in the case of Chronic Cerebrospinal Venous Insufficiency (CCSVI), the controversial treatment for MS popularised in the early 2000s. In 2008, Italian doctor Paulo Zamboni published a contentious paper suggesting that the biological root of Multiple Sclerosis is compression of blood drainage from the central nervous system. He then suggested an equally controversial treatment including the stenting of specific veins as a form of therapy.

This publicy-promoted hypothesis led to increased popularity within Multiple Sclerosis patient communities online at a pace that outstripped the scientific research and clinical trial process of evidentiary testing. Arie Gafson and Gavin Giovannoni found that the digital platforms of YouTube (owned by Google as of 2006) and Facebook played a significant role in the spread of a belief in CCSVI as
the cause of Multiple Sclerosis and the unsubstantiated interventions promoted to address it (2013). Gafson and Giovannoni conclude:

It is clear from the statistics that the social phenomenon of CCSVI outpaced scientific publication. Part of this may have been due to the weak evidence base and the current dogma of MS as an autoimmune demyelinating and neurodegenerative condition. Most neurologists and scientists initially dismissed CCSVI as a viable explanation for MS as CCSVI is not congruent with the current knowledge. It is also undoubtedly the case that scientific publication by its nature can be a slow process from outlining an initial protocol, gaining ethical approval, conducting a study to preparing and submitting a manuscript. However, when the social phenomenon exponentially grew in popularity, the scientific community was forced to engage in a debate when patients began putting themselves forward for a procedure that had little evidence base and was associated with complications that in some cases proved fatal (Rudick, 2010). (Gafson and Giovannoni 2013)

This is a clear example of the ways in which technological platforms like Google create space for new discourses to be created. These resulting discourses are untethered from traditional discourses of medical authority and evidentiary development.

LIMITS WITHIN BIOMEDICAL RESEARCH: OBOKATA AND MACCHIARINI

However, there are recent clear cases specifically within stem cell research that raise questions about the presentation of “proof” and risk assessment. I would like to briefly explore these as they are clear examples of two separate points of misconduct that expose the limitations of traditional models of authority and risk assessment within medical research and clinical practice.

The case of researcher Haruko Obokata draws into question the creation of evidence as part of a claim to authority within basic research. Obokata served as a lead researcher at the Laboratory for Cellular Reprogramming, RIKEN Centre for
Cellular Biology in Japan where she focused on refining the protocols for creating induced Pluripotent Stem Cells. In 2014, her dissertation and subsequent papers were submitted and accepted by *Nature* – one of the foremost peer reviewed medical journals. The papers claimed to document a new and efficient protocol for creating Stimulus-Triggered Acquisition of Pluripotency (STAP) cells – a simple method for inducing pluripotency within any cell found in the body (Obokata 2014). The benefit of a discovery like this would be the ability to easily and efficiently harvest autologous multi-potent cells from any source within the body to be used within clinical treatments. However, it was found that not only could the protocol not be reproduced in other labs, but that the data included in her papers appeared to be falsified (Nature 2014). Obokata was subsequently stripped of her PhD and accreditations and *Nature* ran a full retraction of her work. This example of falsification of evidence is not unique to stem cell treatment but, for my purposes, raises an interesting question of evidence and authority within the field.

On the one hand, this case could be seen as a success as data falsification was caught and the research removed from the public body of evidence on STAP cells. However, it also underlines that the pre-publication peer review – which should enable all published material to benefit from the collective claims to authority created by joint review – failed. Moreover, Obokata’s case may be endemic of wider-spread failure within the creation of evidential knowledge within the medical research and clinical practice community. Questions about reproducibility within published papers in peer-reviewed journals has come to the fore. In 2016, *Nature* surveyed 1,576 researchers connected with the publication and found that 70% of them had tried and failed to reproduce the experiments of other scientists as published within the journal (Nature 2016). Yet the stakes are high for researchers. If proven true, Obokata’s STAP protocol would have made possible advances in regenerative medicine that would have real impacts on clinical treatments for patients suffering from acquired diseases and conditions. If medical research and clinical practice rely on the establishment of a body of evidentiary knowledge supported by peer-reviewed mechanisms to establish claims to authority, then what do we do when these systems fall short?
The second case is that of stem cell researcher and surgeon, Paolo Macchiarini. His work on tracheal scaffolding using autologous stem cells was widely reported within the medical research and clinical practice community as a breakthrough in treatment. In the early 2010s, *The Lancet* – a well-renowned peer-reviewed medical journal – ran a series of articles on Macchiarini hailing his work on autologous organ scaffolding as the “new frontier of regenerative medicine” (Holmes 2012). During his time as a visiting researcher at the Karolinska Institute in Sweden and in coordination with University College London (UCL), Macchiarini performed several operations on patients to provide them with new tracheas based on the evidence he provided in his previous research.

However, these treatments resulted in serious accusations of research misconduct, falsifying qualifications and gross medical misconduct with fatal consequences. Of the reported eight transplant recipients (including some without life-threatening conditions) to receive the tracheal implants, seven died immediately and the eighth was left permanently reliant on medical intervention (Schneider 2018). As of 2016, Macchiarini is under investigation by the Swedish authorities for medical misconduct and manslaughter tied to the operations performed on three patients operated on in Sweden. *The Lancet* – which initially ran articles in defence of Macchiarini’s published works within the journal – finally ran an article declaring the researcher as guilty of misconduct in 2018 and retracted his previous article submissions as containing falsified evidence (Lancet, July 2018).

Here a number of fail-safes in the burden of proof failed. First institutionally with the failure of both Karolinska Institute and UCL to properly vet, provide ethical oversight and review the evidentiary work of Macchiarini before approving, funding and supporting the application of his treatments on living patients (Morgan 2016). Secondly in the peer review process as *The Lancet* – a peer reviewed medical journal – failed to print retractions until 2018 (Lancet 2018). So, while claims of authority to define “proven” and “unproven” stem cell treatments are derived within the medical research and clinical practice community from the mechanisms of checks and balances, we can also see that claims of “indigenous morality” as a guide for risk and benefit analysis is largely left to the individual.
The cases of Obokata and Maccharini explored in this chapter illuminate difficulties within traditional models of authority within medical research and clinical practice. These examples underline a tension that plays out within Google Search results. In my analysis of Google Search results in Chapter 5, 6, and 7, I try to keep in mind the characteristics of these breaking points in traditional models of authority. Namely, the paradox created between widespread recognition of the limitations of biomedical science and the fallibility of scientists and the perseverance of “promissory expectations” towards new treatments and therapies as promoted in digital search spaces. This lends itself to highly legalistically-mediated movement of therapies into healthcare marketplaces and, in particular, their accessibility within highly regulated healthcare infrastructure like the NHS (Mahalatchimy and Faulkner 2017, Rose and Williams 2012).

2.1.1 COMMERCIALISATION WITHIN BIOMEDICAL RESEARCH

Before moving on to the private commercial actors within digital health, I would like to pause and address the issue of commercialisation within the biomedical research community. In separating out commercial actors in this field, I do not want to imply that there is no commercial influence within the research community impacting the dissemination of information around stem cell-derived treatments. That would not only be short-sighted, but overlook the well-established literature on commercialisation of regenerative medicine products and services within the private sector (Banda, Mittra and Tait 2017, French et al. 2013, Sipp 2011).

However, I would argue that the impact of regulation, public funding, and integration into institutional healthcare structures of these routes to clinical application create a degree of separation in the progression of treatments as compared to open markets represented by direct-to-consumer healthcare markets. This does not mean that publicly-funded research initiatives and translational projects are free from market influence. There appears to be an increased focus on accelerating research from lab to clinic through translational funding opportunities.

---

18 In the UK, most commercial developments of regenerative medicine products are made by Small-to-Medium Enterprises (SMEs) reliant on public grant funding rather than private funding (Banda, Mittra, and Tait 2017).
(e.g. European Union Horizon 2020 funding in Andersson 2017). Additionally, several national and international Cell and Gene Therapy Catapults have been established since 2015. These Catapults are aimed at translating publicly funded regenerative medicine products into commercially viable products (Gardner and Webster 2017).

These funding streams privilege clinical (and, therefore, commercial) development over foundational research within regenerative medicine research. In a term borrowed from Andrew Webster, this may create an environment primed for “recursive relations” in which potential treatments are limited by what has been defined as objective for commercial use. Regenerative medicine researchers respond to the wants of a health marketplace while the digital health marketplace responds to available products produced by that research (Webster 2017)19.

Additionally, these mechanisms of regulation and oversight within traditional translational pathways are dynamic in response to public discourse. In the United States, changes to regulation governing access cited clinical trial evidentiary pathways as “slow, laborious, and expensive” and prompted policy makers to reconsider the methods of evaluation for regenerative medicines (Hogle and Das 2017). Inherent here is a responsiveness to shifts in public and commercial conversations around risk and evidence - again, harkening safety and efficacy - in regulating access to emerging medical interventions.

Moreover, it is helpful to underline the connection of regenerative medicine to a narrative of hope and expectation in relation to future treatments and ‘promissory identities’ (Morrison 2012, Gardner, Highham, Faulkner and Webster 2017). This often influences the wording of grant bids and mission statements for publicly funded research projects. For example, the British Heart Foundation Centre of Regenerative Medicine’s £2.5 million pound project uses the phrase “Using stem cells to mend broken hearts” throughout its project documentation - both aimed at the public and

---

19 Webster only uses ‘recursive relations’ to relate to the “makers of” regenerative medicine products. However, taken in hand with the funding priorities expressed by research funders like the European Union, I move this definition one step back in the pipeline of development to empower researchers, not simply makers of commercial regenerative medicine products.
funders. This phrasing suggests an immediacy of treatment despite the foundational nature of the research undertaken with the grant. As a strategy for funding bids, appealing to the positive possibilities of stem cell-derived treatments in the treatment of heart disease is effective. This is despite the disconnect between the language of immediacy and the actual progress of clinical treatments.

However, while there are several accelerated pathways to regulation and reimbursement around Advanced Therapeutic Medical Products (ATMPs) within countries like the United Kingdom, there is limited impact of these acceleration schemes on time to clinics. This may, in part, be due to the need for more robust evidentiary bases for treatments than those supported by accelerated, commercially-driven funding schemes (Mahalatchimy and Faulkner 2017).

Commercial pressures on the biomedical community are real and driving; for institutions and individual scientists themselves. However, for my purposes, there is enough of a difference in the nature of the relationship between the biomedical research community and commercial drivers to privilege the perspective of the biomedical community in my definition of “proven” and “unproven” treatments.

2.2 AUTHORITY, RISK, AND COMMERCIAL ACTORS

Despite the complexity of the underpinning science, stem cells have risen to the fore as the vanguard of public facing, condition-focused research since the early 2000s. Within the public eye, stem cells have moved from an aspect of established treatment protocols (e.g. bone marrow transplants) to standalone therapies applicable to a range for acquired and genetic diseases and conditions (Regenberg et al. 2009). This shift is echoed in the increase of private commercial actors contributing to the space of stem cell information online.

The growth of private for-profit clinics offering myriad stem cell treatments around the world has produced a strong reaction from traditional medical authorities, researchers, and healthcare policy makers. In the face of the rising profile of direct-to-consumer stem cell treatments online, these traditional authorities have published strong warnings on the risks for patients associated with experimental treatments as
well as collective calls for stricter regulation for consumers (e.g. McLean, Stewart and Kerridge 2015, Sipp et al. 2017, Cohen and Cohen 2010, Jian and He Dong 2016, Lysaght et al. 2017, Regenberg et al. 2009). These calls to action have come on the heels of high-profile cases of private clinic closures after patients suffered gross bodily harm or death after undergoing treatment (Mendick and Hall 2011).

Existing research on the proliferation of stem cell treatments as private business have tended to focus on defining the “problem” (namely lack of oversight, regulation, or evidence) and establishing the “scope” of established private clinics currently offering unproven stem cell treatments through digital advertisement (Berger et al. 2016). Like my research, the existing literature tends to look at the impact of these treatments on patients of acquired diseases and conditions – rather than speculative impacts on genetic conditions or individuals impacted by early genetic intervention (Kuriyan, Albini and Flynn 2017, Matthews and Ilitis 2015, Murdoch, Zarzeczny and Caulifield 2018). The reasons for this, it seems, is that the opportunity of exploitation of living patients is increased for those already suffering from acquired conditions who might elect to pay and undergo treatments for which the safety and efficiency has not yet been established. My research deviates from existing work by focusing on the platform through which patients seek information on their disease or condition, rather than on the clinical treatment providers themselves.

GUIDING SOCIOLOGICAL WORK ON PRIVATE-SECTOR STEM CELL TREATMENTS

Two fundamentally important evaluations considered in the development of my perspective of stem cell clinics are the work of Leigh Turner (University of California – Irvine) and Paul Knoepfler (University of California – Davis). Turner and Knoepfler’s survey-based research projects focused on mapping the claims of American private clinics offering unproven stem cell treatments online. Their research provided a roadmap for categorising commercial offerings, while underlying the sheer scale of the private stem cell treatment market. The project mapped existing clinics advertising online within the United States through a thematic analysis of treatment claims on public-facing websites. Turner and Knoepfler focused exclusively on the use of Mesenchymal Stem Cells (MSCs) as a marker of a clinic’s
lack of evidence-based treatments. This is supported by the current literature within the scientific community on the use of MSCs as according to the ISSCR (2018), there are currently no proven safe or effective MSC-derived treatments for any condition.

It should be noted that the majority of the private clinics within Turner’s studies used the terms “mesenchymal stem cell” or, more often, “adipose-derived stem cells”. Adipose – or fat – tissue is a controversial source of mesenchymal stem cells but seems to be valued for its prevalence within human bodies and, as explored in discourse around fat tissue in regenerative medicine (Erikainen, Couturier, and Chan, 2018), its socio-cultural implications. As part of a short project in coordination with Sarah Chan and Sonja Erikanian of the Usher Institute for Population Health Sciences and Informatics, we expanded the coding established by Lee Turner and applied it to an analysis of 71 UK-based private clinics offering stem cell treatments. Although this study included private clinics offering cosmetic treatments alongside treatments aimed at acquired diseases, we also found wide reference to the use of mesenchymal stem cells (Erikainen, Couturier, and Chan, 2018).

The issue of private clinics offering unproven stem cell treatments opens up a question of claims to authority and navigation of risk. Additionally further projects on the increased use of crowd funding sites like GoFundMe.org to raise money for individual access to unproven stem cell treatments offered an evidence-rich, patient/user focus for analysis of the draw of private health care offers (Snyder, Turner, and Crooks 2018). Note that within this literature on the social aspects of unproven stem cell treatments, there is a focus on public use and salience of “stem cells” rather than the larger umbrella of regenerative medicine. In terms of regulation, there have been a few projects focused on mapping country-specific approaches to oversight of private clinics offering stem cell treatments including international comparisons (Lysaght, Kerridge, Sipp, Porter, and Capps 2015) and country-specific mapping of regulatory oversight of stem cell clinics (Sipp and Turner 2012). These papers share a corollary focus on the ethical questions and potential limitations of nationally-centred regulation of stem cell therapies.
The work of Alan Petersen, Claire Tanner, Casimir MacGregor, and Megan Munsie (2017) explores the rise of “stem cell tourism” – the political economy behind the effort to influence individual decisions in favour of often ‘unproven’ stem cell-derived treatments unavailable through local healthcare. Additionally, there have already been a number of research projects into the increasing access to stem cell-derived treatments and the evidence base of treatments and practices in terms of regulation (Snyder et al. 2012, Turner et al. 2016). Most salient for my purposes is their finding that private stem cell clinics surveyed funneled up to 30,000 USD per month back into marketing and paid advertisement mechanisms (Petersen, Munsie, Tanner, and Kotey, 2017).

Crucial to the discussions of stem cell tourism advanced by Petersen, Tanner, MacGregor, and Munsie is the delocalisation of services being sold by private clinics online. Geography - critical in conversations around regulation and reimbursement on national healthcare levels - is made redundant by digital offers that cross national geographic borders. Indeed, the unplaced nature of these offers is uniquely enabled by digital dissemination platforms like Google Search. These allow offers to transcend locally-rooted regulatory, reimbursement, and consensus structures in the compressed space of search results. This is a particularly interesting aspect that I explore through my data analysis and discussion sections.

RISKS OF PREMATURE AND UNEVIDENCED BIOMEDICAL INTERVENTIONS IN THE MARKETPLACE

What happens when premature or unevidenced treatments move into direct-to-consumer commercial markets? In this section, I will briefly outline some of the major examples and areas of concern raised by these movements. By sharing examples illustrated in the digital health literature as well as through regulatory watchdog agencies, I hope to clarify what is at stake here.

Concern around patient risks tops the list of concerns raised by researchers, patient advocates, and regulatory agencies in the face of expanding unproven stem cell offerings. Indeed, the wealth of digital health literature on unproven stem cell treatments, examples of patient exploitation and questionable practices by private

First and foremost, unproven stem cell treatments carry the risk of grievous harm to patients. A 2016 U.S. Federal Food and Drug Administration workshop on stem cell treatments found significant examples of patients suffering unintended side effects and impacts from unlicensed and unregulated clinics offering specifically marketed stem cell treatments (FDA 2016). This included the growth of tumours at the injection site; debilitating harm from the injection of mesenchymal stromal cells in attempts to cure eye disorders, resulting in blindness; and, in several cases, the death of the patient (Knoepfler, 2017). In 2010, a child treated with an unproven stem cell treatment at the XCell Center in Dusseldorf, Germany died because of an internal brain haemorrhage caused by the intervention. The centre was subsequently forced to close and reopened in 2014 in Lebanon (Petersen, Munsie, Tanner, MacGregor and Brophy 2017).

A 2018 study of PubMed and Google Scholar journal articles by Gerhard Bauer, Magdi Elsallab, and Mohamed Abou-El-Enein found 35 fully documented cases of acute or chronic complications or death resulting from an unproven stem cell treatment delivered through private clinics (2018). This study only included documented cases reviewed and accepted in peer-reviewed and audited journals. It should be noted that these findings are potentially underreporting the impact of unproven stem cell treatments due, fundamentally, to the lack of oversight, patient follow-up and transparency of private clinic practices.

These offers are not simply restricted to a patient’s geographic location. As Petersen, Munsie, Tanner, MacGregor and Brophy document, patients are regularly engaging in overseas travel in the search for stem cell treatments not legally available in their home countries (2017). This poses a particularly vexing challenge for regulatory agencies who must contend with questions of unclear medical authority and patient safety.

This type of medical tourism also creates complex problems for healthcare providers and infrastructure. A secondary health risk to patients associated specifically with
stem cell tourism is the inadequacy of aftercare when a patient returns to their home health system. If a patient is faced with after or side effects of their treatments, it is the home healthcare providers and infrastructure who must carry on the patient’s health management. The lack of documentation and clarity on the procedures undertaken results in incomplete patient information if future interventions are necessary. Within the study conducted by Gerhard Bauer, Magdi Elsallab, and Mohamed Abou-El-Enein in 2018, only 4 of the 35 accepted cases of clinical harm or death included documentation of the cell preparation protocols for clinical application.

Indeed, the presence of clinical documentation does not seem to have an impact on the marketing of direct-to-consumer stem cell treatments. A study by Einsidel and Adamson of 23 private clinic websites found that:

…only one of these 23 providers mentioned a clinical trial on the safety and efficacy of stem cell treatment for damaged heart muscle, although information on this was limited to a couple of sentences. Despite this lack of evidence, cost estimates for treatment ranged from $5000 to almost $39,500 USD. (2012)

I will explore examples of this presentation of direct-to-consumer marketing of unproven stem cell treatments on Google Search more in Chapter 5.

A clear example of the entry of treatments into the marketplace based on poor or premature science is seen in the promotion of osteoarthritis interventions for sports injuries. From 2009-2014, researchers Kirstin Matthews and Maude Cuchiara documented prominent players within the American National Football League undergoing publicly reported stem cell treatments for injuries sustained through their sport (2014). These treatments were directly marketed to athletes – both professional and amateur by private clinics both within the home country of the United States and abroad. These treatments were overwhelmingly described as autologous “same-day” treatments through which “stem cells” (noted here in quotes due to the questionable terminology used) were taken from one donor site in the athlete and injected within the injury location. Matthews and Cuchiara note the lack
of clinical support for these interventions and noted reports of severe side effects including one death and one near death after treatment at a clinic in Germany.

Because these “same-day” autologous treatments did not include the manipulation of cells or additional pharmacological intervention, they fell under the FDA’s regulatory guidelines for cosmetic surgery, not medical intervention. It should also be noted that the majority of these interventions stated explicitly the use of adipose-tissue derived mesenchymal stromal cells (MSCs) – another problematic consideration. These interventions of “stem cells” persist both through Google Search and indeed media coverage of athletic regimes. A 2012 study by Timothy Caulfield and Amy McGuire on media coverage of professional athletic undergoing stem cell treatments recorded a strong bias in favour of unsubstantiated interventions:

Most of the articles (73%) do not even touch on whether there was evidence of efficacy. Given that these stories are generally about the use of stem cells as a therapy or training aid, silence on the evidence could be viewed as implying potential efficacy. Although several articles note that the treatment was probably ineffective (or unproven), an equal number (10%) explicitly describe the treatment as being effective. A large number of the articles (42%) mention a specific benefit—again, implying efficacy—and several (13%) quote a stem cell provider in support of the treatment. Only five articles (6%) mention any possible risk or safety concern, such as tumor growth, with stem cell treatment. (Caulfield and McGuire 2012)

The premature marketing of stem cell treatments in this case is influenced by the socio-economic context in which it plays out. The media reporting of the interventions, the player and sports organisation profile, and the uniquely permissive and desirous context of health interventions in professional and amateur sports creates the perfect conditions for unproven treatments to thrive. Here authority of evidence is secondary to the authority of community and social relations in providing support for these medical interventions.
NON-CLINICAL COMMERCIAL ACTORS

Although private clinics are the primary focus of the literature on stem cell treatments online, there are additional commercial actors present exerting influence in this space. I will expand upon these groups within my methodology and results chapters. However, I would like to first describe a few of these voices included in evidence in a page of returned Google Search results of which to be aware.

In addition to primary treatment providers, related medical actors and institutions include biobanks, pharmaceutical companies, private companies offering complementary treatments, and biotechnology companies. These actors differ in their intended audience from patient publics, policy makers, and researchers themselves (in the case of biotechnology companies). I will explore each of these actors in detail in Chapter 4 as they form a significant part of the content producers within Google Search results. It should be noted that the characteristics, targeted audience, and messaging of these actors on stem cell treatments varies across geographical location and commercial aim. However, each of these actors offer a commercial product or service related to stem cell treatment, rather than a stem cell treatment offer itself.

Finally, there are commercial actors within this space focused on the mediation of information around treatments without providing direct offers of intervention (whether primary or complementary). These include content producing actors like aggregate health news sites, lifestyle and non-traditional medicine blogs, and broadcast news outlets. Again, I will map the boundaries of these actors in later chapters. However it is helpful to note that, by and large, these actors are funded by in-page advertisement revenue as an advertiser-targeting service. Additionally, they act as both content producers and informational aggregators in digital discussion spaces on stem cell treatments.

TOWARDS THE PRACTICE OF GOOGLE SEARCH AND PLATFORM MEDIATION

Within this chapter, I first established my approach to defining and situation stem cell treatments vis-a-vis the actors within the field of regenerative medicine and digital
I, then, established the qualifiers against which all proven and unproven stem cell treatments will be judged; safety, efficacy, and plausibility. These characteristics guided the exploration of the history and current state of understanding of stem cells and their potential for treatment in clinical settings, specifically for the treatment of Multiple Sclerosis and Parkinson’s Disease. I then brought these definitions into conversation with the major actors and stakeholders who will be tracked in my data collection and analysis. I explored the shifting relationship of stem cell treatments, expertise, authority and risk as seen by research and traditional healthcare bodies, commercial actors, and patient publics. From this exploration emerges a complex constellation of actors with different motivations for engagement in information sharing and decision-making. Whereas patients may be motivated by personal investment in health options and outcomes, traditional healthcare bodies and commercial actors may introduce new economic considerations and objectives into stem cell treatment discourse.

Deborah Lupton offers helpful encouragement to explore the impact technologies have on the complex arrangement of digital health subjects. In particular, she marks the importance of situating human actors within dynamic systems of technologies and socio-economic power. These networks and interests collide as they seek engagement from lay public attention. This collision brings together a range of disparate actors, including:

... members of the medical profession and allied health professionals, health insurance companies, pharmaceutical and medical technology companies, hospitals, patient support associations, government agencies and digital device and software developers (Lupton 2014).

Lupton calls for the identification of commercial influences on the creation, dissemination, and harvesting of digital data (from users) within digital health (2014). This focus on identifying and mapping the influence of commercial actors within this digital health space carries through my analysis in Chapter 5, 6, and 7.

Now that I have outlined stem cell treatments as the messy objects they are, I can begin to analyse the point of compression that is Google Search mediation through
which these disparate definitions, pressures, and actors are forced together enabling specific constellations of information. However, to analyse that moment, I must first delve into the most influential actor yet uninterrogated in this dissertation, the platform of Google Search itself. This, indeed, deserves its own chapter.
In 2017, I found myself in a sprawling hotel ballroom in Boston along with the representatives of the major international stem cell research networks and institutes at the annual International Society of Stem Cell Research (ISSCR) conference. I was there as representative of the European stem cell network, EuroStemCell whose remit - unlike many of the other research associations present - includes a strong focus on public engagement and education. There was overwhelming consensus amongst discussants with each executive representative of the 50-odd meeting echoing similar sentiments: *We are overwhelmed by the amount of misinformation and the spread of unproven stem cell treatments targeting the public. We do not have the resources to aid the sheer number of patients directly reaching out to our research institutes.* Some possible solutions were promoted but, both personally and professionally, I was shocked that no one addressed the elephant in the room: Google Search.

In this chapter, I contextualise Google Search and its role in the mediation of information about stem cell treatments for publics and patients. I cover the aspects of Google Search that are necessary for contextualised engagement with my core research question; *how does Google Search mediate information that people access online about ‘proven’ and ‘unproven’ stem cell treatments?* These are the histories, borders, and impacts of Google Search that I feel are essential to understanding the role Google Search plays in this constellation of actor interactions.

As in Chapter 2, this chapter does not shy away from the technical details of the subject. Indeed, details of how Google Search operates and the mechanisms at play when we engage in searching the internet are necessary to better track how technical platforms and their processes contribute to the dissemination of information. The deployed mechanisms of search are critical to my analysis and, as
will be explored, can only be uncovered indirectly due to the opacity of Google Search as an algorithmically-mediated platform.

In this chapter I explore the complexities of studying Google Search as an algorithmically-mediated platform and epistemological machine (Manovich 2013). I first situate Google Search within existing STS literature on digital health and work specific to stem cell treatments to create a grounding for the conceptualisation of Google’s potential as a mediating actor. I then map the history and characteristics of Google Search that give rise to its importance as a research subject. I look at a few examples of how digital methods have been employed (to lesser or greater success) for the analysis of outputs created by Google Search. Finally, I discuss the position of Google Search vis-a-vis stem cell treatments and end users to bring us to what is, for my purposes, the most important object of analysis; the return of results.

3.1 SITUATING ALGORITHMICALLY-MEDIATED PLATFORMS WITHIN STS LITERATURE

[T]hese media are not the open, collaborative spaces that they once were as they have become commodified and bent to commercial interests. New forms of power have emerged in the digital age. The Internet empires control the digitized knowledge economy. Users’ interactions, including their search engine queries, are now archived and algorithmically manipulated for commercial purposes, delimiting the choices that users may be offered. (Deborah Lupton 2015)

Within literature on critical algorithm studies, a key question is: how do we study algorithmically-driven platforms when the term “algorithm” is deployed in so many dissimilar ways across disciplines? Within computer science, search algorithms simply refer to functional processes of returning results from a set of indexed sources. Shifting to digital sociology studies, algorithms focused on search and retrieval have important societal significance in that they can ‘condition our very existence’ through the ways they shape language, ideas, and subjects in relation to searchers (Kitchin and Dodge 2011). The logic of these algorithms possesses the “power to enable and assign meaningfulness, managing how information is perceived by users, the ‘distribution of the sensible.’” (Langlois 2012).
When navigating between theoretical discussions and empirical studies of algorithmic impact, an interesting tension arises in the space between perceived algorithmic agency, actual technical capabilities, and their conceptualisations within research (Rogers 2009, Umoja Noble 2018, Chun 2005, 2011, Puschmann 2019). This tension is embodied by what Bucher refers to as “the algorithmic imaginary” – or what users believe an algorithm does (2017). A belief of what algorithms do may be as light as “they answer questions” to the complex “they vet and curate information.” Moreover, this tension creates space for misunderstandings of technological impact. This has produced significant critique focused on the use of the term “algorithms” as a stand-in for larger systems of decision making on platforms. As Ian Bogost (2015) writes:

Concepts like “algorithm” have become sloppy shorthands, slang terms for the act of mistaking multipart complex systems for simple, singular ones. Of treating computation theologically rather than scientifically or culturally.

Specific to my research, this is a particular salient point when navigating the treatment of topics on Google Search. It prompts me to take particular care when considering Google Search not as the clean homepage with a single entry point for queries to be processed by one single “algorithm”, but rather as a portal to a number of algorithmically-mediated mechanisms. This will become clear through my analysis in Chapters 5, 6, and 7 as the “multipart complex” system is observed within a single search engine result page.

Puschmann suggests that in analysing algorithmically-driven systems important to focus on the particular characteristics defining algorithms as a subject. These include public relevance algorithms (Gillespie 2012), networked information algorithms (Annany 2015), and algorithmic personalisation (Puschmann 2019). For my purposes, Gillespie’s work on public relevance algorithms suggests several important characteristics to track in my analysis of Google Search. Public relevance algorithms are those algorithms that perform crucial functions within high-usage algorithmically-mediated platforms like the public-facing platforms of Google Search, Facebook, Twitter, and YouTube as well as platforms that govern civic infrastructure.
like criminal records databases, taxation departments, and banking. The characteristics that Puschmann suggests are important to analyse include “patterns of inclusion” (the choices made by developers on what information is included within a search index), “cycles of anticipation” (the efforts made by platform providers to anticipate the needs and wants of users and that logic’s inclusion within search results), and “the production of calculated publics” (the predetermined logic that renders groups of users as distinct publics and logic that governs what information would be ‘most beneficial’ to them) (Gillespie 2012).

Much of the early sociological critiques of algorithmic governance tended to focus on what Richard Rogers calls ‘culpability politics’: questions of the ownership of results on and through engines and platforms. Essentially, when results are returned, who is responsible for them? Has the user, by rendering specific input through their query and informing the platform through previous user history, any ownership of the results? Within this literature is a clear focus on personalisation as “end-user co-authorship” and the impact of this configuration of user to platform. On the one hand, there is an ‘invisible algorithmic editing of the web’ (Pariser, 2011) while, on the other, the engine and platform become ‘inculpable’ as only partial producer of results (Rogers 2009). This inculpability due to personalisation is disputed by several research projects on the impact of personalisation including Feuz et al. (2011) and Puschmann (2019).

In the context of the search for information around stem cell treatments, inculpability due to personalisation might suggest that the risk of search (the risk identified as exposure to potentially dangerous or exploitative treatments) is held by the user, not the platform. Through this distancing, the responsibility of accurate and credible information becomes nebulous and separate from the goings-on of the platform. What would it mean if harmful or misleading medical information is not the responsibility of a platform to moderate? In this situation, the user – despite lacking access to the algorithmic decision making that mediates query result generation – assumes responsibility for a technology’s output. If as Fuez et al. (2011) and Puschmann (2019) argue, opaque intervention in results by platforms themselves should negate the influence of a user’s query in co-producing results. In Chapter 7 through my data analysis, I engage with this level of analysis through the discussion
of a single results page for stem cell treatment queries and touch on these questions of culpability.

Since 2010, there has been a shift from this epistemological critique towards a critique based on representation, which has further underlined the importance of critically engaging with algorithms and platforms directly, rather than only the user (Rogers 2015). Well-documented cases of platforms and online services actively or passively engaging in discriminatory practices shed light on the impact of bias on user results and interactions. These include racial discrimination in online ad delivery (Sweeney 2013), prejudice within search autosuggestions (Umoja Noble 2018), and price discrimination based on browser (Hannek 2016). Through these examples we see a broader question arise of whether machine learning inherently reproduces societal bias and issues of censorship (O’Neill 2017). These findings push back on the idea of co-authored results and underline the importance of the platform itself in the ebb and flow of information and subsequent mediation of knowledge.

In terms of health information on digital platforms, Astrid Mager’s work on the sociotechnical practices of users and providers of health information online has been particularly informative in providing a theoretical framework for approaching this project (2009, 2015). To characterise platform impact, Mager deploys the idea of algorithmic ideology - or the inherent values and characteristics promoted through the processes deployed on algorithmically-mediated platforms. A major argument within her work is that the accessing of health information online has been established as largely informed by a user’s normal patterns of web navigation (Adams et al., 2006). These patterns may include daily usage, case usage, and behaviour like most visited sites. This suggests that the study of “good” or “bad” sources of health information is less helpful as a research route than looking at the practices that inform the platforms users are most likely to use.

While Mager’s work touches on the impact of Google Search as a mediator, it remains grounded in navigating the relationships between users and information providers through interviews and user-end processes. It does not – or, perhaps, cannot – directly interrogate the platform. It is helpful here to note that many recent works specifically on search struggle with this same limitation that an opaque
proprietary system like Google Search presents. For example, Safiya Umoja Noble’s *Algorithms of Oppression* (2018) only included self-auditing as a methodological approach to data collection. This is helpful to uncover general examples of racism as presented through search returns. However, this approach does not, and cannot, account for more granular differences in results between users that are inherent to an algorithmically-mediated and user-specifying platform like Google Search. These works wrestle with *what* is returned rather than *how* it is returned.

Informed by this body of STS and critical algorithm studies literature, my project shifts focus to the platform itself as a mediator of health care information. By adopting a methodological approach based on ‘Blackbox analysis’, I am able to turn the lens away from the impact on the user and towards the platform by interrogating the actual platform mechanisms at play. Although I do touch on categorisations of marketing of proven and unproven treatments with my data, I am not as much interested in the “goodness” or “badness” of these returns, but rather how they move through Google Search and how different platforms mechanisms shape the final results provided to a user.

### 3.2 A BRIEF HISTORY OF GOOGLE SEARCH – THE UBIQUITOUS AND THE INVISIBLE

My thesis builds on Matthew Hindman’s work\(^\text{20}\) in arguing that Google Search occupies a central position within the study of knowledge online due to two critical characteristics: its ubiquity and its relative invisibility (Hindman et al., 2003, Hindman, 2008, Hindman 2018). In this section, I discuss how Google Search’s proprietary algorithm and market offering has led to this unique and under-interrogated position within discussions of knowledge formation and moulding in digital spaces. Although I would like to make space for future research on the tension between the accessibility and availability of material mediated through Google Search, here I am focusing on Google Search’s mechanisms of mediation. I then discuss how Google’s ubiquity

---

\(^\text{20}\) Hindman’s work centres on Google Search and digital democracy with a particular focus on political monopolisation in digital spaces around political campaigning and democratic institutions.
and invisibility contribute to its importance as an actor within the context of health knowledge mediation online.

As explored in Jonathan Zittrain’s *The Future of the Internet – and How To Stop It*, digital infrastructure and product design has coalesced from the rhizomal, open development era of the 1990s to the monolithic platform funnels of the early 2000s (2008). This, Zittrain argues, leads to a fundamental transformation in the ethos of an interconnected internet, away from open innovation and power dispersal and towards rigid technological monopolies. This is echoed in Robin Mansell *The Information Society* (2009) and Shoshanna Zuboff’s *Surveillance Capitalism* (2019); work in which they argue for more attention to be paid to the influence of commercial entities that have increasingly come to monopolise, monetise and control the digital knowledge economy.

Within this shift from iterative and generative networks, Google Search has both manoeuvred and been manoeuvred into a unique, outsized position as conductor and vehicle for content discovery and flow in the global internet. Bearing Zittrain’s critique in mind, it is helpful to track the circumstances that have contributed to Google’s rise.

Six years before Sergey Brin and Larry Page enshrined their unofficial motto, “*Don’t Be Evil*”, into their corporate code of conduct, Google Search’s 1998 entrance into the market began a process of change in the way nodes of knowledge were ranked within search. During Google’s formative years, the computing landscape was dominated by Microsoft and Apple. Both companies split the market on development of software and hardware that would underpin the capabilities of personal computing. While both embraced different ‘points of entry’ for creating systems that interacted with the development of the world wide web, their focus remained largely on the means of accessing the internet rather than the what – or governing protocols and logics – of the internet (Arthur, 2014). Search engines emerged from this economic tug of war as small-scale web-based products developed as either profit-neutral add-

---

21 As defined by World Wide Web Consortiums definition: “The World Wide Web (WWW, or simply Web) is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI)” (W3C 2004).
ons to existing web technology companies or as standalone products that adhered to a profit model that relied on on-page advertisers.

 Whereas early subject-specific search engines such as Lexus-Nexus sought to recreate the conventions of scientific scholarship, proprietary search engines developed with three general principles. These new search engines were often characterised by a few common traits: the de-professionalization of knowledge, the generation of results based on understanding user-submitted queries, and adherence to private economic models of profit (Seymour, Frantsvog, and Kumar 2011). Applying Astrid Mager’s algorithmic ideology to these common traits suggests that the former characteristics (de-professionalisation and user-interaction mining) serve the latter (economic extraction mechanisms). Indeed, my findings in Chapter 7 support this hierarchy of principles.

 Before 1994, the first search engines – like Excite! – relied on hand-curated databases of web content (Spink and Jansen 2004). Practically, this meant that a search engine captured websites based on basic information which was then indexed and prepared to be displayed when triggered by user queries. The growth of these indexes quickly made clear the need for automated indexing to handle the growing resources offered on the world wide web. This was addressed throughout the mid-to-late 1990s by developments in indexing software. During this time, search engine algorithms ran on a basic approach to data rankings – namely one-to-one query term matching amongst equally weighted indexed pages. Engines like Altavista, Yahoo, Ask Jeeves, and Netscape were programmed to crawl available URLs, scrape meta-data, and create proprietary indexes of pages available on the web. Importantly, resulting query functions treated all page indexes as equal and valued direct query term matching. Any page – regardless of origin or content – would be presented to a user if it contained the most references to the query term. Competition, therefore, was focused on the size of the index and the speed at which a search engine could return results to a term-specific query.

 Additionally, the integration of focused advertising in the form of graphical banner ads impacted results. These advertisement systems were based on traditional print models of wide-net advertising. Regardless of query term, advertisers would buy
exposure (e.g. appearance in a pre-defined advertisement space on the homepage or search result pages). Search result pages would pull simple image and hyperlink ads from the advertising database to be shown alongside query results. This basic model would allow the search engines to return to clients the metrics of advertising exposure, but did not include user targeting (Van Couvering 2008).

However, this search approach produced some critical drawbacks. Results soon became cluttered with spam created by web pages which, recognizing popular search terms, would include additional, invisible text within the html of their sites. This index manipulation was largely a pursuit on the side of content producers of web pages to maximise hits to their page and hide content (for example, pornographic content was often masked by unrelated text). This invisible content-unrelated but search-relevant text would be picked up by the search engine’s crawler and entered their indexes without prejudice. Speed of results also suffered as the software incongruity between the webpage indexes and advertising databases slowed down results on the user’s end. These issues created result pages that were both irrelevant and cumbersome.

Google sought to address these problems by changing the way weighted value was assigned to pages. Rather than focusing on one-to-one query term matching, Google developed a link analysis algorithm known as PageRank (named after Google founder, Larry Page) that proportionally weighted pages within Google’s index based on two factors: the probability of a user clicking on a link and the number and value of ‘in-links’ (Brin and Page, 1998).

It is helpful here to borrow from digital methods the concept of nodes and edges as a way to visualise web structures and relationships (Rogers 2011). In their simplest form, nodes are simply points on a graph – in the case of ranking search, web domains or individual web pages. Edges are links between these pages that form routes of connectivity and allow for the analysis of node influence. If a node (or webpage) has a large number of other websites that link to it, it is determined to be

---

In-links are hyperlinks or references to a specific webpage or domain from other websites. For example, a link from a BBC article to an NHS webpage would be considered a high-value ‘in-link’ for the NHS domain.
important, a hub of relevant information as demonstrated by the high number of ‘in-links’. There is no inherent value placed on the information of nodes; rather considering nodes and edges allows for the mapping of both relationships and relative importance within a digital ecosystem.

Within PageRank, webpages function as nodes and hyperlinks as edges with recursivity in-built to the relative ranking. For example, nodes with links from other pages with high PageRank coefficients would have their coefficients comparatively increased. This created a weighted network within Google’s search index in which value is derived from relative embeddedness within a content-creation network. PageRank also assumes a user’s role within the algorithm based on the theoretical value created by a user clicking through to a link.

Although Google introduced several distinct characteristics of search, targeted advertising and the introduction of bespoke page valuation enabled Google Search to dominate web search by the 2010s. This bespoke page valuation included the structure of both organic results and in-result page additional feature like advertising, promoted content, and additional search topic-specific information to illustrate query results.

Here it is helpful to introduce the basic computational concept of “Probably Approximately Correct” (PAC) learning. PAC learning is a principle in machine learning that supports what algorithms can learn through interactions with user input. For my purposes, the most important characteristic is that PAC learning-driven interactions provide positive success feedback for answers that are deemed probably approximately correct, rather than concretely correct and evidenced (Haussler 1990) Similarly, Google’s self-reporting on PageRank suggest that results are valued and returned based on the potential for “correctness” measured by user interaction (Seymour 2011). The result is that a search query return is considered a “success” if the results generated by the user query results in a click through. The roll-out of PageRank as a driving mechanism behind the production of search results led to the creation of early client-side Search Engine Optimization (SEO).
Early manipulation of PageRank through practices like ‘Google-bombing’\(^{23}\) (Lev-On 2008) exposed the techniques through which results could be manipulated. By optimising web pages for integration within the Google Search valuation algorithm, website owners can increase the prominence of their URLs. This has, in turn, created its own economic market for SEO services from third-party vendors.

Google Search also introduced a change in the inclusion of advertisement as a profit model within search results signalling the beginnings of advanced Search Engine Marketing (SEM). As discussed, early search engines relied on models of advertisement profit generation that avoided the targeting of users and opted for simple, untargeted exposure. Building upon the concepts developed by Bill Gross at Idealab in the early 2000s, Google developed Google Ads (previously named AdWords) – a system through which advertisers could be connected to specific search results (Van Couvering 2008). This project created an economic mechanism through which advertisers could ‘bid’ for their ads to be included on the search results pages for specific search terms relevant to their product.

Finally, the introduction of search personalisation in 2008 shifted the production of ranking and ad targeting towards a co-production model of search returns. By combining the data analytics culled by users’ Google accounts as well as IP-specific information such as geolocation and previous searches, Google Search began to return unique results. The extent of this personalisation is a difficult question to pin down without Google’s direct disclosure. However, as will be discussed later, there are digital methods-focused studies that have attempted to better understand personalisation’s impact. For my purposes, personalisation is of note due to the presumed ‘co-production’ of results between the user and algorithmic engine and its potential for targeted advertising.

It is important to note that since its introduction, PageRank and Google Ads are now among over an estimated 250 mechanisms within Google Search’s proprietary black

\(^{23}\) ‘Google bombing’ is a collective action to manipulate Google Search results related to a specific term originally achieved through the handling of hyperlinks to skew results towards specific returns (Tatum 2005). Many incidents of Google Bombing were intended as tongue-in-cheek social commentary or protest (e.g. the Google Bombing of Rick Santorum by Dan Savage and readers [Gillespie 2017]).
box that determine results (Arthur 2014). However, their introduction changed how value is assigned within digital search platforms as it integrated more complex and sophisticated means to determine relevance. The engine’s market domination establishes the stakes for our analysis of its current contribution to knowledge mediation in digital spaces. Because the Google Search platform is a closed proprietary platform, it is not possible for me to observe and analyse each of these mechanisms. Rather, I have given privilege to the filtering mechanisms that can be identified through what is observable by the public through appearing in search results pages throughout my research; those that are observable to the public and those that appear on search results page.

Before moving on, I would like to note one other subtle development in web architecture that will become important in my later analysis. Within the web publishing standards world and corollary to the continued development of Google Search, the Schema.org project began solidifying web standards for the indexing of web pages and resources across shared digital spaces (Guha et. al 2016). Like Jonathan Zittrain’s characterisation of early cooperative internet development, Schema.org’s ethos reflects an early profit-neutral fraternity of developers committed to creating shared vocabularies for the attribution of web pages and resources. It describes itself as “a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond” (www.schema.org/about, accessed by author 3.12.2021). The project was founded in 2015 by Google, Microsoft, Yahoo and Yandex through a community committee at the major international standards for the World Wide Web organisation, the W3C (Schema.org 2022).

These standards form the basis for web publisher-side attribution markup of web pages for integration into platforms like those created by Google and Microsoft. Put simply, these standards provide a common vocabulary and hierarchy for machine readability to identify pages as having specific identifiers (‘What is type of page is this?’) and characteristics (‘to whom does this page belong?, ‘when was this published?’; ‘what is the associated URL?’). This open standards development is important to keep in mind later on in my data findings and analysis.
3.2.1 HOW WIDESPREAD IS GOOGLE SEARCH AND HOW IS IT USED?

Google Search’s impact on global web behaviour and activity cannot be understated. The platform occupies a unique position of near monopoly in web navigation and engagement not only through the search platform, but also the additional platforms acquired by the corporation.

As of September 2018, Google controlled 72.37 percent of the worldwide search engine market share with over 3.5 billion searches per day (Netmarketshare, 2018). In 2017, £73.2 billion of Google’s £85.1 billion profits were generated from advertisement (Statistica 2018). With the 2006 acquisition of YouTube, Google and its underlying algorithmic valuation infrastructure makes up an estimated 1 out of every 6 minutes spent online worldwide (Pew Internet Research 2017).

A 2012 study of (Purcell, Brenner, Rainie 2012) explored the growing role of search engines on the behaviour of users looking for resources and web navigation online. Purcell, Brenner, and Rainie found that:

A February 2012 Pew Internet survey finds that 91% of online adults use search engines to find information on the web, up from 84% in June 2004, the last time we did an extended battery of survey questions about people’s search engine use. On any given day online, 59% of those using the Internet use search engines. In 2004 that figure stood at just 30% of internet users. (Purcell, Brenner, Rainie 2012)

This dominance translates into real world influence notably through the integration of Google Search into individual decision-making processes.

When it comes to where Americans place their trust as they gather information before making an important decision, a big majority (81%) say they rely a lot on their own research – many more than say they rely a lot on friends and family (43%) or professional experts (31%), according to a 2018
Pew Research Center survey. Some 15% also say they rely on their own research “a little” as they make major decisions. (Turner and Rainie 2020)

When respondents were queried around the definition and methodology of conducting one’s “own research”, over 46% of respondents cited using “digital tools”. This is a significant number more than relied on asking friends or family for advice (25%) (Turner and Rainie 2020). These studies portray Google Search as an outsized influence within the process of individual decision-making superseding even existing modes of information gathering and assessment like real world contact consultation.

Specific to health and information online, a 2006 study found that 80% of American adults have engaged search platforms for information around health issues. 97% of those searches were focused on specific conditions and diseases (McMullan 2006). Hay et. al found that patients with chronic illnesses are more likely to search for information around their conditions than non-affected users. Individuals diagnosed with Multiple Sclerosis were found to search twice as much for information around their disease than other chronically-ill patients (Hay et al., 2008).

This trend in user practice is critically important to my research as it underlines the day-to-day importance of Google Search in healthcare information gathering. Moreover, it is indicative that those who would want to access patients – whether for commercial or non-commercial purposes – recognise the importance of search platforms as points of patient access. This motivation and opportunity for patient engagement – as will be explored in my results in Chapters 5, 6, and 7 – is uniquely enabled by the platform of Google Search itself.

3.3 SEARCH ENGINES AS EPISTEMOLOGICAL MACHINES

As discussed earlier, digital methods scholars who use digital methods tend to focus on the nodes (websites) and edges (links) as a way to navigate and track the rendering of information within digital spaces. Search engines as mediums through which users are exposed to these nodes and edges make for trickier subject matter due to the opaque yet pervasive nature of their underlying functional and
infrastructure. In this section, I outline how search engines can be difficult and elusive research subjects, while underscoring the importance of turning a critical STS gaze on their potential impact on the mediation of knowledge in digital spaces.

Existing work on the active production of knowledge and information explores digital knowledge production as intricately linked to the form and function of the platforms on which they are produced. This includes work on the impact of mechanisms of personalisation on information search (Feuz, Fuller, and Stalder 2011), the influence of the platform Google Scholar on academic knowledge production (Van Dijck 2010), and the impact of filter bubbles on user-specific inclusion and exclusion within result rankings (Pariser 2012). These studies have informed my own research as they identify material mechanisms of platform influence that I looked for when analysing the data I scraped from Google Search. Indeed, identifying these mechanisms is a major contribution of this research to both digital methods and digital health studies.

Search engines can be thought of as “epistemological machines” as they organise, rank, and serve particularly rendered choices for a user to navigate (Manovich 2013). Search engines, therefore, provide ongoing maintenance to the digital rendering of knowledge by perfecting the practice of including relevant sources within its rankings and improving user’s trust in the rank (as indicated by users not looking further for relevant nodes) (Rogers 2013). This is a dance done on a scale beyond the individual user. If one thinks of the search engine as a machine whose purpose to create ‘click-throughs’ to remain relevant and be used, then it must always be reorienting itself based on the perceived wants (determined through collective ‘click-throughs’) of its user base. The Probably Approximately Correct answers are both reinforced through user choice and shifted in response to collective user search behaviour.

This dynamic relationship is addressed in the field of political economy through the concept: ‘Googlization’ (Vaidhyanathan 2011). This term encapsulates two key characteristics: the making available of a free service in return for a human-generated data profile economic model. The increase of user data as a source for content mediation in the case of recommendations based on previous user action. The second can be seen clearly in the integration of “most emailed” or – as the New
York Times website employs – “Recommended for You” features that organise and promote nodes of content based on user activity, either as a group or individually.

The importance of studying a search engine like Google can be traced to two key characteristics that inform this research project: its ubiquity as public-turned-private digital commons and the invisibility of its economic model. In 1996, critical media researcher Herbert Schiller pre-empted a core issue faced by researchers tackling issues of information transfer and knowledge creation in digital spaces (1996). In his analysis of government data released to the private sector in the early 1990s, Schiller argued that the democratically-protected data as the underpinning of the digital public commons became critically endangered. Corporate users are now able to purchase previously publicly held information to create proprietary knowledge demonstrating a shift from the democratising instinct of Web 1.0 development (Coleman and Golub, 2008). Schiller (1996) writes:

> The practice of selling government (or any) information serves the corporate user well. Ordinarily individual users go to the end of the dissemination queue. Profoundly antidemocratic in its effect, privatising and/or selling information, which at one time was considered public property, has become a standard practice in recent years.

Applied specifically to digital health and patients online, Deborah Lupton made a similar argument in her work on the commodification of patient opinion. She notes that for patients digital technologies have both offered new platforms as well as new challenges. She writes:

> …via these online platforms patients' opinions and experiences may be expressed in more diverse and accessible forums than ever before, but simultaneously they have become exploited in novel ways. (Lupton 2014)

Fundamental to my purposes, Schiller identified that for non-corporate users of digital platforms, the privatisation and commercialisation of information has been normalised to the point of being hidden (Schiller 1996). As Greg Elmer (2013) notes:
There would be no Google search engine … without the content, information, and demographic profiles uploaded, revised, updated, and shared by billions of users worldwide.

This begs a question that is central to my research – how much of Google is a data collection platform, how much is neutral, and how much is impacted by profit extracting activity like advertising? This appreciation for the mechanisms of power that mediate information transfer and exchange on digital platforms made invisible underpins my research. By investigating mechanisms like search engine marketing and content guidelines, my research seeks to make their influence visible as part of a larger question of the logic governing search results in relation to contested medical treatments. Moreover, what is the relationship of these mechanisms to what Soshanna Zuboff calls “surveillance capitalism” or the enabling of extractive commercial activity through user’s ongoing interaction with and data monitoring by domestic technologies (2019).

At the same time that invisible processes have become the norm in digital spaces, user traffic consolidation around a few, powerful epistemic machines complicate the notion of digital knowledge dissemination as a neutral web. The idea of a digital rhizome – first introduced by critical theorists Deleuze and Guattari (1988) – has been appropriated and used to conceptualise the internet a rootless network with “limitless expansion, random intersecting points, and abilities of rupture and re-growth”. In this conceptualization, the early internet was a system of networks characterised by “non-hierarchical systems of de-territorialized lines that connect with other lines in random, unregulated relationships” (Best and Kellner 1991).

These analyses were tied up in discussion of the democratising nature of the internet. Howard Rheingold’s influential concept of the internet as a means for reconstituting community gave rise to a swath of scholarship on the role of the internet in influencing and enabling democratic transfer of information and mobilisation (Ivie 2015; Owens and Palmer 2003). This also included discussion of the democratisation of healthcare decision making such as Hardey’s ‘Doctor in the House’ (1999) and McDorman’s work on digital communication within right-to-die advocacy communities (2005). However, the idea of the internet as democratic
distributor of knowledge is complicated by the introduction of search engines and platforms as a primary means for navigation.

Pushback from researchers like Hess suggest that search engines structurally move users into “marketing niches rather than assist in the development of new knowledge” (Hess 2008). The question of why – amidst the glut of search engine platforms in the late 1990s – Google Search became the most used epistemic machine for navigating indexes of information online is discussed in the next section.

It is helpful to briefly look into how nodes and edges are constituted within digital methods research. The discussion of “web epistemology” outside of search engines generally falls within two camps when discussing nodes and edges as subjects: hypertext literary theory and social network theory. Both approaches provide valuable routes to analysing relationships and movement of information in digital spaces. Hypertext literary theory (Zuern 1999, Morita and Fujita 2002) suggests that hyperlinks form distinct pathways for users to navigate through text. Here content flows to content directly to form modes of understanding the queried subjects; a hyperlink neutrally moves a user as if from thought to thought.

Social network theory (Glückler, Lazega, and Hammer 2017, Baldwin – Philippi 2007) grants more weight to the directionality of ties between nodes. The identification of “highly between” actors who operate as hubs between actors due to inbound and outbound links allow for an analysis of the flow and constitution of knowledge. For social network theory, the “politics of association” exert pressure on the navigation of a user through information on the web (Rogers 2014). These approaches, however, fall short when applied to the question of the influence of search engines. For my purposes, how can we study these relationships of users navigating collected knowledge when the associations and presentation of results have already been subject to an unseen process of rendering value?

Despite the clear significance of search engines in social scientific analysis of knowledge mediation online, search engines present practical methodological challenges due to their proprietary nature and their economic model based on free-service-for-user-data, which requires them to operate as black boxes. Even the
gathering of results for analysis can be challenging. As of 2009, digital methods tools that employ automated query submission and scraping of results have been expressly targeted by Google as violations of the search engine’s terms of service (Rogers 2013). Employing tools that scrape Google results like IssueCrawler or the Lippmanian Device (both software that scrapes Google platforms for analysis) do not replicate the rankings conferred by Google or allow one to interrogate results, but rather recreate the basic direct links between nodes. If you combine scraped results, you can analyse networks and connections between nodes and actors, but you cannot render these connections in the same way Google does. This seems to have resulted in a focus within digital methods research on search engines as secondary subjects. Most tools focus on the movement and growth of subjects over time or in relation to other nodes and actors rather than as part of the results product created by Google Search engine.

3.4 MECHANISMS IN GOOGLE SEARCH: DETERMINING THE F IN $F(\text{QUERY})=\text{RESULTS}$

Within the study of computer science and data retrieval, the most basic conceptual principal of any search algorithm is deceptively simple:

$$F(\text{Query})=\text{Results}$$

In the case of Google search, this implies that a user-generated query – most often search terms or questions – are entered and return results (in this case, web links) that are most relevant to the original user’s query. Unfortunately, identifying the $F$ is not straightforward. In fact, it may be the most difficult question within this research due to on-going technological changes and the opacity of proprietary black boxes. However, naming the known$^{24}$ mechanisms of intervention - even if only to note their presence within the computation of results returned - is fundamental to this platform analysis.

---

$^{24}$ By material necessity, I must rely on mechanisms that have been publicly acknowledged by Google (Google 2022).
As established, there are unknown numbers of influences on decision-making within the Google Search proprietary platform. This section will provide context for understanding Google Search as a platform and outline a few mediating mechanisms - both within and without the platform - that I have identified as crucial contributors to the creation of search results. It is important to note that this list is indicative rather than exhaustive and based on the public documentation of Google Search by Google itself (Google 2022). These are iterative processes continually in dynamic flux as Google’s underlying platform is updated regularly.

I. **Indexed Library of Uniform Resource Identifiers (URIs)/Universal Resource Locator (URLs)**

The Index Library of URIs/URLs mediates which web pages and sources are captured for potential inclusion in search results. A common misconception is that, when a query is entered, Google Search searches for results in real time across the web. Rather, before a query has even been entered, Google has developed a structured library of billions of available web pages and information associated with them. While search results will take the form of URIs or URL web links, to establish relevance, Google first indexes possible results using automatic bots known as Spiders or Crawlers to roam URLs to scrape relevant data.

II. **Query Analysis**

Query analysis represents the meeting point between human user and machine through the capacity of a search engine to assess, contextualise, and predict the meaning behind user-input. It determines how queries are understood by the search platform. While Google provides support for Boolean search using advanced symbols (Frants et. al 1999), it also includes Natural Language Processing (NLP) to integrate informal query structure into its analysis (Google Research 2022).

III. **Algorithmic Valuation**

Perhaps the most impenetrable mechanism, algorithmic valuation determines the weight and importance of web page resources and sources within the indexed library which impacts both probability of display and rank within the
search result page. PageRank is an example of the pre-determined networks of value placed on URLs before a query is entered.

IV. Personalisation
Personalisation is the tailoring of search results based on the profile of the user engaging in search. It targets search and advertisement targeting based on geographical location, language settings (both in query terms and based on a user’s IP address) and cookies (saved locally and through Google accounts).

V. Search Engine Optimisation (SEO)
SEO is a core business pursuit within Google centred on the optimisation of web page creation and publishing to align to changing weight and valuation mechanisms within the Google Search platform (Think with Google, 2019). This includes client- or web publisher- side interventions including standardised markup and publishing best practices to integrate into each stage of Google Search index and delivery from web crawl to return. Google promotes SEO adherence itself through its secondary platforms including Google Analytics, Google Business, and Google Retail.

VI. Search Engine Marketing (SEM)
SEM is another core business arm of Google which allows for the management, monetisation, and promotion of web pages and resources through Google Search (Google Developers 2022). This is the most direct monetary intervention in Google Search results which creates an internal platform market for the buying (publisher side) and selling (Google Ads) of space within search results.

VII. Knowledge Graph
Google Knowledge Graph (Figure 1) is a secondary intervention within the architecture of Google Search results page that provides limited factual content creation culled from a primary Google-owned database. Knowledge Graph relies on both human-audited web sources deemed by Google to be trustworthy, and sources indexed through Schema.org (Google Support,
FIGURE 1: SNAPSHOT OF GOOGLE SEARCH FOR THE QUERY “STEM CELL” FEATURING ‘KNOWLEDGE GRAPH’ RETURNS, CAPTURED BY AUTHOR 12.4.2022.

In addition to these constructive mechanisms for the creation of results, there are multiple mechanisms that place limitations on search results. Contrary to the previous examples, these interventions bar inclusion for some web page returns based on company or regulatory-driven limitations.

I. Platform Policy Restrictions

Google platforms (including Search, Ads, and Knowledge Graph) have adopted a policy of preclusion of a number of restricted search terms that are publicly available. At its broadest level, Google Policy excludes the following types of material: child sexual abuse imagery, highly personal information, valid legal requests, spam and malware, webmaster requests, potentially offensive content, and explicit content (Google Support 2022). This content is both algorithmically excluded through internal processes and manually excluded through user flagging.

II. Human review

Google’s human review of results process may play a factor in the weight of search results (Google Support 2022). The use of human review of results is
connected to the original company target of click-conversions (or the likelihood that an individual searcher will click through to a result listed on the first page of results). However, there are currently no primary documents on what criteria human reviewers have employed when reviewing results.

Interesting here, Google also solicits responses from the general public in regard to illegal or inappropriate material returned through Google Search (https://support.google.com/legal). These forms submit questionnaires through which users can report content that is returned through Google Search. Users are prompted to select from a list of “reasons for removal” (see Figure 2). They include:

- “I would like to report malware, phishing or similar issues
- A piece of content I am concerned about has already been removed by the webmaster but still appears among the search results
- Remove personal information from Google under product policies (personally identifiable information, doxxing, non consensual explicit imagery, etc)
- Legal personal information issue: request removal of my personal information from Google’s search results
- Intellectual property issue: report copyright infringement, circumvention, etc.
- Other legal issue: report content for a legal reason not already listed” (Removing Content from Google, 2022)
III. Query-governed Interventions (e.g. Samaritans)

Some search terms result in the return of a query-specific intervention based on the content of the search query. This includes content related to the topics of suicide, self-harm, disordered eating, and domestic abuse. In the case of queries related to suicide, Google Search has created a specific widget that presents the user with a local suicide prevention contact number based on a user’s IP address.

Astrid Mager suggests that these mechanisms enable “war-like” dynamics between content producers and the platform. The push and pull of content producer, user, and platform developer upon these mechanisms contributes to the overall undertaking of the platform as it renders results. She suggests that this mandate to update is predicated on the adoption of capitalist reliance on growth; without ever increasing numbers of users and user traffic, this system falters. This mandate impacts not just the platform developers, but commercial content creators as well. In an effort to maintain or raise the value of their digital domain, website providers...
develop strategies targeted at increasing user attention rates that specifically engage with the search platform’s mechanisms for ranking. By doing so, they aim to appear more frequently and in a higher position with search engine results, thus potentially diluting or skewing the relevance of returned results. The cycle continues as the platform is then forced to find new ways to prioritise relevance to users, rather than flooding results pages with well-integrated but unwanted commercial content (Mager 2012). Interestingly, Mager points out that in the area of health, this cycle has a detrimental impact on smaller, non-profit websites (Mager 2010).

In my initial readings around algorithmic accountability, I found my position vis-à-vis this particularly combative conceptualisation of relations on a platform like Google Search difficult to sustain. I originally considered the platform to be less active a participant in the commercial activities in digital health spaces. To me, the content creators – particularly those offering direct-to-consumer offerings – displayed insistent behaviour towards users (through, for example, SEO manipulation). However, as I began to observe the mechanisms at play in Google Search results first hand, I saw clearly evidence supporting Mager’s point that the platform itself is maintained through aggressive commercialisation. As will be explored in Chapter 7 and 8, mechanisms for commercial interactions persist throughout results pages for searches around stem cell treatments, whether through direct marketing or secondary commercial activities.

Moreover, the “tweaking” of the algorithm and subsequent reaction by content producers indicated a power relationship determined and led primarily by the platform. For my research, this made more clear the importance of focusing on the activities and interventions of the platform as the determining factors of how information is digitally mediated on Google Search. These are the mechanisms - both platform and policy based - that I will bring into conversation as they converge at the point of search returns. What does this deliberate content intervention look like? As Google Search is an iterative, dynamic platform and as this research demonstrates, it is difficult to track minute changes in the presentation of results. However, the process is briefly visible when it goes wrong or fails.

3.4.1 GOOGLE SEARCH, NEUTRALITY, AND TRUST
In 2016, Guardian writer Carole Cadwalladr raised alarm that Google Searches for the query “Did the Holocaust happen?” returned top results from Neo-Nazi website Stormfront.org (Cadwalladr, 2016). The content included Holocaust denial material refuting the evidence for the mass genocide of Jewish and other minority groups during the Second World War. These anti-Semitic results were documented as being displayed to users located in both the United States and United Kingdom. Google became the target of public outcry against this spread of malicious information through the search platform.

Following the article’s publication, Google initially responded that it would not remove the link, issuing the statement that:

… tinkering with results by human intervention was not in line with Google's belief in the neutrality and objectivity of its engine. Editing search results would call into question the impartiality of the products of its supposedly unbiased algorithms and thus harm company values. (Kerssens 2017).

This response caused significant public outcry leading to Google Search manually removing what it classified as “non-authoritative information” two weeks later (Peyser 2016).

Throughout this intervention, an important structural message was broadcast from Google’s corporate structure: intervention into the Google Search results was not and would not be carried out by human programmer directly excluding or including specific results. Rather, Google stressed that any intervention had to be made through reprogramming the platform’s mediating algorithms. This is a crucial policy distinction that dispels the possibility of human oversight or intervention in favour of technology-determined inclusion and exclusion.

This moment of intervention suggests two noteworthy discoveries. First, that when faced with a mandate to intervene, Google validates only technological intervention while limiting human-initiated interventions. This has interesting implications for my discussions of interventions in Chapter 7 and 8.
Secondly, this moment suggests that the adopted strategy is part of a larger, deliberate tactic of promoting Google Search as a positivist technology. Positivism here is the suggestion that algorithmically-mediated processes and logics, beyond the influence of human input, operate neutrally and reflect an objective reality (Reeves and Ney 1992).

Instead of delving into a critique of positivism (which is well-trodden territory and arguably the founding conceit of STS25), I would like to instead take seriously the strategy Google has adopted in positioning the Google Search platform vis-à-vis neutrality. What assumptions are made in the conscious position of Google Search as a positivist technology? What is the subsequent role of neutrality on Google Search? What happens to our understanding of Google’s strategy of rendering results when we take seriously a commitment to perceived positivism?

In his work on digital media and communication through algorithmically-mediated platforms, Niels Kerssens of Utrecht University applies technological positivism to specifically search platforms thusly:

Algorithmic search is entangled with a positivist ideology biased towards the assumption that neutrality can only be provided when search is performed by computational processes while shielded from human agencies. (Kerssens 2017)

I would like to push these assumptions further as the basis for future critique in Chapter 7 and 8. To address the questions I’ve posed, I have parsed a few assumptions that Google must adhere to in its strategy for Google Search. It is crucial to note here: these are not assumptions I am arguing are true but rather theoretical assumptions that are made by Google through its governance and development of Google Search. I would summarise the assumptions made by Google as:

- Algorithmic decision-making is neutral by nature.

25 See Fuller 2013, Reeves and Ney 1992, Feenberg 2017 to name but a few.
- There is no difference in the practice of enquiry across search.
- Search is separate from both opinion and common sense.
- Google searches are value-neutral and can only be judged by the algorithmically-mediated logic it applies.

Kerssens’ astute observation has implications for the types of information and knowledge Google Search itself creates. Astrid Mager argues that this adoption of machine-as-neutral positivist ideology is not naïve, but rather a conscious choice made to position Google as objective through mechanisation (2012).

Kerssens takes the argument further to interrogate the design of user interfaces adopted by search engines as part of this strategy. Menu interfaces provide limited input options for users, normally a simple query box which returns a page of results. The Google landing page is an example of one such menu interface. Database filtered interfaces allow for the application of filters to shift results according to user interests. He argues that menu interfaces – rather than database-filtered interfaces for search results – encouraged this positivist assumption of algorithmic search through two fundamental changes.

The first is the dissociating of what Kerssens calls the “democratic service function at the front-end” from the “editorial function in the back-end” (Kerssens 2017). The “democratic service” rendered through user-friendly access menus on results pages are rendered. The editorial functionality is then shifted, exclusively, to the algorithmically-mediated backend. In doing so, the menu interface becomes what Kerssens characterises as a “façade”; the illusion of interaction with pre-vetted and designated material produced by the now un-interrogatable backend.

The second change is the discouraging of intellectual labour input in curation of search results. Kerssens points to the work of librarians as human input into the iterative process of information curation. As seen in the Stormfront example, Google has made a conscious effort to distance human-intervention (whether lay or expert) from the practice of Google Search. Kerssens’ interrogations are particularly useful for my purposes as they centralise the front-end architecture of search platforms as specifically transformative.
I use this privileging of visible mechanisms of influence as a central focus in my research. Functionally, this means that in the setup of my data scraping, I look for visible clues within public-facing results of the mechanisms at work when search results are rendered. This underlines the importance of what is made available to end users (through either the structure of the results page which includes both advertisement, organic results, and other interventions).

Google’s strategy of publicly promoting algorithmic positivism has increased public trust in the platform. The more than algorithmically-mediate platforms are seen to be adapting, growing, improving, the more their handle on the information and knowledge they handle appears to be more intelligent and reliable for users. Alexander Halavais submits that search engines “have become an object of faith…trusted more readily than they ought to be” (Halavais 2009). This is supported by research on user trust vis-à-vis search engines and the information they return. A 2012 Pew charitable trust study found that of American search engine users, around “two-thirds of searchers (66%) say search engines are a fair and unbiased source of information” (Purcell, Brenner, Rainie 2012).

The public positioning of Google Search as an algorithmically-mediated positivist technology and the trust that strategy imbues on public-platform dynamics has implications throughout my research. In Chapter 5, 6, and 7, I test this neutrality through my data collection. This leads to a critique of these stated assumptions and their impact in my conclusion in Chapter 8.

3.5 GOOGLE POLICIES OF INTERVENTION AND HEALTH RISKS

Within Google’s policy documentation, particular focus is paid to the area of health information and the potential for risk exposure through the act of search. This focus underlines an ideological acknowledgement of both the risk presented by and the potential for exposure to health endangering information on Google’s platforms. Specific to medical information, Google policy sets out an explicit statement worthy of analysis. In regard to medical information returned by Google Search across platform mechanisms, Google policy dictates:
When we highlight information on medical topics across Search, we strive to show information that reflects scientific consensus and evidence-based best practices, since we consider this content high quality. To this end, if this highlighted information runs contrary to general scientific consensus, we reserve the right to correct or remove the information from the feature. Please note that medical information surfaced should not be used to provide medical advice, diagnosis, treatment, or provide medical or counselling care. (Google Support 2022) [emphasis added]

Interestingly within this policy statement are a few terms that overlap with the discussions of authority and consensus building within the scientific community explored in Chapter 2. First, the explicit appeal to scientific consensus and evidence-based best practices suggests that Google engages in some form of oversight in line with best practice within the scientific community. It does not, however, go into detail as to how this consensus is reviewed and assessed as this information is not made publicly available. Without the documentation of how this orientation towards scientific consensus building is carried out, this could be cynically interpreted as a Google performing the process of due diligence. However, it can also be viewed as a guiding principle on medical ethics of risk – the effectiveness of which I will explore in the results Chapters 5, 6, and 7.

Secondly and impactful within the context of health information mediation, Google states that the information “surfaced” by engagement with the search platform is not equitable to medical advice. This is to say that Google Search - the discovery platform on which these movements and exchanges of information are occurring as documented across digital health discourse - explicitly positions itself as a non-authoritative, neutral voice. Google Search, according to Google itself, should not take part in the decision-making process made by individuals and their healthcare providers.

This echoes a concept digital methods researcher Richard Rogers touches on around the question of responsibility and ownership of results generated by technology platforms. Rogers points to the tendency of platforms to use the term “co-
production” of information by users as a means to create distance between the technology platform and content it produces (Rogers 2013). In terms of the risk assumed through interventionist health information, distance is similarly helpful. Not only is Google Search positioned as neutral, but any negative outcomes can be attributed to patients through the supposition of co-ownership of content.

This is a provocative self-positioning of the platform which will require further investigation. I will explore these implications in Chapter 7 and in my conclusion in Chapter 8. While it may be the official policy of Google to position itself as a neutral service provider, this may not observably be the role it plays. Before testing this through my data analysis on stem cell treatments, I would like to provide a brief overview of a striking example of intervention on a health-related topic on Google Search.

3.6 SEARCH RESULT PAGES: MOMENT OF COLLAPSE AND PRACTICE

As contextualised in this chapter, Google Search as a platform casts an enormous shadow within digital health discourse and infrastructure on the shared web. The platform’s development and ingratiation into the flow of information and users on the web sets the stage for a unique confluence of motivations. This conflict brings together in one space - Google Search - information publishers from health authorities, regulators, patient organisations, and profit-seeking actors each with their own missions in competition for the ‘click-throughs’ of users. Google, through its possibility permitting and limiting mechanisms, interacts with each of these publishers according to its own logic and drive.

As suggested in my introduction, the product of interaction between user and platform when query is entered generates a snapshot of this conflict through the search result page. In analysing the single search result page and all its components, my research is able to gain insight into Google Search’s practice on stem cell treatments and their implications for broader topics in digital health. However, before I can explore this platform influence, I must first conceptualise users as critical actors within this moment of interaction.
Before going into my methodology, it is important for me to outline different approaches to and understandings of the ‘user’ engaging in health-related Google Search. The user can be conceptualised in many ways. Some particularly prominent and, indeed, dominant ones are patient as user, patient as citizen, and patient as consumer. These provide valuable insight into the proposed behaviour and power dynamics within digital health moments of exchange from the perspective of the query asker. However, when engaging in search around stem cell treatments, patients perhaps transform into a fourth conceptualisation; patients as products.

As I will discuss later in this section, the conceptualisation of patients as products suggests that searchers are first and foremost rendered as commercial products when engaging in search. This orientation is vital to understanding not only why search results are structured as they are, but how information is mediated and where risk is assumed when searching for health information online.

Within this research project there lies a difficult question of where to place those impacted by the major diseases or conditions for whom direct-to-consumer marketing of unproven treatments might be relevant. How should the person on the other end of the keyboard engaging in search be classed? As a user? A citizen? Or should they be seen as a consumer? Or – perhaps cynically – as product?

As the focus of this project is not patient perspectives, but analysis of Google Search as a mediator, I struggled with how best to position those targeted by marketing of unproven stem cell treatments. I do not intend to solidify my conceptualisation of patient/users/consumer within this section, but rather outline current approaches that might be applied in my future analysis of my data donation results. Indeed, without getting ahead of myself, I intend to set the stage for the novel findings around the conceptualisation of patient/user analysis that I uncovered through my research.

This section briefly looks into different ways that patients have been conceptualised in terms of risk and agency. I am particularly interested in discussions around digital
health rather than traditional methods of patient/user/consumer integration. It also addresses Charities and Non-Governmental Organisations (NGOs) as both constituted by and constituting bodies impacting patient participation and access to information on stem cell treatments.

First, in order to explore the question of proof and citizen/user/consumer attitudes, I would like to explore a few examples of these groups interacting with claims to authority as well as a brief discussion on risk and benefit analysis from these group perspectives. For patients/users/consumers, are objects like “unproven stem cell treatments” promissory future treatments or concrete current offers for involvement in treatment-focused research? How is knowledge of new treatments constructed? And how is risk and benefit weighed? What of their relation to health economics discourse and commercial intent?

AS USER

When attempting evidence-based enquires with technical platforms, I found it useful to begin with the conceptualisation of patients as users. This conceptualisation brings immediately into focus the functionality of a user interacting with a technological system. Perhaps it does not matter why and to what end a patient clicks on one link or another. It only matters that they, a user, clicked.

Within information management research, there has been a movement to add in social agency to users to complicate their interactions with information and communication technology. In fact, Roberta Lamb and Rob Kling specifically call out STS-founded critique of user conceptualisations within computer science, including Woolgar (1991) and Westrup (1997). They first point out the adoption of “an interpretive, constructionist perspective” of end users through which agency and structure (much like power and discourse within the Foucauldian framework) engage in a process of mutual constitution (Lamb and Kling 2003). They note that:

Technologies, particularly ICTs [Information and Communication Technologies], are integral to these interactions and so shape identity and institutions. In use, ICTs are an extension of practice and also a part of
structure—having dual effects and creating unintended outcomes.” (Lamb and Kling 2003).

In her work “Interpellating Patients as Users: Patient Associations and the Project-Ness of Stem Cell Research”, Henriette Langstrup provides a direct application of this conceptualisation of patients within stem cell research and treatment delivery. She argues that the “project-ness” of stem cell research as an outcome of technology-based treatment development prompts patients to envision themselves as future users, rather than recipients of stem cell interventions.

It can be extrapolated that this translates through to patient/user information discovery and decision-making through the content produced by researchers and providers alike in digital spaces. The way stem cell treatments are linguistically and conceptually situated within “project-ness” predetermines the way they will be engaged within digital search. In the following chapter, I will explore this conceptualisation of patient/users in more depth from the perspective of Google Search as a platform.

AS CITIZEN

Digitalisation has had a marked impact on the ways through which patients can express, identify, and engage with the process of healthcare. The boundaries and possibilities for patients appears characterised by conversations around biocitizenship and patient citizenship as well as their unique relationship to the rise of digital health. While patient agency and proactive action on the part of patients is a vital feature of the patient as citizen, there is an equally important emphasis on rights-based claims. This grants patient as citizens a strong connection to their political identities and contexts. Within the biomedical context, this deepens the conceptualisation of how patients can express their agency – not simply as actor within a medicalised context, but as citizens within specific civic system.

Classical ‘biocitizenship’ is demonstrated in the example of researcher Steven Epstein’s exploration of the construction of lay expertise within HIV/AIDS activism in the 1990s. This work lays the foundation for understanding the potential of patient
involvement in the creation of knowledge around disease and treatment (1996). Additionally, there are long established central themes within several patient-focused STS writings of the “lived experience” of disease (Jasanoff 2002) and the experiential knowledge held by patients (Marent et al. 2013).

Jessica Mulligan provides a useful and succinct definition of biological citizenship and its related terms and themes. She writes:

Biological citizenship—also called medical citizenship; biocitizenship; health citizenship; therapeutic citizenship—describes forms of belonging, rights claims, and demands for access to resources and care that are made on a biological basis such as an injury, shared genetic status, or disease state. (2017)

Inherent in this definition is a constant flux and struggle as a subject is situated in circumstance – both biological and socio-economic. John Powell outlines the call to action this implicit dynamism elicits within biological citizenship driven by digital technologies, writing:

In an influential 2014 article in the Harvard Business Review, Jeremy Heimans and Henry Timms argued that new digital technologies are causing a shift in the business sector from a model of ‘old power’ which is principally a consumption model grounded in what a few individuals or organisations know, own or control, to a model of ‘new power’ which is characterised by a participatory approach of peer coordination and harnessing the agency of the crowd. If the business of healthcare is to harness the benefits of new power, to be truly patient-centred and participatory, then it is vital that we embrace open debate about the rights, responsibilities and expectations of future digital health citizens. This debate needs to happen not only amongst healthcare professionals and patients within the NHS, but also within society as a whole. (2016)

This, interestingly, links back to the tension explored by Collins and Evan’s around the Problem of Extension (2002). The call for expanded engagement in decision-
making in the context of biological citizenship is impacted by the digital technologies through which these engagements are made possible in ever increasing speed. As suggested in my discussion in Chapter 2, any move towards technologically-enabled participatory healthcare through biological citizenship faces both opportunities and challenges through the technology that facilitates it.

There are two works that I would like to focus on in the way they inform my analysis as it applies to stem cell treatments: “'No one here's helping me, what do you do?': addressing patient need for support and advice about stem cell treatments” by Tanner, Petersen, and Munsie (2017)26 and “The shifting politics of patient activism: From bio-sociality to bio-digital citizenship” by Petersen, Schermuly, and Anderson (2018). Tanner et al. provided crucial evidence for the difficulties patients faced in navigating information and decision-making around stem cell treatments. Although not a key finding, this work underscored the agency of patients simultaneously empowered by digital methods for treatment discover and unmoored by the separation from traditional vehicles for healthcare information specific to treatments.

This work leads to Alan Petersen et al.’s secondary conceptualisation of health citizenship - or ‘bio-digital citizenship’ - (2018) rooted specifically in the pathways for patients made possible by digitisation of healthcare. Risk within this construction of patient-citizen is self-mediated based on the individual’s determination. Relation to economic intent for these patients is mediated by the economics of the healthcare system in which they find themselves.

Specific to stem cells, an emphasis has been placed on these patient perspectives of self-seeking and receiving stem cell-derived treatments (Scott 2007, Petersen, Seear and Munsie 2014, Langstrup 2011, Prasad 2017). Corollary to this, there has been considerable focus made on stem cell tourism – or the seeking of treatments across borders – and its impact on patient outcomes and experience (Petersen, Munsie, Tanner, MacGregor, and Brophy 2017).

26 Disclosure: I have collaborated with Claire Tanner, Alan Petersen and Megan Munsie on a number of projects and symposia on issues of digital health and stem cell treatments with funding through the Brocher Foundation’s (https://www.brocher.ch/) and my work at EuroStemCell (www.eurostemcell.org).
In *The Cultivation of Digital Health Citizenship* by Dimitra Petrakakia, Eva Hilberg, and Justin Waring provide a crucial critique to Petersen et al. 2018. They note that this interpretation does not include the larger techno-social and bio-social context and interactions made possible by digital health spaces. Essentially, it is not simply the mechanisms of interaction made possible by the platforms, but the peer and community level interactions which inform patient bio-citizenship. Importantly, Petrakakia et al. underline the importance of ‘algorithmic nudges’ in health decision making within this contextualization. For Petrakakia et al., digital health citizenship is defined as:

…a set of rights and responsibilities that emerge through the use of digital technology (health apps and platforms) to meet health-related purposes such as to log the side effects of a treatment; to seek consolation in an online patient community and to provide feedback to a healthcare provider. (2012)

This approach promotes a useful conceptualisation of biocitizenship in which patients exert influence and are, in turn, influenced upon within a techno-social sphere. The shift of involvement is mirrored in mechanisms within medical research and treatment standards development - both through literature and stated mandates. Public Patient Involvement (PPI) and Public Patient Engagement (PPE) have expanded within the research community to accommodate this shift. A study by Caron-Flinterman et al. (2007) found that 34% of patient organisations are engaged in direct funding of non-translational research and that 2/3rds of patient organisations include mandates to follow ongoing research. I should also note that there is literature on the impact of patient/researcher interaction outside of STS literature. Most notably within the area of science communication and the development of the dialogic approach to patient empowerment (Holliman et al. 2009, Barfoot 2015).

**AS CONSUMER**

Tied closely to conversations of health economics, ‘patients as consumers’ is a particularly salient (yet contested) conceptualisation when tracking the influence of commercial interests in digital health. Patients-as-consumers primarily situates
individuals in relation to the economic forces that both limit and promote medical interventions. This conceptualisation implies a rational patient consumer who makes balanced choices based on economic factors. It also dictates the way in which institutions structure technological moments of interaction with patients.

Here patient choice becomes an important characteristic of agency. In Annemarie Mol’s *A Logic of Care*, choice is presented as both part of the patient-as-consumer (or customer) and also, in some ways, a foil to the idea of care within healthcare. Mol writes:

… patient-customers are not necessarily on their own. They may organise themselves. Like other customers, patients can have their products tested or share their experiences without professional go-betweens. They can collectively acquire detailed knowledge of the niche market for the devices tailored to their needs. Websites and patient magazines may gradually collect all the relevant information. This is one of the creative innovations made possible by the market: as organised consumers, patients help each other to choose. (Mol 2008)

An example of this adoption of consumer conceptualisation can be seen in the language used by healthcare portals. In 2010, the move towards patient-facing digital platform mediation of care through healthcare portals in the United States and United Kingdom indicated an institution-level evolution of patient care towards consumer care (Ranerup 2010). Williams et. al refers to this process as the “pharmaceuticalization of society” (2011) in which risk and authority are measured through commercial factors. This includes the shift in practice towards the direct-to-consumer marketing of medical interventions and the social integration of these direct marketing tactics in community and peer conversations (Deshpande et al. 2004).

Daniel Costa et al. explore this adoption of “consumer” and “client” terminology across health provision across healthcare literature (Costa et al., 2019). This scoping exercise is useful in noting the rise of these commercially-influenced terms across healthcare settings including within both clinical practice (as shown through reporting
on clinical practice) and in medical journals. However, the paper also notes that healthcare recipients still largely prefer the term “patients”.

The person behind the keyboard accessing information on stem cell treatments through Google Search is perhaps best described as a shifting character; a citizen, a user, a consumer at different moments within their search journey. Which actors they interact with through the act of search will impact which of these identities comes to the fore at any given point along this progress. I will return to this concept throughout my data analysis as a way of re-grounding the relationship between information producers and disseminators and their intended audiences.

**AS PRODUCT**

The conceptualisation of patient as product first gained traction within community health studies in Allen Spiegel and Florence Kavaler’s 1985 work on Prospective Payment System (PPS), which are reimbursement programs rolled out in the United States (Spiegel and Kavaler, 1985). In their critique of these programs, Spiegel and Kavaler argued that the impact of treating patients as products resulted in a diminished focus on patient wellbeing and increased centring of economic drivers of medical care. This was then expanded on in the context of the corporatization of medical practice in the United States by Lu Ann Aday in 1987. Most importantly, patients as products have characteristics unique to those seeking medical care. This includes specific motivations, abilities, and interactive lifespans which are all factored into the economic extractive activities of the healthcare providers with which they interact. Abstractly, this might be observed as the inverse of the patient as consumer. While the patient as consumer foregrounds the agency and choice of the patient, patients as products centres the actors offering healthcare to them.

For my purposes, the conceptualisation of patient as product described by the health research community does not go far enough when applied to issues of digital health within Google Search. To create a useable and informative approach, it is useful to bring into conversation the strategy within digital political economics to view the user as a product. In for-profit digital platforms like Google Search which operate cost free for the end user, the commercial ask is on the side of the content publisher like
advertisers and service providers. Content publishers - through one platform mechanism-enabled route or another - pay the platform for access to eyeballs, or, in this case, patient users. I must stress here that users as products should not be seen as passive. A patient user as product has specific characteristics, needs, motivations, and online lifespans influenced by the role they inhabit as a recipient of medical care.

Hyunjin Kang and Matthew McAllister suggest that the users of free-to-access platforms like Google Search are subjected to a process of commodification. Kang and McAllister point to five unique aspects of Google’s process of user commodification. They propose:

- Google practices much more direct audience commodification than traditional media. By not relying on estimates of audience use patterns based on sampling from a third-party like Nielsen, commodification can be instant and more cost efficient.
- The targeted advertising strategy of Google intensifies the commodification of audiences. Google uses behavioural information provided directly by its users of attention to advertising. Therefore, it can be said that each individual, as opposed to a speculative and constructed aggregate, is a commodity of Google.
- Google commodifies diverse components of its users, including private information, online behaviours, and cognitions. The activities through which people willingly or unwillingly give their personal data while they use the Internet and their aggregated consciousnesses (which is represented by the popularity of keywords) are also commodities of Google. Thus, it seems that Google may acquire information about users’ social networks and other lifestyle and financial indicators. This will result in the commodification of social relations as well.
- In the commodification model of Google, there is even less distinction between leisure and labour time, given especially the multi-function and multi-entertainment nature of the Internet.
- Google’s audience commodification model, if applied to digitized versions of traditional media such as television, could raise
expectations about audience information collected from those media and the level of integration between advertising and media content.
(Kang and McAllister 2011)

This commodification of the user constitutes, in turn, a surplus value beyond the profit generated through its primary economic activity of advertisement on its platforms. This commodification does not simply enable the extraction of capital directly through advertisement sales and data collected on individuals, but through the scraping and selling of what Kang and McAllister refer to as the “aggregated consciousness” of users. This “aggregated consciousness” includes the combined and synthesised information gleaned through group behaviour on the platform. This information can then be sold back to advertisers as a secondary product (Kang and McAllister, 2011). For Google Search subjects like stem cell treatments, this might be similar conceptualised as aggregated discourse; the particular set of user community-generated boundaries and relationships that coalesce around search terms.

Furthermore, Google Search encourages a deeper and unforeseen search process. To borrow from Ned Rossiter and Soenke Zehle’s work on anonymity online - the positioning of users as products is radically impacted by shifts in the way information is shared online and user agency (2013). Users no longer search for destinations - like medical authorities or reference guides - but rather for terms, phrases, ideas. This is enabled by the information-gathering streams encouraged by Google Search platform as a central component of their commercial pursuit of user eyes to sell. This positioning - although subtle in its differences - may provide useful insight into the moment of collapse within the single results page created through queries around stem cell treatments. While a ‘click through’ is a success, so too is extended engagement by the patient user as it contributes to the platform’s commercial offerings to advertisers.

Given this exploration of Google’s rise to ubiquity, its mechanisms for mediation, and the exceptional configuration of actors it engenders, I can now return to that room in Boston. The disconnect between the worried heads of research networks and the influx of individuals appealing for information for their medical cases is partially
explained by the unmentioned spectre of Google Search. In parallel to the development of emerging treatments in the field of regenerative medicine, Google Search slowly but surely integrated itself within digital health behaviour as both a mediator and motivator for patient information gathering. By the time the heads of research networks gathered in that ballroom, the public landscape for health information was already being driven by third party actors outside traditional public health infrastructure. In the next chapter, I will describe the methodological approach I adopted to capture this moment of conflict.
Placing Google Search under a microscope vis-a-vis its output on specific topics is a difficult task that requires both methodological rigor and flexibility. To understand the choice and development of my methodology, first it is crucial to understand a few factors that contribute to the challenge of research on Google Search. These challenges can be broken up into a) difficulties endemic to researching Google Search; b) challenges for STS researchers specifically; c) the difficulties of studying controversial subjects like stem cell research online. Note here that these subjects are controversial not necessarily because of their content (as in ethical considerations), but rather the potential for controversy around their regulation, access, and potential for risk.

This chapter addresses my secondary research question: *How can Google Search – as a proprietary, algorithmically-driven search engine – be studied empirically? What methodologies can be employed and why is this decision important?* To achieve this, I first set out the challenges presented by Google Search and key questions that arise when addressing it as a subject. I then explore my chosen strategy for tackling Google Search as borrowed and adapted from algorithmic accountability studies. I then describe the development work I undertook in collaboration with my colleagues at the Algorithm Accountability Lab at the University of Kaiserslautern. This includes the breakdown of my data collection plugin, study design, participant recruitment and setup (in the case of my virtual donors), ethics, and finally my approach to coding returns. Throughout this, I also include the challenges faced and competencies I acquired through this long process of digital method development. Finally, I address the tools of analysis I employed to handle the large dataset my data collection returned.

The result of this effort is a data-rich digital ethnography of emerging therapies on a difficult to study platform and a novel contribution to digital methods. Throughout the
development of my initial research design I found myself motivated by Deborah Lupton’s approach to ‘lively’ data. She describes lively data thus:

First, these data are about life itself. Second, they are dynamic, constantly being configured and reconfigured as people interact with online technologies and circulated and repurposed by a multitude of different actors and agencies. Third, these data are a key part of the global knowledge economy, contributing to commercial, managerial, government and research enterprises (‘livelihoods’). And finally, these data have become an influential part of everyday lives, affecting beliefs and behaviours and increasingly, people’s life chances via the assumptions and inferences that are developed from algorithmic analytics.” (Lupton 2015)

When I read this, it set the tone for my approach to my data collection; I wanted to wrestle with this type of data as difficult as it might be. As I discussed in Chapter 3 around risk and culpability, conversations around Google Search are often side-lined by narratives around anecdotal evidence of the behaviour of the platform verses large scale observation of behavioural trends. Lupton also wrestles with the idea of these data as “digital data assemblages” specifically produced at the interaction of technologies and human (2015). For my purposes, the data I am aiming to create will constitute this same dynamic, intersectional approach.

Google’s Search platform occupies a significant and well-integrated position within online practices. By design, most users will only witness the output of the algorithmic mechanisms that bring a user from input of query to output of results. This, however, leaves the path from user input to result output beyond the scope of observable research. Therefore, most research on Google Search has been limited to attempting to draw observations from results rather than observe the mechanisms of algorithmic source gathering and decision-making directly (see Thorton and Handley 2011, Bar-Illan 2005, Speretta and Gauch 2005). Even when researchers treat produced results as data points, there are several factors that further complicate the study of search.
Firstly, Google’s business model functions only as far as it can be mobilised to connect advertisers to customers (Rutz and Bucklin, 2011). In practice this means that multiple algorithmic input/output systems are at play in every search. Unlike early search engine results of the 1990s, where search indexes functioned essentially as large databases with vocabulary limited to search terms, there is no single list of results. Rather, there are multiple systems in use on top of the already opaque organic search results. These additional indexing mechanics include results generated from Google Ads, Google News, and potentially more.

Google Ads – as explored in Chapter 3 – is an ad exchange and auction platform targeting content (or website) producers. It allows for the purchasing of space in presented results on Google Search as well as other Google-mediated platforms like banner ads on hosted websites. Google News is a compiling index of news stories pulled from major identified new outlets and associated media including the Associated Press, news blogs, and syndicated news networks. This news indexing is localised based on the user’s location, localisation (like browser language) and other methods of personalisation. It also operates time-sensitively with newer articles appearing at the top of the results page. However, there is still an aspect of return mediation through the ranking choice based on the term “relevance”. Each of these systems have their own logic that influence the output of results.

As with the nature of Google Search as a proprietary search platform, the outputs produced by Google search are opaque and include results originating from those different systems. Researchers observing these outputs must themselves parse the different sources to distinguish, for example, ad results from organic results (Chan et al., 2012).

Secondly, Google employs multiple systems that - whether by design or by effect - interfere with large scale studies of results (Rogers 2019). Anti-automated failsafes like Captcha and server request limits are ostensibly in place to prevent the exploitation of Google’s platform or protect from targeted automated attacks like Distributed Denial of Service (DDoS). DDoS attacks are commonly deployed tactics for disrupting access to webservers achieved through overloading servers with requests for access. However, they also limit researchers’ ability to scrape data in an
organised (and often bulk) manner, undermining both data collection and data continuity (Wilson 2019).

Geographic location of both user and server impact search results. Google Search aims to provide results that are both localised and relevant for users through decentralised hubs spread throughout the world (Krafft, Gamer, Zweig 2018). This means there is default input of information regulated by both the user’s geographic location and the location of the server through which the query is handled. User location is informed by determinants like IP address, default user language, browser location services and local files on the user’s device. Server geographical inputs are determined by the physical location of the server hub nearest to the user. These Google server hubs are located throughout countries and provide an additional input of information into search result generation.

Finally - and perhaps most frustratingly for researchers, Google Search is ever changing. That is to say, the algorithmic constellation that makes up the platform is not constant but is reviewed and refined as an ongoing process. We can infer some of the strategies for change adopted by Google from the release of new features on Google Search, the documented acquisition of smaller companies focused in specific sectors of algorithmic processing, and public statements by Google vis-a-vis both corrective measures and expansive functionality. However, there is also the risk of overemphasizing the fluidity of these structures. As my research will evidence, there is often a disconnect between the stated decision making of Google the company versus the change implemented on the platform.

These factors, which are intrinsic to Google Search itself, present obstacles to social science researchers attempting a rigorous study of search. Social science researchers also face several logistical challenges, including technical learning curves and the requirement of flexibility.

4.1 BALANCING METHODOLOGY AND STS CRITIQUE

While Google Search is an observable subject that plays a major role in the STS actor-network analysis of digital infrastructure, it is also one that requires a modicum
of technical knowledge of networks and digital infrastructure to accurately and reproducibly study (Bandy 2021). This either requires a researcher to enter with the required understanding of networked systems and algorithmic structures or to enlist the help of expert opinion to guide the development of any methodology and analysis strategy.

Deborah Lupton provides a particularly astute reflection on current blindspots within digital health and associated STS research. She suggests that the “working practices and mentalities of the commercial interests involved in the development” of technologies underpinning digital health are currently under examined in general and more specifically in relation to digital health. From the software development, business practices, production, and marketing, Lupton points to a dearth of information on the influence of commercial interests outside of conversations around so-called “Big Data” (Lupton, 2014).

She argues that critical engagement with commercial analysis of development and deployment within these technologies is particularly crucial because of the implications for public health and medicine. In particular, she notes the impact of the resulting technology in relation to privacy and ethical issues around the handling of highly-sensitive patient data.

Important to my research, she also points out the under interrogated impact of algorithmic decision making on the formation of health and patient categorisations in producing subjects and bodies. Lupton suggests that the growing value of personal data must serve as an impetus for social research analysis with an eye on commercial intent and influence. She writes:

…people’s prosumption activities are now frequently commodified in ways of which they may be unaware. This is a crucial area of research in a context in which digital data have become increasingly valued and commodified and will require greater attention in future critical digital health research. (Lupton 2014)

With Lupton’s guidance in mind, I began to consider my research question in terms of how I might meaningfully capture the complex moment of techno-social
interactions created by health-related Google queries. I began by reaching out to researchers within algorithmic accountability with particular focus on digitally-native methodologies.

Their guidance led me to my first methodological realisation: *studying Google Search requires flexibility*. As previously mentioned, Google employs multiple mechanisms that channel users towards certain specialised results (from a geographic to a user level) as well as defence mechanisms to block organised scraping of data for the purpose of study. Additionally, the dynamic nature of Google’s Search platform may result in collected results not conferring with the anticipated result types, forms, and structures (Reber, Krafft, Krafft, Zweig, and Couturier, 2020).

These factors required me to adopt a flexible strategy; first in collecting data and second in the analysis thereof. This required me to quickly upskill my competencies in data collection, handling, and analysis of large-scale data samples. Here it became clear to me how the technical competencies required to undertake these digitally-native methodologies might discourage social science researchers.

The Google Search platform is updated often and irregularly. This means that the format and content of data delivered *might* change across the research period. I had to account for this platform instability through planning of oversight and the possibility of ongoing changes to the data collection software.

Finally, Google Search’s treatment of query subjects differs drastically depending on context and content which may create difficulties for researchers looking to apply single methodological approaches. This includes variable frequency of content types (including organic Google results, advertisements, News items, images, and Knowledge Graph inserts). This required me to conduct a basic initial survey of the types of queries and their subsequent results across my targeted user profiles (like geographic location). This allowed me to have a general understanding of what results might be produced in my data collection. As explored in my discussion of obfuscated prescription treatment websites in Chapter 5, I still encountered unexpected data return types.
For my purposes, terms like “stem cell treatment” and “stem cell therapy” have different meanings and relationships to power as they move between disciplines and contexts. Therefore, researchers must see the case study focus as separate entities. In my case, this means making clear the difference at the point of analysis between stem cell treatments the dynamic and discourse-dependent messy objects, and “stem cell treatment” the query.

As discussed in Chapter 2, stem cell treatments are messy boundary objects whose characteristics shift according to context and actors. For example, the boundaries of stem cell treatments in the context of patient decision-making is different than when discussing in terms of translational research regulations. Queries around stem cell treatments do not reflect these nuanced discourses of stem cell treatments that emerge when in the context of biomedical research, patient journeys, or commercial enterprise.

Rather, when transformed into query inputs in the context of Google Search, these stem cell treatment discourses are collapsed into a single space as rendered by Google Search. This constructed entity has different modalities, biases and borders and must be treated as an independent object. Without this sensitivity to the rendering of stem cell treatments as a query subject, there is a risk of carrying over definitions and limitations of stem cell treatments from the field of regenerative medicine into the discussion of algorithmic practice (Lindvall and Hyun 2009). Put simply, what a stem cell treatment is to a biomedical research or clinician is not necessarily the definition of a stem cell treatment that is depicted through Google Search’s rendering of results around them as query subjects. They are separate ontological entities.

Within my data collection stage, I was not interested in the complexity behind stem cell treatments within biomedical research but rather in capturing how Google Search – through its practice – represents stem cell treatments. To address this, I made the choice to treat query terms separately from the messier objects of stem cell treatments. This enabled me to keep focus on Google Search’s mediation practices throughout the data collection period, rather than the contentious discussions around stem cell treatments in other fields. By making this distinction, I
qualified the data that I collect as distinct from the more contextual and discourse-specific analysis of stem cell treatments vis-à-vis digital health discussions. This separation enabled me to return to the more complex discourses around stem cell treatments and fairly gauge Google Search’s impact and implications.

4.1.1 SUICIDE AND GOOGLE SEARCH

In my initial research into health information online, I found it particularly helpful to look at search queries that might have some overlap in terms of the potential risk presented to users. In particular, the question of Google Search and information relating to suicide online has received attention from medical sociologists as a complex exemplar of digital possibilities and restrictions (Recupero et al., 2008, Biddle et al., 2008, Sakarya et al., 2013, Westerlund et al., 2012, and Wong et al. 2013). As these studies focused on search queries identified as immediately harmful to users, I assumed that they might showcase limitations placed upon Google Search returns. Moreover, they provided an initial starting point for how this mediation might look in regard to user behaviour and information retrieval.

For my purposes, these studies were relevant as they focus on search as a key component of research (rather than content analysis of blogs or forums, for example) and specifically focus on a controversial topic that could cause harm to the searcher. These studies scraped and analysed the types of material returned on queries related to suicide using a variety of query modifications. This section will take a closer look at one of the most recent studies to better understand how researchers are currently employing digital methods to aid their analysis of the role of search engines in sensitive health related topics online as well the limitation of those methods. Reviewing these medical sociology studies also allowed me to refine my methods and inform my study while keeping one eye on discussions of the impact of search on individual health and well-being.

Thornton, Handley, Kay-Lambkin, and Baker (2017) sought to contribute to a general discussion on the types of results returned for queries related to suicide. Building upon the work of Biddle et al. and Sakarya et al., Thornton, Handley, Kay-Lambkin,
and Baker posed the question: “Is a person thinking about suicide likely to find help on the internet?” This project deviated from previous studies that focused on patient interviews in their approach to query creation. Previous studies focused on either single “pro-suicide” or neutral query terms, leading Thornton et al. to suggest that this left a hole in understanding of search results. Thornton et al. (2017) write:

It remains unclear how commonly potentially harmful resources appear in generic or ‘suicide preventive’ searches, such as when an individual uses the Internet to search for assistance with suicidal thoughts.

The study used a range of neutral and suicide prevention-focused query terms with a particular emphasis on the ease of access to online treatments for mental health issues for nonprofessional users. The authors established a set of 13 search terms and an additional set of the same terms paired with the qualifier “help”. Between August 11th and 29th 2014, a total of 520 search results were retrieved. The researchers noted that the browsers used were cleared of cookies after each search and did not include sign-in to a Google account. The results were then analysed based on the following criteria: their page rank, relevance/irrelevance to the topic of suicide, the URLs country of origin, source type (commercial, health profession, governmental), information provided, links, and “attitude towards suicide” (Thornton et al. 2017).

Focusing solely on the methodology rather than the results, the study has limitations for extrapolation of the state of information on suicide found through Google. For example, if geolocation impacts results, how does locality impact results by a research browser based in Australia? These limitations are not limited to the 2017 study. Results from the 2008 Biddle et al. UK study and the 2013 Sakarya et al. Turkish study also do not consider the impact of geolocation or personalised filtering within their analysis of results, but rather make broad statements on the general availability of “harmful” or “helpful” links. Additionally, while the research scraped results from the first two pages of query returns, they did not indicate whether this included targeted ads in the results. How, too, could Google’s own regulations (through Google Ads and content censorship) impact the results of any query related to suicide – pro or neutral? What is the relevance of time (in the case of Thornton et
al. within one month in 2014) given Google’s integration of query trends into the presentation of search results?

The methodological limitations shown by this well-cited study underlies how search results are often a moving target within research. Yet all the studies included phrasing that indicated universal findings. Additionally, the application of these searches underlines the relevance of the following question posed by Richard Rogers (2013): for potential users (such as people inputting suicide-related queries or people searching for information on stem cell therapies) are search engines as socio-epistemological machines fundamentally research machines or consumer appliances? Does the awareness of the function of search engines change the behaviour of the user seeking particular knowledges?

I address this first question through my interrogations of the search engine result page in the case of stem cell treatment queries. However, the second question – more reliant on the secondary impact on user behaviour – is one I hope can be addressed in future work that takes on these rich approaches to data collection through data donor engagement.

4.2 BLACKBOX AUDIT ANALYSIS – CROWDS AND SOCK PUPPETS

Within my research, I am using a definition borrowed from algorithmic accountability literature in which the methodology was developed. Blackbox analysis is an analytical approach

“whereby an opaque system is scrutinised by analysing observable in- and outputs, deducing the inner mechanics that transform the former into the latter and approximating the inner workings with models” (Reber, Krafft, Krafft, Zweig, and Couturier 2020).

Having adopted this definition, it is worth noting that the term “Blackbox” has been widely used within STS as a metaphor (Woolgar 1979, Winner 1993). Most notable and pertinent in the context of digital health, Trevor Pinch (1992) used the term in his exploration of the sociology of science. He uses the term to explore the relationship
between the act of “performing” science and the role of scientists as investigators. While the critique Pinch forwards is useful in my analysis of authority and stem cell treatments, it is not the perspective I am taking in this chapter. As suggested by my adopted definition, Blackbox – at least in the context of my data collection – designates a more material object. I explore this tension between Google Search as a material and metaphorical Blackbox throughout my analysis in Chapter 7 and 8.

I settled on pursuing a methodological approach that would enable me to collect the cleanest data-driven snapshot of Google Search’s offerings on stem cell topics and create a flexible pool of data points from which to conduct a deeper analysis; crowd sourced and Sock-Puppet Blackbox audits. Crowd sourced audits include participation by public participants. In Sock-Puppet audits, researchers use software as stand-ins to simulate users. Here I will explore the strategies and mechanisms for crowd sourced Blackbox analysis, discuss why this strategy is a suitable one for answering my research questions, and outline the steps I took in order to equip myself with the resources necessary to undertake it.

Technically, this project builds on the work of others. From its inception, this crowd sourced Blackbox audit was built upon the work undertaken by the Algorithm Watch in their audit of Google News in relation to the 2017 Bundeswahl (Algorithm Accountability Lab 2017). In the spirit of transparency and collaboration, the Algorithm Accountability Lab published the data donation git-repository for fair use. Despite the changing landscape of Google’s algorithmic mechanisms, this ensured that there was a roadmap for creating a plugin capable of carrying out large-scale data donation and scraping of Google results.

Before I detail the methodology, I first need to explain a few conceptual definitions I have borrowed from digital methods and information science. These concepts are central to the logistical development of the EuroStemCell Data Donation plugin and provide the strategic path for the type of donors included and data this project was able to generate.
CODE AUDIT

A code audit might be used when the underlying code of a platform is available (either publicly or through direct agreement with the platform provider). In this type of Blackbox analysis, researchers can interrogate the code directly to map the mechanisms at play between input and output. This strategy paints the clearest picture of the practices of an algorithmically-mediated platform through transparency. Researchers can document directly the logics and mechanisms used to determine result inclusion and rank. Unfortunately, these audits are limited in their possible deployment as they rely on full transparency and access granted by the platform provider. For most proprietary algorithmically-mediated platforms like Google Search, this type of audit is not possible.

SCRAPING AUDIT

A scraping audit is a secondary type of Blackbox analysis that tracks observable interactions with algorithmically-mediated platforms to infer platform practices. This type of analysis stems from the known limitation of proprietary code; if we cannot interrogate the code directly, we can measure differences in output when manipulating inputs to the closed software (Sandvig et al. 2014). Scraping audits are useful methodological tools for interrogating proprietary platforms – although their use often goes against the User Agreement policies of proprietary platforms (Rogers 2014). Within this type of audit, there are three crucial sub-types which I will explore: Normative Users Audits, Crowd-Sourced Audits, and Sock-Puppet Audits.

To emulate real-world use cases, the source of these scrapes in these studies is known as the “Data Donor”. Data donors – or individuals submitting the queries – provide the software context for the outputs of results generated by the Blackbox platform. These data donors may be real individuals access the algorithmically-mediated platform through their computers or they may be virtual stand-ins tasked with result retrieval through virtual environments.
Normative User Audits are a type of scraping audit in which a researcher may use their own hardware and software to gather personally-specific observations. In these audits, a researcher scrapes their own search results generated through their browser to create a dataset for analysis. Thus, the data donor is the researcher themselves.

This approach is widespread and has been used effectively within STS discourse. For example, Safiya Umoja Noble’s work on algorithmically-enabled racism on Google Image Searches in *Algorithms of Oppression* relies methodologically on this approach (2018). These audits are particularly helpful in generating anecdotal support for wider themes within STS and algorithmically-mediated platform discussions. Researchers can quickly set up audits of results for self-selected queries. Moreover, they are also more easily accessible than other types of Blackbox analysis as they require fewer technical competencies to access and can be run from a researcher’s own computer.

However, there are inherent drawbacks to adopting Normative User Audits as core methodology. These include sampling, sample size, and self-reporting biases (Sandvig et al. 2014) that result from capturing results generated from researcher or researcher-adjacent browsers. As Google Search seeks to localise – if not by personalisation, then through geographic influences – these biases limit findings generated by this type of audit.

Crowd Sourced and Sock-Puppet Audits attempt to address these problems of bias and sample size by decentralising the distribution of data donors. These approaches recruit real or virtual participants respectively to use their own access to the algorithmically-mediated platform to scrape results. This expands the profile of the accessing and query-submitting user away from the researcher themselves.

In the case of my research, the opaque system is Google Search platform, the inputs are search queries made from donated browsers, and the outputs are the first page
of returned results composed of both ads and organic search results. In this case, scraped data would include the ads and organic results generated from the first page of results returned on a set list of queries. These audits create an opportunity to construct a snapshot of results generated by the manipulation of inputs and sources.

The mechanism for facilitating the data scrape is a browser plugin. A browser plugin is an extension or add-on to a web browser - often Mozilla Firefox and/or Google Chrome - which adds functionality. Specifically, it extends the browser’s capability to automatically search for and snapshot specific returns generated by query inputs. This methodology requires the development of a study-specific browser plugin.

Whether a real human or virtual, the owner of the browser from which the scraping audit is run adds another variable. This project combines two strategies for data donor recruitment: crowd-sourced audits and Sock-Puppet audits.

Crowd-sourced audits recruit real human subjects as data donors who opt to take part in the study by contributing data (Mittelstadt 2016). These human data donors (henceforth, simply data donors) become anonymised producers of data through automated collection of data using the donor’s pre-conditioned digital environment. This allows for approximation of natural search results. Recruitment is required in this approach as well as vetting for specific donor profiles (for example, region, language or relationship to a query topic). Recent studies that have used crowd sourced audits of Google Search include audits of partisan bias in Google News (Robertson et al. 2018), proprietary image tagging on Google Images (Kyriakou et al. 2019), and analysis of autocomplete of queries (Robertson 2019).

I was surprised and encouraged by the enthusiasm for human participation as a data donor displayed by potential control participants (i.e. data donors without a stated connection to one of the identified diseases), patient groups and researchers in digital health. On the one hand, limiting my data entirely to virtual data donors would grant me greater control over this methodological experiment. However, I began to see my research as an opportunity to engage with conversations around search and health. Given the ubiquity of Google Search and the way it has worked its way into everyday digital life, engaging in this type of participatory yet anonymous research is
potentially empowering. Noortje Mares describes living as “collective practices of researching social and cultural change, as engaged in by actors who do not necessarily identify themselves as ‘social researchers’” (2012). So here too we have individuals engaging in larger practices of research and knowledge interrogation through quotidian and unobtrusive opting-in.

In *Sock-Puppet audits*, human donors are replaced by automated, virtual data donors. *Virtual Private Servers* or VPS function as virtual data donors (henceforth, *VPS data donors*) to stand in for live human subjects. These VPS can be set up remotely in partitioned servers by location and recruitment is handled by the researcher themselves. VPS data donors create clean digital environments for scraping results from Google Search and allow for real-time adjustment of settings without compromising human donor privacy or time commitment. Other recent studies that deploy *Sock-Puppet* audits include the study of political personalisation in Google Search (Le et al. 2019), racism within online housing markets (Asplund et al. 2020), and the reach of political news on Google Search (Umarova 2020).

It became clear to me early on that a combined approach of both crowd-sourced and *Sock-Puppet* auditing would offer valuable potential to address my research questions while acknowledging my limitations in time and resources. I gave particular attention to the following considerations: the potential to provide a snapshot of result data for analysis, data flexibility, the breadth of data the data collection would return, the data collection’s technical feasibility, and its value as a proof of concept for STS research.

My mulling over of strategies took on new urgency with the announcement of Google’s ban. The announcement from Google regarding cessation of advertisement in relation to stem cell treatments increased pressure on me to adopt a methodological approach in time to capture the results. It forced me to consider how quickly I could set up the data collection and the potential for rich result capture.

After looking into various methodological options, I decided on both a crowd-sourced and *Sock-Puppet* audit designed to capture the transition of search results pre-and-post advertisement ban. The combination of the approaches would enable me to
collect unbiased and wide-reaching documentation of the efficacy of Google’s policy change. Additionally, the deployment of the virtual servers enabled the capture of results across a wide breadth of users and geographic locations. This assisted my subsequent analysis of both ads and organic search results for quality and context.

Data flexibility in *Sock-Puppet* Blackbox audit was a particular selling point as the fluid and opaque nature of Google Search held the potential for unforeseen findings. By setting up multiple sources of data collection and automating the collection over a period of months rather than weeks, the plugin enables the researcher to respond to changes as they develop while also allowing for the mining of specific topics within the data set. Due to the difficulties in both creating the plugin and maintaining the data collection and donor input, I worried early on about limiting the scope of data donor eligibility, geographic location, and time too soon. If any limiting factor such as low donor uptake or region-specific policy development were to intervene, this might result in insignificant or inaccurate findings. Indeed, since recruitment numbers and retention could not be measured until after deployment, any development would have to take into consideration a route for data collection even without live data donors. Rather, by using a broad source approach to data collection, the crowdsourcing Blackbox methodology allows for specification in data parsing to be undertaken *after* collection has already taken place from within the larger data set.

Finally, undertaking a crowd sourced Blackbox audit as a STS project has clear value in contributing to the quality and technical specificity of future research. I reasoned that there is value in developing an STS-guided proof of concept around crowd sourced Blackbox audits and by documenting the steps it has taken to develop and deploy a novel digital method. It has become clear that deploying a methodology like this might go far in encouraging other STS researchers to undertake technically rigorous methodological approaches in the name of creating data-driven reflections.

These factors made clear the case of adopting a crowd sourced Blackbox audit as the best suited methodology to address my research questions. However, it still left open the question of how I, specifically, could develop and deploy it. To adopt this course of study, I took a number of steps to expand my research capabilities. These
included taking stock of my own technical and background experience, assembling a team of expert contributors, securing funding, coordinating with external partners, and setting up fallback mechanisms to ensure a robust data collection.

4.3 THE LOGISTICS OF DATA DONATION PLUG-IN DEVELOPMENT

I came to this project from the perspective of a digital project manager for a researcher-driven public information project on stem cell research, EuroStemCell. Much like many of the other Blackbox analysis studies, it was my suspicion and singular observation of the treatment of a topic within and through Google Search that sparked an interest in documenting and analysing how Google Search rendered information. Although I have extensive experience working within some digital mediums, I had not developed digital methods software or browser plugins before. This was a clear limitation that I needed to address. While I might be able to offer insight on the context and constellations of digital health information and patient involvement, technical expertise would need to be supplemented.

Noting these limitations of my research capacity, I assembled a team with the support of Dr. Professor Katharina Zweig of the Algorithm Accountability Lab at the University of Kaiserslautern in Germany. I was joined by computer science PhD researcher Tobias Krafft and Masters students Martin Reber and Roman Krafft. Under my guidance, Martin Reber undertook the software development of the data donation plugin for his Masters thesis with ongoing discussions with the larger team.

Additionally, I sought advice and support from a number of stem cell research, patient organisations, and clinical research bodies in order to support the development and deployment of the crowd sourced Blackbox audit. This includes support from the Centre for Regenerative Medicine, the Anne Rowling Clinic, Stem Cells Australia, Parkinson’s UK, Spotlight YOPD, and the Parkinson’s Research Interest Group and the Edinburgh Research Interest Group. These groups offered both support in recruiting donors as well as reflective input on the research. In order to support the development of the plugin, I secured funding through both the European Commission funded project, EuroStemCell (£3,000) and the Graduate School of Social and Political Science at the University of Edinburgh (£500 and
This funding went towards the development of the plugin through the time of Martin Reber, the funding of server space and VPS servers in the US, Canada, UK, and Australia as well as funding for an onsite meeting of myself, Tobias Krafft, Martin Reber, and Roman Krafft in Edinburgh in 2018.

In collaborating with more technically-focused researchers and developers, I found it worthwhile to adopt the Scrum strategy for software development during our meetings to focus our efforts on creating a discrete and workable plugin quickly and effectively. I was fortunate to have established experience working in Scrum-guided development through my previous professional work in digital development with the web development system, Drupal. Scrum development focuses on an initial meeting outlining the minimum viable software needed to achieve the goal (in my case, the basic data donation plugin capable of scraping results). It then shifts to focus on individual roles with specific deliverables that contribute to an iterative final software. As Rising and Janoff characterise:

Scrum is a process for incrementally building software in complex environments. Scrum provides empirical controls that allow the development to occur as close to the edge of chaos as the developing organization can tolerate. (Rising and Janoff 2000)

Adopting this approach gave me and my colleagues a sense of direction and deliverables in a short period of time. This was particularly useful in allowing us to adapt to changing requirements from the Google Search platform as well as our own individual academic responsibilities.

I should note that the urgency created by the introduction of the ban required quick action on behalf of my collaborators at the University of Kaiserslautern, particularly on the part of the student research developer who adapted the Algorithm Accountability Watch browser plugin, Martin Reber. His quick action was crucial in getting the data collection up and running by the deadline of October 1st, 2019.

Finally, in order to secure contingencies in the case of difficulties with data collection through native data donors, I set up funding and deployment of 21 Virtual Private
133

Servers (VPS) to act as clean data donors using five different server hosting companies located in the UK, US, Canada, and Australia. These VPS donors would run data collection for the length of the study to create fallback data sets with which to construct my snapshot. These are the steps I took to expand my capabilities to successfully deploy a crowd sourced Blackbox audit.

4.4 EUROSTEMCELL DATA DONATION PLUGIN

This planning culminated in the development and launch of the EuroStemCell Data Donation Plugin (Figure 3). In the following section, I will outline the functions of the plugin, the deployment, a walkthrough of the data donor experience, and the data storage and VPS conditions. Each aspect of the plugin’s design was the result of conscious choices driven by concerns for feasibility, privacy, and research value.

The EuroStemCell Data Donation Plugin (henceforth referred to as ‘the ESC plugin’) is a browser extension or ‘add-on’. Extensions, by definition, are small packets of software which allow for the limited expansion of functionality of web browsers. Before installation, donors were asked to fill out an anonymous survey for more information about their relationship to Parkinson’s Disease and Multiple Sclerosis as
well as generalised radio button questions based on predefined, mutually exclusive answers which allowed for the inclusion of donors into discrete study groups.

The functionality added by the ESC plugin is limited exclusively to conducting an anonymised data scrape of www.google.xx results for a set list of queries on a set schedule. The process of this model is depicted in Figure 4. Upon reaching a scheduled time interval, the extension issues a command to open a new browser window, navigate to the homepage www.google.xx (xx according to the computer’s geographic pre-set), search for a defined query, capture the page source results, and transfer these results as a data file to the servers located at the Algorithm Accountability Lab. The ESC plugin then conducts the same scrape for all queries on the list. The process takes roughly 10 seconds from initiation to completion.

The plugin was deployed through the Mozilla and Chrome browser stores during the study period. This process required us to share with both Firefox (Mozilla) and Google (Chrome) the code used by the ESC plugin and seek safety approval from both stores for inclusion. Additionally during the time, the plugin was public, ongoing monitoring was required to make sure that the ESC plugin continued to send accurate data back to the central servers and that the plugin was not targeted by Google by anti-automation mechanisms.
From the user’s end, the donor process is as follows: A data donor is directed either directly or via a landing page (https://www.eurostemcell.org/datadonation as seen in Figure 5) to the Firefox Mozilla or Google Chrome store according to their browser preference. There the data donor is granted access to the research aims, terms and conditions as well as privacy statement of the project. If they choose to download the plugin to participate, they are prompted to enter an anonymised survey (Figure 6) to designate study grouping. Once submitted, this survey generates an anonymous install ID code which groups together the survey and donated results. No private information (according to the GDPR) was viewed, collected, or saved. This means that identifying information – including but not limited to IP address, name, email, precise geolocation, or search history – is not accessed by the plugin.
Deployment for *Sock-Puppet* VPS donors was much the same as for live data donors with the exception that there was no need for a survey. Rather, the VPS results were sorted into their own clean study group differentiated by geographic location. Four VPS were deployed in the UK, Canada and Australia with 8 deployed in the United States due to the large number of Google Server hubs located throughout the United States. These countries were chosen for a few key reasons; their establishment of English as the primary language of search, their diverse healthcare models and geographies, and their population’s high usage of Google Search as a primary search engine platform. I also considered placement of VPS servers in terms of their closest cities and – subsequently – data hubs through which Google Search routes search traffic. I aimed for high population centres to explore the most widely applicable results.
Once installed and sorted via survey, the ESC plugin operates every 4 hours if the donor’s computer is on, the browser application open, and computer plugged in. At the completion of the data donation or if at any time the donor wished to cease participation, donors can uninstall the plugin and all data associated with the plugin from their browser application. No software was added to any part of the donor’s computer outside of the browser application and no personal information was able to be accessed through the plugin. Data storage was kept in a secure server located at the University of Kaiserslautern accessible only to myself and my stated collaborators in the Algorithm Accountability Lab (Figure 7).

**FIGURE 7:** SKETCH OF THE PLUGIN-SERVER-COMMUNICATION OF THE EUROSTEMCELL DATA DONATION, COURTESY OF MARTIN REBER, UNIVERSITY OF KAISERSLAUTERN 2019.

**SCRAPED DATA**

The type of data collected had a significant impact on the scope of the conclusions that I draw out in my analysis. To better understand the type of data collected by the ESC plugin, I would like to briefly dissect the information made available through the Google Results page (Figure 8) and how the ESC plugin constructs a database of results. The ESC plugin relies on the source code rendered to users when they access results pages. This code is machine-readable but includes valuable information reflecting the same content presented in menu-format to users.
These results include all of the displayed information as well as some critical additional information that may or may not be seen by the user. For in-page ads, this includes the demarcation of the result as an advertisement, the link to its relation to GoogleAds services, the ad title, associated URL, and additional content relevant to the query. The ad is marked as by the html tag `<event>` and all associated content information is all contained with the corresponding html tag `<div>` (see Figure 9).
The markup for organic search results can be similarly parsed as an `<event>` and include information within the associated `<div>` including title, URL, source ID, and a short snippet of subject-specific content.
The anchors and tags within this source code allow for a machine-readable scraping of the results page (Figure 10). This is more conducive to the creation of a results database as information can be parsed without referencing front-end user display.

The raw results data was then separated out into individual submissions (e.g. individual displayed results like the Cells4Life.com or NHS.uk results) and associated with bundled metadata. This bundled metadata most relevantly included the donation number, the associated study id, the id of the plugin (i.e. the anonymised donor number), associated keywords, the search date, title, content, position within results, the specific URL, the URL domain, the region, associated condition, and whether or not the donor was a VPS.

Once this database was established, I separated the results into study groups and added an additional three categories: type of result, commercial intent, and inclusion of direct-to-consumer marketing. These characteristics required manual coding, which will be discussed later in this chapter.
Although potentially powerful in its ability to snapshot Google Results pages, the ESC plugin has a structural limit on the information it can collect. Specifically, it only collects the content displayed on the Google Search results page at the time of scraping. It does not follow through the result links to capture the content pages themselves. The decision to limit the data scraping to just the information made available through the source code was a conscious one due to the time and resource limitations of my research and my colleagues at the Algorithm Accountability Lab. As will be discussed in my conclusion in Chapter 8, future studies may find increased value in a secondary stage of capture wherein a snapshot of the linked content page is also retained.

4.5 QUERIES

To create a steady and comparable data set, I developed a specific list of queries to run during the data collection process. These queries would form the input values within our Blackbox analysis, and, because of this, both their content and construction were vitally important.

In deciding on the search terms to be used as inputs, I erred on the side of casting a wide net in addressing the question of stem cell-derived treatments. Keeping in mind the shifting modalities of the terms around stem cell research (including the use of “treatment”, “therapy”, and “cure”), I opted to include all three query constellations. This, I reasoned, would allow for a broad snapshot of the state of information across these modalities as well as leave open the possibility for internal comparison of results. To capture results based on specific conditions, condition-specific queries were enabled by the substitution of the relevant \{condition\} within the study group.

Finally, it was important for my purposes to include the open-ended query “Can stem cells help me?” Because this query does not include any anchoring information such as condition or a call to specific treatment, I hoped it might pick up on some nuanced decision-making made by Google Search’s algorithm on behalf of the searcher. I was interested in capturing how the platform would handle the subjective term “help” alongside the user-driven call to personalisation (“me”). I hoped that, although it
might be a dead end, this open-ended opportunity for Google Search to fill in the blanks might provide valuable insights.

Due to the functional nature of text-based search, the Boolean construction of these queries required consideration. In its simplest form and in the context of search platforms, a Boolean query is the use of symbols, words, or other special characters to limit or expand search results. These may include the use of written qualifiers such as “AND”, “OR”, or “NOT”, the use of quotation marks (“ “), negative (-) or positive symbols (+). These Boolean qualifiers may shift results as they provide additional information to the search platform through the construction of the query itself (Frants et. al, 1999).

With these content and construction considerations in mind, I established this list of queries to be conducted in each search:

```
"stem cells"
"stem cells cost"
"stem cell treatment"
"stem cells cure"
"stem cells therapy"
"can stem cells cure {condition}?"
"can stem cells help me?"
"{condition} cure"
"{condition} therapy"
"{condition} treatment"
{condition} stem cells cost
{condition} stem cell treatment
{condition} stem cells cure
{condition} stem cell therapy
```

These queries form the backbone of the data set scraped by the plugin and provided consistency for analysis across the study groups.

Finally, to encourage transparency within the project and provide feedback at each data collection for the live data donors, an optional page was integrated into the plug-ins design. This page, displayed by clicking the extension icon upon completion
of a scheduled data scrape, included a brief message of thanks alongside a few examples of scraped results that could be viewed by the donor. Additionally, the links to more information including the privacy statement, frequently asked questions and how to leave the study were prominently displayed (Figure 11).

![Image of a data scrape page]

Thank you for your generous contribution.

After your first donation, you can always review your last submission below or by clicking on the plugin's icon in the top right corner of your browser in the navigation bar. If you want to learn more about the project, our motivation and the progress we are making, visit our website.

Your last donation was submitted on 2020-2-14 17:23:55 following a search for 'parkinson's cure'. See below for more Details.

| Search results | Advertisements | Top Stories |

**FIGURE 11:** “THANK YOU” AND EXAMPLE RESULTS PAGE PROVIDED TO DONORS AFTER SCHEDULED DATA SCRAPE, PROVIDED BY AUTHOR 2019.

### 4.6 ETHICS AND PRIVACY

In approaching ethics within this project, I employed several strategies to ensure transparency and limit the possible impact on live data donors. I envisioned from my research’s inception that I would not rely on mining or recording any personal data of individuals who opted to take part. Indeed, the presence of previous history and relation to the query topics would be restricted to basic Boolean indicators and no personal identifiers would be recorded. I am not interested (nor should my data collection strategy) reflect an interest in the individual experiences or results of human data donors. Rather, I am interested in how Google Search itself treats users in more general terms.

First and foremost, my decisions on privacy and recording of information were driven by a commitment to anonymity on the part of the live data donor and functional limitations on the capabilities of the data scraping software. This manifested itself in three main areas: the development of the plugin, the privacy statement, and the use
of Virtual Private Servers (VPS) donors. These decisions allowed me to collect data for analysis without compromising the privacy of any individuals who took part.

During the initial development of the plugin, I made several functional decisions that opted for live data donor privacy over fine-grain personal information. The goal was to be as unobtrusive and anonymous as possible without losing research value. This meant that the plugin was developed to only collect the following information: a general survey (to be discussed later), the donor’s language preference, the time of donation, the Google account login status (a yes/no function rather than any personal information), and the scraped results. No other information was collected, and each donor was anonymised at the point of collection with the assignment of a donor ID with which to group the results. All information was stored on servers at the University of Kaiserslautern and University of Edinburgh made only accessible to directly involved researchers. However, for the purpose of transparency, it is my intention to make this anonymous material available freely after publication.

This decision to develop the software to purposefully screen out the collection of personal identifying information (PII) was a conscious choice to err on the side of anonymity. This resulted in limited findings in terms of how finely data donors could be grouped together for analysis and avoided the collection and storage of any personal information determined by the GDPR (including IP address, email, or name). This was a deliberate decision as I determined that the general findings would be more helpful than those created by personal information. Additionally, this bypassed the need for intensive onboarding and review of personal information oversight.

Live donors were presented with the following Privacy Agreement; agreement to which was required to install the data donation plugin. Data donors were free to leave the study at any point and opt-out by simply uninstalling the plugin. No additional donation data would be sent, and no software would remain on the donor’s computer. Additionally, upon the completion of the study, donors were prompted to uninstall the plugin as it would no longer be collecting any additional data scrapes of Google search. Upon uninstall, donors were presented with an option to provide an exit survey although uptake was limited.
Privacy Statement

The purpose of this statement is to ensure that you have read and understood the information about the study and are fully aware of your rights should you decide to take part. If you would like to take part, please indicate this by reading the following questions. Consent is required in order to download and install this plug-in.

Please note: since data is anonymised at the point of contact, we cannot retroactively withdraw any data collected before uninstallation.

Declaration of Consent:

With the installation of the plugin I confirm that the first 10 search results and ads of Google on Google.com (in "All") including the above mentioned additional data (plugin ID, time, exact query, approximate location) on my browser will be made available to the public under a CC-0 license for analysis.

I understand that the plugin regularly searches for all healthcare related queries listed above.

I agree that the following data will be collected, processed and published:

- A general location, derived from the IP address, corresponding in precision to approximately your postal code.
- The plug-in ID
- The exact search query
- The time of the search
- The result of the search (everything on the first page of the search results)
- The language setting of the browser
- Whether you are logged in as a user with Google.
I can disable or uninstall the plugin at any time. No further data is sent with the deactivation of the plug-in or with its de-installation.

By downloading the plugin, I confirm that I have read and understand the privacy statement, www.eurostemcell.org/datadonation, for the above study. I have had the opportunity to consider the information and ask questions which have been answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

I understand that any personal data collected during the study will be treated with confidence and handled in accordance with the Data Protection Act 1998 and GDPR.

I agree to take part in the above study.

I am deliberately including the details of the consent form here to underline two aspects of the data donation process. First, as seen here, there are clear and self-imposed limitations of data collected by the data donation plugin which limits personalisation of the resulting data set. This was a conscious choice on my part to adhere to GDPR guidelines and patient data best practice. This statement outlines the commitment within the plugin development to privacy and anonymisation at the point of collection.

Secondly, I want to acknowledge the data privacy information given to participants who chose to participate. I noted throughout my conversations with data donors from patient communities (who often had previous experience with research engagement) that they viewed this engagement as different than other research involvement. Common in my conversations were concerns and questions about privacy and data protection. Additionally, data donors were interested in how much of their data would be accessible by me as a researcher through the plugin software. This privacy statement makes clear the limits of the plugin’s functionality and data collection.
There were some limitations imposed by my focus on data privacy throughout the development for participants. Due to the nature of the anonymisation of results at the point of contact, it was not possible to provide an option to retroactively opt-out of the study once a donation had been made. Any donations that had already been captured would continue to exist in the study results. However, as the opt-out uninstall suggests, no further donations or connection to the study would persist after uninstalling the browser plugin. This configuration was decided upon early as it favoured the overall anonymity of live donors.

Before installation, live donors were asked a series of general survey questions to sort them into study groups based on country location and relationship to the listed conditions (Parkinson’s, Multiple Sclerosis, or Diabetes Type 1 or 2).

Question 1. “Are you or someone close to you impacted by {condition}?”
   • I am a patient.
   • I am a carer.*
Hyperlink: “A carer is anyone, including children and adults who looks after a family member, partner or friend who needs help because of their illness, frailty, disability, a mental health problem or an addiction and cannot cope without their support. The care they give is unpaid. (NHS England)”
   • No.

Question 2. “Are you a stem cell researcher or medical professional?”
   • Yes
   • No

Question 3. “What is your country of residence? *Note: at this point we are only studying the impact of Google advertising on the following English speaking countries.”
   • United Kingdom
   • United States
   • Australia
   • Canada

Question 4. Age:
   • 18-29
   • 30-39
   • 40-49
   • 50-59
   • 60-69
   • 69+

Question 5. Gender:
• Male
• Female
• Other
• I prefer not to say

Question 6. “How often do you use your computer?”
• Daily (More than 2h a day)
• Daily (Less than 2h a day)
• Weekly
• Monthly

Question 7. “How often do you use Google Search?”
• Daily (More than 2 times a day)
• Daily (Less than 2 times a day)
• Weekly
• Monthly

Question 8. “Have you ever paid for or enquired about stem cell treatments?”
• Yes
• No

If Yes: We’d like to hear about your experience. Please contact us here (Hyperlink to Data Donor contact page)

These questions were intended to provide a minimum threshold of context of who the donors were without providing identifiable personal information. It should be noted that the donor’s relationship to a condition or disease was not limited to those directly diagnosed with these conditions. Rather, it includes individuals who either were unpaid carers for individuals diagnosed with the conditions (per the NHS definition) or those who professionally engaged with work around these conditions. The assumption here is that those who are either personally or professionally invested in information about the conditions may have previously searched around the condition, thus impacting their results.

The last question – “Have you ever paid for or enquired about stem cell treatments?” – was included as an open-door call if any participants wanted to share personal anecdotal evidence from their own healthcare and decision-making journeys. I did not expect – and indeed did not receive – any participant response to this question, but was encouraged that the opportunity for engagement was made possible. Had I received more feedback from personal experiences with private stem cell treatments, it would have influenced my understanding and analysis of user behaviour and desires when engaging with search around treatments. However, as
this research project is not primarily focused on patient/user experience, the lack of response did not impact the final data collection or analysis.

The only required responses were Questions 1, 2, and 3 as they determined the study group into which the donor was routed. This was an unfortunate oversight as it meant a missed opportunity to collect comparable data for other indicators. However, the decision at the time was in favour of avoiding forcing participants to disclose more information than they might otherwise want. Because this process is informed by best practices within digital ethnographies, this was a conscious choice to acknowledge the fact that opting-in to study participation did not include one-to-one consultation with the researcher to inform and gain consent for secondary personal information sharing. Instead, the digitally-native intervention (making response optional) was my strategy to address this disconnect between researcher and participant.

To make certain that results could be generated without possible impingement of live donor privacy, I made the decision to deploy a number of Virtual Private Servers (VPS) to function as Sock-Puppet “clean” donors. Mimicking human-user behaviour, these servers provided regular queries to Google Search and returned on-going results. One unique aspect about these Sock-Puppet donors is that they engaged with Google Search without any search or browser history. This allowed me to capture any disparities between users with normal varied search behaviour and those “clean” computers without historical data.

To capture geographical divergences, I established servers located in the United States (Dallas and Miami), United Kingdom (London), Canada (Toronto), and Australia (Melbourne). By including these servers as donor sources, I was able to generate a baseline of search results without needing human donors.

Using VPS donors as a baseline is a tool used within algorithmic accountability studies and digital methods to create clean findings created by a Blackbox without influence of user variability. As stated, these VPS access the Google Search page without user histories; each access recreates the first time a new user accesses the platform without input of personalisation based on history or other characterisations
based on user-specific metadata. This is valuable for three reasons: it creates a baseline of evidence to which to compare specific users groups, it recreates a specific character of real user, and it allows for the ease data collection beyond human participants.

Before I can analyse the impact of user identity (in my case, relationship to a condition) contributes to the rendering of results, I must establish a control group in which the data donor is not impacted by user history. This could be done using human control groups. However, given the challenges of participant recruitment, I decided that VPS donors would provide appropriate base comparisons.

It is also reasonable to suggest that the clean VPS results have value in and of themselves. As Chapter 3 indicates, Google Search integrates several unique and unseen mechanisms of influence in its rendering of results to individuals. “Clean” VPS donors should create results similar to those seen by human users without content-specific histories and therefore provide a solid evidentiary base for subsequent analysis.

Finally, there is an argument that the human engagement in this data donor process has necessarily the potential to impact future search results for those who take part. I recognise that in dealing with a Blackbox platform focused on user identification, there are potential unseen impacts that might lead to future identification of users as associated with the topics of queries the data collection required. However, I would push back against this on a few points. First, there is no direct risk of any corroborating data association from the data collection plugin itself as the anonymisation at the point of collection separates the data collection practice from any identifying user information.

Secondly, if the concern is around future treatment of human donors by Google Search personalisation due to search history, there is a presumption that search query history is the privileged mechanism through which Google identifies user groups. This is not a conclusion that I can reach from my research and, indeed, lies beyond what we are able to know about this proprietary Blackbox. In the future, this
might be an area of investigation for those pushing forward human contributions to data donation methodology.

Both the design ethos of functional privacy and an on-going reliance on non-human data donations as a baseline was a conscious strategy to protect any individuals who took part in the study. With this strategy deployed, I submitted and was approved for ethics oversight from the University of Edinburgh School of Social and Political Science Ethics Review board. This approval was granted in March, 2019.

4.7 GROUNDING DATA ANALYSIS

In addition to the strategy for data collection, I also developed a guiding strategy for the analysis of collected data. This approach is largely based on the insights of Nicholas Diakopoulos’ work on algorithmic governance and Blackbox analysis (2013). In particular, I developed the analysis on measuring the “atomic decisions that algorithms make” through prioritisation, classification, association, and filtering (Diakopoulos 2013). In addition to these, I have included an additional criterion of relationship to risk in my analysis. These aspects guide the subsequent analysis and create the lens through which the returned results have been scrutinized.

Prioritisation within algorithmic processes is the ranking logic through which certain sources and nodes of information are granted weight. This manifests itself in the numerical order of appearance within the first page of search results. From the algorithmic controller standpoint, this is a mechanism by which relevance and authority is established. We can infer from the relative rank (1-10) of the first page results the importance placed upon a result as determined by the algorithm itself. This allows for two forms of meta-analysis: the analysis of the result itself in relation to the query and analysis of the relative rank of the result amongst the other top 10 results. Because we do not have access to the mechanisms within the algorithm that confers this ranking, we can only analyse the importance of certain sources over others by result output.

Classification is a mechanism through which results are grouped within search engine results. From a structural viewpoint, this can be simply the classification of
results as categories such as ads, news items, images, or organic results. This results in different treatments of the results within the search page (e.g. the placement of ads at the top of results with an included small disclaimer). This classification gives us insight into the different types of results that have been connected and displayed in response to user queries. It also allows us to better understand how the space of the results page is carved up to display different types of results.

*Association* is the marking of the relationships of results and sources to each other. This criterion requires a more in-depth analysis of results as it is not made explicitly clear on the first page of results analysed in this project. However, it is important to keep in mind throughout the analysis that the relationships between sources and subjects exist outside of the first result pages analysed here. Where possible, I have tried to follow up these connections to make clear the context in which results relate to each other.

*Filtering* is the inclusion or exclusion of results based on criteria imposed by the algorithmic controller. This is perhaps the most interesting criteria for analysis in this project’s purpose and serves as an important lens through which to analyse the results. There are two aspects to consider here: the content of inclusion/exclusion of results and the rate of change of inclusion/exclusions of results amongst donor groups. By including our “baseline” VPS donors as separate datasets, the analysis included in this project should be able to give a small indication of the extent of intervention of algorithmic responses across donor groups. Indeed, this strategy will also allow us to address the basic question of whether ad inclusion persisted after Google’s policy intervention.

Finally, I have included an additional criterion in my analysis of results due to the nature of unproven stem cell treatments as real, potentially invasive medical interventions; risk. This strategy is guided by a “stoplight” system of classification that has been developed in part by advice from researchers at the Centre for Regenerative Medicine at the University of Edinburgh. This classification ranks results based on their potential for harm to patient users. This classification scheme includes tags for potential for bodily harm and potential for financial harm. The
classification of potential for bodily harm of the patient/user is informed by the work of EuroStemCell on proven effective and safe stem cell treatments. It is my strategy that any proffered treatment that is not included in the EuroStemCell list of approved and proven treatments has the potential for harm to patients. This includes - most importantly for the purposes of categorization of results - any treatments that are marketed as mesenchymal stem cell, mesenchymal stromal cell, adipose stem cell, or fat cell treatments.

I have made the conscious decision to include commercial intent as a classification as there are several knock-on effects to the integration of profit-driven private medical and direct-to-consumer offerings. I recognize that this may be controversial as it paints with a broad brush a potentially negative tint on the operations of private medical practices. However, as the context of stem cell tourism and medical access inequality illustrate, economic risks work together with more immediate risks of bodily harm and have potential for long-reaching socio-economic impacts.

Diakopoulos’ criteria of prioritisation, classification, association, filtering, and relationship to risk drive the analysis of the datasets produced by this project. When applied to the crowd sourced Blackbox audit, these lenses of analysis should grant insight into the quality and limits of results generated by queries related to stem cell treatments on Google Search.

4.8 PARTICIPANTS: REAL AND VIRTUAL DONORS

Relying on recruitment through the EuroStemCell digital network, personal outreach to patient advocacy groups like Parkinson’s UK and Spotlight YOPD, and collaboration with research centres like the Anne Rowling Clinic, I was able to recruit 139 individual data donors to take part in this study. Additionally, we automated 24 sock-puppet VPS donors distributed across each geographic territory. The 139 human donors contributed in varying patterns as donation only occurred when a donor’s computer was on and browser application open. VPS donors, however, donated consistently at 4 hour intervals throughout the study period. As will be explained in the next section, an initial period of data collection fluctuation (due to
human and VPS donor onboarding challenges) preceded a stable collection period of over 2 months.

<table>
<thead>
<tr>
<th>Data Donor Type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>139</td>
</tr>
<tr>
<td>Virtual Personal Server (Sock-puppet)</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>163</strong></td>
</tr>
</tbody>
</table>

**TABLE 1: BREAKDOWN OF DONOR TYPES**

As suggested earlier, the limitation of data collection deployment to the UK, US, Canada, and Australia was due mostly to English being the main operating language of user populations. I considered including Ireland – particularly given the potential for tracking the impact of European Union regulation impact. However, I deemed the population size to be too small for this study. Future work would benefit from including Irish data.

Additionally, as this was a first attempt at deploying this specific methodological tool in a health care-specific context, I could not guarantee the number of participants. Indeed, disparities between the human donor recruits pushed me towards the limiting of my analysis to specific geographical locations.

Geographical breakdown skewed towards the UK and US. This is perhaps due to two factors. The first is the relatively large number of human volunteers recruited from the UK through local Parkinson’s and Multiple Sclerosis patient organisations. As I am embedded within the UK medical research community, it was relatively easy to on-board local data donors. However, outreach through external patient and research networks in the US (and, to a lesser extent, Australia) through personal outreach and EuroStemCell’s media presence allowed for the collection of a sizeable number of returns. Additionally, we ran into significant difficulties with our virtual donor configuration in Canada which led to extremely limited returns captured. I made the decision from this to cut Canada and Australia as part of my major data
exploration. Instead, I have focused my data exploration on the UK and US findings as most indicative of larger themes.

<table>
<thead>
<tr>
<th>Country</th>
<th>Returns Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom (UK)</td>
<td>116,780</td>
</tr>
<tr>
<td>United States (US)</td>
<td>66,586</td>
</tr>
<tr>
<td>Australia (AUS)</td>
<td>27,077</td>
</tr>
<tr>
<td>Other</td>
<td>15,904</td>
</tr>
<tr>
<td>Canada (CA)</td>
<td>252</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>226,599</strong></td>
</tr>
</tbody>
</table>

TABLE 2: BREAKDOWN OF DONOR RESULTS BY COUNTRY

The breakdown of those human donors who identified as impacted by either Parkinson’s Disease or Multiple Sclerosis (as either primary patients or carers) was heavily skewed towards Parkinson’s patients. However, the inclusion of 19 individuals impacted by Multiple Sclerosis across the study period provides a helpful, if slightly weaker, data set from which to judge trends within Google Search results. All but two of these Multiple Sclerosis patients were in the UK which created a rich geographically-bound data set to investigate.

<table>
<thead>
<tr>
<th>Stated Condition Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parkinson’s Disease</td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
</tr>
</tbody>
</table>

TABLE 3: BREAKDOWN OF CONDITION ASSOCIATION OF THOSE IDENTIFIED AS IMPACTED BY DISEASE

In the next section, I will explore the preparation of the data collected from both human and virtual data donors and the analytical tools I used to create a detailed capture of this moment in stem cell treatment information on Google Search.

**4.9 PREPARING AND STATISTICALLY ANALYSING COLLECTED DATA**
Between October 1st, 2019 and February 2nd, 2020, the data donation campaign collected a total of 226,599 single donations of Google Search engine result pages from 163 participants (including 24 virtual personal servers). Of the 163 participants, 102 showed active engagement and regular donation. This resulted in over 1 million individual result links and associated metadata. Data donations stabilised in November 2019 after an adjustment period with consistent rates of daily donation (Figure 12).

![Data Donations](image)

**FIGURE 12: DATA DONATIONS. FIGURE COURTESY OF REBER, KRAFFT, KRAFFT, ZWEIG, AND COUTURIER 2020.**

With the support of Martin Reber and Roman Krafft, we prepared the raw data for analysis; parsing out unreadable and immaterial metadata and creating a clean dataset upon which I applied the coding described in this chapter. Once this coding and data preparation was complete, I began the process of adapting the dataset for large scale statistical analysis and trend tracking. To handle the large-scale data files and preparation, I utilised the proprietary data handling software, CSV Explorer (https://www.csvexplorer.com/).

After experimenting with a variety of statistical programs capable of meaningful, large-scale qualitative and quantitative data extraction and analysis of this particular dataset, I settled on Graphpad Prism (Prism 2021); a commercial graphing and statistical software package. Although Prism is primarily used in bioinformatics, I found that it straightforwardly and effectively created opportunities to interrogate the
dataset to address my questions. For raw text analysis, I used Voyant (www.voyant-tools.org), an open-source text analysis software from the University of Alberta, Canada.

It is important to note that I have included data from the full data collection period in my graphical analysis in Chapter 5 on Advertisements. However, this means that the graphs include the stabilisation period between September and early November 2019. As Figure 12 indicates, return collection stabilised after November 4th. While my written analysis throughout works from this point of stabilisation onward, I have included the full data set for the purpose of documentation.

4.9.1 A SHORT ASIDE ABOUT STATS

To extract the most meaningful and relevant findings out of the mass of data generated from the data donation campaign required me to venture into the world of large-scale statistical data analysis. Luckily, I discovered that a few selected concepts are all that is necessary to grasp nuanced and illuminating themes that run through the dataset. These concepts applied to the data allow us to graph changes over time and plot relational values that reflect the changes in output following the implementation of Google’s policy change. It also allowed me to grasp the weight of actors supplying ads to the Google SEM platforms and how their positionality vis-a-vis Google Search end-users shifted across the study period.

This analysis relies on the statistical testing model, null-hypothesis significance testing (NHST). NHST is a testable statistical hypothesis that establishes significance from an assumed deviation of null change. Although there is significant criticism of NHST for fine grained analysis (Gliner et al. 2002), the model fits for the purpose of simply establishing the presence of change over time. Within this testing model, there are one key concept to be understood as a hook of my analysis: p-value.

P-value indicates the level of significance or difference as compared to no effect. It is a helpful tool for indicating significance (rather than causality). However, recently it has come under criticism for too high a threshold of inclusion leading to a greater
chance of finding significance in large datasets (Vidgen and Yasseri 2016).
Traditionally, one would use Fisher’s established measure point of $p=0.05$ to indicate whether a result is significant or not significant as this is two standard deviations away from the mean for a normal distribution (Fisher 1934). However, to offset for the critique of $P$-values, I adopted the more stringent measure of $p=0.01^{27}$. A $P$-value lower than 0.01 indicates a strong likelihood that the hypothesis that the change is zero is unlikely.

Correlation and causation in the larger sociological analyses are important to keep distinct. It is important to note that we cannot determine from NHST analysis alone the specific mechanisms of how or why any changes in output occurred within the Blackbox of Google Search. One can only note that a change has occurred and dig deeper within the situated analysis to estimate causes. I will examine the implications of these findings in Chapter 7 and 8. This, fundamentally, is the core challenge of analysing Blackbox technology in a social scientific context. However, equally important is the stated assumption that Google’s stated policy change will have an effect. This analysis is therefore based on the supposition that a change observed over the period will be in part due to a change in algorithmic governance by Google.

4.10 CODING RETURNS

To capture a useful and informative snapshot of the over 1 million results for stem cell treatment queries captured, I first needed to create scalable and meaningful qualitative coding of ad host sources upon which I could apply my statistical analysis. This required the creation of unique coding criteria through which the ads and organic results could be understood and contextualised. This included the characteristics of the source actor who produced the result content; the results’ positionality vis-a-vis research questions of unproven stem cell treatments, commercialisation, and target audience; and any novel data anomalies produced by the data collection itself.

---

27 Some statisticians suggest a $P$-Value of $p=0.001$ to reduce chance of false significance (Vidgen and Yasseri 2016). However, given the size, granularity, and nature of my dataset, I felt that a value of $p=0.01$ was appropriate.
These criteria were then translated into numerical representations for analysis. Before I explore these coding categories, there are two points I would like to cover. Firstly, one might ask why I am including this full list in the chapter text rather than an appendix. To that I would argue that in my own experience of reading through large-scale statistical analysis, one often loses sight of the end-result. It is important to establish a qualitative sense of what is being displayed in these ads and organic results to invest in the findings. It is because these characteristic buckets will form the backbone of future findings that I have included examples of each category and in-depth descriptions of each.

Secondly, it should be noted here that at this point of analysis I am relying on my own experience as an “expert” in the field of stem cell research gleaned from my work at EuroStemCell. As I have worked for over 6 years on behalf of the European Commission, University of Edinburgh, and the MRC Centre for Regenerative Medicine, I recognise my experience may colour my categorisation of some ad sources. To mitigate the bias that I might impose on the categorisation of these actors, I have run these categories by neutral colleagues within and outwith the stem cell research community.


Further to these categories, I added four additional criteria:

‘Does the host source include promotion of unproven stem cell treatments?’
‘Is the host source primarily a commercial company?’
‘Is the result content targeting patients and/or carers?’
‘Is the result content primarily targeting medical professionals or researchers?’
The following section gives detail to the categories of ad and organic result types that I applied to the dataset with description, reasoning for specific adoption, and any incidental characteristics outlined as necessary. These categories and the secondary criteria form the backbone of the statistical analysis that follows.
Biobanking is the long-term storage of genetic material from patients and donors for the purpose of future research or treatment (Figure 13). Traditionally within the research community, biobanking is used as a fundamental aspect of medical research in order to provide tissue samples for laboratory testing (Hofman et al., 2013). However, private biobanking - or the collection and storage of genetic material as a paid service on behalf of individual patients - has grown as a commercial enterprise since the early 2000s. These collections largely take the form of cord blood collected at birth and are often advertised to new parents as a speculative investment for future treatments. These private biobanks occupy a potentially difficult space between commercial enterprise and speculative medical treatment which relies heavily on a consumer-as-future-patient model (Turner, 2013).

Within the advertisement data set collected, all biobanking sources were limited to the latter examples of commercial biobanks targeting non-researcher publics.

Meta Content: “We store umbilical cord cells at our advanced UK lab, for decades to come. View our competitive prices. 0% APR finance options! Low rate finance. Interest Free Payments. Present In 75+ Countries. UK’s Largest Family Bank. 190,000 Samples Stored.”
Biotechnical companies represented a large proportion of ad returns within the data set while having the distinction of being one of the few categories to target medical and research professionals rather than the general public or patients. Biotech ad and organic returns included advertisements for products in support of research such as clinical and diagnostic tools for use in laboratory research settings (Figure 14).

CHARITY/NGO

Returned ads and organic results within this category included ads from charitable or non-governmental organisations operating in support of condition-specific patients and research. These results did not include any offers of paid treatments or medical services and restricted commercial activity to organisational fundraising and research support. They included organisations such as Parkinson’s UK, the Michael J. Fox Foundation, and the MS Trust (Figure 15).

Meta Content: “With your help we will work to change attitudes and find a cure. If you're affected by Parkinson's, we will help you find the right support. Make A Donation. Become A Volunteer. Multiple Payment Options.”
CLINICS

Ad and organic returns in this category represent private medical providers of biomedical treatments and interventions for patients (Figure 16 and 17). These clinics offer medical interventions aimed expressly at alleviating causes, symptoms or side effects of clinical diagnoses in patients. This category includes small-scale, treatment-specific clinics, rather than large scale medical practices like hospital systems (most prevalent in the United States). It should be noted that inclusion in this category does not explicitly link an ad or organic source to the propagation of unproven stem cell treatments. That delineation will be discussed later as a secondary criterion. However, this category has a clear relationship to commercial intent and is addressed directly by Google’s terms of advertising which will be addressed in Chapter 5.

Diabetes, Alzheimer's disease.”
Ad and organic returns in this category included recruitment websites for clinical trials as well as databases of clinical trials currently recruiting patient involvement (Figure 18). These advertisements included both recruitment ads and organic results aimed at patients as well as those targeting medical research professionals for the purpose of aiding future clinical trials. Although some clinical trials are pay-to-participate on the part of the patient, this commercialisation aspect was not addressed in this category. Rather, this distinction was applied within the secondary criteria.

Meta Content: “Participate in a clinical trial for symptoms of nOH associated with Parkinson’s.”
COMPLEMENTARY TREATMENT

Complementary treatment ads (Figure 19) and organic results (Figure 20) include website advertising non-invasive, non-medical interventions for the managing of conditions and diseases. These ads and organic results do not include curative treatments, but rather promote the sale of products for the management and alleviation of symptoms caused by conditions like the queried Parkinson's Disease and Multiple Sclerosis. The commercial intent of these sources was determined by the additional criterion.

FIGURE 19: SCREENSHOT FROM MENDIRF.ORG. CAPTURED BY AUTHOR 2.7.2021.

FIGURE 20: SCREENSHOT FROM THERACYCLE.COM. CAPTURED BY AUTHOR 5.7.2021.
Meta Content: “Proven Therapy for Parkinson’s Disease with our unique motor-assisted Therapy Bikes.”
LIFESTYLE AND NON-TRADITIONAL ADVICE

Ads and organic results in this category promote websites aimed at holistic or non-traditional health advice (Figure 21). They include blog-style advice and recommendations, but do not include recommendations for major medical interventions like stem cell treatments. Information and product catalogue aggregators which compile health-oriented promotional material are included here.

NEWS AND INFORMATION

Ads and organic results returned in this category included websites strictly focused on information and news dissemination on health and regenerative medicine topics (Figure 22). These include ads and organic results from science news aggregators and do not include paid promotional material specific to the query diseases or associated treatments. Any additional commercial intent of these sources is determined by the secondary criterion.

FIGURE 22: SCREENSHOT FROM SCIWORTHY.COM. CAPTURED BY AUTHOR 01.7.21.
Meta Content: “Sciworthy.com Provides New Science Articles For People To Better Understand Science.”
Ads and organic returns in this category included direct marketing of pharmaceutical medicines as treatments for the queried diseases labelled ‘prescription treatment websites’. Within ad returns, this category overwhelmingly included obfuscated ads generated by alternative ad-injection services like Google Lead Services platform and Amazon.com’s ad exchange. These include obscured links to sites such as amazon.com and info.com. While this does not allow for the interrogation of the host site, the project’s data collection was able to capture the associated ad text which grants some insight into the ad’s content and links (Figure 23).

Figure 23: Example image of ‘prescription treatment websites’ unreadable ads within the dataset. By author 20.7.2021.

Meta Content Example: “Ask your doctor whether this new Parkinson’s disease therapy would benefit you. Learn how this medicine is unique. View prescribing and important safety information.”

These obfuscated pharmaceutical ads indicate the inclusion of secondary ad return systems within the Google SEM ecosystem as links are provided by the ad exchange platform (such as ad exchanges services by amazon.com or google.com or ad aggregators like seatlesbest.clickfunnel.com).

This category was particularly influenced by geographic origin of the donor (whether human or VPS) as direct-to-consumer advertising of pharmaceuticals is not permitted in the UK, Canada or Australia. Additionally significant in this category were returns on proprietary brands of Levadopa, the primary pharmacological treatment for Parkinson’s Disease. The regulatory implications of this category will be discussed in Chapter 7 and 8.
This category included direct ads and organic results produced by research institutions or organisations engaged in fundamental research in the area of stem cells and regenerative medicine (Figure 24). Returns included ads from organisations such as the New York Stem Cell Foundation (nyscf.org) and the Australian Regenerative Medicine Institute at Monash University (armi.org.au).

FIGURE 24: SCREENSHOT FROM NYSCF.ORG. CAPTURED BY AUTHOR 22.2.2021.
Like the Pharmaceutical category, this category encompassed ad and organic returns that fell outside the aforementioned categories and largely included obfuscated ads generated by alternative ad-injection services like Google Lead Services platform and Amazon.com's ad exchange. While some of these links were unrelated to the input query, a majority of these obscured links returned associated text that indicated a query-specific return, many of which included claims to stem cell treatments for the target diseases of Parkinson’s Disease and Multiple Sclerosis (Figure 25).

**FIGURE 25: EXAMPLE IMAGE OF ‘OTHER’ UNREADABLE ADS WITHIN THE DATASET. BY AUTHOR 20.7.2021.**

Meta Content Example: “Treating Parkinson’s using stem cells & natural therapies”
In order to develop a clearer picture of the aims of actors and potential risks of advertisements and organic results, I included four additional criteria to evaluate each ad return. These additional criteria allowed me to interrogate the data in more depth by assigning values of characteristics that could then be analysed over time and in comparison. Each of these criteria could apply to all, some or no ad returns.

**DOES THE HOST SOURCE INCLUDE EFFICACY CLAIMS ABOUT UNPROVEN STEM CELL TREATMENTS?**

In order to meet this criterion, the linked ad and organic webpage must include the promotion of unproven stem cell treatments - either within the associated website URL or captured associated text. As discussed in Chapter 2, I have marked the delineation of proven and unproven stem cell treatments through documentation from expert research institutions and networks; the ISSCR, EuroStemCell and the Australian Stem Cell Network as well as the discussions of STS critique within Chapter 2. In many cases, the categorisation was determined by the site’s inclusion of offered stem cell intervention services for the treatment of conditions and diseases for which there is a strong lack of evidence and history of misapplication (e.g. Autism, see *Figure 26*).

![FIGURE 26: SCREENSHOT FROM PRIVATE CLINIC ADVERTISING AN UNPROVEN STEM CELL TREATMENT, EMCELL.COM. CAPTURED BY AUTHOR, 15.12.2020.](image)

By cross-referencing the treatments advertised or promoted with the list of proven safe and effective treatments, I assigned a binary value of 0 (no unproven claims or null) or 1 (unproven claims). This additional criterion enabled the distinction between generalised clinics and medical providers who do not include unproven stem cell treatments in their advertised services (such as hospitals) and clinics that explicitly offer unproven stem cell treatments as paid services (Figure 27).

![SwissMedica Ad](image)

**FIGURE 27:** SCREENSHOT FROM PRIVATE CLINIC ADVERTISING, SWISSMEDICA. CAPTURED BY AUTHOR, 5.3.2021.


**IS THE HOST SOURCE PRIMARILY A COMMERCIAL COMPANY?**

This criterion was applied to determine whether an ad or organic result primarily targets the user for the purpose of commercialisation. This includes the sale of products and services to any of the targeted audiences. However, it does not include the solicitation of donations in the case of non-profit entities like charities. This criterion assigned a binary value of non-commercial (0) or commercial (1) to the host source.

**IS THE RETURN TARGETING PATIENTS AND/OR CARERS?**
This criterion was applied to determine if the target audience of a displayed ad or result was patients and/or their carers. For this (and throughout this project), I have deployed the NHS definition of carer as “anyone, including children and adults who looks after a family member, partner or friend who needs help because of their illness, frailty, disability, a mental health problem or an addiction and cannot cope without their support” (NHS, 2021). This criterion assigned a binary value to indicate whether or not the ad targeted an audience of patients and/or their carers.

IS THE RETURN TARGETING MEDICAL PROFESSIONALS OR RESEARCHERS?

This criterion was applied to determine if the target audience of a displayed ad was medical professions and researchers in regenerative medicine. This included medical practitioners, laboratory scientists, and other expert professionals in either patient-facing clinical practice or in basic research. This criterion assigned a binary value.

4.11 METHODOLOGICAL CHALLENGES

Through the practice of undertaking algorithmically-mediated platform auditing, several challenges became apparent. Once I identified a Blackbox audit as my preferred methodology for approaching my research questions, it then became a question of finding the resources necessary to adopt it. Developing and deploying the search engine audit methodology requires time, funding, and specific competencies.

These resource challenges posed significant risks to my project throughout its timeline and were often rooted in the difficulties of engaging with labour-intensive and iterative technologies. They included the challenge of recruiting help in adopting existing browser plugin software to my specific study, the funding of developer time and hardware support (including the setup and monitoring of VPS servers), participant recruitment and training, the storage of large volumes of scraped data, and access and competencies in data analysis software. Here I would like to go briefly into some of these obstacles for future guidance.
The capture of off-Google Search result content through their links might improve the value of the dataset collected in future studies. As discussed in this chapter, the EuroStemCell Data Donation plugin only collected HTML markup as rendered through the Google Search single results page. It did not capture the pages that were linked. This means that any changes in content on these linked pages outside of the study period was not captured. This also required quick follow-up data handling on my behalf to apply the study-specific coding to the datasets before they became irrelevant. Adding this additional capability of off-platform content retention would require further software development outside of the plugin whereby the linked content might be captured. Immediately, this provokes two potential challenges: the potential for undue stress on the accessing machine of the donor and the need for increased data handling, indexing, and storage on the side of the data controller. These are not unmanageable challenges, but would need to be addressed to enable future research.

Throughout the data collection period, I encountered several issues with the ongoing upkeep of the VPS data donors. The cost of maintaining the VPS ranged from £80-100 per month of operation. I was able to secure funding for this through the EuroStemCell. However, this may be a barrier for other researchers without access to secondary funding. The VPS also required daily monitoring for any down time during which they might have missed a data collection window. I was aided here by the automatic tracking services of some of the virtual server hosting services.

Google Search is in a constant state of flux. This posed a particular challenge as any change in the markup of Google Search had the potential to disrupt the data collection approach. The data donation plugin relies on markup within the public source code returned in Google Search results to capture and bookend each entry. Further obfuscation or shifts in this tagging on the part of Google would render this data collection unusable. I was particularly luck that during the data collection period no significant changes were made to the Google Search result page. However, this a concern of which future research should be aware and integrate into any maintenance strategy.
As mentioned in this chapter, the Google Search platform automatically flags for signs of automated data scraping as protection against both DDoS and other malicious automation of platform interaction. This meant that my VPS often encountered CAPTCHA checks or blocks imposed by Google that had to be remedied before data collection could continue. This required constant supervision of the data scraping process during the study period.

Participant recruitment – although relatively successful with the support of the EuroStemCell project – could have been more successful. This relative success was due to my proximity to patient advocates and organisation for Parkinson’s Disease in the UK and access to the EuroStemCell online presence to recruit for Parkinson’s Disease, Multiple Sclerosis, and Diabetes. I initially included Diabetes as well in my research. However, low turnout and involvement as well as an issue with my VPS servers collecting diabetes queries led me to drop the condition in my final analysis.

The methodology itself also proved challenging for human donor uptake. This was not due to lack of interest. Many who I spoke to were interested in having discussions around their experiences with Google Search and their healthcare decision-making process. However, this did not always translate to becoming data donors. Misunderstandings of the methodology also posed a difficult challenge. Questions of user data handling meant that I spent a significant amount of time communicating and presenting documentation on how the plugin worked and the limits of the information that the plugin was able to collect.

Additionally, the targeted patient groups often had difficulties adapting to the digital approach to data collection. Outside of a large group of UK-based early-onset Parkinson’s patients, most of the data donors impacted by Parkinson’s Disease were over the age of 60 years old. The computer literacy required to install the data donation plugin required additional resources. Additionally, many potential participants mentioned that they did not use their personal computers to access Google Search, but rather favoured tablets. Unfortunately, this iteration of the data donation plugin was unable to accommodate tablet browsers.

In the future, I would recommend focusing exclusively on potential user groups and
engaging in further on-boarding work to make the process of engagement easier. As a first approach, I am pleased with the engagement the data donation campaign garnered, but note that there is significant room for improvement.

While I was pleased to work in anonymity at the point of collection, data handling of the scraped results still proved difficult. Each browser’s submission of scraped ad and organic results every 4 hours created significant strain on the storage servers housed at the University of Kaiserslautern. Additionally frustrating with data handling was the unstructured nature of the raw data collected. This was – in part – due to the limitations of development time and funding to create automated data cleaning at the point of storage. Google’s iterative process of publishing also contributed to this instability and the requirement for the raw data to be made readable by hand. Instead, I cleaned the raw data into digestible and manageable databases. In the future, it would be helpful to develop automatic scripts to clean and organise captured data. However, as noted, this would need to be monitored according to iterative changes in markup on the platform.

The size of the databases created presented a particular challenge for my research due to the memory and processing power needed to handle database manipulation. I was fortunate to have access to high-end hardware with which to carry out the database analyses. However, I ran into a number of progress-hindering challenges of hardware limits in trying to render the large datasets. This difficulty might be mediated in the future through cloud-based dataset processing.

In terms of access to data analysis software, I was fortunate in my proximity to the University of Edinburgh’s Centre for Regenerative Medicine and the software made available to researchers through Clare Blackburn’s lab. This includes access to the license-required statistical analysis software, GraphPad Prism. For analysis of the raw data, I used the proprietary CSV analysis software, CSV Explorer. I was able to employ open-source text analysis tools like Voyant free of charge.

With these difficulties in mind, it would be useful to cultivate cooperatively developed and maintained open access tools for data scraping and auditing of this type. This is already underway at the Digital Methods Initiative at the University of Amsterdam.
However, access is strictly limited to students working on specific projects due to user agreement considerations as well as developer time. As technology develops, it would be useful to explore open access tools for data scraping that might get around these institutional and software-specific limitations.

4.12 METHODOLOGY OVERVIEW

This chapter – as both an exploration of my adopted methodology and a finding in itself – answers my secondary research question: How can Google Search – as a proprietary, algorithmically-driven search engine – be studied empirically? What methodologies can be employed and why is this decision important?

Addressing the difficulties of studying Google Search within social science research often feels like an overwhelming challenge created by the sheer size of this opaque and ubiquitous platform. The balance of establishing clear logic to scrape clean data, integrating data collection into an ever-evolving, proprietary and unreceptive platform, and managing donors - both real and virtual - can stop research before it is undertaken. However, the insights I was able to glean from the results generated by the EuroStemCell Data Donation plugin provide unique insight into this critical area of digital health.

In the following chapters, I will present the findings of the 226,559 individual data donations (resulting in over 1 million results) collected through this methodology over 4 months in 2019/2020. In Chapter 5, I focus on commercial advertisements returned by Google Search and document the impact of Google’s policy change towards stem cell treatment advertisement. In Chapter 6, I map the characteristics of organic search results generated by Google Search. In Chapter 7, I explore the search results pages in their entirety with a particular focus on mechanisms for commercial activities, Knowledge Graph and additional markup found in the process of my data collection. These chapters not only shed light on the impact of the Google Search platform on stem cell treatments but also provides a unique application of novel methodology for digital health.
In September 2019, Google posted an entry on its advertising policies help page entitled, “Update to Healthcare and medicines policy (October 2019)” (Figure 28). Without much fanfare, the post outlined that Google’s Healthcare and medicines advertising policy would be updated to explicitly “prohibit advertising for speculative and experimental medical treatments” to come into effect on October 1st, 2019. It stated that products and services promoting “stem cell therapy, cellular (non-stem) therapy, gene therapy, and similar forms of regenerative medicine” would not appear in search engine results after the policy implementation deadline.

Update to Healthcare and medicines policy (October 2019)

In October 2019, the Google Ads Healthcare and medicines policy will be updated to prohibit advertising for speculative and experimental medical treatments. This policy will apply globally. After the policy update, ads for speculative and experimental medical treatments will no longer be allowed.

New Speculative and experimental medical treatments policy

This policy will prohibit advertising for products or services that promote speculative and/or experimental medical treatments.

Examples of products and services that will be prohibited (non-exhaustive)

- Stem cell therapy, cellular (non-stem) therapy, gene therapy, and similar forms of regenerative medicine,
- platelet rich plasma, biohacking, do-it-yourself (DIY) genetic engineering products, and gene therapy kits

(Posted September 2019)

FIGURE 28: SCREENSHOT FROM THE GOOGLE HEALTH POLICY BLOG, CAPTURED BY AUTHOR 02.09.2019
In Mountain View, California this blog entry (Figure 28) may have marked an iterative update to a living policy. In Edinburgh, Scotland, I began to move quickly to put into place the technical oversight and donor recruitment needed to capture this algorithmic shift in action by October 1st, 2019. In discussions with my colleagues and collaborators, there was speculation that this shift in policy would make my planned capture of ad returns on Google Search obsolete.

This was not the case. From September 30th, 2019 to March 5th, 2020, I was able to capture over 25,000 unique advertisements displayed to both human and virtual data donors returned on queries related to stem cell treatments for Parkinson’s and Multiple Sclerosis across four countries.

In this chapter, I answer my secondary research questions in regards to Google Search advertisements: How is the mediation of information on stem cell treatments shaped by structural factors within the platform of Google Search? What impact did Google’s policy change have on paid search advertisements relating to stem cell treatment queries? Does the information found differ depending on the user’s or query’s characteristics? When? And, hopefully, why? In answering these questions, I address important, underlying issues around platform mediation of advertisements with potential risks; marketing strategies of both for-profit and charity organisations; the potential for vulnerable user targeting; and commercial activity around health and wellbeing. I will show how a policy change announced quietly on a blog resulted in real changes in algorithmic outputs on a platform used by millions of individuals daily. In this chapter. I provide a snapshot of algorithmic mediation between content-producing and content-consuming actors and illustrate what happens to this interplay when changes or modifications are made to the algorithmic rules that govern users’ exposure to content.

In this chapter, I explore what was my initial motivation for undertaking this research; to analyse the prevalence of advertisement within search engine results returned on queries relating to stem cell treatments on Google Search. In particular, I use the period after Google’s October 1st, 2019 policy change and implementation as a
springboard for discussion on the state of Search Engine Marketing (SEM) and stem cell treatments, their prevalence and impact, and the potential for future regulation.

What I will demonstrate is that, while the display rates of ads explicitly offering unproven stem cell treatments decreased across the research period, Google did not implement a blanket prohibition of advertising connected to queries of stem cell treatments within Google’s ad exchange environment. As such, problematic ads persisted. Indeed, problematic advertisement continued to be captured by both virtual and human data donors across the study groups and geographic locations in the 5-month data capture period post-policy change and beyond. While impact of personalisation on the display of ads was potentially limited (Krafft et al., 2020), there is evidence that human donors were significantly more likely to be presented with ads than their virtual donor counterparts.

Additionally, it appears that the change in policy coincided with a shift in both the rate of advertising and language used by problematic advertisers seeking to avoid implication in the ban while still maintaining access to users. Finally, at the end of the chapter I raise further questions about the oversight framework established by Google in the case of advertisements of sensitive healthcare information as well as the types and motivations of actors who engage in paid SEM around the topic of regenerative medicine.

5.1 FINDING 1. BLANKET BAN EFFECTIVENESS

Google’s policy shift was not a blanket ban on advertisement around stem cell treatment queries as advertisements continued to be displayed to Google Search users after the ban was enacted on October 1st, 2019

The first and most obvious finding was that the ban did not halt the inclusion of all paid advertisements on results pages on queries related to stem cell treatments and Parkinson’s Disease and Multiple Sclerosis after October 1st, 2019. Indeed, the data donation campaign captured a total of 25,951 paid targeted advertisements from a series of 285 host sources across the study. However across the study period, the frequency of ads declined for some ad return category types. It should be noted that
these ads were not all for private treatments, but an array of categories that will be explored later in this chapter. This is important because it indicates that any intervention algorithmically-mediated mechanism within the Google Search platform did not entirely exclude advertisement capabilities within search results despite the shift in policy.

Figure 29 indicates that the general rate of all advertisements tended down\textsuperscript{28}. However, the rate of ad returns did not reach zero at any point across the study period. Adjusted to compensate for stabilisation of donation rates in early November 2019, Figure 29 shows a consistent downward trend with a clearly significant \textit{p value} of <0.0001.

\textsuperscript{28} As mentioned in Chapter 4, I have included data from the full data collection period for documentation purposes.
It is important to note that these returns were skewed towards searches around Parkinson’s Disease (Figure 30). This is due to the largest recruitment of human participants into the UK and US Parkinson’s studies which then maintained the highest level of consistency in human data donation throughout the study period. As explored later, the inclusion of human data donors had an impact on frequency of ad returns.
Interestingly, although the sample size of US donors and returns was smaller than the UK, the frequency of ads included in search results was markedly higher at the beginning of the collection period (Figure 31). This may indicate a greater integration of ad mechanisms within US-initiated search queries on a structural level within the Google Search platform.

![Ad Frequency by Query](image)

**FIGURE 32: TOTAL AD RETURNS BY QUERY**

Queries returned advertisements at different frequencies with queries around Parkinson’s Disease and general stem cell treatment terms resulting in the highest occurrences (Figure 32). Noting the breakdown of ads returned by study and the demographic of data donors explored in Chapter 4, I am hesitant to state that these rates are influenced by the query content. Rather, there are more returned advertisements around these queries because there were more data donors impacted by Parkinson’s Disease than other stated diseased and conditions. The data set is simply larger.
Of the advertisement host sources, there was a clear population of significantly engaged content producers (Figure 33). While 80% of advertisers appeared less than 50 times in the data, the top 10 advertisers constituted 44% of the ad returns.

Each of the category types were represented to varying degrees of significance within the ad returns (Figure 34). Top advertisers had significant representation of Pharmaceuticals, Clinics, Charities/NGOs, Biotech and Biobanks. Constituting over 50%, the two largest categories of returns were Clinics and Charities/NGOs. Ads displayed from these source categories tended to link web pages with calls to actions for either paid treatments (Clinics) or to donations (Charities/NGOs).
This indicates that the ad ban relied on the platform-mediated identification and quarantining of specific advertisement content. This finding does not indicate how that intervention was achieved. Indeed the continued return of commercial ads from private clinics, indicates that the intervention was only partially successful.

The benefit of the doubt must be given and the question must be asked; does the downward trend of ad returns captured by the data donation campaign indicate that Google will in time complete its implementation of the ban on promotion of unproven stem cell treatments? Is this data a snapshot of a reality in Google Search that no longer exists? Have these ads disappeared in the period since the data donation campaign?

As of July 8th, 2021, over a year after this data collection period, they have not. Instead, problematic advertisements explicitly banned by Google’s own policy continue to be presented to users (Figure 35). Consequently, this suggests that content producers and Google can engage in economic activity around unproven stem cell treatments despite being at odds with internal company policy.
Categories of ad returns were not impacted uniformly across the study period.

Over the study period, the frequency of ad display varied across the source categories. Some categories experienced significant downward trends (Biobanks, Clinics, Clinical Trials, Pharmaceutical, and Research) while the display rate of others were insignificantly affected and remained consistent (Biotech, Charity/NGO, Complementary Treatment, Lifestyle and Non-Traditional Advice, News and Information, Other) (Figure 36). This indicates that the implementation of Google’s ad policy was targeted at the specific ads themselves rather than ads as a whole. This also supports the theory that the ad ban would not be implemented across all categories of sources, but would be taken on a case-by-case basis (Table 4).
The disparity in impact becomes particularly clear in the comparison of ad returns for Clinics vs. Charity/NGO (Figure 37). It is clear that there is a sharp drop in Clinic ads.
getting through to users from the beginning of November 2019 with a downward trend continuing through the study period. At the same time, ads from Charities/NGOs remain relatively stable throughout. It is apparent here that a shift in governance has occurred on the side of the ad exchange platform which distinguishes between the types of ad sources.

![Charity/NGO vs Clinic Ad Returns](image)

**FIGURE 37: AD RETURNS OF CHARITIES/NGOS VS CLINICS OVER TIME**

As in Finding 1, this reveals platform intervention in some form to distinguish content qualitatively. The mechanisms for this intervention is not made clear in these findings. However, the consistency in appearances of Charities/NGO might indicate that these sources have been identified as acceptable sources of information and have, therefore, been whitelisted through the Google Search advertisement mechanisms. The web presence of these organisations are largely established and authenticated through cross connection (e.g. as outwards links on other websites). In Finding 7, I will discuss this specific relationship between Charities/NGO and Google Search in more detail.

5.3 FINDING 3. IMPACT ON DISEASE-IMPACTED DONORS

Human data donors impacted by one of the associated diseases or conditions were most likely to be presented with ads.
Tracking the rate of ad return across time between human data donor groups and the VPS reveals that human data donors were significantly more likely to be presented with ads than the VPS data donors. This suggests that Google’s search platform can distinguish between users based on information gathered about the user based on previous metadata (Krafft et al. 2020). As the VPS data donors had no previous search history and no linked accounts, they would not include any metadata with which to be targeted.

![VPS vs Human Data Donor - UK Parkinson’s](image)

**FIGURE 38:** VPS VS. HUMAN DATA DONOR AD RETURNS OVER TIME IN UK PARKINSON’S STUDY

Looking further into the breakdown of results within the UK study groups for Parkinson’s Disease\(^{29}\) revealed a further significant finding (*Figure 38*). The UK Parkinson’s Diseases study groups comprised the most balanced recruitment and consistent donations of three types of data donors; donors who indicated they were impacted by the diseases as a patient or carer, a control group unimpacted by the disease, and clean VPS donors. In analysing the results, it became clear that affected data donors were more likely to be presented with ads within their collected search results than both their unaffected data donors and VPS counterparts (*Figure 29*).

\(^{29}\) Unfortunately due to low participation of individuals impacted by Multiple Sclerosis in the US and UK, I was not able to capture meaningful comparison between human and VPS virtual donors for queries around Multiple Sclerosis.
The breakdown of the source types reflected in these advertisements were consistent with the general breakdown discussed in Finding 2.

FIGURE 39: FRACTION OF ADS IN TOTAL DONATIONS PER PARTICIPANT, FIGURE COURTESY OF REBER, KRAFFT, KRAFFT, ZWEIG, AND COUTURIER 2020.

MEANS: AFFECTED 0.76, CONTROL 0.10, VPS 0.05
MEDIANs: AFFECTED 0.57, CONTROL 0.07, VPS 0.01

This graph indicates that ads – as a fraction of both ad and organic returns on a single result page – made up a more significant portion of returns for those who self-identified as impacted by one of the target diseases. When compared to non-affected human donors and the virtual donors, these disease-impacted donors saw a greater number of advertisements per search engine query. This suggests that there is some mediation of search results towards advertisements as a result of the identification of users as impacted by a condition or disease.

As stated previously, there is a limit to what this data can tell us due to the limitations of Blackbox analysis. From this data, we cannot determine the details of how or why these patients/carer groups are being targeted. However, these findings do support the hypothesis that users impacted by conditions and diseases are acknowledged and handled differently than non-affected users on the Google Search platform. This may be achieved by the platform through aggregated grouping users based on metadata like previous search history or another means of user sorting (Krafft et al. 2020).
This result has social and ethical implications for digital health and safety of patients. In particular, it is crucial to recognise that here targeted users are – through some mechanism – enacted as a discreet and identifiable patient group through their connection to disease. As introduced in Chapter 3, the rendering of patients (as users, consumers, or products) influences the kinds of information and knowledge to which they are likely to be exposed. This finding suggests that the rendering of this targetable patient group includes the introduction of Google Search platform’s commercial mechanisms (specifically, exposure to advertisement). Though this, the user has first been identified as a product through the backend of the advertisement platform. They are then rendered as a consumer (through the presentation of advertisement) in the presentation of search results. I will explore this transformation more in Chapter 7.

5.4 FINDING 4. CONTINUED APPEARANCE OF PROBLEMATIC ADS

Problematic ads for clinics promoting unproven stem cell treatments continued to appear. However, their frequency was reduced over time.

To capture the potential for promotion of unproven stem cell treatments, I needed to take a closer look at Clinics. With 27% of ads including appeals for patients in clinics, it was essential to parse the ads by the treatments offered in order to differentiate the level of risk promoted. I was interested in tallying the types of treatment and stem cell source published as part of the commercial offer of these clinics.
<table>
<thead>
<tr>
<th>Clinic Sub-Category</th>
<th>Hosts</th>
<th>Ad Count</th>
<th>Share of Clinic Returns (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved/Traditional treatments or Complementary Treatment for Parkinson’s Disease (PD) or Multiple Sclerosis (MS).</td>
<td>3</td>
<td>648</td>
<td>11.0%</td>
</tr>
<tr>
<td>Alternative/Non-Traditional (non-stem cell) treatments for PD or MS</td>
<td>3</td>
<td>464</td>
<td>7.9%</td>
</tr>
<tr>
<td>Stem cell-derived sports medicine, muscular, or chronic pain treatments</td>
<td>10</td>
<td>463</td>
<td>7.9%</td>
</tr>
<tr>
<td>Stem Cell Treatments for PD or MS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embryonic Stem Cells</td>
<td>4</td>
<td>226</td>
<td>3.8%</td>
</tr>
<tr>
<td>Mesenchymal Stromal/Stem Cells (MSC)</td>
<td>12</td>
<td>1100</td>
<td>18.8%</td>
</tr>
<tr>
<td>HSCT</td>
<td>3</td>
<td>984</td>
<td>16.8%</td>
</tr>
<tr>
<td>Unknown</td>
<td>16</td>
<td>2748</td>
<td>47.0%</td>
</tr>
<tr>
<td>Other (suspect)</td>
<td>3</td>
<td>203</td>
<td>3.4%</td>
</tr>
</tbody>
</table>

**TABLE 5: BREAKDOWN OF CLINICS BY TREATMENTS OFFERED, AD COUNT, AND SHARE OF RETURNS.**

As *Table 5* indicates, only 11% of *Clinic* ad returns included promotion of approved treatments or complementary treatments for Parkinson’s Disease or Multiple Sclerosis. The majority - whether marketed as stem cell treatments for the named diseases or unrelated stem cell treatments - constituted unproven gene and cell treatments in opposition to Google’s stated policy.

Over time, it appears that ads including unproven treatments dropped significantly compared to other ads from November 2019 onwards. However, these problematic ads did not disappear entirely and continued to be displayed to users at the end of the collection period. Moreover, after the initial drop in return rates of in November

---

30 These coding categories were initially established under my advice by Frederik Maximilian Stegner of the University of Kaiserslautern as part of his Bachelor’s Thesis in computer science in which he analysed parts of the dataset collected in the EuroStemCell Data Donation campaign for which I gave permission.
2019, problematic returns mimicked the return rates of advertisements overall for the remaining period of study. This suggests that the effectiveness of the policy change as a full ban of advertisement of unproven stem cell treatments was moderate rather than complete.

This has significant implications for my central question of Google Search as a mediator of information. It suggests that there is a limit to the control Google – the proprietor – can exert on its Google Search. As explored in the example of Holocaust denial content in Chapter 3, Google has explicitly positioned itself against direct intervention in the algorithmic governance of Google Search and the results it produces. The inability of the ban on stem cell treatment advertisements to fully embargo problematic results is a clear indication of the limitations of policy intervention. As Figure 40 indicates, the problematic ads were reduced through what I must assume was an algorithmically-driven intervention. However, they were not entirely expunged. Perhaps this illustrates a baseline of generalised returns under which algorithmically-driven intervention cannot functionally intervene – regardless of policy.

![Ads Including Unproven Treatments](image)

**FIGURE 40:** RETURNS OF ADS INCLUDING EXPLICIT REFERENCE TO UNPROVEN TREATMENTS VS ALL OTHER CATEGORIES OVER TIME

5.5 FINDING 5. TARGETED AUDIENCES
Target audience frequencies were varyingly impacted across the study period.

Here I would like to pull out a few secondary criteria for which there was a significant change in frequency over time and indicate an unequal distribution of impact over the data collection period. These unequal impacts give support to the supposition that Google’s policy change focused on specific ad sources while continuing to encourage ad revenue from others.

Across the study period, there was a significant decrease in advertisements targeting patients/carers in comparison to ads targeting medical professionals or researchers (Figure 41). This mirrors the decrease across categories with calls to action relevant for patients like Biobanks and Clinics. The largest category of ads targeting medical professionals and researchers, Biotech, remained consistent throughout the study period indicating that its engagement in SEM around the relevant queries was not impacted by policy implementation.

Target Audience: Patients vs Medical Professionals

While both public/non-profit and commercial ad returns decreased in frequency over time, there was a starker change in commercial returns across the study period (Figure 42). This finding suggests that commercial ads that included monetary calls
to action of users were potentially more heavily impacted by the implementation of the policy change than public or non-profit-sourced ads.

5.6 FINDING 6. SHIFT OF AD CONTENT

Some Clinics promoting unproven stem cell treatments appear to have shifted their ad content language over the study period to omit the term “stem cells”.

Ads originating from Clinics promoting stem cell treatments declined over the study period. The supposition is that this reduction is a direct result of the policy change implemented on October 1st, 2019. While we can observe this change in frequency, I was curious as to whether or not the ad sources observed that change as well and shifted their practices in order to circumvent the ban. To observe this, I compared ad text from the largest private clinic advertiser Swiss Medica from the beginning of the data collection period to that at the end. With over 1,300 individual ads, Swiss Medica stands out as prolific in its practice of SEM throughout the study.

“Stem Cells Cure | Regeneration with Stem Cells”

“Latest therapy with stem cells. Higher succes [sic] rate. No side

Example of Swiss Medica Ad return text captured 03.08.2019

After November 13th, 2019, text of the ad content shifts and no longer includes the term “stem cells”.

“Clinic for Innovative Therapy | Swiss Medica Clinic”


Example of Swiss Medica Ad return text captured 13.11.2019

From this period onwards, these clinic advertisements do not include any reference to stem cells or regenerative medicine within the ad text. The linked websites, however, do continue to display ads promoting unproven stem cell treatments. This has a functional implication for the idea of Google-as-mediator. It suggests that Google’s ability to qualitatively judge the content of linked websites is limited to what is made available through its indexing process. That might include the metadata supplied by content producers themselves, but not – as it appears – the dynamic content on the websites themselves. Significantly for my research, Google Search’s mediation is qualitatively limited by the data and metadata the platform collects. In other words, the quality of information presented by Google Search around stem cell treatments is impacted by the functional information Google Search collects on the content sources.

5.7 FINDING 7. NON-COMMERCIAL ACTORS AND SEM

Charities/NGOs form a significant percentage of content creators engaging in paid SEM around condition-based queries.
The significant share of ad returns produced by Charities/NGOs is an unanticipated finding (as displayed in Figure 37). It indicates that there are a number of non-profit public organisations engaging with the SEM commercial market. This means that these public and semi-public bodies have made the strategic decision to devote funding into Google as an SEM platform, above and beyond any Search Engine Optimization (SEO) strategies that they may employ to drive engagement with their charitable work.

Analysis of the ad content from the Charity/NGO ad returns indicate that many of them are direct links to donation drives within the organisation. This underlines a curious cycle; these non-profit organisations purchase ad space within the Google ad exchange to drive public donations to the organisation, a portion of which is then siphoned once again into the ad exchange. The outcome of this cycle is a commercial avenue for Google to access public funds channelled through charities in exchange for search return prominence.

The impact of this siphoning is significant for larger conversations around digital health, non-commercial actors, and emerging treatments. It follows that the more non-commercial discourses are influenced by commercial intent and activities around stem cell treatments, the more commercial considerations are emphasised within non-commercial content.

Although it only made up 2% of returns, Research content producers serve as a noteworthy example of this influence of commercial activity. As discussed in Chapter 2, funding opportunities within biomedical research towards translational medicine includes a shift towards “cure” centred language. The commercial pressures exerted by Google Search can be only read as an additional pressure on non-commercial research entities to adapt to the shifting mores of health commercialisation.

---

31 I would also argue that Research content producers may organise public engagement through the activities of Charities/NGOs. This might dilute their presence in direct SEM activities and increase the significance of Charity/NGO engagement.
While I do not yet have data on the potential shift in strategies undertaken by these actors within the area of search and digital health, this shift may be observed simply in their self-promotional language and communication. For example, a promoted press release from Leiden University entitled “First patient in Netherlands successfully treated with stem cell gene therapy” strikingly encapsulates this shift (Figure 43, Leiden University, 2022). The press release covers the treatment of an under one year old child with the severe and rare congenital immune disorder, Severe Combined Immunodeficiency (SCID).

Firstly, the headline itself is potentially misleading. Here we see a reduction of condition and application-specific “stem cell gene therapies” to the singular “stem cell gene therapy”. This suggests a generalised treatment as opposed to a finely-specified condition and patient medical intervention. I might have read this as simply an example of misrepresentative public relations without relevance to a larger discussion on stem cell treatment information if it was not for the subsequent content of the press release itself.

Within the five subsections of the press release, two highlighted subsections stand out: “Your Own Donor” and “Price Tag”. “Your Own Donor” conveys a relational
position of the treatment to “you”, the reader, who is presumably an adult individual. This stands in contradiction to the profile and biological reality of the patient who underwent the treatment, an infant with a specific genetic disorder. While I recognise that this may simply be a way of communicating colloquially that the treatment is autologous, it has hard to ignore the disconnect between the targeted patient group and the language used. “Price tag” applies an unusually idiomatic term to apply to a clinical trial outcome to insinuate the market availability of the treatment. When taken in context with the headline which touts the singular “therapy”, this suggests a singular, soon-to-be available, cost-applicable medical treatment targeting the reader.

The finding that Charities/NGOs make up a significant percentage of advertisement returns on Google Search illustrates a specific digital economy in which non-commercial actors are engaged in fundamentally commercial activities to the benefit of the Google Search advertising platform. I will explore the interplay between commercial and non-commercial entities within Google Search in Chapter 7.

5.8 SUMMARISING ADS

The findings in this chapter highlight the cat-and-mouse game of documenting and analysing the types of information content and actors that coalesce around advertisement and stem cell treatments on a Blackbox technology like Google Search. While the dataset of donated search results gives us a snapshot, it is clear from the findings that this is a living negotiation between myriad ad creators, users, and the platform that mediates interaction.

Like Jeannete Pol’s domestication of technology (2012), there is a constant unleashing and taming by all participating actors to get more out of the interaction around search. Whether that be for more information on the part of a patient, more new paying patients on the part of commercial clinics, increased donations to charities, or click-throughs for the search platform, each actor has unique motivations competing within one digital space. Despite these ever-moving factors, the findings in this chapter offer indications of impact on that space following Google’s October 1st, 2019 policy change deadline. From this capture, I can return
to my initial research sub questions: *How is the mediation of information on stem cell treatments shaped by structural factors within the platform of Google Search? What impact did Google’s policy change have on paid search advertisements relating to stem cell treatment queries? Does the information found differ depending on the user’s or query’s characteristics? When? And, hopefully, why?*

Although there has been an effect on ads displayed since the implementation of the ban on October 1st, 2019, it does not appear consistent and has not resulted in an outright ban of ads around stem cell treatments for Parkinson’s Disease or Multiple Sclerosis. This data set clearly supports the assertion that Google has not responded to the raising of concern around potentially dangerous marketing practices in the same way it has around other risk-fraught topics like suicide (as discussed in Chapter 4). While query topics like suicide, and self-harm appear to be exempt from Google’s ad exchange platform, queries around controversial and unproven medical treatments remain subject to the commercial marketing practices. Fundamentally, the lack of transparency of how Google search mediates ad insertion within Google search leads to more questions when a change in that policy can only be observed by its outputs.

The continued display of ads with a downwards trend of frequency indicates that the change implemented by Google’s policy decision was mediated on an ad-hoc basis. This resulted in varying impacts on categories of ad sources. It remains to be seen what impact this will have on *ad creator* strategies going forward. Will *Clinics* like Swiss Medica continue to evolve their marketing strategies to bypass restrictions? More controversially, will *Charities/NGOs* continue to funnel funding into SEM as a ‘necessary’ commercial investment?

Additionally, evidence presented here that suggests human data donors were subject to increased ad display has ethical implications beyond this study. On the one hand, Google has indicated that it is moving away from platform-based personalisation (Krafft et al. 2020) as ineffective. However, recent developments suggest that Google is moving towards browser software-based data collection and personalisation (Wired, 2021). This shift towards data mediation from browser software like Chrome itself would indicate a move towards increased meta data
collection and filtering of web-accessed material. The ethical and societal implications of this opacity of these data mediation and personalisation practices for healthcare information and decision-making in digital spaces is fraught.

As explored in Chapter 2, patients are enacted through their interaction with Google Search – as users, consumers, and products. The return of advertisements is further support for the conceptualisation of patients as product at the point of query. Before advertisements are displayed to users to interact with as consumers, Google has already rendered the user a product. They have already been identified and sold as a target audience to content producers like private clinics and charities. The resulting advertising results are thusly mediated through Google Search’s own commercial mechanisms.

Like many algorithmically-driven system administrators, Google appears to operate with a strategy of reactive issue solving; it responds retroactively to implement changes to its algorithmic governance rather than proactively (Tsamados et al. 2021). This allows for a slower response to ownership and responsibility of ethical quandaries that inevitably arise in an area of interaction like information about emerging medical treatments. In 2019 alone, over 75 new gene and cell therapy treatments entered international markets with a hundred more in the following year (McKinsey 2020). Without a transparent communication of how and why ads around the topic of emerging medical treatments will be included or excluded, Google Search’s inclusion of material that presents potential harm to the public nears inevitability.

Currently Google Search is attempting to be all things for all actors; it sits in multiple modalities as a tool for information discovery for users, of new clients for ad creators, and as a source of profit for itself. The act of adopting a ban on unproven stem cell treatments as policy indicates that there is a limit to the company’s willingness to enable all practices within these modalities; there is a point at which profit should be hemmed in favour of user safety. Yet as these findings suggest, these policy changes are not enough to address the potentially volatile assemblage of actors who meet at the intersection of search and medical treatments.
CHAPTER 6:

MAPPING ORGANIC SEARCH RESULTS

Between October 2, 2019 and March 2nd, 2020 and with the help of 163 human and virtual data donors, I collected over 220,000 unique organic Google Search results on queries related to stem cell treatments. My findings allow us to see Google search platform as an active actor and observe its mediation of knowledges in real-time. The data results collected constitute a uniquely-curated index of expertise created by Google’s search platform outside of the explicitly stated market intervention of targeted advertisement. In this snapshot of informational offerings, we are provided with evidentiary building blocks with which to investigate Google search’s logic of indexing knowledges and its practices in displaying them within organic search results. As an attempt at a ‘material turn’ (Reckwitz, 2002), this chapter answers my secondary research question: How is the mediation of information on stem cell treatments shaped by structural factors within the platform of Google Search?

Unlike Chapter 5 which addressed the research question of impact of Google’s policy change on paid advertisement, I did not approach my analysis of organic search results with the assumption that I had captured changes in content over the collection period. Rather, my analysis of organic search results provides an evidentiary basis for answering the question of how mediation of information is shaped on the Google Search platform. My focus on organic search results collected requires emphasis on several key indicators. These include mapping the hierarchies of sources within the results and their potential variations across geographies, query structures, and user groups. I am also interested in the discrepancies between the breakdown of types of content as compared to those displayed in advertisements. In examining these indicators within the organic search results captured in my data collection, I am working to both reverse engineer the mechanisms of influence used by Google Search and map the constellations of content producers presented to users searching for stem cell treatment information.
Within my dataset, I found that the collected datasets of organic search results were characterised by a narrower range of source return types from a smaller number of sources than those collected within the search engine advertising returns. Within organic results, returns from Research source types (characterised by institutes carrying out or supporting primary clinical investigations into stem cell research) were consistently the largest source type followed by News and Information and Charity/NGOs. The primary make-up of results by category type remained largely consistent across the data sets.

It is also clear from the findings that the algorithmic practice creating these returns are influenced by platform logic and variations within the algorithmic Blackbox, leading to diversity across results in any one set of search results. In this chapter, I investigate a number of factors within the dataset with the purpose of determining their potential impact on end results.

The findings in this chapter builds towards Chapter 7’s exploration of search engine results pages for stem cell treatments as a whole. Understanding the range and characteristics of organic results that are returned in response to stem cell treatment queries as well as advertisement returns gives us a full picture of the moment after a user clicks “search”. From there we can engage in an informed discussion of the landscape of information available on Google Search and how the search engine itself can shape the public discourse on regenerative medicine.

In the first section of this chapter, I explore the top results for the dataset through the results generated by clean Virtual Personal Server (VPS) donors within each geographic study. I have made the conscious decision here to focus my analysis to the US and UK Parkinson’s Disease study groups as they provided the largest and most consistent data collection for both human and virtual data donors. In terms of statistical significance and impact, these studies have the potential to render the strongest evidentiary base. However, I will, when appropriate, bring in insight from the Multiple Sclerosis studies and the complete dataset (geographically unrestrained).
These results provide a broad overview of the Google Search engine result page sources for queries around stem cell treatments for Parkinson’s Disease. We can extrapolate from these high level results the general characteristics evaluated and returned by Google’s organic search algorithm. In subsequent sections, I address the top sources according to distinguishing variables within the study groups including the differences between human and virtual data donor and query structure. In these sections, the impact of algorithmic decision-making provides insight into the logic through which these variations in returns are presented to end users.

6.1 PRACTICAL ANALYTICAL CONSIDERATIONS

Before diving into the results, there are a few factors in the practical analytical approach within this chapter’s data presentation and analysis to take into consideration when reviewing this chapter’s data and analysis. As discussed in previous chapters, Google’s search results pages have been structurally developed to privilege the first Google Search engine result page which comprises roughly 10 returns per page. In its primary computational task, Google’s search engine is designed to deem the specific user interaction of “finding” a result as a success; the platform’s core function is completed successfully upon user click-through (Prakash, 2020). Therefore, the mechanisms of indexing that occur within the platform Blackbox encourage returns that are most likely to result in user interaction (a click-through), rather than the most objectively correct answers. This is a subtle yet important distinction of what constitutes a desired result from the perspective of the search engine. Furthermore, this has larger implications for the way we analyse Google search vis-a-vis epistemological understanding of regenerative medicine online, which will be discussed later.

This is an important distinction for the theoretical grounding I am employing here because it ties directly to the work of Mager on algorithmic ideology (2012) and Pols approach to the process of collective assemblage building (2012). The logic of the Google Search algorithm are the parameters through which value is assigned to indexed sources as potential returns to queries (i.e. why is certain content included within the output list of search returns). This is an unseen logic as it is part of myriad proprietary mechanisms within the algorithmic Blackbox. However, for organic
search results, this is intricately tied to the platform’s stated task of prompting a user *click-through* and the consideration of that action as a *success*. The *practice* of Google Search algorithm in this instance is the subsequent output of specific content to which that logic has been applied. I observed this through analysis of which content is included with the output list of search returns. Indeed, I contribute an observation of *practice* and inference of *logic* performed by Google Search as a critical contribution to STS and digital health literature more broadly.

However, it was helpful for the purpose of my analysis to accept this structural privileging of the top 10 results to a query for three reasons. First, focusing on the top 10 results recreates a limit of information exposure with which one would expect the end-user public to engage. To put it simply, Google’s search platform and I both expected most end-users to engage only in one page of the Google Search engine result pages in their search for information around stem cell treatments. Second, I expected that this practical philosophy would have a reciprocal influence on the platform’s design strategy; we will see more of the platform’s intervention within these results as they are essential to “successful” completion of its task. This reasoning allowed me to gain insight to both the thought process behind Google’s decision-making (whether stated strategy or created as a by-product of its algorithmic indexing processes) and the public-facing knowledge created. Finally, limiting analysis to top 10 results and source position within the first Google Search engine result page allowed for accessible statistical analysis of the over 200,000 results generated by my data donation collection.

As I anticipated the results would include fewer instances of direct promotion of unproven stem cell treatments, in this chapter I focus on the result sources themselves - be they charities/NGOs, public resource arms of institutional healthcare systems, private news sources, research centres, or private companies. These larger categories will carry over from my analysis of the Google ad returns. However, unlike in the analysis of ads, I have placed less emphasis on the importance of result frequency over time. As Google’s policy change in September 2019 was limited to Search Engine Marketing (SEM) results, I did not anticipate a temporal impact of results across the research period.
Additionally, although the content of the results is important, I have privileged here the dimensions of the content producer - or source - itself. Unlike ads, where the presence of content - including the stated promotion of unproven stem cell treatments - was vital for analysis vis-a-vis the role of the platform, organic search results reflect a different set of platform logic. More prescient for the discussion of platform logic are the questions; Who has been given the greatest access to the public on queries around stem cell treatments through Google search platform and, I will speculate, how, why and when? For my purposes, the factors for organic search results that gain most relevance are the characteristics of the source, the rate of display within the Top 10 results (as representative of the first page of results), and the average position placed within those top 10 results.

Finally, and keeping in mind these two preceding considerations, I am limiting the analysis of organic search results to those studies within my dataset that had the highest levels of engagement both by my virtual data donors (VPS) and human data donors; Parkinson’s Disease in the United Kingdom and United States. These study groups returned the most consistent donations across the dataset and study period and provided the most evidentiary-based opportunities for documenting the shifts in algorithmic logic in the delivery of health information on Google search.

6.2 PAINTING A PICTURE OF RETURNS WITH OUR VIRTUAL SEARCHERS

In this section, I map the rough borders of who and what are included in stem cell treatment search engine result pages for Parkinson’s Disease in the United Kingdom and United States. This provides some baseline insight into the generalised logic used by Google Search in providing the information most likely to result in a click-through. The purpose of this data exploration is to create a control baseline of organic search result returns within which I can then look for variations that might point to different practices of mediation.

To this end of better understanding algorithmic logic within these searches, there are three characteristics of source returns that I would like to emphasise within this chapter: type, frequency, and average position. These indicators help illustrate
Diakopoulos’ strategies of analysing algorithmic processes as discussed in my Methodology in Chapter 4 (2013); prioritisation, classification, association and filtering. By tracking these indicators within my organic data collection, I am able to build a rich digital ethnographic picture of trends and themes within a large body of data.

*Type* - As with the analysis of advertisement returns within Google Search results pages collected through this data collection, I have applied the categorisation types across these organic findings. The purpose of this is to allow for cross comparison with the advertisement results (Chapter 5) while providing a clear baseline for analysis of the types of source-instigated knowledges presented to end-users in organic Google searches.

*Frequency* - This variable documents how often a source appears within the dataset by raw value. This characteristic has two implications. The first is that the Google Search algorithm has placed a positive value on the source as a content producer within its task to present the “most likely to prompt a click-through” content to the user. Secondly, it indicates that the source - as an actor - has created content that fills this role within the point of information exchange of search. This indicates that the content provider has successfully leveraged Google Search technology to become prominent in the displayed results.

*Average Position* - This characteristic is tied to value placed on a source and its content. Within Google Search engine results pages, there is a relationship between the placement on the search engine result page and the value attributed to the source by the process of Search Engine Optimization (SEO). The lower the value, the higher the content appears on the Google Search engine result page and the more likely it is for a user to reach that crucial click-through interaction.

A quick reminder about the VPS donors used in this study. Using the VPS donors, I established a continual data donation engagement throughout the study period using neutral browsers to collect Google Search engine results page data. Functionally, this allows them to return results each donation period without the influence of any personalisation that might occur on a human data donor’s browser. The only
distinguishing features that should be considered within the Google Search Blackbox are the geographic language and location (Krafft et al., 2020). This is to say that there will be variation based on the default browser language (English - UK or English - US) and the geographical location of the VPS. This, in turn, determines the data centre through which the search is routed (Le et al., 2019).

6.3 ORGANIC RESULTS

The following sections explore the data produced by the data donation campaign. Included here are specifically organic search results. It is important to note that organic search results constitute results upon which there is no direct Google-mediated platform for ranking influence (as compared to the direct influence of Google’s SEM advertisement platform). The commercial aspects of organic search result ranking (through SEO and SEO-enabling commercial activity) will be explored in the next chapter. In the following data (as with my ad return results), I am prioritising the study groups that produced the longest and most sustained engagement from data donors. This has resulted in outsized emphasis on the Parkinson’s and Multiple Sclerosis studies from the United Kingdom and United States. The limitations of my methodological deployment will be explored in my conclusion in Chapter 8.

6.3.1 ORGANIC RESULTS FOR PARKINSON’S DISEASE QUERIES IN THE UNITED KINGDOM

Between October 1st, 2019 and March 2nd, 2020, I used the VPS donors located within the United Kingdom to collected 549,134 individual returns on all search queries32 related to stem cells and Parkinson’s Disease. These results indicate that publicly funded research institutes held the largest share of returns (Table 6). This category was followed by News and Information, Charity/NGOs, and Biobanks. Clinics took up a smaller share of returns than in the Search Engine Marketing (SEM) datasets. Biotech and Clinical Trials made up only a small percentage of returns for the entire dataset.

32 The queries used here are referenced in Chapter 4.
Within the Google Search engine result page dataset, the same top categories also appeared on average in the top 5 positions within the results page. This suggests that categories such as Research, News and Information, and Charity/NGO sources benefited from some privilege within the Blackbox logic of returns. At an average position of 7.8, Clinics appeared to have less pronounced visibility within the Google Search engine result page for the organic results dataset (Table 7).

The frequency of Research and News and Information content types indicate that sources characterised by biomedical authority and information-gathering gained prominence in the organic search results across this dataset. This further indicates that the algorithmic logic on returns for these Parkinson’s Disease-related queries

<table>
<thead>
<tr>
<th>Source Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>220558</td>
</tr>
<tr>
<td>News and Information</td>
<td>126908</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>90249</td>
</tr>
<tr>
<td>Biobank</td>
<td>14557</td>
</tr>
<tr>
<td>Clinic</td>
<td>5377</td>
</tr>
<tr>
<td>Biotech</td>
<td>3607</td>
</tr>
<tr>
<td>Clinical Trials</td>
<td>110</td>
</tr>
</tbody>
</table>

**TABLE 6: VPS ORGANIC RETURNS FOR STUDY 6 - PARKINSON’S DISEASE IN THE UNITED KINGDOM BY ‘TYPE’**

<table>
<thead>
<tr>
<th>Type</th>
<th>Avg. Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>4.71</td>
</tr>
<tr>
<td>News and Information</td>
<td>5.41</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>5.12</td>
</tr>
<tr>
<td>Biobank</td>
<td>6.14</td>
</tr>
<tr>
<td>Clinic</td>
<td>7.80</td>
</tr>
<tr>
<td>Biotech</td>
<td>3.57</td>
</tr>
<tr>
<td>Clinical Trials</td>
<td>8.50</td>
</tr>
</tbody>
</table>

**TABLE 7: VPS SEARCH RETURN AVERAGE POSITION BY TYPE FOR PARKINSON’S DISEASE IN THE UNITED KINGDOM**

The frequency of Research and News and Information content types indicate that sources characterised by biomedical authority and information-gathering gained prominence in the organic search results across this dataset. This further indicates that the algorithmic logic on returns for these Parkinson’s Disease-related queries
are friendly to non-commercial sources at least within the organic search results. Diving into the specifics of the top sources, a more complex picture of the actors within the top categories emerges.

**NEWS AND INFORMATION**

*News and Information* source, Parkinson’s News Today (parkinsonsnewstoday.com), appeared as the most frequent return with an average position of sixth within the top source results of the dataset. Other top generalised information sites included (in descending order of frequency) WebMD, Wikipedia, MedicalNewsToday.com and Medium.com. These sources are diverse in their characteristics as aggregators of information and secondary content producers. Additionally, these sources represent a diverse array of commercial and non-commercial approaches to content publishing and site ownership.

In particular, content included within *News and Information* often included secondary direct-to-consumer marketing of stem cell-related products and treatments through secondary advertising (e.g. in-page clinic advertisement on some privately owned news aggregation sites). While the *News and Information* content may not have included links to private treatment sites, the advertisement platforms on the websites often included dynamic advertisement links to private clinics offering stem cell treatments. This is not content that would be captured by Google Search, but does indicate another route of access to potential patients by commercial interests.

**CHARITIES/NGOS**

Parkinson’s-specific charities shared a large percentage of returns including (in descending order of frequency) Cure Parkinson’s UK, the Michael J. Fox Foundation, and Parkinson’s UK. These content sources would be considered non-commercial in so far as they tend to operate on a non-profit basis. However, there are significant funding arms of all of these organisations, often with the express goal of raising financial capital for ongoing research projects aimed at the expansion of condition-specific research and clinical development. It follows that many of these websites feature prominent calls-to-action for private donations and donation-driven merchandising.
I note that the position of charities and NGOs are both as representatives of patient interest groups and independent actors within the discourse of stem cell treatments online. The impact of Charities/NGOs on individual patient decision-making and information on treatments largely depends on factors including relationship to patient advocacy, funding strategy, and patient type. In general, the non-profit status of Charities/NGOs has enabled higher levels of trust in shared information (Marshall and Williams, 2006). Further to this, Charity/NGO relationship to commercial intent for NGOs is varied by mandate and by regulatory geography.

RESEARCH INSTITUTIONS

Inter-institutional research networks and associations also shared a large percentage of the returns including (in descending order of frequency) EuroStemCell, the International Society for Stem Cell Research (ISSCR via www.acloserlookatstemcells.org), Hopkins Medicine, the Mayo Clinic, Neuro Central, and the California Center for Regenerative Medicine (CIRM via Ispcell.com). These sources are by and large partially or completely publicly funded and operate with the oversight of funding bodies whether governmental, academic or civic. While these organisations may have direct fundraising calls-to-action on their websites, they tended to be targeted to professional membership approaches or institutional collaboration.

GOVERNMENTAL AND REGULATORY

Representation of governmental and regulatory organisations include (in descending order of frequency) the National Institute of Health (United States), the Federal Drug Administration (United States), the National Health Service (United Kingdom), and websites from the state government of California (United States). These returns indicate that there is geographic overlap in returns which enables non-location specific sources to appear to users outside of their jurisdiction. This is a particularly interesting point to which I will return. These sources function as non-commercial governmental entities and derive their funding from national monetary policy and strategic civic investment.
Only one Biobank appeared in the list of top sources in the United Kingdom: Cells4Life (www.cells4life.com). Cells4Life - as the sole but frequent Biobank return source in the dataset - operates on a for-profit basis as an independent commercial entity.

The dearth of category types that encourage direct patient intervention in medical treatments (like Biobanks and Clinics) is potential evidence that the logic governing organic search results is less friendly to commercial sources. There may be a number of reasons – both technically-encouraged and locally determined – for this tendency away from commercial medical offerings. However, I suggest that the relatively low rates of returns for these types of results may be due to the influence of a strong central medical infrastructure and regulation of access within the UK through the NHS. Private healthcare and patient-initiated treatment pursuit is a relatively new aspect of healthcare in the UK and remains heavily regulated through centralised points of access for care. If privatisation of healthcare increases within the UK away from the NHS, we might see an increase in result source types encouraging private, commercial targeting patients directly.
This constellation of source returns - consistent in its breakdown of type across the data collection period - points to a steady practice by Google Search in its organic search return logic. I can say with some evidential certainty that content produced by research institutions, news and information aggregators, and charities/NGOs focused on condition-specific advocacy and practices make up the overwhelming majority of organic search results for queries around Parkinson’s Disease in the United Kingdom. However, does this observation only hold for the United Kingdom and the logic applied to users in its locale?

6.3.2 ORGANIC RESULTS FOR PARKINSON'S DISEASE QUERIES IN THE UNITED STATES

Although the UK Parkinson’s control results provide a good demonstration of the types of organic search results returned by stem cell-related queries, I decided that it
would be helpful to compare these results to the similarly populated dataset produced by VPS data donors in the United States. This allows my discussion of the types of return sources and logic governing Google search in areas of stem cell treatments to hold relevancy outside of one geographic location.

Between October 1st, 2019 and March 2nd, 2020, the VPS donors located within the United States collected 389,872 individual returns on all search queries related to stem cells and Parkinson’s Disease. The breakdown of the types of sources collected within this dataset was similar to that of the United Kingdom study (Table 9). However, there was an increased frequency of returns originating from sources within the Clinic subcategory. Additionally, there were a few returns (although statistically small) which did not appear at all in the United Kingdom study (Other, Lifestyle and Non-Traditional Advice, and Complementary Treatment).

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
<th>Avg. position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>183399</td>
<td>4.41</td>
</tr>
<tr>
<td>News and Information</td>
<td>117295</td>
<td>5.39</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>68274</td>
<td>6.36</td>
</tr>
<tr>
<td>Clinic</td>
<td>17484</td>
<td>6.41</td>
</tr>
<tr>
<td>Biotech</td>
<td>3573</td>
<td>4.55</td>
</tr>
<tr>
<td>Biobank</td>
<td>428</td>
<td>7.49</td>
</tr>
<tr>
<td>Other</td>
<td>366</td>
<td>5.28</td>
</tr>
<tr>
<td>Lifestyle and Non-Traditional Advice</td>
<td>32</td>
<td>7.62</td>
</tr>
<tr>
<td>Complementary Treatment</td>
<td>19</td>
<td>7.89</td>
</tr>
<tr>
<td>Clinical Trials</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**TABLE 9: VPS ORGANIC RESULTS FREQUENCY AND AVERAGE POSITION FOR STUDY 9 - PARKINSON’S DISEASE IN THE UNITED STATES BY ‘TYPE’**

Once again, Research results comprised the largest share of returns followed by News and Information, and Charity/NGO with an increased presence of Clinics. Biobanks as a type are reduced to near statistical insignificance compared to the UK returns (outside of the single outlier of Cells4Life). This may indicate a slight preference within the geographically-localised algorithmic logic for commercial returns (as further supported by the increased presence of Clinics and Biotech).
The Top 20 sources remain comparable with the top sources from the United Kingdom with the addition of a few US-specific sources (*Table 10*). Interestingly, unlike in the United Kingdom returns, which included six explicitly non-UK based sources, only *two* non-US sources (EuroStemCell and Cure Parkinson’s UK) appeared within the top 20 results. This geographic power imbalance in content mediation will be discussed later in this chapter. For now, it is important to note that this relationship of international information origin appears to favour local content sources within US-based Google Search engine result pages, while results pages within the UK appear to include more non-local sources.

This suggests that Google Search introduces an element of geographical warp to local users in regards to the types of information presented. What information is offered is in part determined by the local actors engaged in content creation around stem cell treatments. However, it also indicates that the blurring of regulatory and geographical boundaries of stem cell treatment information internationally is amplified by engaging in search.
<table>
<thead>
<tr>
<th>Source Host</th>
<th>Avg. Position</th>
<th>Frequency</th>
<th>Min. Pos.</th>
<th>Max. Pos.</th>
<th>Source Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>parkinsonsnewstoday.com</td>
<td>5.94</td>
<td>195942</td>
<td>1</td>
<td>13</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>michaeljfox.org</td>
<td>7.55</td>
<td>167517</td>
<td>1</td>
<td>15</td>
<td>Charity/NGO</td>
</tr>
<tr>
<td>apdaparkinson.org</td>
<td>5.71</td>
<td>133588</td>
<td>1</td>
<td>15</td>
<td>Charity/NGO</td>
</tr>
<tr>
<td>eurostemcell.org</td>
<td>5.44</td>
<td>122763</td>
<td>1</td>
<td>14</td>
<td>Research</td>
</tr>
<tr>
<td>hopkinsmedicine.org</td>
<td>4.94</td>
<td>116065</td>
<td>1</td>
<td>14</td>
<td>Research</td>
</tr>
<tr>
<td>webmd.com</td>
<td>4.99</td>
<td>112436</td>
<td>1</td>
<td>14</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>closerlookatstemcells.org</td>
<td>3.19</td>
<td>83550</td>
<td>1</td>
<td>14</td>
<td>Research</td>
</tr>
<tr>
<td>cureparkinsons.org.uk</td>
<td>7.7</td>
<td>80753</td>
<td>1</td>
<td>13</td>
<td>Research</td>
</tr>
<tr>
<td>en.wikipedia.org</td>
<td>6.51</td>
<td>68971</td>
<td>1</td>
<td>11</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>nih.gov</td>
<td>3.74</td>
<td>64063</td>
<td>1</td>
<td>15</td>
<td>Research</td>
</tr>
<tr>
<td>health.usnews.com</td>
<td>6.04</td>
<td>62747</td>
<td>2</td>
<td>11</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>medicalnewstoday.com</td>
<td>2.5</td>
<td>59544</td>
<td>1</td>
<td>15</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>mayoclinic.org</td>
<td>2.43</td>
<td>50647</td>
<td>1</td>
<td>12</td>
<td>Research</td>
</tr>
<tr>
<td>neuro-central.com</td>
<td>8.35</td>
<td>43932</td>
<td>4</td>
<td>15</td>
<td>Research</td>
</tr>
<tr>
<td>fda.gov</td>
<td>3.43</td>
<td>41688</td>
<td>1</td>
<td>14</td>
<td>Research</td>
</tr>
<tr>
<td>ca.gov</td>
<td>3.43</td>
<td>41330</td>
<td>1</td>
<td>11</td>
<td>Research</td>
</tr>
<tr>
<td>ajmc.com</td>
<td>8.63</td>
<td>36086</td>
<td>3</td>
<td>14</td>
<td>Research</td>
</tr>
<tr>
<td>stemcells.nih.gov</td>
<td>4.07</td>
<td>32821</td>
<td>1</td>
<td>14</td>
<td>Research</td>
</tr>
<tr>
<td>ucsf.edu</td>
<td>4.82</td>
<td>30469</td>
<td>1</td>
<td>10</td>
<td>Research</td>
</tr>
<tr>
<td>cellmedicine.com</td>
<td>9.13</td>
<td>29485</td>
<td>6</td>
<td>15</td>
<td>Clinic</td>
</tr>
</tbody>
</table>

**TABLE 10: TOP 20 VPS RESULT SOURCES FOR PARKINSON'S DISEASE QUERIES IN THE US BY FREQUENCY AND POSITION**

Additionally, while no *Biobank* sources were counted amongst the Top 20 sources as in the US returns, a Panamanian private clinic (cellmedicine.com) did appear over 29,000 times (*Figure 44*). This clinic website includes advertisements for stem cell treatments for conditions including Autism. These treatments are currently unproven (Marks and Hahn, 2020).
6.3.3 GENERALISING RESULTS ACROSS DATASETS

Placing the UK and US returns into direct comparison, it becomes clear that there is a consistent logic applied to the mediation of organic search results for queries around Parkinson’s and stem cell treatments (Table 11). This logic results in practice by search algorithms that display content from Research, News and Information, and Charity/NGO source types most often. Relative to the share of returns within search engine advertising, commercial entities like Clinics, Biotech, and Biobanks are less likely to appear to users.
In response to the question of commercial call-to-action, ‘how many clicks away is a user from direct-to-consumer marketing of stem cell products?’, this evidence suggests that sources tied to non-directly commercial knowledge sharing maintain a high level of integration and average positioning with Google Search engine result pages. Although News and Information include sources that rely on digital advertising revenues in their commercial model, these are secondary.

Can this picture of knowledge mediation be made a bit more complex? Are there variables within the datasets that could grant more insight into the subtle fluctuations in logic and practice exercised through search algorithm mediation? The following sections will explore these differential factors in detail and, in so doing, will map out a few complex areas that characterise the role of search algorithms as an actor.

6.4 VARIATIONS: VIRTUAL DONORS VS. HUMAN DONOR?

It was clear within the dataset on advertising that users who identified themselves as being impacted by a condition as either a patient or carer were more likely to receive advertisements within their search results. As discussed, this is indicative of mechanisms within the Google search Blackbox to identify and adjust display on the search results page by users’ previous activity. Importantly, VPS donors, potentially due to their lack of search history, were less likely to display advertisements - problematic or otherwise. This raises the question in mapping organic search results: *is there a measurable difference in the makeup of types of sources displayed*
between “clean” VPS data donors and condition-impacted human data donors? If not, what does this imply about mediation of results by Google Search platform?

To investigate this, I needed to bring the VPS baseline into conversation with results only produced by human donors who self-identified as belonging to the group of users impacted by one of the named conditions addressed by in the query sets. Here, again, I am going to rely on the geographic and conditional dataset with the largest, most consistent input from both VPS and human data donors: Parkinson’s Disease in the United Kingdom. With over 100 human participants, this data set provides a solid foil to the clean VPS baseline for organic search results related to queries around stem cell treatments and Parkinson’s Disease.

The result was statistically clear; both human donors and clean VPS donors were presented with the same relative breakdown of source types within organic search results (Table 12). Within the UK, both sets of data returned statistically comparable shares of each source type across the study period.

<table>
<thead>
<tr>
<th>Type</th>
<th>UK VPS Donors</th>
<th>UK Human Data Donors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>47.81%</td>
<td>47.33%</td>
</tr>
<tr>
<td>News and Information</td>
<td>27.51%</td>
<td>26.60%</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>19.56%</td>
<td>19.66%</td>
</tr>
<tr>
<td>Biobank</td>
<td>3.16%</td>
<td>3.42%</td>
</tr>
<tr>
<td>Clinic</td>
<td>1.17%</td>
<td>1.08%</td>
</tr>
<tr>
<td>Other</td>
<td>0%</td>
<td>1.08%</td>
</tr>
<tr>
<td>Biotech</td>
<td>0.78%</td>
<td>0.80%</td>
</tr>
<tr>
<td>Clinical Trials</td>
<td>0.02%</td>
<td>0.02%</td>
</tr>
</tbody>
</table>

| TABLE 12: PERCENTAGE SHARE OF ORGANIC VPS RESULTS BY ‘TYPE’ - UK VPS VS UK HUMAN DATA DONORS |

Interestingly, although the Top 20 sources remain largely the same, there is an increase of both frequency and average position for the one commercial source, Biobank Cells4Life (see Table 13). However as this falls within statistical margin of error, it would be inappropriate to attribute this change to an established practice of personalisation within search results. This is to say that I cannot attribute the increase in visibility of the source Cells4Life – a for-profit commercial entity – on
specific SEO strategies targeting individuals who might be most interested in their services. This is an area that could be explored by future research on the targeted SEO practices of commercial enterprises.

<table>
<thead>
<tr>
<th>Source Host</th>
<th>Average Position</th>
<th>Frequency</th>
<th>Min. Pos.</th>
<th>Max. Pos.</th>
<th>Source Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>parkinsonsnewstoday.com</td>
<td>6.95</td>
<td>308533</td>
<td>1</td>
<td>12</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>cureparkinsons.org.uk</td>
<td>5.73</td>
<td>176207</td>
<td>1</td>
<td>10</td>
<td>Charity/NGO</td>
</tr>
<tr>
<td>nih.gov</td>
<td>5.45</td>
<td>147776</td>
<td>1</td>
<td>16</td>
<td>Research</td>
</tr>
<tr>
<td>eurostemcell.org</td>
<td>3.82</td>
<td>126556</td>
<td>1</td>
<td>13</td>
<td>Research</td>
</tr>
<tr>
<td>closerlookatstemcells.org</td>
<td>3.27</td>
<td>125712</td>
<td>1</td>
<td>10</td>
<td>Research</td>
</tr>
<tr>
<td>webmd.com</td>
<td>5.13</td>
<td>119456</td>
<td>1</td>
<td>20</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>parkinsons.org.uk</td>
<td>3.35</td>
<td>103115</td>
<td>1</td>
<td>20</td>
<td>Charity/NGO</td>
</tr>
<tr>
<td>medicalnewstoday.com</td>
<td>3.04</td>
<td>93278</td>
<td>1</td>
<td>19</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>cells4life.com</td>
<td>5.97</td>
<td>86644</td>
<td>2</td>
<td>13</td>
<td>Biobank</td>
</tr>
<tr>
<td><a href="http://Www.nhs.uk">Www.nhs.uk</a></td>
<td>4.57</td>
<td>74446</td>
<td>1</td>
<td>20</td>
<td>Research</td>
</tr>
<tr>
<td>mayoclinic.org</td>
<td>2.61</td>
<td>72285</td>
<td>1</td>
<td>8</td>
<td>Research</td>
</tr>
<tr>
<td>neuro-central.com</td>
<td>8.18</td>
<td>69895</td>
<td>5</td>
<td>15</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>michaeljfox.org</td>
<td>8.07</td>
<td>63351</td>
<td>3</td>
<td>19</td>
<td>Charity/NGO</td>
</tr>
<tr>
<td>fda.gov</td>
<td>6.42</td>
<td>62233</td>
<td>1</td>
<td>20</td>
<td>Research</td>
</tr>
<tr>
<td>en.wikipedia.org</td>
<td>4.36</td>
<td>61737</td>
<td>1</td>
<td>18</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>mstrust.org.uk</td>
<td>7.1</td>
<td>58173</td>
<td>2</td>
<td>10</td>
<td>Charity/NGO</td>
</tr>
<tr>
<td>hopkinsmedicine.org</td>
<td>8.37</td>
<td>52959</td>
<td>2</td>
<td>18</td>
<td>Research</td>
</tr>
<tr>
<td>ca.gov</td>
<td>4</td>
<td>49912</td>
<td>1</td>
<td>13</td>
<td>Research</td>
</tr>
<tr>
<td>medium.com</td>
<td>6.32</td>
<td>43161</td>
<td>1</td>
<td>18</td>
<td>News &amp; Information</td>
</tr>
<tr>
<td>stemcells.nih.gov</td>
<td>3.54</td>
<td>35538</td>
<td>1</td>
<td>14</td>
<td>Research</td>
</tr>
</tbody>
</table>

TABLE 13: TOP 20 HUMAN DATA DONOR RESULT SOURCES FOR PARKINSON’S DISEASE QUERIES IN THE UK BY FREQUENCY AND POSITION

The parity in results across data donor type is indicative of an absence of user profile dependent variations on organic search results (or direct personalisation). This also indicates that the Blackbox logic that governs organic search returns is different from that which governs advertisement integration (which was impacted by user profile). This practice shift between search engine mechanisms within single Google Search engine result pages will be discussed in the following chapter. However, it is clear that user-side differentiation based on previous searches is not evident within these organic search results.
This result mirrors findings by previous researchers within Algorithmic Accountability studies using similar methodologies to my own. A 2013 study by Hannak et al. of 200 Google Search users found that on average only 11.7% of results deviated due to personalisation. However, this varied widely depending on the query, type of result (e.g. Google News or Organic results) and rankings returned. The study also found that the mechanisms of personalisation within organic search results were most likely to be IP address (geolocation) and logged-in status (whether or not a user’s browser was actively linked to a Google account) (Hannak et al. 2013).

Interestingly, personalisation did influence the display of results within the Google News results generated by Google Search and showed that previous interactions through Google Search reinforced political echo chambers around the 2014 elections (Robertson, Lazer, and Wilson 2018). This is further supported by Le et al.’s Sock-Puppet audit-informed work on Google News search results as reinforcing partisanship within collected results (2019).

Taken in context alongside my findings that users with relevant histories were more likely to be presented advertisements, this organic result finding indicates that there are multiple mechanisms of influence enacted within the search result page. Users are treated differently even within the single page based on the practice of the mechanism (ad, organic, news) that mediates the results displayed. As one moves from result to result on the page, users also move between logics.

I should stress here that this finding does not mean that platform ideology and commercial activities do not play a role in the display of results. As will be discussed in the following chapter, secondary SEO practices introduce commercial activities into organic search results. These reintroduce questions of platform ideology expressed through secondary platform mechanisms beyond direct interaction with users (as seen in advertisements). These secondary mechanisms will be discussed in Chapter 7.
So far consistency in the characteristic makeup of organic search results appears to hold up across data donor categories. However, perhaps variation can be seen between queries - both in their structure and content. Here I am looking for differences in the “query” part of the Blackbox audit. Does shifting the language and structure of the command input to the search platform demonstrably change the output?

I also questioned if the language of query might have an influence on the emphasis of personalisation expressed through in the difference between Human and VPS donors. Could it be that the variations of influence became clearer as I dug more closely into the query-specific returns? To address this, I briefly interrogated the deviation of all query results based on source type and donor type. As a bellweather, I chose to interrogate the human data donor and VPS results for the Parkinson’s study in the United Kingdom as this is the study for which I had the best parity between donor groups and contributions. As Table 12 suggests, the deviation between VPS and human data donors was negligible.

At this stage, I also checked to ensure that my data collection had, in fact, collected comparable quantities of results across queries to ensure accurate data spread. As Table 15 suggest, a comparable number of returns were collected. Variations in the returns may be linked to the order of search and aborted attempts to collect data. For example, when the VPS donors were targeted by Google Search DDoS safeguards, it interrupted the process of data collection. This may have led to a drop-off in collections of query results positioned at the end of the protocol. Future studies might offset this risk by randomising the order of query input through the plugin software.
The assumption here is that use of natural language (in this case, demonstrated by the grammatical use of a question structure and personal subject “me”) may be subject to varied logic of mediation within Google search (Table 16). Within the UK, there is a large increase in Research source returns and a smaller increase in returns for Clinics. All other statistically significant (p > 0.001) returns show a marked decrease in return share when compared to the percentage of returns for the dataset in its entirety.

<table>
<thead>
<tr>
<th>Type</th>
<th>All Queries</th>
<th>‘Can stem cells help me?’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>47.56%</td>
<td>66.10%</td>
</tr>
<tr>
<td>News and Information</td>
<td>27.03%</td>
<td>18.84%</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>19.61%</td>
<td>7.26%</td>
</tr>
<tr>
<td>Biobank</td>
<td>3.29%</td>
<td>1.92%</td>
</tr>
<tr>
<td>Clinic</td>
<td>1.12%</td>
<td>5.26%</td>
</tr>
<tr>
<td>Biotech</td>
<td>0.79%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Other</td>
<td>0.56%</td>
<td>0.62%</td>
</tr>
</tbody>
</table>

Table 15: Search return share by type for “Can stem cells help me?” queries vs.
In the United States, this relative increase and decrease is equally noticeable. In particular, returns for Charities/NGOs decrease to a statistically insignificant percentage (Table 17). This indicates a disconnect determined by the search algorithm logic between the natural language question and condition-specific content created by Charity/NGOs. This is perhaps due to a lack of condition-specific language within the query.

<table>
<thead>
<tr>
<th>Type</th>
<th>All Queries</th>
<th>‘Can stem cells help me?’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>46.77%</td>
<td>63.77%</td>
</tr>
<tr>
<td>News and</td>
<td>30.18%</td>
<td>24.15%</td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>17.50%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Clinic</td>
<td>4.44%</td>
<td>11.90%</td>
</tr>
<tr>
<td>Biotech</td>
<td>0.94%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Biobank</td>
<td>0.09%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Other</td>
<td>0.06%</td>
<td>0.06%</td>
</tr>
</tbody>
</table>

TABLE 16: SEARCH RETURN SHARE BY TYPE FOR “CAN STEM CELLS HELP ME?” QUERIES VS. ALL QUERIES IN THE UNITED STATES

The use of natural language prompts a particular increase in the return of private clinic results, indicating an alignment of personalisation and content promoting private intervention. Here, again, I interrogated the human data donors vs VPS donors in the United Kingdom Parkinson’s study. Again, I found negligible differences between the two donor types (see Table 18). This leads me to conclude that the mechanisms for personalisation within organic search result ranking may not be user-specific (e.g. search history), but potentially tied to more wide ranging characteristics like geolocation. This is supported by the wider margin of difference between the category makeup of results by natural language query between UK and US donors (see Table 17).

Similar to the natural language query of “Can stem cells help me?”, queries including the term “cure” might present opportunity of content shift. I initially thought that an emphasis on “curative” treatments rather than condition maintenance or treatment
might be used within direct-to-consumer marketing of treatments. This might prompt similar reactions in the makeup of results to favour commercial results. However, this was not what I found. In the UK, Clinic results were near negligible with the majority of results split between Research, News and Information, and Charity/NGO results (Table 18).

<table>
<thead>
<tr>
<th>Type</th>
<th>All Queries - All Donors</th>
<th>&quot;Cure&quot; Queries - VPS</th>
<th>&quot;Cure&quot; Queries - Human Donors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>47.57%</td>
<td>41.87%</td>
<td>41.71%</td>
</tr>
<tr>
<td>News and Information</td>
<td>27.05%</td>
<td>31.13%</td>
<td>30.49%</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>19.61%</td>
<td>25.02%</td>
<td>24.52%</td>
</tr>
<tr>
<td>Biobank</td>
<td>3.29%</td>
<td>1.93%</td>
<td>2.18%</td>
</tr>
<tr>
<td>Clinic</td>
<td>1.12%</td>
<td>0.04%</td>
<td>0.03%</td>
</tr>
<tr>
<td>Other</td>
<td>0.93%</td>
<td>0.00%</td>
<td>1.07%</td>
</tr>
</tbody>
</table>

TABLE 17: SEARCH RETURN SHARE BY TYPE FOR ALL QUERIES VS “CURE” QUERIES IN THE UNITED KINGDOM (PARKINSON’S DISEASE)

Results for Parkinson’s queries and “cure” queries in the United States showed similar breakdowns as the United Kingdom study. These included a majority of returns as Research, News and Information, and Charity/NGO source types. Notably, “cure” queries included a significant reduction of Clinic type results as compared to all results. This difference appears to have been reallocated to Charity/NGO results (Table 16).

<table>
<thead>
<tr>
<th>Type</th>
<th>All Queries</th>
<th>&quot;Cure&quot; Queries - VPS</th>
<th>&quot;Cure&quot; Queries - Human Donors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>46.77%</td>
<td>39.61%</td>
<td>39.56%</td>
</tr>
<tr>
<td>News and Information</td>
<td>30.18%</td>
<td>34.88%</td>
<td>34.60%</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>17.50%</td>
<td>24.51%</td>
<td>24.76%</td>
</tr>
<tr>
<td>Clinic</td>
<td>4.44%</td>
<td>0.68%</td>
<td>0.91%</td>
</tr>
<tr>
<td>Biotech</td>
<td>0.94%</td>
<td>0.19%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Biobank</td>
<td>0.09%</td>
<td>0.09%</td>
<td>0.14%</td>
</tr>
<tr>
<td>Other</td>
<td>0.06%</td>
<td>0.09%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Lifestyle and Non-Traditional Advice</td>
<td>0.01%</td>
<td>0.03%</td>
<td>0.02%</td>
</tr>
<tr>
<td>Complementary Treatment</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

TABLE 18: SEARCH RETURN SHARE BY TYPE FOR ALL QUERIES VS “CURE” QUERIES IN THE UNITED STATES (PARKINSON’S DISEASE)
My initial theory that curative language might encourage direct-to-consumer offers to users overlooked the importance of language in regulation. As regulation enforcement hinges on promotion of unproven claims, using a term like “cure” might be avoided by commercial actors as too far-reaching a claim. Content creators like Research actors might have a different relationship to curative claims as contextually part of larger biomedical research undertakings. Since these do not include direct-to-consumer marketing, they do not need to avoid discussing wholesale cures of conditions or diseases.

**DOES THE INCLUSION OF THE SEARCH TERM “COST” CHANGE THE CHARACTERISTICS OF THE ORGANIC SEARCH RESULTS?**

In investigating the general proximity of a user to direct-to-consumer marketing within organic search results, it could be suggested that a direct use of a commercial term like “cost” might encourage the platform’s practice to shift the characteristics of returns. Indeed, this is borne out within the UK results where there is a marked shift towards commercial source types when limiting results to queries that include the term “cost” (*Table 20*).

<table>
<thead>
<tr>
<th>Type</th>
<th>All Queries</th>
<th>‘Cost’ Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>47.56%</td>
<td>38.60%</td>
</tr>
<tr>
<td>News and Information</td>
<td>27.03%</td>
<td>24.47%</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>19.61%</td>
<td>15.37%</td>
</tr>
<tr>
<td>Biobank</td>
<td>3.29%</td>
<td>10.79%</td>
</tr>
<tr>
<td>Biotech</td>
<td>0.79%</td>
<td>5.86%</td>
</tr>
<tr>
<td>Clinic</td>
<td>1.12%</td>
<td>4.73%</td>
</tr>
<tr>
<td>Clinical Trials</td>
<td>0.02%</td>
<td>0.18%</td>
</tr>
</tbody>
</table>

*Table 19: Search return share by type for “cost” queries vs. all queries in the United Kingdom*

In the United States, this shift towards commercial source types is dramatic. Clinics are nearly 20% more likely to appear in search results sparked by queries including the term “cost” while Research, News and Information, and Charity/NGO sources decreased significantly (*Table 21*).
<table>
<thead>
<tr>
<th>Type</th>
<th>All Queries</th>
<th>‘Cost’ Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>46.77%</td>
<td>30.07%</td>
</tr>
<tr>
<td>News and Information</td>
<td>30.18%</td>
<td>26.44%</td>
</tr>
<tr>
<td>Charity/NGO</td>
<td>17.50%</td>
<td>13.29%</td>
</tr>
<tr>
<td>Clinic</td>
<td>4.44%</td>
<td>23.36%</td>
</tr>
<tr>
<td>Biotech</td>
<td>0.94%</td>
<td>6.65%</td>
</tr>
<tr>
<td>Biobank</td>
<td>0.09%</td>
<td>0.13%</td>
</tr>
<tr>
<td>Other</td>
<td>0.06%</td>
<td>0.06%</td>
</tr>
</tbody>
</table>

**TABLE 20: SEARCH RETURN SHARE BY TYPE FOR “COST” QUERIES IN THE UNITED STATES**

This increase across cost queries suggests that there are geolocation-dependent practices that translate commercial access within organic search results. The stark increase of commercial results within “cost” queries in the United States might point to increased activity around direct-to-consumer marketing of stem cells to American patients specific to the healthcare system in the country as compared to the United Kingdom.

**FOR WHICH QUERIES WERE PRIVATE CLINICS OFFERING COMMERCIAL MEDICAL TREATMENTS MOST LIKELY TO APPEAR AS RESULTS?**

To address my central concern of the promotion of unproven treatments, I parsed directly out clinics offering unproven stem cell treatments by query. This presented a broad overview of which types of queries around stem cells are most likely to return potentially problematic results to users.

In the United Kingdom, the queries of “stem cell cost”, “Can stem cells help me?” and “Parkinson’s stem cell treatment” were most likely to return content containing direct-to-consumer marketing of unproven stem cell treatments. These top results included both condition specific and general queries (Table 22). This may suggest that there is not preventative mediation of results based on conditions. That is to say, there does not appear to be direct intervention within the organic search results to exclude commercial results (as explored in the advertisement in Chapter 5).
In the United States, this display was significantly higher with the top results returned by the queries “stem cells cost”, “Parkinson’s stem cell cost”, “Can stem cells help me?” and, interestingly given my previous findings, “stem cells cure”. The increased presence of private clinics offering unproven stem cell treatments may here again be influenced by the geographical location and economic healthcare context the user is searching from (Table 23).

These findings indicate that there are, indeed, mechanisms through which a shift in practice (i.e. algorithmically-determined calculations of ‘best answer’) shift the share
of returns amongst source types. This suggests that while the logic of success (a “click-through”) remains, variables like the use of natural and commercial language shifts the results. This may be due to additional considerations of user needs and wants that are inherent in the use of language variables. As language and structure shift towards or away from variables like natural language or inclusion of commercial calls-to-action, there appears a shift in the makeup of source types returned and the types of information and knowledge most likely to be access by the user. This is modulation in practice and is evidence that how search is conducted by the user will impact the types of information Google search displays and how far its practice deviates. This indicates potentially that the performance of a user as patient (expressed through natural language) may reinforce their rendering as a patient in search results. In effect, how one searches determines how one is rendered and the results returned.

6.6 REVIEWING ORGANIC RESULTS

So after all this data-diving, what do we know, what can we infer, and where are we headed?

Documenting stable algorithmic practice is a crucial piece of the puzzle when analysing digital health platforms and the silos of knowledge they produce. Consistency in practice is important to document as it creates an evidentiary base for a more complex understanding of how the Google Search platform mediates information, where these moments of deployment for logic-driven decision-making within the black-box platform, and how the platform as an actor impacts the larger constellations of discourse and power.

From this chapter it is clear that there is a relatively consistent logic valuation applied to sources returned for condition-specific queries related to stem cell treatments. Moreover, Google Search privileges sources that fall under the categories of Research, News and Information, and Charity/NGOs. Although these categories have variation internally, by and large we can identify them as largely focused on public commons with input from direct-to-consumer commercial offerings only through some sources contained within News and Information.
Additionally, it is clear that while user profile is not necessarily a determining factor for variations within the returns, query structure and content *is*. This elevates the importance of *what* is being asked of the search algorithm, *how* that query is asked, *which* logic the algorithm deploys, and *what types of content* sources publish.

My findings in this chapter suggests that the practices deployed in displaying organic searches are not tied to direct functionality or primary mechanisms for commercial activity. Unlike the mechanisms for ads as seen in Chapter 5, organic search results do not have a *direct* system for content creators to influence ranking. This seems to be part of the balancing act Google must play in mediating results; maintaining relevancy of service to the user through non-commercial consideration of content and the injection of user-targeting or user-rendering mechanisms through which Google can draw profit. To be cynical, the implementation of commercial influences into organic search may be a matter of time. I will discuss this possibility in Chapter 7 and 8.

Moreover, this does not mean that there are no commercial means for influence of organic search results around health interventions. Importantly, the evidence explored in this chapter leads to my final results chapter in which I will take into consideration Google search and stem cell treatments as a whole.

It is now possible to discuss with evidential certainty the larger picture of search as a mediator of knowledge on emerging medical treatments. We can now have a deeper discussion of my central research: *how do search platforms mediate information on stem cell treatments?* In the following chapter, I will explore an additional aspect of search results display and the influence of the secondary market of Search Engine Optimisation. This will lead to a broader conversation about the implications of my findings and what it suggests are the mechanism of mediation at play on Google Search.
In this final results chapter, I will explore the fundamental question of what *mediation* means in regard to stem cell information on Google Search. The previous two chapters dissected the results of my data donation campaign for the purpose of mapping the landscape and uncovering the mechanisms of influence introduced by the Google Search platform on advertisement and organic search. Taking Astrid Mager’s articulation of platform politics as my starting point, in this chapter I address the holistic impact of Google Search on returned results. This chapter focuses on platform influence in its entirety using the data generated. Specifically, this chapter addresses the complete impact Google Search’s indexing mechanisms have on the presentation, dissemination, and access to information around stem cells and stem cell treatments.

To briefly return to the question of Blackbox analysis and its value for STS research, it is important to once again recognise the limitations of the data collected and its temporality. The results I captured are snapshots of Google as it was, so we can only *infer* algorithmically mediated processes between the point of query and output. Despite the insights gleaned from the data captured, the mechanisms deployed on search results pages are still frustratingly opaque and dynamic. The data points gathered - from source frequency to metadata to top results - are indicative more of trends and tendencies within results rather than concrete, immutable digital realities. This being said, the inclusion of direct data from Google provides important empirical evidence that I use to unpack the key characteristics of this specific form of platform mediation. The mediation I will describe is as dynamic as the platform itself.

The methods I adopted enabled me to gain insight into the structure and processes underpinning Google Search through dynamic snapshots of results. By combining the findings of the ad and organic search returns and the major actors within stem cell-related searches on Google, I am now finally equipped to answer the
overarching question of my thesis: how does Google Search (i.e. its governing algorithm and content-mediating mechanisms) and associated practices such as Search Engine Marketing (SEM), mediate the different sources of knowledge that people access online about stem cell therapies? And, crucially, why does this mediation matter? In this chapter I break this question down into three parts; what can we learn from the data donation campaign results about Google Search’s role as mediator? What does the data indicate Google Search’s impact to be when applied to stem cell treatments? What are the implications of this mediation for stakeholders in digital health?

In this chapter I argue that the indexing and display within Google search creates hierarchies of knowledge based on logics of success and vehicles for economic extraction. As discussed in Chapter 3, these include click-through rates-as-success, internal markets like Google Ads, and secondary markets of SEM. These hierarchies converge at the moment of search to create unique returns constituted by silos of digital health knowledge and discourse. Obfuscation plays a vital role as evidenced by hidden metadata and secondary economies contained within single result pages. These mechanisms play critical roles in Google’s Search’s process of mediation. For those searching for information on stem cells and emerging stem cell treatments, this has implications for health information access, risk, and the decentralisation of misinformation.

Throughout this chapter I am confronted with the question of whether or not Google Search steers or merely reflects the digital reality of information on stem cell treatments in its practice. Do the findings of Chapter 5 and 6 suggest a passive representation of the field of stem cell treatment information online or does it shape them?

To answer this, I return to Shove et. al’s suggestion that “practices consist of elements that are integrated when practices are enacted” (2012). It becomes clear through each step of this process of mediation that Google Search’s practice acts both as an active participant and enabler of secondary economies of health information online. Through its practice of returning search results, Google Search is both representative and performative of the health information it returns.
In this section, I discuss the mechanisms of influence through which hierarchies of knowledge are created within the search results page as a whole - both ad results and organic. I then address the factors that influence the ranking of results and the impact these rankings have on the regulation of emerging or unproven stem cell treatments within the larger context of stem cell information. Throughout this section, it becomes clear that out of the three practice logics (advertisement, “click-through success”, and a policy commitment to accurate representation of medical science), Google’s policy commitments to scientific accuracy are the least supported by the site’s own mechanisms of result ranking and presentation.

Herein I find one of the most compelling impacts of Google Search’s mediation on stem cell treatment information online; while stated corporate commitment to accurate information can create impetus to mediate search results away from unproven treatments and results with commercial intent (Lewandowski 2011), the structure of result pages themselves are overwhelmingly oriented towards profit mechanisms - whether to the benefit of content creators or the platform itself. This is both a core structural component of Google’s business operations and a challenge for search as a de facto cornerstone of public health information gathering.

Robin Mansell’s ‘new’ forms of scarcity in digital economies is relevant here (1999), as space within search results pages becomes a resource commodity. Mansell suggests that rather than digital technologies creating ever more opportunities and access, they can be impacted by monopolisation which then allows for manipulation of resources (be that material resources as in the case of online ticket sales or more nebulous resources like user access and information).

Throughout her work, Mansell has chosen to refer to digital technologies as intermediaries, rather than mediators. Indeed, I would argue that while Mansell anticipates the monopolisation by resource (or content) producers, it is the platforms themselves – as mediators – that potentially control the ebb and flow of scarcity.
through their very structure. In introducing space for ads, organic results, and Knowledge Graph returns within sensitive search results like those for stem cell treatments, it is creating the very scarcity-based markets that it will go on to rely on for commercial activities. While there may have been a market for unproven stem cell treatments before the invention of Google Search, there was no market for direct-to-consumer advertising of them through search until it was created by the platform.

Applied to my research, this scarcity is reflected in the space and access limitations within the search result page. *Who* can see *what* when is limited by the mechanisms Google deploys within the search result page (i.e. ads, organic results, and Knowledge Graph interjections). This scarcity is manufactured by the Google Search platform not as a bug, but a feature. The market for user's eyes is created, enabled, and disabled by the algorithmically-mediated structure of the Google Search platform.

In this chapter, I break down the spaces made available by Google within search result pages in terms of the mechanisms that regulate access to their self-created scarce resources. These resulting characteristics of search results pages include implications that will be explored for user groups, content creators (including private and public health care providers and research organisations), and for the platform of Google Search itself.

As revealed in the previous chapters, search results are Blackbox outputs regulated by three umbrella factors; what the platform wants users to see based on their advertisement and regulatory logic, what the algorithmically-mediated logic deems *probably, approximately correct* and leads to “successful” click-through, and a policy-mandated commitment to “accurate” information from the scientific and medical community. These logics operate in sync to fulfil the proprietary purpose of the platform. Namely, to be relevant to users, a profitable investment for its advertisement partners, and - at least as reflected in Google’s stated policy - reflective of the state of the science (Google ‘Knowledge Graph’, 2022).
I find it useful here to pause and analyse single search result pages for stem cell treatment related queries to make concrete the mechanisms at play (Figure 45, 46, 47, and 48). This gives form to these competing logics as reflected in a search result page’s architecture. In the following examples, advertisements (green), organic results (orange), and ‘Knowledge Graph’ (purple) returns are outlined as presented to a user.\(^{33}\)

---

\(^{33}\) These search result page images were captured using the protocol for Normative User Audit (Diakopoulos, 2017). This type of audit relies on the researcher’s own browser and creates anecdotal rather than broad evidentiary support. Regardless, for the purpose of demonstration in this section, I find it helpful to include as it clearly illustrates natural search result structures.
FIGURE 46: SNAPSHOOT OF GOOGLE SEARCH FOR THE QUERY “STEM CELLS” HIGHLIGHTING ADS (GREEN) AND ORGANIC (ORANGE) RETURNS, AUDITED BY AUTHOR 1.10.2019.

FIGURE 47: SNAPSHOOT OF GOOGLE SEARCH FOR THE QUERY “STEM CELLS MULTIPLE SCLEROSIS” HIGHLIGHTING ADS (GREEN), ORGANIC (ORANGE) AND ‘KNOWLEDGE GRAPH’ (PURPLE) RETURNS, AUDITED BY AUTHOR 4.16.2018.
Advertisements appear at the top of the search results page with a small indicator of its type (‘ad’). As discussed in Chapter 5, these are indexed and displayed through the Google Search ad platform which includes different algorithmically-mediated mechanisms. These include direct targeted advertisement based on advertiser-determined fields including search term and user type as well as automated auction systems like Google Ads which automatically connect advertisements to users. Advertisements may appear at top and bottom of the first page of search results. As the data donation campaign data showing a decrease (but not disappearance) of promotion of unproven stem cell treatments post-policy change suggests, there is potential for corporate policy-led intervention within this results-display mechanism.

Organic results appear after these top-level results. As explored in Chapter 6, these results are determined by a number of factors guided principally by the concept of Probable Approximate Correctness. Because success for the platform is grounded not in the objective answer to queries, but in the subjective behaviour of the user, we can infer that results can and most probably will change overtime. This shift is guided not only by what the individual user wants and needs, but by the collective behaviour of users interacting with the search platform.
The structural intervention of Google’s ‘Knowledge Graph’ information provides additional functionality on the search result page. As demonstrated in the audit snapshots, these inventions include frequently asked questions, direct messaging, and widgets for recommended results including similar searches and Google Scholar links\(^{34}\). In many instances, ‘Knowledge Graph’ information does not appear at all\(^{35}\). The content for this information is drawn from Google’s 2018 launched library of indexed information called ‘Knowledge Graph’ (Fensel et al., 2020). As discussed in Chapter 3, ‘Knowledge Graph’ is an opaque database of core facts dynamically compiled since 2012\(^{36}\) integrated with Schema.org’s open source web schema framework. Schema.org, it should be noted, is a non-profit collective project associated strongly with the World Wide Web Consortium, the main online international standards organisation (Schema.org, 2022). The impact of Schema’s particular focus on medical metadata will be discussed later in the chapter.

This mechanism perhaps reflects Google’s attempt at filling the role of a more judicious aggregator of primary content. Here also is the most opportunity for intervention from the policy team in terms of endorsed information. As the content presented is collected from Google’s own curated database of knowledge, Google plays the role of both librarian and library. In 2016, Google published that its ‘Knowledge Graph’ held over 7 billion facts (Noy et al., 2019) compiled from “a variety of sources that compile factual information” and licensed data to “provide information such as sports scores, stock prices and weather forecasts” (Google

\(^{34}\) Perhaps interesting to note that Google Scholar only appeared in my audits when I used a clean browser in which I was logged into my Google account. As a Normative User Audit is only anecdotal evidence of my results, it is difficult to claim directly that the Google platform has profiled my account as one held by a researcher who would be interested in results from their proprietary academic search platform, Google Scholar. However, it would not be unreasonable to assume that some user characteristics - including an interest in academic literature - has been associated with my account and, therefore, influences my results.

\(^{35}\) I did not collect ‘Knowledge Graph’ results during the Data Donation audit as they did not consistently appear as a singular field. Because they have a variety and dynamic output fields (e.g. FAQs, Recommended Search Queries, Google Scholar, etc.), it would have complicated the development of the data collection plugin beyond our available resources.

\(^{36}\) Google first coined the term ‘Knowledge Graph’ in 2012 as an ontological concept of the ‘real world made digital’. This led to the ‘Knowledge Graph’ database and platform integrated into Google Search platform in 2018 (Fensel, 2020).
The motivation for this self-administered database of knowledge put succinctly:

Our goal with the Knowledge Graph is for our systems to discover and surface publicly known, factual information when it’s determined to be useful. (How Google’s Knowledge Graph works, Google 2022).

I reviewed the documentation on the development and rollout of ‘Knowledge Graph’ through both publicly available documentation from Google (Google documentation, 2022) and reviewed literature on its development (Fensel et al. 2020, Vang 2013, Chah 2018). Both the documentation and literature on the algorithmically mediated platform suggest that the economic intent of its development is in its inherent proprietary value by means of deployment, rather than in-built functionality for profit-generating mechanisms. That is to say that the value of Knowledge Graph comes from its contextual applicability and usefulness to users, rather than through direct commercial activity. It has value because it becomes valuable to users. This is contrary to the platform structure used in the Google Ad platform in which the value is in selling advertisement space to content creators, rather than the sale of content itself. However, I will return to this idea of ‘Knowledge Graph’ as neutral later in the chapter.

So what does this mean for the balance of results returned for queries around stem cell treatments? Here I draw on some of our data donation campaign evidentiary base to conclude that organic results tend to favour charities, NGOs, and healthcare providers alongside news outlets.

Returning to the single results page, I was struck by the diverse mechanisms immediately deployed each time I searched. Algorithmically-mediated ads, organic results and ‘Knowledge Graph’ interventions shape this package of links to information around the complex topic of stem cell treatments. Each of these interventions are faced with their own internal pressures for content mediation that come directly from the structure and design of Google platform itself. However, these visible interventions are not the only mechanisms for mediation at play in our audited results pages. In the next section, I will pick apart the...
mechanisms for mediation hidden within single search results pages and track their impact on stem cell treatment information.

7.2 OBfuscAted IEnTercATIONS - MEdATA AND MiCRoDATA

In this section, I discuss the hidden mechanisms that impact the production and distribution of material for inclusion in Google Search results and the role obfuscation plays in this secondary economy within Search. Dissecting a single Google Search result page uncovers hidden mechanisms for intervention through which the Google platform shifts results according to its logics. I will explore two of these mechanisms - metadata and schema microdata - as examples of structural intervention into the creation of search results as a form of Blackbox auditing.

The first structure is analysis of metadata attached to each search return. This additional kernel of data creates space for a secondary discourse around the returned results. The second obfuscated mechanism is hidden markup used in the indexing of ads, organic search, and ‘Knowledge Graph’ results. Finally, I explore secondary economies present within a Google Search results page including previously discussed advertisement platforms and paid-services of Search Engine Optimisation (SEO) targeting organic results. These structural mechanisms for mediation information on stem cell treatments contribute to the discourse on Google’s active or reactionary impact on digital health information.

7.2.1 SCRAPED METADATA CONTENT - RICH TEXT SNIPPETS

The wealth of content-specific metadata collected was an unexpected find in the data donation campaign. Whereas I had anticipated only measuring the scraped visible data displayed on both ad and organic search results data, automated HTML scraping of rich text snippets made possible the extraction of additional data alongside each search result. This ‘content’ field included text excerpts scraped either automatically from the website by Google’s indexing bots or generated by the content producers themselves through metadata fields (demonstrated by the purple highlighten text in Figure 49).
Metadata generally is additional descriptive information packeted alongside a search return. In the case of my scraping results, these results were additional text either scraped from text included on a website’s landing page or hidden in the page’s code to be automatically picked up by indexing trawlers. Metadata packeted alongside the data collection queries included succinct messaging on the behalf of the content creator. As the following analysis appears to show, this scraped content does not appear to be included in Google’s automated curation as some messaging may fall outside the bounds of accepted marketing claims.

Before that, I would like to note a few observations about the information contained within the metadata content of sources labelled NGO/Charities and Research/Governmental. Often this included information on what a user should expect to find on the linked result page, a brief overview of resources available on the site, and key terms. Some metadata content included calls to action for funding and charity.

---

37 According to Seattle-based Digital Marketing Agency, Portent’s research, roughly 63% of summary content is changed by Google Search upon display from the original text provided by a web site’s metadata (Portent 2020).

This free fact sheet is as a [sic] concise introduction to Parkinson's disease. It provides details on the causes, diagnosis, symptoms and treatments of the condition. Contact a nurse. Call the helpline. Download a free booklet. Amenities: Neurological charity, booklets and fact sheets. (Brainandspine.org.uk, Organic Metadata)

{KeyWords:UK Stem Cell Foundation}. Develop The Most Promising Area Of Medical Research. Providing Funding For Invaluable Research Into Stem Cell Treatments. Make A Donation Now! Treatments & Technologies. Clinical Trials. Funding Research. (UKSCF.org, Ad Metadata)

Other metadata content addressed the targeted disease directly and included messaging around potential treatments and options available to patients. The language used within emphasises management of disease and result expectations.

While available therapies can treat some symptoms, people with Parkinson's urgently need better treatments to improve quality of life and slow or stop disease … (Michael J. Fox Foundation, Organic Metadata)

There's currently no cure for Parkinson's disease, but treatments are available to help relieve the symptoms and maintain your quality of life. supportive therapies, such as physiotherapy. medication. surgery (for some people) (NHS.co.uk, Organic Metadata)

In the case of returns from clinics advertising unproven stem cell treatments, associated metadata often included text that otherwise would be flagged by Google's ban in both advertisement and organic returns. The following metadata content was scraped from returns from Clinics with the highest frequencies of appearance across the data donation dataset.
Throughout these associated texts are several themes not evident in other result types. Patient experiences (‘Kevin Weeks’), costs of treatment and insurance considerations, availability and immediacy of access, and claims of treatment success are consistent throughout these metadata. These are all direct appeals to patients impacted by the targeted diseases.
As noted in Chapter 5, the implementation of Google’s ban on advertisement of stem cell treatments did result in the overall reduction of clinic advertisements. However, Clinic returns continued throughout the studies without noticeable change to the potentially problematic metadata content included.

This calls into question how Google’s mechanisms for content overview reviews this type of metadata content when implementing policy-driven interventions into search results. Consensus within SEO communities suggest that the use of metadata is not used in the process of ranking results on Google Search. However, they are used as part of larger presentation purposes (Miller 2022). The inclusion of potentially problematic – by which I mean not scientifically proven – text within the collected results of my data donation dataset suggests that rich text snippets are not reviewed by Google’s oversight review mechanisms. They are, however, still displayed to users.

If content scraped and displayed to users by Google skirts and crosses the line for approved messaging around the promotion of treatments, then we must conclude that policy-implemented content filtering is limited in its ability to implement content-specific bans. For stem cell treatments, this allows for the inclusion of direct marketing appeals to patients despite Google’s ban.

7.3 HIDDEN INFORMATION - MICRODATA AND WEB SCHEMA

Assessing obfuscated influences on Google Search mediation requires an interrogation of content websites and their hooks into Google’s indexing mechanisms. Here I shift my focus to analysis of HTML markup on oft appearing web pages and content producers seen within the data donation dataset. This analysis uncovers how the behaviour of web content producers aiming at hooking into Google’s indexing process - whether in ads, organic results, or ‘Knowledge Graph’ - is part of a larger dynamic relationship between Google and content producers towards the production of knowledge around search queries.

38 When generated from producer supplied content, rather than automatically by Google
I mentioned earlier the importance of Schema.org - a collaborative effort strongly supported by both Google and W3C for the development of best practice schema for structured data across the web and particularly deployed within Google Search (Kärle et al. 2016). This reference website is part of a semantic project to standardise content and source mapping across content creators. Essentially, it provides a common word book and sentence structure through which websites can be understood and related to one another.

Not only does Schema.org’s work form the underpinnings for Google’s ‘Knowledge Graph’, it also informs webpage markup for the purpose of indexing across the public internet. The organisation sets the standards for meaningful digital ontology – that is the fundamental organisation and codification of web content and structures – and provides best practice markup to allow website creators to codify their content for legibility for both humans and machines (Mika 2016). This shared vocabulary is critical for establishing a workable ontology of a common internet (Patel-Schneider 2015). All this is achieved through Schema.org’s library of microdata markup.

Before we get ahead of ourselves, what is ‘microdata’? Microdata is an HTML tool used to house metadata - or additional structured data taxonomy, source, descriptive information - alongside visible information within a website document (WHATWG, 2022). Microdata includes more information than simple tagging; it includes hierarchies and ontological relationships between terms alongside the packaging of content-specific information. Microdata uses a data model based on types and properties (Schema.org, 2022). In their broadest terms, this allows pages to be tagged with large umbrella terms associating the content with a particular thing (e.g. an Action, Organisation, Medical Entity, Person, Place, Product, etc.) and associated characteristics (e.g. address, domain, description, image, relation).

This HTML markup (often written in JavaScript Object Notation or JSON) is primarily focused on machine readability. That means it is not visible without inspecting a web page document and is aimed at integration into automated systems of aggregation like those used in Google Knowledge Graph and indexing. Microdata was introduced alongside HTML5 in 2011 with the intent of - amongst other concerns - creating hooks through which content published on the web might be integrated into existing
structures of knowledge and information gathering, mediation and display (Ronallo 2012). The schema includes microdata markup with descriptive fields that sit in a webpage’s public - yet not visible - code.\(^{39}\)

Including Schema.org markup on a webpage provides a number of benefits for a website vis-a-vis integration into indexing within the larger web processes. These include most importantly improved search integration and indexing as well as marking of relation between web pages and content (Patel-Schneider, 2014).

Salient for its impact on stem cell-related searches, Schema.org includes a particular focus on “Health and Medical Types” and definitions of medical actors - including hospitals, organisations, and practitioners - within the schema development. This includes structured data around the descriptions of conditions and their treatments as well as requirements for tertiary information around content publishers claiming medical authority. It is important to note here that Schema.org works to remain agnostic towards medical terminology. Their official documentation for Health and Medical Types notes:

> Note as well that this schema is not intended to define or codify a new controlled medical vocabulary, but instead to complement existing vocabularies and ontologies [sic]. (Schema.org, 2022)

To envision what this microdata markup looks like, let us take a look at one of the pages from my data donation dataset returned in high volume for search queries around Parkinson’s Disease from the Mayo Clinic (Figure 50). This page includes information for patients and carers about the disease of Parkinson’s alongside lists of symptoms, treatments, and signposting to related resources both within the Mayo Clinic digital informational ecosystem and beyond.

\(^{39}\)To note: best practice and historical norms in web publishing pushes strongly back against hidden text within web documents (Seymour et al., 2011). Schema.org markup is not considered ‘hidden text’ in the traditional sense although it is not visible during normal web viewing. This can be largely attributed to the integration of Schema.org into publishing best practices through the evangelism of the ontological approach by Google, Bing, and other index-focused actors online.
To observe microdata, it is necessary to view the webpage’s source code (Figure 51). Fortunately, the structure of Google indexing enabled me to do this. In order for a webpage to be indexed into Google Search, it must first be crawled. Crawling is the capturing of web information by bots using the information contained within a website’s source code to codify, sort, and manage found content on the web (SEO Starter Guide - Google, 2022). Because inclusion of data within the indexing process requires structural open access to any observer, the source code is only obfuscated within a website’s public source code, not inaccessible to users.
By cross referencing a website’s public source code with Schema.org’s code validator (https://validator.schema.org/), it is possible to note where microdata integrated into Schema.org is at work. Returning to the example of the Mayo Clinic’s page on Parkinson’s Disease, there is a clear example of Schema.org’s microdata for health information. The Mayo Clinic’s website on Parkinson’s includes three scripts referencing Schema.org, including the organisational reference, a list of included content, and the following:

```html
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "MedicalWebPage",
  "about": {
    "@type": "MedicalCondition",
    "name": "Parkinson's disease",
    "associatedAnatomy": [
      {"name": "Nervous system"},
      "signOrSymptom": [
        {"name": "Dieulafoy's lesion"},
        {"name": "Dysarthria"},
        {"name": "Parkinson's disease psychosis"}
      ],
      "audience": "http://schema.org/Patient"
  }
}</script>
```

Evident here is a best practice use of Schema.org's Health and Medical Types schema. This script allows any machine reading the website to understand that the content is a "MedicalWebPage" addressing the “MedicalCondition" of Parkinson's disease. It additionally provides the information that the disease impacts the Nervous system and lists some associated symptoms (“Dieulafoy's lesion”, “Dysarthria”, and “Parkinson's disease psychosis”). The intended audience - or user group - of the webpage is patients as indicated by the link to Schema.org's reference for patient groups.

I then applied this observation to my largest and most complete datasets for Multiple Sclerosis and Parkinson’s Disease. An audit of the most frequent sources within the Parkinson’s Disease and Multiple Sclerosis data across the data donation dataset revealed that a majority included some form of Schema.org markup within their source code (Figure 52). This indicates widespread ‘opting in’ by sources concerned
with stem cell information into Schema’org’s structure and hooks for web integration as this is not an automatic feature of web publishing.

![Figure 52: Top 100 webpages for multiple sclerosis and Parkinson’s disease across dataset with inclusion or exclusion of Schema.org markup.](image)

I then wondered if I might find interesting microdata breadcrumbs within the websites of private clinics explicitly promoting unproven stem cell treatments who appeared regularly within my data donation database. This led to two additional questions; one I felt able to answer using basic data analysis and the other an open-ended discussion. How were these clinics using Schema.org markup to integrate into Google’s indexing processes? And, on a fundamental level, why would content producers engage in this specific type of SEO?

![Figure 53: Screenshot of www.stemcellmexico.org, captured by author 29.12.2021.](image)
I audited the source code of one such private clinic, Stem Cells Mexico (www.stemcellmexico.org), a private clinic offering stem cell treatments for Parkinson’s Disease and Multiple Sclerosis amongst other major conditions and diseases (Figure 53). These offered treatments fall within my categorisation of “unproven” as no information made available on the website indicated that they correspond to any stem cell derived treatment currently approved as safe and effective by the MHRA, FDA or international research community as represented by the ISSCR and European Medicines Agency.

Within the webpage’s source code, I noted two instances of Schema.org markup - both of which aimed at demarcating the webpage as belonging to the schema type “Hospital” and “MedicalOrganization”. These categorisations were supplied alongside the clinic’s postal address, geolocation, and domains.

While the distinction of MedicalOrganisation is arguably broad enough a category type to include private clinics, I questioned the inclusion of the source type “Hospital” as these entities include specific types of medical infrastructure, regulation and
oversight by local government. I cross-referenced the provided addresses against the addresses of hospitals within Tijuana certified by the government of Baja California, Mexico and no hospital was found at the stated addresses (Gobierno de Baja California, 2022). This indicates that the clinic falsely attributed credentials and authority to their web presence.

Now, of course, this is just a small example of schema misuse. However, it suggests that standard markup used for the purpose of creating common taxonomy, identification, and authority conferring over the web has the potential for misuse within the context of information on medical treatments online. Moreover, Schema.org markup appeared in over 96.3% of source codes from the top advertisers across the data donation dataset (excluding the unscraped returns auto-generated by Google Ads). This was above the average (57.4%) for all return types.

Why does microdata matter? Since it has been explicitly developed for and integrated into Google Search and other keystone platforms on the web, this markup allows for exploitation of the automated systems and algorithmically mediated platforms that create the knowledge about stem cell treatments. As discussed, schema markup provides the basic hooks of information through which results on search platforms are coded, indexed and returned. This is particularly salient for Google Search given the company’s algorithmic reliance on Schema.org specifically for the indexing and subsequent mediation of returns. Moreover, abuse of schema markup is hidden in plain sight within the public code of content competing for views within Google Search results. While Schema.org’s validator allows individuals to check a website’s markup for correctly written documentation, style, and format, it does not serve as oversight for the content included therein. The result is that the self-declaration mechanism within schema.org without secondary confirmation creates a space for targeted misidentification leading to inclusion of misrepresented material in sensitive searches.

Addressing this “mishandling” of Schema markup poses a particularly difficult challenge for Google. Fundamentally, a markup scheme developed for web-wide usage relies on self-regulation on the part of content producers and web publishers. Herein lies the problem. If it is advantageous for a content publisher to include
markup that will increase the chances of content being organically included in search results, there is little incentive to exclude it – even if it violates the principle of the schema. Add to this the fact that this markup is publicly visible but obfuscated behind page source code and the accountability collapses.

If Google were to address the misuse of markup, it would require oversight and regulation of the schema’s use. This would require some third-party or Google side authentication and certification mechanism through which the categorical claims assigned to web pages were confirmed. Currently, Google has only mirrored Schema.org’s approach to self-authentication through its developer’s tool that allows for testing of rich text included in structured data (*Schema Markup Testing Tool*, Google 2022). This tool does not provide authentication or certification of the content included in web pages, but simply tests for machine readability.

7.4 MEDIATION THROUGH SECONDARY ECONOMIES - KNOWLEDGE GRAPH

As established in my chapters on Ad and Organic returns, profit extraction mechanisms are the *raison d’être* of ad inclusion and rest as secondary marketplaces atop the mediation of organic search results. While search engine ads include Google’s controlled marketplaces and exchanges, secondary economic activity around SEO is hidden within the business of web development. Two different mechanisms of result mediation and two different approaches to economic activity.

However, we still have one outlier for activity enabled by Google’s platform structure with direct economic activity; ‘Knowledge Graph’ returns. Indeed, the documentation on Google’s ‘Knowledge Graph’ appears economically agnostic and idealistically driven. In the project’s 2012 launch statement, Amit Singhal ties ‘Knowledge Graph’ to established informational guides as a grounding for ontologically-comprehensive database:

Search is a lot about discovery—the basic human need to learn and broaden your horizons.
Google's Knowledge Graph isn’t just rooted in public sources such as Freebase, Wikipedia and the CIA World Factbook. It’s also augmented at a much larger scale—because we’re focused on comprehensive breadth and depth. It currently contains more than 500 million objects, as well as more than 3.5 billion facts about and relationships between these different objects. And it’s tuned based on what people search for, and what we find out on the web. (Singhal 2012)

While it is not stated in the press material for the project, the introduction of ‘Knowledge Graph’ is also the functional introduction of Schema.org’s approach to structured data and integration into Google’s algorithmically mediated indexing of web content. As evidenced earlier in this chapter, we know that Schema.org’s markup of taxonomies and associated characteristics are used by content creators within the area of stem cell queries to signpost their content. This is a primary interaction between the content producer and Google Search platform for the purpose of promotion; it is a direct one-to-one relationship. However, do we have any evidence that secondary digital economies play a role in the indexing and mediation of results as related to ‘Knowledge Graph’? To answer this, I would like to return to microdata.

During my on-going audit of web pages included within the data donation dataset, I noticed that some schema.org markup included script indicating deployment as part of secondary SEO activities. That is to say, markup included indications of being part of a proprietary Software-as-a-Service (SaaS) model of SEO. This demonstrates that a content producer (or search result source) has paid a secondary SEO company to specifically include schema.org markup as an SEO service.

To understand this better, let’s pull apart one example of this secondary schema markup. Returning to the prolific stem cell clinic SwissMedica, a scraped snapshot of the public source code includes the following markup (emphasis added):
I noted a number of schema markups within these private clinic web pages included the term “Premium Yoast SEO plugin” which would indicate investment in a higher cost package. Yoast offers a number of Software-as-a-Service (SaaS) plans priced between £99 to £229 per year (Yoast SEO, 2022) which grant access to an SEO plugin that publishes schema.org markup alongside additional metadata (Figure 55). According to Business Insider Nederlands, Yoast SEO published profits of €10 million in 2018 and is active on over 7.9 million websites (Business Insider Nederlands, 2018).

![Yoast SEO for everyone](https://example.com/yoast.png)

**FIGURE 55:** SCREENSHOT OF YOAST.COM HOMEPAGE, CAPTURED BY AUTHOR 22.2.2022
Of note within the SwissMedica Yoast markup are two aspects. First, markup includes self-promotion of the Yoast SEO plugin as an active service on the website. This includes a link to the plugin’s documentation on the Yoast website alongside the active plugin version (v16.9). Secondly, the plugin script includes explicit content description that includes medical claims to efficacy explicitly targeting the treatment of Multiple Sclerosis and Parkinson’s Disease and Diabetes. This means that the SEO metadata scraped by Google Search is attempting to provide information on the efficacy and accuracy of treatments outside of traditional medical authorities and without link to local regulations or standards in medical practice. Additionally, in including this schema markup, the source – a private clinic – becomes self-identified as a source of medical expertise within the epistemological structures of Google Search’s information indexing. The specific Yoast plugin seen on SwissMedica’s website appeared in seven of the highest frequency private clinic returns included in the data donation dataset and over 15,000 scraped results for clinics within the dataset.

I must stress here that the implications of schema.org markup as SEO has significant knock-on effects for the epistemological organisation of information and expertise created through Google Search’s indexing of results for both organic search results and its fundamental intelligence library, Knowledge Graph. As discussed earlier, schema.org markup does not simply provide hooks for indexing of organic results within Google’s algorithmically mediated search platform. It also integrates into Google’s ‘Knowledge Graph’ from which the “factual” knowledge base returned in search is mediated. This provides an entry point for economic intent into what is presented as an agnostic mechanism for informational indexing.

7.5 ANALYSING THE IMPACT OF GOOGLE’S SEARCH’S MECHANISMS OF INFLUENCE

In analysing the mechanisms for result delivery, it is challenging to keep track of all of the dynamic interactions at play when delivering search results. With that in mind, I think it is helpful here to return to this chapter’s initial grounding object before moving on to apply the findings uncovered here, the single search engine result page (as exemplified in Figure 56). To me, this is the most important point of
discussion: the moment of contact between the platform and a user in which all economic and policy-based pressures on the platform are made concrete.

Now that we’ve pulled apart the three main contributing types of results that make up the page - ads, organic results, and ‘Knowledge Graph’ returns - the characteristics of platform mediation becomes clearer. Each facet of a single search result page is impacted by structures and mechanisms which place the end-user in the role of either buyer or product. By its very nature, the configurations of organic and Knowledge Graph search results extends the “warlike” relationship between content producers and platform Mager observed in search engine marketing. These mechanisms are what I term mechanisms of influence.

This relative positioning vis-a-vis economic intent is obfuscated on the page on either direct ad exchange platforms, within the markup on a page, or by the presence of secondary economic markets like SEO. Each facet includes a self-reinforcing reliance on Google’s platform structure and mechanisms to enable this economic activity. In this way, while Google’s policy oversight has some impact as seen in the reduction of clinical ads, the platform itself generates and favours primarily
economically-driven mediation in which the end-user is realised as a consumer or product.

Herein lies the tension. On the one hand we have users - potentially patients - who have integrated Google Search into their behavioural habits of information gathering around medical decision making and knowledge. On the other hand, we have content producers who - regardless of their mandates - are competing as if display to users are a resource subject to scarcity and subject to market economic mechanisms for competition. If we know now that the platform itself is structurally geared towards profit producing activities, then discussions of health online must be - first and foremost - a question of economics, rather than information dissemination.

For whom are these profit producing activities enacted? For Google Search, the product being sold is user access across mechanisms of influence (ads, organic results, Knowledge Graph). For content producers engaging in SEO and SEM, user are consumers. This is a fundamental transformation. Google Search – through its platform practices – transform users (including user as patient) to products to consumers.

Part of the impetus for my research has been to provide evidence that it is not simply sprawling concepts like polarisation, filter bubbles, and echo chambers that impact information dissemination around stem cell treatments online. These concepts are rooted in the behaviour and activities of the user, rather than the behaviour and activities of the platform. Indeed, I sought to pull at the strings of digital phenomena I observed first-hand through my work with EuroStemCell and determine the impact of the platform itself in shaping the type, quality, and characteristics of results generated when searching for stem cell information. As this chapter’s evidence indicates, mediation of information on Google Search is conducted through concrete mechanisms for indexing and returns - all of which are driven by economic intent yet importantl, different actors with distinctive intentions and economies. This creates a platform on which - while policies may indicate conceptual lines of demarcation of acceptable and unacceptable content – it is the profit extraction infrastructure that determines the platform’s activities and the returns it generates.
Once again I would like to bring attention to my point that Google Search both represents and is performative of the information it returns on stem cell treatments. As this chapter illuminates, organic results feed into the status quo until user-qua-public-discourse shifts towards a new ‘correctness’ uniquely enabled by Google’s mechanisms of influence. Here I am adopting the Foucauldian approach to *discourse* by which I mean regimes of knowledge impacted by the shifting interplay between language and power. I do not mean simply the language and vocabulary used when describing stem cell treatments. In particular, I point to the material aspect of discourses as “practices that systematically form the objects of which they speak” (Foucault, 1972).

As established in Chapter 2, the terminology of “stem cell treatment” (or, indeed, any of the query terms within my data) function as messy boundary objects moving through contexts – medical research, patient decision-making, private medical clinic offerings. As they do so, the discourse around them shifts. “Stem cell treatments” are the objects produced when we speak about them and this is materially influenced by the speakers, the contexts, and the interplay of power exerted on that object at the time of speaking.

I argue that the single search result page for a “stem cell treatment” query is discourse made material; each different interpretation of the term, each carried meaning, relationship to power, and relationship to the user is made machine readable and captured in this moment. The shift in discourse is potentially two-fold: across queries and across time. As the platform mechanisms of influence shift – like localisation, user targeting, platform policy restrictions – so too does the discourse shift as displayed in the search result page.

Another essential element uncovered through the data collection of Chapter 5 and 6 is the impact of geolocation and the de-localisation of stem cell treatments. Just as each result displayed on Google Search carries particular dynamics of discourse, so too do they carry geographical considerations. These considerations may at once be
enabling and restricting. Whereas commercially-motivated results open up possibilities of engagement with patients beyond nationalities, other results like clinical trials may be limited to specific national patient profiles. Despite the disparity of geographically-relevant access, these results are condensed into the single result page.

The findings of Chapter 5 and 6 are particularly relevant here. They demonstrate that two individuals Googling “What is a stem cell?” will return different results based on user-side specifications like user location and history as well as platform-side interventions like platform policy and regulation. Similarly, comparing one discourse-comprising search result page for a single query across time periods offers insight into the shift in discourse over time. This is demonstrated in Chapter 5 in the decrease in display of clinics offering stem cell treatments after the implementation of the ban on ads.

7.5.2 RENDERING CONSENSUS ON GOOGLE SEARCH

As discussed in Chapter 2, discourses of consensus on the efficacy and safety of emerging treatments are vastly different processes. Within medical science, consensus building is fraught, yet structured processes of evidentiary review that requires functioning oversight within and between medical research actors to progress local regulatory approval. It includes both regulatory review processes by medical authorities and inter-collegial negotiations between researchers. While there is extensive discussion around the efficacy of consensus building within the medical research community, established strategies for structured research consensus building include nominal group process, consensus development panels, and the Delphi technique (Waggoner et. al 2016). However, I am not going to expand much further into this; this simply isn’t the place for these debates. This process of consensus building within the medical community is a structured – not to suggest perfect – process. Sufficient to say, when at its best, the guiding principles of evidentiary review and substantiation within the pipeline from lab to clinic should be the baseline guide of medical scientific consensus. We can infer that this process often works as intended if we use documented patient outcomes from improved treatments as a bellweather. In the case of emerging treatments and, particularly,
those with running clinical trial histories, consensus may be a question of timing and communication.

For the public, consensus discourse in the case of stem cell treatments may entail a focus on evidentiary support and emphasis on common audit of experience and exposure (Downey and Geransar, 2007). As discussed in Chapter 3, behaviour of users - particularly patients and those impacted by conditions and diseases queried - is often an exercise in information gathering for personal use vis-a-vis the availability and potential of stem cell treatment intervention. Opinion based on personal research and decision-making presents a bell-curve of consensus and terminology adoption. This includes search as a review of options within one’s medical decision-making process. There is a tension here between gathering information on treatments offered by local health infrastructure and information on any stem cell treatment offered without geographic or healthcare infrastructure limitations.

In the case of stem cell treatments, this is where clinical possibilities and hype around emerging treatments may become blurred. Governmental regulations, like the ban on direct advertisement of medications in the United Kingdom, establish legal norms that Google must adhere to when making available its platform in different regulatory contexts. Therefore, Google provides geographically mixed results. However, the discourse (specifically reflected of dynamic relationships of power and language around stem cell treatments) transcends these regulatory borders. This broadens the range of the discourse to include medical interventions not just limited in availability, but also potentially outside local health care oversight structures that endorse an intervention’s efficacy and safety. This is not necessarily a problem in terms of patient rights and the right to information. However, it does create an unbounded discourse on the part of users which, in turn and overtime, creates feedback that shifts results generated through search platforms.

One might connect this information gathering process within the playing field of a search platform to Herbert Simon’s ‘bounded rationality’ in which expectation, desire, and preference impact the process of decision-making. However, within my research, I am more interested in the role the platform assumes rather than user-as-patient behavioural paths.
Examining the popularity of search terms related to emerging treatments over time here supports the idea of Google Search queries as a reflection of interest by the public. In these graphs, there is a clear upward trend overtime for the queries “stem cell treatment” and “stem cell treatment cost”. The following Google Trends depicted in Figure 57 and 58 illustrate this movement.

Both graphs include a noticeable dip in 2020. This may be related to the constriction of health care services, travel, and non-emergency interventions during the COVID-19 pandemic.

Instead of raw numbers of searches over time and region, Google Trends documents “popularity” of a search query over time defined as search interest over time relative to the highest point of interest for a given region. “A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.” (Google Trends, 2022)
FIGURE 58: WORLDWIDE POPULARITY OF SEARCH QUERY “STEM CELL TREATMENT COST” 2004-2022, SCRAPED FROM GOOGLE TRENDS, RETRIEVED BY AUTHOR 01.2022.

Here is writ large the public-qua-user interest in stem cell treatments and their availability. Combined with the understanding that Probably Approximately Correct logic and “click through” success filtered back into the algorithmic governance of Google Search platform, there is a strong indication that public interest and demand will influence information mediation in this case.

This is another demonstration of a structural logic through which Google intervenes as mediator. In the case of organic search results, SEO, as discussed in this chapter, introduces economic intent into the creation and dissemination of information.

7.5.3 ACCESS TO INFORMATION AND AUDIENCES

These findings indicate that in rendering both users and information, Google Search actively creates secondary economies around health information dissemination. These are supported by structural mechanisms specific to both the platform and – more granularly – the query submitted. These include traditional SEO interventions as seen in the adoption of Schema.org markup as well as interventions through neutrally-presented mechanisms like Knowledge Graph.
For commercial actors, the presence of platform-specific markup suggests that there is an understood cost of entry to recruiting patient customers through Google Search. Moreover, as this chapter indicates, commercial actors like private clinics are willing to engage on the terms set by Google Search through hooks like schema markup to reach potential customers. This includes not only investment in advertisements on Google Search but in secondary SEO including SEO targeting schema markup. The cost associated with these engagements can be used as evidence that these commercial actors see strategic value in paying towards Google Search’s commercial practice.

For non-commercial actors, the implications are perhaps a bit more concerning. On the one hand, the absence of SEO markup on most non-commercial pages suggests that non-commercial actors are not engaging as deeply with Google Search’s commercial activities. This may indicate that they have not yet identified the importance of these activities (in particular, schema markup) or they have chosen to rely on other means of positive weighting to raise the profile of their returns in stem cell treatment related searches.

Conversely, the willingness of some non-commercial actors to engage in SEM (as seen in Chapter 5) indicates that there is some strategic consideration of Google Search as a worthwhile investment for reaching patients and publics. This indicates that some public funding – such as that earmarked in support of charities, NGOs, and research organisations – may be being funneled into Google Search. This raises ethical questions about the validity of paying for access on proprietary digital platforms with public funds.

Interestingly, the uptake of schema markup by commercial, but not non-commercial actors, undermines the relative authority of the metadata indexed. If only commercial actors are using the markup and if that markup is being misused, what value does it bring to the platform’s indexing of results? This may be an area that shifts as the Google Search platform continues its iterative development.

As seen in Chapter 5 and 6, geography contributes to the characteristics and makeup of returns presented to users. In some cases, this may be due to differences
in regulations. This was seen in the presence of Google Ad exchange returns featuring the promotion of direct-to-consumer pharmaceutical interventions in the US. These returns were not found in the results generated from the UK, Canada, and Australia due to local regulations prohibiting pharmaceutical marketing to patients.

In other cases, these geographical differences are due to relevance and localisation of services. For example, relevance of the NHS returns was expected and consistent within UK results. It should be noted that the inclusion of geographically-specific results across study groups suggest ignored borders. This includes returns across all studies granting weight to US-based charities like The Michael J. Fox Foundation (US), MS Society (UK) and stem cell research organisations and institutes like the California Institute for Regenerative Medicine (US) and EuroStemCell (UK/EU). Additionally, the increased presence of private clinic advertisements withing US returns when compared to UK, Canadian, and Australian returns suggest the platform notes proximity of private stem cell clinics within the United States.

Structurally, the importance of geographic relevance is reflected within Google Search in the ability to target location on Google Ads and within SEO. As I will discuss in my conclusion, the role of localised regulatory oversight may be limited in influencing the platform’s practice outside of user exclusion from functionality.

In my conclusion, I will explore what these findings might mean for stem cell treatment information and digital health more broadly. Calling on a wide spectrum of STS, sociological, and digital health literature, I will not only reflect on the process of data collection I undertook here but find new significances for digital health researchers, considerations to address the issues raised by my research, and potential for future work.
As Michel Callon wrote of obligatory passage points, “to translate is to displace” (1986). The practice of Google Search translates stem cell treatments out of their medicalised contexts. In so doing, they are displaced and made “messy” as Andrew Webster suggested. However, keeping in mind the self-rendering and perpetuating system of Google Search, I would suggest that Google Search does not simply displace messy objects, but renders them in forms specific to its platform structure and motives. This extrapolation places me in a position to address my core research question: how does Google Search mediate information that people access online about ‘proven’ and ‘unproven’ stem cell treatments?

Just as coding makes text machine-readable, so too does Google Search render query topics like “stem cell treatments” understandable through the context of Google’s lens. “Stem cell treatments” are simultaneously medical interventions, biomedical concepts, and – most importantly – potential products to be offered. In applying this rendering, Google Search absorbs the query “stem cell treatments” into its ideology and practice. This, then, impacts how, when, and where it can be interacted with (whether as an information content producer or user).

Within this system users are only guaranteed access to information that serves Google Search’s framework of commercial activities. While SEM (through Google Ads) provides direct mechanisms for content producers to pay for access to users, SEO’s secondary economies provide ample space for a variety of profitmaking interventions in the hopes of raising content rank. Even Knowledge Graph – although presented as neutral – is connected to commercial intent through two functions; its integration of SEO-impacted schema markup and through the project of encouraging further user engagement. The practice of encouraging further user engagement through the presentation of information primes users to return to the platform and be rendered products in future queries.
What of content producers? Bearing in mind the results presented in Chapter 5 and 6, it is clear that all content producers take part in some fashion in the commercial activities of the Google Search platform. From clinics to charities/NGOs, whether directly (SEM) or indirectly (SEO), all content producers contribute to the effectivity and undertaking of the platform. This means that economic capital is being spent by diverse contributing actors captured in the results, all of which contribute to the project of Google Search’s commercial practices. Contextually this has wide-reaching implications.

If everything returned on Google Search around stem cell treatments is either explicit advertising or the result of a process structurally influenced by economic intent focused on targeting users, then what does this say about digital health knowledge on this, the largest platform for patient health information discovery? What are the implications of this mediation for patient and healthcare stakeholders in digital health? *Cui bono* from this invisible, churning algorithmically-mediated platform’s behaviour?

To answer this question, I found myself informed by an unexpected source. As demonstrated in the example of Holocaust-related searches explored in Chapter 3, moments of system disruption often provide critical insight into the inner workings of opaque structures of organisation. In 1978, the Czech poet, playwright and future statesman Vaclav Havel penned the essay *The Power of the Powerless* as an exploration of the post-totalitarian Soviet governance system in crisis. In this essay, Havel explored the fundamental practice of this system exposed by the moment of crisis he observed. He wrote “while life ever strives to create new and ‘improbable’ structures, the post-totalitarian system contrives to force life into its most probable states.” Havel surely had not envisioned the digital health context in which I read his words nearly half a century later. And yet his observations of a system seeking influence of both power and discourse bear significant application to the algorithmically-mediated technologies that have infiltrated everyday digital life. Havel continued:
The aims of the system reveal its most essential characteristic to be introversion, a movement towards being ever more completely and unreservedly itself, which means that the radius of its influence is continually widening as well. This system serves people only to the extent necessary to ensure that people will serve it.

(Havel, 1978/2018, pg. 19)

My mapping of Google Search results – through the analysis of advertisement results, organic results, and the search return page itself – paints a picture of a similar technological system. Its fundamental adherence to *Probably Approximately Correct* computing to encourage user click-throughs is itself an act of pushing users into “probable states”. It’s policy-applied yet platform enabled mediation of advertisement results consolidates its power in determining legitimacy (in my case, approved and unapproved stem cell treatments). The complete search result page is platform structure attended by content creators (who adapt their web pages to better serve the platform) and users (who serve the platform as product). The difference here is that this system is not yet in crisis; it operates, is maintained, and expands as intended.

In this conclusion, I reflect further on my supporting research questions that have empowered my exploration of information around unproven stem cell treatments on Google Search. This is an exercise to reflect on my efforts in deploying my methodology, the findings they returned, and their impact on STS and digital health literature. It is also an exploration of potential further studies, open questions and challenges, and considerations for future regulatory approaches to health information on Google Search.

8.1 POWER AND THE DEFINITION OF “UNPROVEN”

*How has “unproven stem cell treatment” been defined and by whom? What practices of power are performed through these definitions?*

Chapter 2 explored the question of “proven” and “unproven” and found that terminology around stem cell treatments is traditionally defined largely through
proximity to cycles of regulatory approval. However, as Chapter 3 investigated, the implementation of this traditional definition on Google Search is often referenced in policy, but not made explicit on the platform of Google Search. Importantly, my research found that the adopted definition of “unproven” or “unproven” was less important in searches around stem cell treatments than the mechanisms of influence on Google Search in presenting content to users.

As the wide range of content types returned and reviewed in my results chapters suggests, the definition of “proven” and “unproven” stem cell treatments on Google Search is in constant flux. It is Google Search’s algorithmic ideology, itself, that mediates the definition of stem cell treatment – providing what Havel describes as “both excusatory legitimacy and an inner coherence” (Havel, 1978/2018). Havel’s consideration empowers Mager’s algorithmic ideology to break away from simple strategic governing tenets and becomes power itself. Google Search’s practice expresses this ideology: *Google Search provides information only in so far as it serves the overarching project of commercial intent – whether that commercial intent be the direct sale of advertisement or the configuring of users as products.*

If Mager argues that Google’s algorithmic ideology is not naïve, I argue it is fundamental and purposefully hidden behind policy-declared positions of neutrality. It is not through the intentions of Google – the company indicated by its internal policies – in which this ideology finds purchase. In fact and as discussed in Chapter 3, Google has conscientiously promoted a narrative of the positivism of Google Search; its algorithmically-mediated practices serve as a shield of objectivity for its undertakings. This is deliberate in its self-construction vis-à-vis user engagement and trust, but it is not the expression of the platform’s true ideology. Rather, it is the mismatch between the perception of Google Search as a neutral arbiter of knowledge and the real power deployed by Google Search as a commercial machine through where this algorithmic ideology is able to flourish.

Mager frames the deployment of this strategy in terms of capitalist enterprise and policies supporting extraction. She writes:
First, it is essential to understand that privately owned search engines benefit from our marketing strategies, consumer desires, ignorance, compliance, innovation fetish, politics of privatization and, most of all, globalized capitalism that increasingly escapes local socio-political cultures and frameworks. It is important to see that our own actions and willingness to be seduced by search engines and their convenient services help to stabilize search engines and the commodification of information. (Mager 2012)

Mager recognises the interplay between user and system that hints at a crucial idea: users would not be searching for information around stem cell treatments if, in many cases, they did not want, hope, or desire something. This search would not be prompted without the cycles of expectation and promise perpetuated by biomedical research that are attached to the development of emerging treatments. This is a fundamental circumstance that enables Google Search.

As seen in Chapter 5, Google Search’s two competing practices of self-regulation – policy and commercial – are often in conflict. While a policy may be implemented as a point of governing philosophy, that policy may not be carried out to the letter by the platform itself due to the conflicting mission of the platform’s commercial intent, selling user views to content creators. Fluctuating yet pervasive returns of policy-banned content indicates that these lines of acceptability are struggled within internally. Google Search’s de facto defining of proven and unproven stem cell treatments through its internal mediation is an act of flexing this peculiar, but very real, strength.

Power – signified by capital, resource allocation, the organisation of human pursuits like the collective actions of patient groups and publishing of stem cell treatment-addressing websites – serves the ideology. This is a subtle but significant consideration that impacts how solutions might be envisioned. Power within this system always contributes a portion of its activities to maintaining and expanding Google Search platform. This, in turn, always serves the algorithmic ideology that underpins the platform: its commercial intent.
This control shifts non-commercial content creators (like research institutes, regulators, and governmental healthcare agencies) to a secondary and referential position determined by their size, offline influence, and economic ability to engage in the Google Search economy. They are included in so far as they support the line of demarcation that the algorithmically-mediated (rather than policy-driven) components of Google Search – internally – determine acceptable from unacceptable content. The moments of intervention by human actors to regulate inclusion is limited in the face of the platform’s commercially-driven functions. This is observed in both the limitations of policy decisions implementation and user-dependent content reporting. The integration of Schema.org markup across commercial actors indicates similarly a prioritisation of adherence to the platform’s commercial intent rather than policy.

As discussed in Chapter 7, scarcity plays a significant role in this configuration of Google Search’s power. By enabling only limited space for user access across the platform, Google can control both the way content creators and users interact while encouraging future contributions to the project of Google Search as a monolithic platform. Scarcity creates impetus to engage as Google Search’s influence grows.

What does the imposed scarcity on and power expressed by Google Search mean for stem cell treatments and digital health information down the line? For those concerned with digital health and stem cell treatments, this coalescence of power and discourse is potentially worrying. The limitations of Google Search’s ability to self-regulate away from commercial content paints a picture of an algorithmically-mediated platform beyond traditional means of intervention. Although I hesitate to suggest an out-of-control Artificial Intelligence terrorising humanity, there is something particularly troubling about the findings in Chapter 5. They suggest that Google’s employees are limited in their ability to intercede in the functioning of the platform when human ethics have determined certain results to be inappropriate.

New research should focus on “shifting the focus of attention from impacts search engines have on society towards social practices and power relations involved in the construction of search engines” (Mager 2012). This might include work on the integration of library sciences back into Google Search platform (as suggested by Kressens 2017) and further research in Blackbox auditing to provide a form of
oversight-through-research. These approaches may shed further light on how platform power impacts the boundaries of biomedical definitions like “proven” and “unproven”.

8.2 LESSONS LEARNED FROM STUDYING GOOGLE

*How can Google Search – as a proprietary, algorithmically-driven search engine – be studied empirically? What methodologies can be employed and why is this decision important?*

Blackbox analysis applied to Google Search offers deep and meaningful routes for digital health sociology and STS-grounded research by providing rich data for lively digital ethnographies. As my work indicates, grounding research in technologically rich and data intensive methodologies create opportunities for interrogation that could lead to meaningful wider research impacts. The significance of Google Search documented within my research is reason enough that tackling algorithmically-mediated platforms *where and as they are* is crucial for a number of reasons.

I find myself returning here to Deborah Lupton’s discussion of digital data assemblages of lively data. Future studies may find a completely different character of results for digital health topics as processes, oversight, required regulatory standards and technological ability shift over time. I would argue that the value here in integrating directly-observable, yet dynamic data is the moment of truth it contains. The search platform may change, the discourse around stem cell treatments may change, the medical treatments offered may change but, in that time and place of data donation, this captured result was the output given to a user. The themes, trends, and practices displayed in the data I collected are meaningful outside of their specific content because they provide evidence for the platform structures that influence them. Tracking this overtime through robust data will grant insight into long term evolution of information and mediation over time. This represents an important aspect of the power of auditing in support of social science research (Krafft 2021).

Firstly, as my research suggests, Google Search cannot be ignored as a uniquely influencing actor in any query subject constellation in which it is used. It warps the
language, relationships, and movement of information across discourses of any query subject. Moreover, through its platform structure, it “displaces” (Callon 1986) and renders query subjects through its own lens which privileges commercial intent. This is a subtle, but crucial, impact on the trajectory of any other actor’s behaviour and opportunities for engagement.

As discussed in Chapter 3, the movement towards monopolistic platforms on the world wide web is increasing (Rantilla 2020). The more time is spent by users on average on these platforms – either inclusively or as a jumping off point for their online activities – the more these platform’s importance grows. Wrestling with large datasets and dispersed users is a significant challenge of observation presented to future STS and digital health researchers.

I would argue that relying on Normative User Audits (as discussed in my Methodology in Chapter 4), while easier for technically-limited researchers, creates a power imbalance in research published. As mentioned in Chapter 3, sprawling platforms like Google Search have often relied on arguments of result “co-production” by users to avoid scrutiny. Adopting crowd sourced and Sock-Puppet audits as a methodological tool creates richer, more significant results from diverse user perspectives. Importantly relying on larger and more diversely generated datasets, researchers avoid the possibility for research to be dismissed as case (or user) specific.

Secondly, questions like those raised by Richard Rogers about the accountability of proprietary technology platforms in serving in critical public sectors (2009) must be addressed by observation. Crowd sourced and Sock-Puppet audits offer researchers a way to place Google Search returns under scrutiny. The opacity of algorithmically-mediated platforms must not be a barrier to study due to their importance in shaping digital health realities. Indeed, avoiding tackling Google Search with evidentiary rich methodologies further entrenches the platform’s position as interrogatable.

Astrid Mager’s line-in-the-sand call to irritate platforms (2014) finds a methodological foothold in Blackbox auditing. By deploying a trawling net of data scraping, researchers can create snapshots of digital life and discourses that may go beyond
their own research applications.

Should not our methodologies be equally expansive and dynamic? Even when failing or creating limited scope data, the act of auditing technology platforms (as well as other ways of performing Blackbox analysis not explored here) creates literature that acknowledges the secreting of information by technology platform owners. Addressing this – even through failure to break through – may create the academic, social, or political capital needed to prompt technology companies to make their platforms more accountable. This is particularly important in areas where private technologies moderate subjects of public concern like health information.

8.3 WRESTLING WITH THE THEORETICAL IMPACT OF GOOGLE SEARCH

I would first like now to return to my secondary research question - How is the mediation of information on stem cell treatments shaped by structural factors within the platform of Google Search? What impact did Google’s policy change have on paid search advertisements relating to stem cell treatment queries? Does the information found differ depending on the user’s or query’s characteristics? When? And, hopefully, why?

Not only does it appear that Google Search – through its structural practice and algorithmic-ideology – actively mediates information, but through this practice extends novel opportunities for engagement that expand its own influence. It is Google Search and those that engage in commercial activities who are most enabled by the platform. Because Google Search has become a de facto public platform through its proliferation, digital health information for treatments like stem cells cannot help but be guided first and foremost by commercial interests.

Chapter 5, 6, and 7 provided evidence that Google Search enables the mediation of information through platform-specific structures of information indexing and display. This led to different modalities of filtering and privileging being applied to each section of the single result return page that shifted the types of information presented. The results suggest a platform system overwhelmingly driven by an
algorithmic ideology that incentivises and privileges commercial intent. This occurs even when contradictory to Google’s corporate policy (as demonstrated in the implementation of the ban on direct-to-consumer marketing of stem cell treatments which resulted in a reduction, not elimination, of problematic content).

These chapters also indicated the influence of user targeting – based on location, local regulation, user history, and targeted condition or disease – in the display of both advertisements and organic search results. This influence is subtle and difficult to capture, yet clearly present. Chapter 7 also suggests that the very practice of website indexing, and display has become part of a particular information economy through the adoption of Google developed and promoted schema markup. Fundamentally, this is due to the inseparability of Google’s informational and economic infrastructure. Therefore, the commercial influence happens as a number of stages within the ecosystem of Google Search, including at the point of search return to the user as well as during website development on the side of the content producers.

The characteristics of ubiquity and opacity of Google Search observed by Matthew Hindman (2008) are not static but growing characteristics. Growth here is indicated by the expansion of Google’s influence on technical web infrastructure development, the hidden integration of more specialised algorithmic mechanisms for mediation, and the developing secondary economies of influence budding around Google Search. As more health services move online, this suggests that platform ideology will begin to influence the fundamental characteristics and expression of healthcare offline.

Through these mechanisms of influence, Google becomes the reference, the referrer, and reference point for all information on stem cell treatments on the web. Stem cell treatments – therefore – cannot be understood as messy boundary objects without considering the filter of commercial practices applied to them at the point of any query search including them. From the perspective of the public who use Google Search as a primary mode for health information gathering, this is a potentially critical consideration. For these actors, stem cell treatments become objects that have been pre-rendered vis-à-vis their economic relationship to the searcher. This is
a fundamental shift in the character of the information made available to Publics-in-
Particular like patients. As Chapter 5 and 6 suggest, the information presented to
users condenses conversations of authority, risk, and access into singular
presentations within the search result page.

Direct commercial intent may not be included in the content contained within search
results as indicated by Chapter 6 and 7. However, the influence of secondary
markets of SEO and integration of SEO into schema markup for baseline Knowledge
Graph data suggests that non-commercial results presented to users may still be
tertiarily related to capital creation on behalf of the Google Search platform. If the
commercial intent is not direct between seller and consumer, Google Search still
engages in commercial practices through its rendering of users as products. The
more invested users are in finding health information through Google Search, the
more Google Search expands its influence, maintains its hold on the flow of
information, and recruits future users-as-products for its ad services.

There are both explicit and subtle mechanisms for encouraging Google Search’s
tendency towards introversion or its “movement towards being ever more completely
and unreservedly itself” (Havel 1978). Refusing to engage with Google Search’s
platform environment may result in the exclusion of content creators both
commercial and non-commercial. For example, failure to adopt Google’s approved
schema for web indexing results in greater difficulty in content indexing. Web
developers must adhere to these additional markups to have their content indexed.
Soft power pressure to engage with Google Search on its terms include expansive
focus on the importance of SEO for content discovery and the marketing of SEM as
an engagement strategy. All of these mechanisms encourage the movement of user
and content creator engagement into the terms and spaces dictated by the Google
Search platform.

8.4 TOWARDS ADAPTIVE REGULATION

As the areas of regenerative medicine and digital health information evolve, it is
crucial to interrogate the regulatory framework for algorithmically-mediated platforms.
To interrogate these frameworks, I would like to briefly discuss the oversight and the
rule of regulations from the perspective of Google and governmental oversight. Finally, I engage with existing literature on potential solutions addressing Google’s supremacy over health information online including anti-trust proposals and the potential for a rendering of Google as a public utility as a possible future framework.

As Chapter 5 and Chapter 7 suggests, there is an internal tension within the oversight and internal governance of the Google Search platform between policy and structural function. Whereas Google may wish to intervene in search results based on policy decisions like its ban on stem cell treatment advertisement, the structural function of the platform is designed to resiliently offer this content. Intervention on the part of Google within the logical functions governing these mechanisms of influence may not be adaptable to regulatory requirements.

I would like to return briefly to a section of my dataset that was particularly fascinating when compared across geographic donor origins. As seen in Chapter 5 and 6, there are some situations in which certain user groups have been excluded from Google Search mechanisms of influence because of adhering to local regulatory standards. Users within the United States were presented with targeted advertisements for prescription pharmaceuticals enabled through Google Ads exchange and auction. This was displayed in the results database as specific types of obfuscated code. This functionality was not extended into search results presented to users in the other countries of origin – UK, Canada, and Australia – due to stricter regulations on the direct-to-consumer marketing of pharmaceuticals to consumers.

However, I would argue that this inclusion and exclusion of certain results based on compliance with local regulations does not indicate a successful example of invention. The mechanism of Google Ad exchange was not influenced in its logic to halt the display of certain results. It was simply turned off for certain users groups (e.g. the lack of advertisement returns around suicide as discussed in Chapter 4). Taken in consideration with the findings on unproven stem cell treatment advertisements from Chapter 5, this might indicate that the management of Google Search platform simply cannot control the platforms outputs in its current iteration. Future reworking of Google Search’s algorithmically-mediated software would need
to address this issue to create a system that could respond to regulatory intervention.

Assuming user access to a fully inclusive algorithmically-mediated platform, what role might future regulation play in the oversight of a platform like Google Search? Here I find it useful to look at current governmental approaches to algorithmically-mediated platform regulation. Governmental regulation of information on stem cell treatment and Google Search is currently light. Internal regulation on technology platforms largely leaves the onus on content producers to apply non-digital limits to medical intervention regulation to digital spaces. This means that general regulation of access to medical interventions is assumed to be applied online.

I must be careful here to note the difference between the wealth of literature on the regulation of stem cell treatments compared to that addressing stem cell treatment information. There is significant (and locally specific) work on regulations and oversight of stem cell treatments which rely on legislation at the level of the applicable healthcare system as well as standards for product commercialisation. As discussed in Chapter 2, regulation (particularly so-called Right to Try legislation) has considerable impact on the geographic distribution of medical treatment facilities and the interventions they are able to offer.

However, for digital health information preceding these legislatively-approved routes for new treatments, active legislation is less pertinent. Outside of issues of data privacy, direct marketing of pharmaceuticals, and general guidelines for legal and illegal content related to health topics (as discussed in Chapter 2), there is little direct legislation addressing digital health information on platforms like Google Search. Instead, discussions around anti-monopolistic legislation may be particularly relevant in discussions of possible regulatory interventions.43

---

43 The European Union has been ahead in applying anti-trust regulations to technology platforms that handle user data as well as data privacy laws across platforms (like the General Data Protection Regulation). Interestingly, these efforts have largely applied anti-trust legislation to address issues of anti-competitive behaviour.
Firstly, this is interesting because it applies a regulatory framework in which platform activities are judged vis-à-vis access to fair markets, rather than other civic considerations. Concerns are not over what is presented to users through platforms by rather if the process enables fair competition despite the proprietary nature of the platform. Inherent here is a suggestion that these platforms perform and constitute unique public services in digital environments that require fairness. It is difficult to maintain this position while defending the platform’s status as a private digital property. I will return to this idea of digital public services shortly.

Secondly, within these legal considerations, there is a clear line of demarcation between the handling of data. Personal data – as addressed by regulation like the GDPR\(^{44}\) and as discussed in Chapter 4 – is defined as identifying and privileged information like name, contact information, IP address as well as “special categories” of information like criminal conviction (ICO 2022). Non-personal data is defined as information collected “automatically without human involvement” (Rusche 2019).

This legal framework begs the question for future research: at what point does the automatic data collection and analysis performed by algorithmically-mediated platforms constitute the collection of personal data? If algorithmically-mediated platform like Google Search can identify and cater results to specific user groups based on previous user history and relation to conditions and diseases (as seen in Chapter 5), does it matter that this protected information was ascertained without explicit user self-identification? When does regulation on private data kick in as algorithmically-mediated platforms get better at their proprietary functions? What does this imply for sensitive gathered information like diagnosis status? These are questions that will need to be addressed by future STS and digital health researchers.

To envision solutions to this outsized influence, I would like to consider proposals explored within anti-monopolistic regulatory discussions of technology platforms. In *Antitrust and Platform Monopoly*, Herbert Hovenkamp outlines the legal perspective on solutions-driven trustbusting specific to technology platforms. He points to several

---

\(^{44}\) These definitions are commonly held across the UK, US, Canada and Australia.
examples of anti-monopolistic regulatory strategies when technology platforms have moved into a punitive phase of review. He writes:

In many situations, a better approach would be to re-structure management rather than assets, which would leave the platform intact as a production entity but make decision-making more competitive. A second option to breaking up firms would be to require interoperability -- and in the information context, mandate the pooling of valuable information. These measures could promote competition and simultaneously increase the value of positive network effects. (Hovenkamp 2020)

The results of Chapter 5 suggests that the former solution – to restructure management of the Google Search platform – presents limited opportunities for intervention and oversight. As Chapter 7 illustrates, every aspect of the Google Search result page is functionally tied to its commercial activity and mission, even if simply as far as they contribute to the stability and supremacy of the platform. This indicates that changing management as an antitrust strategy for moderating Google’s influence on the platform is limited and may not yield enough oversight to meaningfully change the impact of Google’s Search’s commercial practice.

The secondary solution – to require interoperability – is more interesting in the context of Google Search and digital health. It is important here to reimagine the term “competition” outside of a profit-driven context. Interoperability offers the possibility for the Google Search platform to enable competition of authority – that is to say, space for checks and balances on evidentiary support of health information presented. One might imagine a separate search platform only for health information in which content is coded in coordination with medical regulatory agencies.

However, this is a difficult solution for several reasons. Firstly, it requires a proprietary company to hand over management of their digital property to third parties unaligned with the platform’s algorithmic ideology. In this case of digital health information and Google Search, the intension of the platform and management would be mismatched; the former designed to carry out a commercial practice, the latter presumably guided by localised healthcare regulation and ethics.
Secondly – and perhaps as my data suggests, more difficultly – this solution still requires technical solutions to allow the platforms to functionally operate within set restrictions. The problem of how intervention is achieved within the machine-run part of the platform still must be solved.

The impulse to treat proprietary technology platforms like normal commercial products aligns with many of the regulatory frameworks already in place in countries like the US, UK, Canada, and Australia. However, as the proposals for anti-monopolistic intervention suggest, regulation quickly devolves into specialised consideration due to the importance of these platforms. Perhaps it is more valuable to adopt a drastic reconsideration of algorithmically-mediated platforms like Google Search. This might look like the regulation of Google Search as a public utility.

Mark Jamison argues that Google Search (at the time of his writing in 2012) should not be presented as a public utility as it is “not monopolistic in nature, does not preclude rivals from competing against it, does not rely on grant of a franchise as does a utility, and does not take control of rivals content or services” (2012). This, I would argue, is an overly simplistic view on the nature and trajectory of Google Search within the critical area of public health. If we keep in mind both the active obfuscation of Google’s practices and the impact of those practices on digital health information as supported by the evidence gathered in Chapter 5, 6, and 7, then it is harder to suggest that Google Search does not occupy a functionally monopolistic or quasi-monopolistic position globally. This being said, Google Search as a public utility is a difficult solution to imagine as it contradicts the very fundamental structure of the platform as a vehicle for profit extraction. Not least of all, it would require a significant investment of political and social capital to carry out.

8.5 RECONSIDERING HUMANITY IN SEARCH

In reflecting on my findings, it becomes clear to me that the strategy of Google Search within health is to place the onus of responsible information discovery on the users. This is present in the very structure of Google Search. As Chapter 5, 6, and 7
indicate, the rendering of patients as products permeates Google Search’s practice of mediation. The platform’s corporate reliance on direct-to-consumer and user-as-product rendering places the onus of analysing personal risk squarely on the shoulders of the patient.

Yet at the same time, this rendering of users is part of Google Search’s project of obfuscation. While configuring users as products through its technical mechanisms of influence, Google Search promotes the message of searchers as independent, self-regulating entities who hold the power of choice. This underpins Google Search’s self-position as a neutral provider of information.

Deborah Lupton immediately complicates the notion of the independent patient/user by stating that context, power dynamics and resulting behaviour cannot be divorced. She writes:

Those who advocate the ideal of the self-responsible actor fail to recognise the socioeconomic and political factors that influence people’s use of digital health technologies. (Lupton 2014)

As Chapter 5 revealed, those who are impacted by conditions and diseases may be those who are rendered most vulnerable to practices of commercial targeting on the Google Search platform. Understanding Google Search as an economic platform and de-neutralising its space to the public is the first step towards accountability in digital health. If, as suggested in Chapter 3, Google Search renders patients as products through this practice, then what solutions might we seek out to address this potential for exploitation?

Before exploring some of the proposals I find most valuable, I would like to briefly engage with solutions from the field of market-based interventions. Several recent anti-trust cases against the online retail company, Amazon, included a peculiar approach to consumer welfare. In her research, Lina Khan points to weak rulings against the monopolistic behaviour of Amazon justified through the idea of market-based solutions for protecting consumers. The consumer welfare theory applied in these cases suggest that digital monopolistic behaviour is tolerated if consumers
benefit. Here benefits are defined as decreased prices or better products (Khan 2017).

Unfortunately, here the conceptualisation of patients as products fails to translate to human actors in digital health. Or rather, it may translate, but horrifically. One might be tempted to apply solutions from product safety discourse to the question of how to protect patients on Google Search. However, once a patient is treated, they are uninteresting as a product as they can no longer be engaged with as a target for a commercial offering of medical intervention. This is, of course, unless the patient/product sent into a pipeline of continued, sustained, and extractive treatments for which Google Search might provide the direct marketing. This is, quite obviously, a repugnant idea and antithetical to the basic standards of public health. These solutions still rely on the rendering of end users as independent actors. Although available market options may shift, it is users themselves who continue to carry the responsibility of choice.

At the same time, I am wary of the durability of any solution that does not fundamentally empower end users. As Andrew Webster’s approach to RRI suggest, end users (or, in the case of biomedicine, patients) must be involved in meaningful ways. This conceptualisation for engagement pushes back on the unidirectionality of information from “expert” sources to lay patients. Surely a solution must be able to strike a balance between ethical information mediation and the empowerment of individual choice.

In contrast to either paternalistic or potentially exploitative rendering of users, I found Niels Kerssens’ insightful reflection on the value of ethics in search engines that enable ever widening gulfs of agency. Kerssens states that:

… more humanistic values warrant our attention in a positivistic search culture where the editorial gap between the haves (algorithms) and the have-nots (users) is widening with the former becoming increasingly refined and complex, while user activities are progressively pampered and formatted by search engine interfaces. (Kerssens 2017)
Kerssens then points a source of human intervention best practice that I had not previously considered; library studies. Kerssens writes:

In a nutshell, library-oriented information scientists substantiated the idea that searching for information, whether online by computer or manually, was a fundamentally human process and by its very nature a form of skilled human behaviour necessitating what Julian Warner described as selection labour (2010)). In their view, information searching involved complex human behavioural and cognitive processes that could be enhanced by the use of various techniques, strategies and tactics, but which could never be fully mediated by computers. The aim of Bates and others devoted to the psychology of searching was to theoretically support the centrality of human thinking processes in the search for information, and to force a shift in attention ‘from a focus on the machinery, the information technology, to the brain that is running it’ (Bates, 1979, p. 205). (Kerssens 2017)

This perspective was utterly eye opening for me. Integrating the rich history of library sciences and the power relations it creates through the act of information gathering has significant potential to change the relationship between mediator and recipient. With this final reconfiguration of mediator/user relations through library science, I feel there is potential for future research on a way forward. Future explorations may find value in bringing into conversation Mager’s call for restructuring, Andrew Webster’s approach to RRI, the interoperability proposed by anti-trust regulation, and library sciences-as-meaningful-user/creator-collaboration. Future work within STS and digital health might address in more detail the ways patients integrate Google Search into their decision-making process in regards to stem cell treatments. This is important because – although Google Search is a key gateway into health information line – it is not the only one. Its outsized influence is also diminished once people are enabled to engage on other platforms and information sources they trust and rely on, such patient forums or groups.
It’s three a.m. You’ve been recently diagnosed with a life-altering condition. You pull out your phone or laptop and begin to Google. What do you find?

Early internet information architecture pioneer Peter Morville famously stated, “what we find changes who we become” (2009). As I reflect on my outcomes, I cannot help but be struck by the feeling that this adage reflects only half of the equation. It is not simply what we find but how we find that changes our future selves and lived realities. The way we are presented with knowledge in an age of bombarding informational input must be considered as important as the knowledge itself. Or, at the very least, this knowledge cannot be separated from the means of its acquisition.

It is because of this observation that my findings around digital health leave me with a sense that Google Search is a perversion of the act of searching. For an activity so closely tied to the experience of being human, there is shockingly little humanity within it. More said, there seems to be a conscious effort on the part of the platform and its operators to divorce humanity from the act of searching for information to create collections of knowledge. The result is shadow play of information discovery. Google Search presents a collection of user-targeted ads, commercially-optimised organic results, and market-filtered subject content to users masquerading as choice.

For stem cell treatments and other unproven biomedical treatments as messy, but critical objects of emerging biomedicine, this obfuscation of commercial intent creates an unequal and potentially exploitative gauntlet run by users and patients. Using the unique methodology of a crowd-sourced and Sock-Puppet Blackbox audit, my research captured a critical moment of informational mediation by Google Search around stem cell treatments. It included the capture of Google Search’s policy change impact around direct-to-consumer advertisement as representative of its potential for intervention, the indexing of organic results, and invasive and commercially-influenced markup within SEO. These findings contribute valuable methodological and evidentiary findings to larger discussions around the
dissemination and search of information about emerging medical treatments. My research provides a strong argument for the consideration of search technology itself as an outsized influence on critical issues facing STS researchers such as data protection and privacy, public education, and civic commons online.

When I return to my 3 A.M. Rule, I realise that I have implicitly and unexpectedly imbued it with an appeal. We must find ways to make the health information we find in moments of vulnerability better, safer, more equitable. This requires addressing the fundamental influence of Google Search as a mediator of information passing through its platform architecture. It is my hope that my work included here might provide a material starting point for that next conversation.


Caplan, A. I. (2017). Mesenchymal Stem Cells: Time to Change the Name! Stem Cells Translational Medicine, 6(6), 1445–1451.


EuroStemCell, (2016). *What diseases and conditions can be treated with stem cells?* Retrieved 4 April 2022, from https://www.eurostemcell.org/what-diseases-and-conditions-can-be-treated-stem-cells


Foucault, M., & Sheridan, A. (1972). *The archaeology of knowledge ; and, The discourse on language*. Retrieved 20 June 2022 from https://www.semanticscholar.org/paper/2dba3a05f5f9f2e1db7b158eb7f7f1aaa18920eb6


Hovenkamp. (2020). Antitrust and platform monopoly. Yale LJ. Retrieved from https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/ylr130&section=39&casa_token=K3m64Yo-fQAAAAAAA:sZuL-hAACGams5oSp2ukF6a5yWdOa5Us6VrRSrxz3Mh6_qgepNnbkKplFZAI6B-b9_dpQLOgPA


Pols, J. (2012). Care at a Distance. Amsterdam University Press.


Tanner, C., Petersen, A., & Munsie, M. (2017). ‘No one here’s helping me, what do you do?’: addressing patient need for support and advice about stem cell treatments. Regenerative Medicine, 12(7), 791–801.


Turner, E., & Rainie, L. (2020). Most Americans rely on their own research to make big decisions, and that often means online searches. Retrieved 26 June 2022, from https://www.pewresearch.org/fact-tank/2020/03/05/most-americans-rely-on-their-own-research-to-make-big-decisions-and-that-often-means-online-searches/


Webster, A. (2017). Regenerative medicine and responsible research and innovation: proposals for a responsible


