Beyond Rare Disease Patients: Exploring Machine Learning Interventions To Support People Experiencing a Diagnostic Odyssey

Emily E. Nielsen
[redacted]

Submitted to Swansea University in partial fulfilment
of the requirements for the Degree of Doctor of Philosophy

Swansea University
Prifysgol Abertawe

Department of Computer Science
Swansea University

17th June 2024

Declaration 1

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed Emily Nielsen (candidate)
Date 17th June 2024

Declaration 2

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed Emily Nielsen (candidate)
Date 17th June 2024

Declaration 3

I hereby give consent for my thesis, if accepted, to be available for electronic sharing.

Signed Emily Nielsen (candidate)
Date 17th June 2024

Declaration 4

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed Emily Nielsen (candidate)
Date 17th June 2024

Abstract

People with a rare condition face several hurdles throughout their odyssey to obtain a diagnosis. This odyssey lasts several years and involves frequent referrals and misdiagnoses, often resulting in permanent and severe consequences on patients' health. In addition, patients feel unheard by their healthcare providers and isolated from their peers who 'just don't understand'. The UK Strategy for Rare Diseases states that patients can play a significant role in their diagnosis if given suitable resources. However, patients with rare diseases feel that they lack the support they need. This thesis explores the role that technology can have in addressing this gap in support.

Within this context, this thesis spans a range of topics, from human-centred design approaches to generating data and presenting a new methodological approach. Through a human-centred approach, we **characterise the needs of rare disease patients**, thus opening the research space to include previously unmet support needs. In addition, we identify limitations with existing measures of success and highlight the importance of a reduction in the time of diagnosis for rare disease pre-diagnostic technology. This provides the basis the **simulation-based methodological approach** that we develop. The simulation-task aimed to mirror the information seeking tasks that rare disease patients undertake. To do this, we **curate data that is representative of a rare disease patient's perspective**, both in terms of the terminology used and the stage in which symptoms and clinical findings are discovered. In addition, we **curate a pre-diagnostic patient matching prototype that is designed around rare disease patients' needs** and demonstrate that (in comparison to two search engines) our application shows greater potential to: aid clinical experiences; facilitate empathetic support networks; and provide better facilitation of information-seeking. All of these contributions stem from a critical examination of the experiences that rare disease patients go through on their journeys towards diagnosis and aim to pave the way for future research within this area.

Acknowledgements

First and foremost, I want to thank my amazing team of supervisors who have all contributed so much to my PhD journey. I specifically want to thank Tom for the support and kindness when I was unwell; it was over and above what I expected and it gave me the peace of mind that I needed during my recovery. I would count myself lucky just to have a first supervisor as good as you, but Matt Roach and Alan Dix have also been instrumental in my journey. You have all inspired and supported me greatly, each in different ways, and I cannot express how grateful I am for your support.

Thank you to Amicus Therapeutics UK and the EPSRC for funding this project. Thank you to everyone from the Amicus team who supported me throughout my PhD. I would like to specifically recognise the support from Solomon, Laura, Lisa, and Gerard. Thank you so much for all your help and encouragement.

I also want to thank my family for putting up with my stressed, busy PhD life. You have been so supportive and never judged me for struggling to balance everything, even when I found it hard to make time for you. I love you so much, and couldn't have done this without you. Thanks Mum for listening to me ramble on when it's really not your thing.

Ray, I don't know what I would have done without your amazing mentoring during my PhD. Thank you for keeping me calm when everything became overwhelming; helping me to work with my neurodivergent brain; and teaching me to be kinder to myself.

Thank you to the CDT, and everyone who has supported me as staff or students. Thank you to all my friends who made my PhD life so much fun. I particularly want to thank Anna and Beth who were always there to chat after a bad day, or celebrate after a good day.

Finally, I want to thank my partner, Connor, for his support during my PhD. Thank you for loving me despite the chaos of my neurodivergent brain. I love you so much.

Thank you to everyone else I haven't been able to mention. I am so grateful to have so many amazing people in my life.

Contents

Chapter 1

Introduction

> "Feeling alone when faced with a diagnosis of a rare disease is common. That feeling of loneliness actually connects us."
>
> *Tamron's story - NORD*

1.1 Background

For the purposes of this thesis, we will use the UK's definition of a rare disease (a condition affecting less than 1 in 2,000 people [1]), however, there is no globally consistent definition of a rare disease. Many countries have significantly different definitions for a rare disease, by standardising the definitions to every 100,000 people, we see that a disease is said to be rare if it affects less than: 10 in Russia [2]; 11 in Australia [3]; approximately 37 in Singapore [2]; 40 in Japan [3]; 50 in the UK and EU [1, 4]; approximately 75 in the US [2] and 65 according to the World Health Organisation [3]. The lack of a consistent definition has been suggested to contribute to a number of issues which arise with rare diseases [5], including diagnostic difficulties, a lack of effective treatment as well as limited knowledge and resources for both clinicians and patients.

A rare disease is characterised purely by its prevalence and can vary significantly in terms of disease progression and symptom manifestations, however, people living with rare diseases share similar experiences in their health journeys [6]. In particular, obtaining a diagnosis is a significant hurdle faced by people with rare diseases [7–9]. In the process of obtaining a diagnosis, the average rare disease patient waits four

years, attends five consultations, and is misdiagnosed three times [7]. Furthermore, it is estimated half of the 30 million people living with a rare disease in Europe remain undiagnosed [8]. This conveys only a small insight into the challenges faced by rare disease patients who are seeking a diagnosis. A rare disease patient's journey towards diagnosis is often described as a diagnostic odyssey [7,9], since patients must embark on a long and difficult journey to simply attain a diagnosis.

Incorrect and delayed diagnoses result in delayed and undue treatments, including surgeries. This can have irreversible detrimental impacts on the health of rare disease patients [6]. Moreover, untreated health conditions can make people unable to work, resulting in a loss of pay, or participate in social activities [6]. Once a correct diagnosis is obtained, many rare disease patients still have significant unmet medical needs [10] as over 90% of rare diseases do not have effective treatments [11]. The challenge of diagnosis is a significant contributing factor since a lack of people diagnosed also means a lack of people available to participate in the clinical trials required to identify new treatments [10]. The diagnosis of rare diseases also presents a financial burden, both for patients and for the healthcare system. Many of the costs incurred (e.g., paying for incorrect treatments) are unnecessary and do not help the patients [12].

Beyond physical or financial impacts, rare disease patients face significant emotional and psycho-social challenges [13]. Many rare disease patients feel neglected and disenfranchised from the healthcare system [6,7]. This raises a greater need for self-advocacy since *'no one is taking ultimate responsibility'* of their diagnosis or health [14], but also increases the emotional burden of seeking a diagnosis. Furthermore, patients often feel undermined by their doctors and their peers, as symptoms are disregarded as laziness or character flaws [6], in turn causing doubt and a greater desire for the validation that comes with a diagnosis.

Clearly, rare disease patients face significant physical, social, emotional and financial impacts due to the challenging nature of their diagnosis. The importance of well-being beyond the physical impacts of a condition is recognised by the World Health Organisation who defines health as *"a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."*[1]. This shows that the wider impacts of patients are an important part of their health and should not be overlooked. Therefore, it

[1] https://www.who.int/about/governance/constitution

is important to consider how to support individuals with the many challenges associated with seeking a rare diagnosis - not only for their needs as patients but also as people.

However, rare diseases are often overlooked due to their lack of prevalence [6]. This may be because utilitarian-based thinking may lead to the disregard of rare conditions simply because they represent a minority. The World Health Organisation states that, *"Every human being has the right to the highest attainable standard of physical and mental health."*[2], but health authorities, healthcare providers, patients, and patient organisations assert that individuals with rare diseases are deprived of this fundamental human right [6].

Moreover, we later demonstrate in Chapter 3 that rare diseases receive even less attention when it comes to pre-diagnostic patient-facing technology. This may be due to several challenges within this research space. Firstly, the lack of patient perspective data may present a significant hurdle for researchers. Secondly, rare disease patients inevitably present a limited pool from which to recruit participants, which is heightened by the lack of social, collective identification. Thirdly, the challenge of evaluating technology over-diagnosis presents hurdles. Long-scale deployments are not feasible given that it can take 30 years for an individual to receive a diagnosis of their rare condition [6].

As we can see, rare disease patients experience significant challenges that ultimately result in the denial of their human right - to obtain sufficient healthcare. The UK Strategy for Rare Diseases [15] recognises that patient input may have a significant role in diagnosis decisions, but only when they have access to suitable resources. Existing technology used by patients prioritise popularity, so they will not cater for rare diseases. Clearly, there is a need for pre-diagnostic patient-facing technology which supports rare disease patients on their journey towards diagnosis. However, there exist significant barriers to conducting research in this space, thus hindering progress in this area. Therefore, motivated by the difficulties faced by patients seeking diagnosis and the need for greater support, this thesis aims to explore potential avenues to address these barriers within the research community in order to pave the way for future research in this area.

1.2 Research Agenda

Let us now explore the approaches taken within this research. This PhD (and an integrated Master's degree) was proposed jointly by the author, supervisory team and

[2]https://www.who.int/news-room/fact-sheets/detail/human-rights-and-health

Amicus Therapeutics UK, a pharmaceutical company for rare diseases. The author's focus on rare diseases was motivated by the significant challenges that patients face when seeking a rare diagnosis and the belief that research into patient-facing technologies have the potential to provide much needed support to help people facing a challenging and rare diagnosis. Founded by a father of two children with a rare neuromuscular condition (Pompe disease), Amicus aims to address unmet treatment needs for several rare conditions, predominantly focusing on lysosomal storage disorders which are multi-systemic and are often life-limiting and progressive. As such, the context of a patient experience and the challenges associated with diagnosis were a key driver for this research. Quarterly meetings with Amicus Therapeutics were held throughout this project in order to provide support and guidance, including facilitating conversations with other rare disease organisations to aid the contextual understanding of the diagnostic odyssey of rare disease patients. This includes conversations with the founder of Swii.ch Health – a specialist agency, focused on rare disease patient engagement, the CEO and the data science team at Aparito – a drug development company designed for rare diseases which aims to empower patients in the drug development pathway and the CEO of Medics4RareDiseases – a charity which aims to change the attitude towards rare diseases amongst individuals training to become healthcare professionals.

This PhD included an integrated Master's degree which focused on the potential negative impacts of online health information, particularly focusing on the risk of false positives, self-diagnosis and health anxiety[3]. Of course, the master's component is not an examinable part of the thesis, but it may offer insight in these areas and leads into the formulation of the PhD direction and research questions.

It is important to acknowledge that, for any research project, the researcher is not an objective, neutral party, and as such, we must consider the influence of the researcher's positionality and reflexivity on the research. The research presented in this thesis was conducted by Emily Nielsen with support from her supervisors as well as from Amicus Therapeutics, other rare disease organisations, and other stakeholders involved in rare disease research and diagnosis. This includes human-computer interaction (HCI) researchers, machine learning researchers, clinical experts and patient advocates. In addition, it is important to note that Amicus Therapeutics was founded by a father of rare disease patients, and that the researcher has a lived experience of a diagnostic

[3] swansea.ac.uk/computational-foundry/epsrc-centre-for-doctoral-training/msc-projects/emily-nielsen/

odyssey for an currently undiagnosed (but suspected rare) condition. As such, the numerous stakeholders involved in this research are committed to improving the experience of people seeking a diagnosis. Therefore, we recognise that these personal and professional experiences have influenced the research approach, agenda and the lens in which we analyse this work.

Epistemologically, this thesis is based on a constructionist perspective, which aims to understand the truths of individuals, acknowledging that knowledge is context-dependent and socially constructed. In particular, we draw from interpretivism, recognising that knowledge is subjective and shaped by the researcher's own experiences and biases and thus identifying the strengths of the positionality of the researcher. As such, we explore the participants' perspectives and experiences, rather than imposing our beliefs and preconceptions on the research.

This lens aids the research by challenging the research paradigm. In particular, we critically examine the support needs of rare disease patients and the meaning of success in technology within this area. As a result of this, we expand the scope of pre-diagnostic technology for rare disease patients to provide empathetic support, not just informational support that has typically been explored, recognising that rare disease patients are people first, and patients second, as has been recognised within HCI approaches for common conditions [16]. In addition, whilst the existing research primarily focuses solely on the accuracy of pre-diagnostic technology we argue that the timeliness of diagnosis is a crucial factor to assess the potential benefits of pre-diagnostic technology for rare disease patients.

We maintained a clear focus throughout this PhD on the needs of rare disease patients. As such, the specific research directions at each stage were shaped by touch points with rare disease patients. Hence, this approach remained centred around the patient throughout, rather than setting a static objective from the start which would not allow the research approach to be fully led by the needs of rare disease patients. As such, our research approach evolves from the experiences of people with rare diseases and the contextual implications throughout this project. In particular, we adopt a patient-focused and context-specific approach as a key driver throughout this research in order to centre our work around the needs and challenges faced by rare disease patients. This resulted in the research outlined in this thesis covering several different approaches and

research areas in order to best meet the gaps within the research space (i.e., human-centred approaches and low fidelity prototyping, implementation and evaluation of prototype web applications; curation of methodological approaches, machine learning, and synthetic data generation). In addition, we adopt a collective approach to explore the needs of rare disease patients as a collective, rather than disease-specific needs. We describe these approaches and why these choices were made in more detail below.

1.2.1 Patient-Focused Approach

The majority of the research in this area focuses on clinically-based interventions, however, rare disease patients consult with numerous specialists along their journey towards diagnosis. As such, the consulting clinician is constantly changing, so rare disease patients are the only constants on their diagnostic journey. Moreover, patients are the ones living with their condition each day and know their symptoms intimately in a way which cannot be conveyed in a 10-minute consultation. The importance of patient's experiential knowledge has been increasingly recognised within healthcare and research [17]. Moreover, rare disease patient often become experts in their condition and have to explain it to their healthcare providers [18]. It follows that a patient-focused approach will facilitate a greater appreciation of the diagnostic odyssey. Moreover, a patient-focused approach can explore the wider issues beyond the direct or physical impacts of diagnosis that patients experience.

In addition, the role of the patient may be far more significant for rare conditions [6,15]. 94.6% of clinicians believe that they have insufficient or very poor knowledge of rare diseases [19] and lack time to research rare diseases themselves. In contrast, it is common for patients to research their health, utilising resources such as ChatGPT, Google or Facebook [20–24]. It follows that patients may be able to contribute significantly to consultations with their healthcare providers. Indeed, The UK Strategy for Rare Diseases states that patients can play a significant role in diagnosis and treatment decisions *if given suitable resources* [15].

However, patients with rare diseases feel that they lack the support they need [25–27] which may be why patients resort to technology which is not specifically designed for health. In addition, these popular technologies are designed with the common interest in mind, and as such, prioritise popularity in their results. As such, these applications will not cater for rare disease patients; information that is relevant to a rare condition

is inevitably irrelevant for the majority. For example, people with Fabry disease may experience a number of non-specific symptoms, such as flu-like symptoms or pins and needles in their hands and/or feet. Given that only 855 people are affected by Fabry disease in England[4] in comparison to 15 million people affected colds and flu each year[5], it follows that information about Fabry disease will be far less relevant for the majority of people than information about colds and flu.

This suggests that rare disease patients do not have the resources required to play an active part in their health. Therefore, there exists a need for patient-facing pre-diagnostic technology which caters for the needs of people with undiagnosed rare diseases.

Therefore, we explore the area of pre-diagnostic patient-facing technologies. In particular, we adopt a human-centred approach to explore the design space of pre-diagnostic technology with a focus on the needs of rare disease patients. Clearly, the identification of individuals with a rare condition who remain undiagnosed presents several challenges, so instead we recruit individuals who have obtained a diagnosis to reflect on their experiences of seeking a diagnosis. As such, we want to empower patients to advocate for their health investigations; facilitate greater conversations during consultations with their healthcare provider; and support patients with the challenges they face during their diagnostic journey.

The promise of machine learning within health has been recognised throughout academia [28, 29] and the NHS[6]. In particular, machine learning approaches can utilise data from numerous years and locations. In contrast, the lack of prevalence of rare diseases make it unlikely that a clinician will have experience of a given rare condition before diagnosing a patient. Therefore, machine learning approaches show significant potential to aid the rare disease diagnostic odyssey, so we explore the potential of machine learning based pre-diagnostic technology for rare disease patients in this thesis.

1.2.2 Context-Specific Approach

Maintaining the context of a rare disease diagnosis was a key drive for this project. In particular, we aim to identify the technological implications of the lengthy and challenging diagnosis of rare conditions. By exploring the contrasting needs of rare

[4] nice.org.uk/guidance/hst4/resources/migalastat-for-treating-fabry-disease-pdf-1394900887237

[5] waht.nhs.uk/About-The-Trust/News-and-Media/Campaigns-and-Projects/Get-Ready-For-Winter/Colds-and-Seasonal-Flu

[6] england.nhs.uk/long-read/artificial-intelligence-ai-and-machine-learning/

disease patients with those of common conditions, we identify unique requirements from technology to establish a design space which caters for these unmet needs. A context-specific approach has been recognised as a crucial component of the design of technology [30], and that the potential of digital health can only be realised with a context-specific approach [31]. In particular, by using a context-specific approach, this thesis identified the wider needs and unique challenges of rare disease patients.

As such, this PhD was shaped and re-directed throughout by our findings around the context and the patients. This drives the design of supportive pre-diagnostic technology rather than a self-diagnosis intervention. Not only can self-diagnosis interventions have detrimental impacts on a patient's health [23], but they would not meet the wider support needs of rare disease patients. In contrast to self-diagnosis interventions which encourage patients to diagnose themselves, our research aims to empower patients with relevant information in order for them to have greater consultations with their healthcare providers. That is, self-diagnosis interventions do not account for the clinician's role in diagnosis; in contrast, pre-diagnostic technology may facilitate greater team work between patients and their healthcare providers in order to obtain a diagnosis. Hence, our research aims to explore the design space of pre-diagnostic patient-facing technologies to provide suitable support for individuals with a rare condition alongside their clinical investigations.

In addition, we identify the differences in the evaluation of common conditions in comparison to rare ones. The information support of pre-diagnostic technology for common conditions may be evaluated by the performance at a distinct time. In contrast, we argue that a positive outcome in the context of rare diseases is in the reduction of the time taken to achieve diagnosis, and not simply the performance at a single point. As such, patient-facing pre-diagnostic technology for people with rare conditions has greater information support outcomes if relevant information is provided early on in the diagnostic odyssey. In addition, pre-diagnostic technology has greater outcomes if it caters for the unmet support needs of rare disease patients which include the need for: empathetic support; advocacy in clinical consultations; and the validation of symptoms and lived experiences. Therefore, this context-specific approach is also a vital component in identifying a suitable method of evaluation. In particular, we examine the impact of pre-diagnostic technology on individuals with rare diseases, both as patients seeking health information and as people with wider support needs.

1.2.3 Collective Approach

Given that there may be over 10,000 types of rare diseases [5], it would require thousands of research projects to attempt to cater for each condition individually. As rare diseases are defined only by their prevalence, their manifestations may vary significantly. However, rare disease patients face similar challenges due to the lack of prevalence of their condition [6]. In particular, many healthcare professionals lack knowledge on rare diseases [19]. This results in substantial delays for patients to obtain a correct diagnosis; frequent consultations with different clinicians; and multiple misdiagnoses along the way [6,7]. Therefore, we adopt a collective approach to support rare disease patients through these collective experiences.

Moreover, whilst rare diseases are of course individually rare, they are common when we consider the many different types of rare conditions as a collective. More specifically, rare diseases affect 1 in 17 people in Europe [32], 1 in 10 in the US [33], or between 263 and 446 million people globally [34]. To put this into context, as shown in Figure 1.1, the highest number of active COVID-19 cases at any one time was less than 63 million people globally, far less than the number of people affected by rare conditions. Hence, millions of individuals are affected by a rare disease, but many remain undiagnosed [6] and over 90% of rare diseases do not have effective treatments [11]. As such, we can consider rare diseases to be a public health crisis [35,36] and shouldn't underestimate the impact that rare disease research can have.

1.2.4 Accommodating a Clinically Vulnerable Population

The COVID-19 pandemic was recognised to be a public health crisis and impacted millions of people worldwide. This was an extremely challenging time for everyone; many people struggled to stay or feel safe and experienced severe impacts on their well-being. However, the pandemic was particularly challenging for people who were affected by a rare condition[9]. Many rare disease patients were at an increased risk from COVID-19 and as such were shielding long after the end of COVID-19 legal restrictions.

[4]Data source: https://www.worldometers.info/coronavirus/

[8]Long-term cases include an estimated number of cases with long-covid and number of covid-related deaths

[9]https://covid-19.geneticalliance.org.uk/wp-content/uploads/2020/07/Covid-19-Rare-Reality.pdf

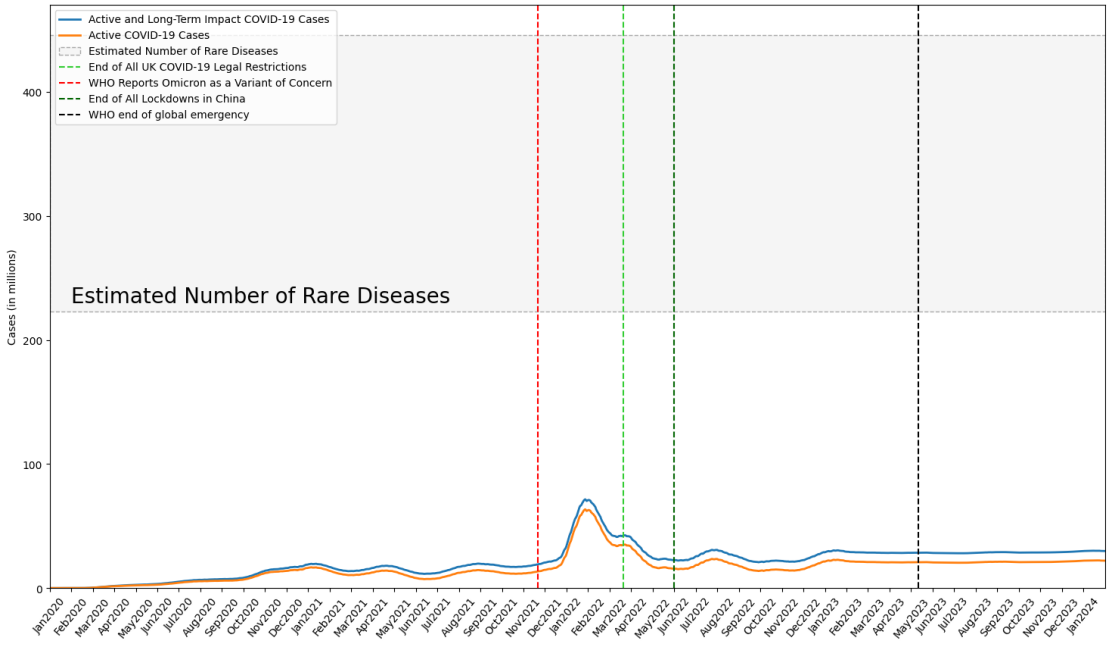


Figure 1.1: COVID 19 active cases compared with rare diseases [7,8]

Several individuals with rare conditions voiced concerns about the dangers of COVID-19, even after they were advised to stop shielding, as shown by Queen's University Belfast's webpage of quotes[10], where one person felt *"We are still very vulnerable and I will continue to stay at home as much as possible"*, whilst another stated *"I am not comfortable sending my very vulnerable child to school in the current circumstances"*. As a result of this, it was not deemed ethical to conduct in-person studies for the majority of this PhD research. However, the widespread use of video conferencing technologies as a result of the pandemic facilitated a wider geographical recruitment area, which otherwise would probably not have been feasible.

1.3 Research Questions

This thesis investigates potential avenues to support rare disease patients and facilitate future research in this area. As such, the five separate research questions we explore in this thesis all relate to the following overarching research question *"How can we*

[10]https://www.qub.ac.uk/sites/RareDisease/COVID-19/Shielding/

support and facilitate further research into supporting rare disease patients on their journeys towards diagnosis?".

Firstly, **RQ1** and **RQ2** examine the context of a rare disease diagnosis. Considering both the clinical experiences and the support needs of individuals with rare conditions, we identify opportunities for improvement. This provides the basis to explore the design space for pre-diagnostic technology.

Following our exploration of **RQ1** and **RQ2**, the importance of the remaining research questions became apparent. In particular, the lack of suitable methods of evaluating pre-diagnostic prototypes for rare disease patients presented a significant hurdle. As such, we explore potential evaluation approaches for pre-diagnostic prototypes for rare disease patients in **RQ3**. In this exploration, we propose a new methodological approach which facilitates the discrimination of prototypes based on the timeliness of patient knowledge discovery.

In addition, the lack of patient-perspective data for rare disease patients and the need for technology to meet the wider needs of rare disease patients, catering for empathetic support and information-seeking needs. Hence in **RQ4**, we explore the data requirements for patient-facing technology and create an open-source dataset to facilitate future development of patient-facing pre-diagnostic prototypes. In addition, we curate a recommendation system to provide empathetic support pairings between rare disease patients and to facilitate knowledge discovery. To evaluate this recommendation system, we utilise our methodological approach identified in our exploration of **RQ3** to assess whether and to what extent our recommendation system meets the needs of rare disease patients in comparison with two other patient-facing pre-diagnostic prototypes **RQ5**. In addition, this enables us to evaluate the suitability of our methodological approach.

RQ1. What are the typical experiences of people with rare diseases during their journey towards diagnosis?

a) What are the typical clinical interactions that people with rare diseases face when seeking a diagnosis?
b) How are people with rare diseases affected by typical clinical interactions when seeking a diagnosis?
c) Where are there opportunities for improvement within the rare disease diagnostic odyssey?

RQ2. What are the implications of these experiences on the design of pre-diagnostic technology for rare diseases?

a) What are the implications of clinical interactions on the design of technology to support diagnostic decisions?
b) To what extent is the design space of pre-diagnostic technology for rare diseases being met?
c) What are the implications of the opportunities for improvement identified on the design of technology?

RQ3. How might we evaluate prototypes to account for the lengthy rare diagnosis?

a) How can we evaluate whether prototypes promote faster knowledge discovery for rare diseases?
b) How can we evaluate whether technology facilitates better accommodations for the support needs of rare disease patients?

RQ4. How can technology facilitate better accommodation for the support needs of rare disease patients?

a) What are the data requirements for patient-facing technology for people with rare diseases?
b) How can we generate data which is representative of terminology and information discovery from the patient's perspective?
c) How can machine learning interventions support the information-seeking needs of people with rare diseases?
d) How can machine learning interventions identify pairings between people with rare diseases which offer empathetic support?

RQ5. How can a symptom-based patient matching approach facilitate better accommodations for the support needs of rare disease patients?

a) How can Peer Matching facilitate greater support for the information-seeking needs of people with rare diseases?
b) How can Peer Matching facilitate empathetic support through pairings between people with rare diseases?

Figure 1.2: Research Questions

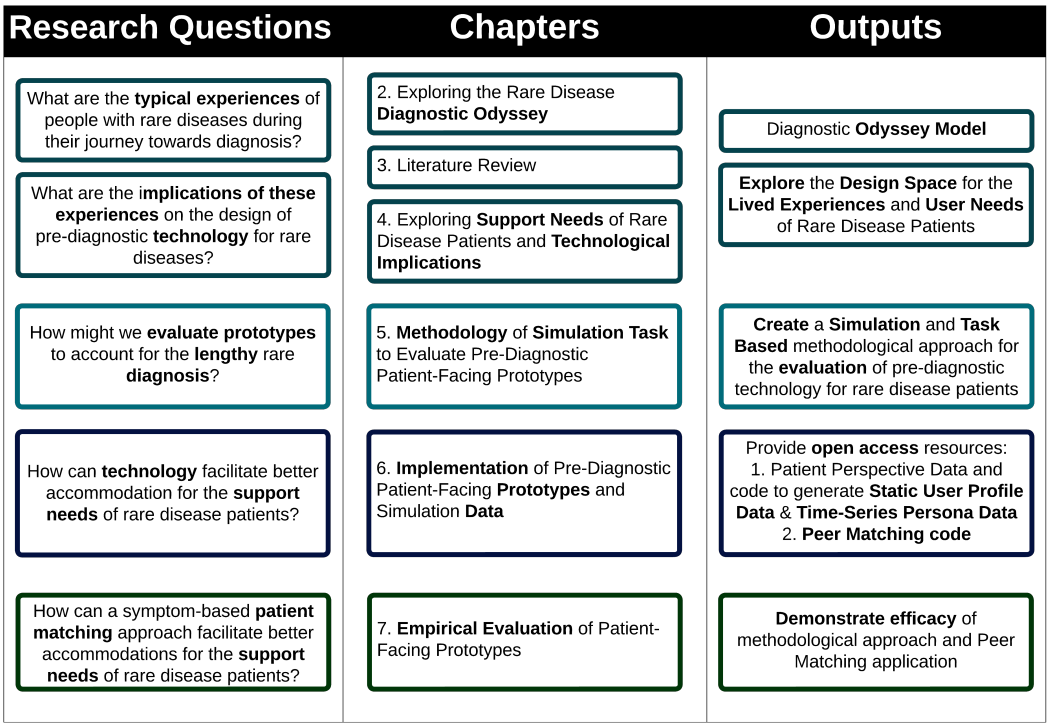


Figure 1.3: Diagram linking research questions to each chapter and the contributions

1.4 Research Contributions

Figure 1.3 shows the specific outputs and contributions for each of the research questions which we explore in this thesis as well as the corresponding chapters. This shows that Chapter 2, Chapter 3 and Chapter 4 each contribute to answer **RQ1** and **RQ2**. **RQ1** explores the challenges faced by undiagnosed rare disease patients to identify opportunities for improvement. The exploration of **RQ1** does not limit this exploration to technological implications to prevent limitations of scope due to perceptions of the role of technology. **RQ2** explores the implications for technology by building upon the identified challenges and opportunities for improvement through a human-centred design approach and comparing to existing approaches within the literature. By separating these research aims, we identify the need for greater empathetic support, a previously unexplored area of pre-diagnostic technology for rare disease patients. Thus, our approach establishes the lived experiences of rare disease patients and forms a comprehensive exploration of the design space for pre-diagnostic technology for rare disease patients. In addition, we demonstrate in Chapter 3 that there is a lack of data

and suitable evaluation methods for patient-facing pre-diagnostic technology for rare disease patients and argue that this may be why there is little focus on this area within the research community. This leads us to our next three research questions.

More specifically, as a result of our exploration of the above research questions, we present four core contributions relating to the research space of designing and evaluating pre-diagnostic technologies for rare disease patients. In particular, we aim to pave the way for future research by:

- Exploring the design space for both patient-facing and clinically-based pre-diagnostic technologies for rare disease patients
- Establishing the lived experiences of rare disease patients
- Providing a methodological approach which facilitates the evaluation of patient-facing pre-diagnostic technologies
- Curating open-source materials for patient-facing pre-diagnostic technologies

All of these contributions facilitate future work within this area by: providing direction for future research by exploring the design space; facilitating meaningful evaluations in a practical time frame; and providing required resources for prototype curation.

1.4.1 Exploring the design space

Our first research contribution lies in the exploration of the design space for both patient-facing and clinically-based pre-diagnostic technologies for rare disease patients. By taking a patient-focused approach and critically reflecting on the unique attributes of a rare diagnosis, we identify key implications in technology development, not only for patient-facing technology but also for interventions within the clinical space.

In particular, we first present a comprehensive and rigorous overview of the clinical interactions of rare disease patients on their journeys towards diagnosis. This provides crucial contextual information in which to interpret the contribution that technology can have on the diagnostic odyssey of a rare disease patient.

Following this, we then perform the first systematic approach to critically review the literature around the typical clinical interactions of rare disease patients. This identifies key gaps within the literature to provide greater support along the rare disease diagnostic

odyssey, including the need for human-centred design approaches for pre-diagnostic technology to support rare disease patients with their unmet support needs; data which represents the patient's perspective; and for evaluation approaches which assess pre-diagnostic patient-facing technologies for their information seeking capabilities (i.e., the relevancy and timeliness of information support) as well as the supportive capabilities (i.e., the provision of the unmet support needs of rare disease patients).

1.4.2 Establishing the lived experiences of rare disease patients

Building upon our exploration of the design space, we identify a previously unexplored area of design in pre-diagnostic technology for rare disease patients by exploring the lived experiences of rare disease patients. That is, we adopt a human-centred approach to understand the unmet support needs of rare disease patients. In particular, in addition to the information-seeking support typical of pre-diagnostic technology, we highlight the need for support with information management, clinician communication and empathetic support to alleviate feelings of isolation for rare disease patients who are *"all alone in the world"*[11] and *"wish ... [they] could find someone that could share in what ... [they] go through"*[11].

Therefore, by establishing the typical clinical interactions and unmet support needs of a diagnostic odyssey of a rare disease patient, we present a comprehensive exploration of the design space for both patient-facing and clinically-based pre-diagnostic technology to accommodate the clinical and explore the lived experiences of rare disease patients, beyond their clinical experiences, to establish the human side of the diagnostic odyssey for rare disease patients.

1.4.3 Develop a new methodological design

The third core research contribution of this thesis is the development of the first simulation-based laboratory study to facilitate the empirical evaluation of pre-diagnostic technology for rare disease patients in a shorter time frame. We argue that pre-diagnostic technology can have greater outcomes when relevant information is provided as early on in the diagnostic odyssey, since this may facilitate greater conversations with healthcare professionals early on in a patient's diagnostic odyssey. However, it is not feasible to conduct real-world deployments over the diagnosis periods of rare disease patients as

[11] Participant quotes during an interview (Chapter 7)

this could last 30 years and requires the identification of rare disease patients before clinical experts are able to identify them. Therefore, we use a simulation-based design to recreate the informational journey of a rare disease patient. This was done through patient personas, consisting of phenotypes (i.e., symptoms and clinical findings) in the order in which a patient would discover them. This facilitated empirical evaluations of pre-diagnostic technology which discriminated between design prototypes based on their performance for information-seeking tasks as well as their supportive capabilities. In particular, to evaluate performance on information-seeking tasks, it assesses, how quickly pre-diagnostic technology can consistently identify relevant information for a rare disease patient persona. To evaluate supportive capabilities, it qualitatively assesses how many support needs are met and the degree to which these support needs are fulfilled.

1.4.4 Generation and Publication of Materials

The final core research contribution of this thesis lies in the generation of prototypes and materials to facilitate ongoing rare disease research in human-centred digital interventions.

Firstly, we published open-source software code for Peer Matching[12], a recommendation system to pair rare disease patients in order to provide empathetic support. We demonstrate that Peer Matching performs better than Google and a rare disease specific search engine for information-seeking tasks. In particular, it facilitates the consistent identification by non-experts of relevant information earlier on in the information discovery journey of a rare disease patient. In addition, we demonstrate that Peer Matching has the most promise in providing empathetic support, promoting self-advocacy, and empowering patients to have a stronger voice during consultations.

Secondly, to address the lack of data representing the patient's perspective (as highlighted in Chapter 3), we create our Patient Perspective Dataset based on a clinical knowledge base, which represents information discovery from a rare disease patient's perspective. We publish an open-source dataset on GitHub [13]. This dataset provides the base data from which we generate the Time-Series Persona Dataset which facilitates our method of evaluation of pre-diagnostic technologies (personas that simulate the informational journey of a rare disease patient) and the Static User Profile Dataset (the

[12] github.com/902549/peer-matching
[13] github.com/902549/patient_perspective_data

user base for Peer Matching). The Patient Perspective Dataset along with the code to generate the Time-Series Persona Dataset and Static User Profile Dataset are published as open-source resources on GitHub.

By creating publicly available materials for pre-diagnostic technology for rare disease patients, we facilitate future research in this area. In particular, the lack of data may present a significant hurdle for the curation of technology in this area.

Contributing Publications

The following publications describe the work covered in this thesis.

[P1] Emily Esther Nielsen, Tom Owen, Matthew Roach, and Alan Dix. 2023. A Patient Centred Approach to Rare Disease Technology. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23).* https://doi.org/10.1145/3544549.3585826

[P2] Emily Esther Nielsen, Tom Owen, Matthew Roach, and Alan Dix. 2023. Simulating the Rare Disease Diagnostic Journey. In *Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '23 Companion).* https://doi.org/10.1145/3596454.3597188

[P3] Emily Esther Nielsen. 2023. Motivations of Technology Use in Undiagnosed Rare Disease Patients. In *CHI'23 workshop on Intelligent Data-Driven Health Interfaces (IDDHI '23).* https://iddhi2022.create.aau.dk/wp-content/uploads/2023/03/iddhi22_paper_4677.pdf

All of the work in these papers was conducted by Emily Nielsen, except for editing and guidance provided by supervisors: Tom Owen, Matthew Roach, and Alan Dix. Publication [P1] relates to the research presented in Chapter 4; [P2] relates to the research presented in Chapter 5 and Chapter 6; and [P3] related to the research presented in Chapter 2-4.

Chapter 2

Exploring the Rare Disease Diagnostic Odyssey

The context of a rare disease patient experience and the challenges associated with diagnosis were a key driver for this research. Therefore, throughout this project, we take a patient-centred and context-focused approach to explore pre-diagnostic technology which support rare disease patients before they receive a diagnosis. Hence, we need a thorough understanding of the context of the journey that a rare disease patient experiences when seeking a diagnosis before exploring the literature more broadly. This enables us to frame existing approaches in this space around the context of rare diagnoses, and as such, we explore how approaches alleviate the challenges of a rare diagnosis as well as the gaps, and specific challenges when using technology to aid rare diagnosis. Therefore, in this chapter, we examine the context of rare disease patients by examining the surrounding literature.

In particular, we explore and illustrate the common shared obstacles faced by rare disease patients during their diagnostic journey [6]. These obstacles contribute to the lengthy and arduous process of obtaining a diagnosis. For rare disease patients, this is often the most challenging barrier [13] as well as the first significant hurdle they face and a vital step in accessing suitable care [6].

To increase awareness and understanding of the challenging experiences that are typical of a rare diagnosis, some organisations have conducted quantitative studies of typical diagnostic experiences [6, 7, 12] and have developed diagrams (Fig. 2.1) which illustrate the diagnostic odyssey. These diagrams were created by two patient

organisations: Eurordis, an organisation for all people living with rare diseases in Europe that aims to improve their quality of life; and the National Organisation for Rare Disorders (NORD), based in the United States. These diagrams were used to illustrate the long odyssey to attain diagnosis, emphasising the difficulty and frequent stages along the way. The key similarities between the diagrams include a high number of consultations with specialists and repeated genetic testing.

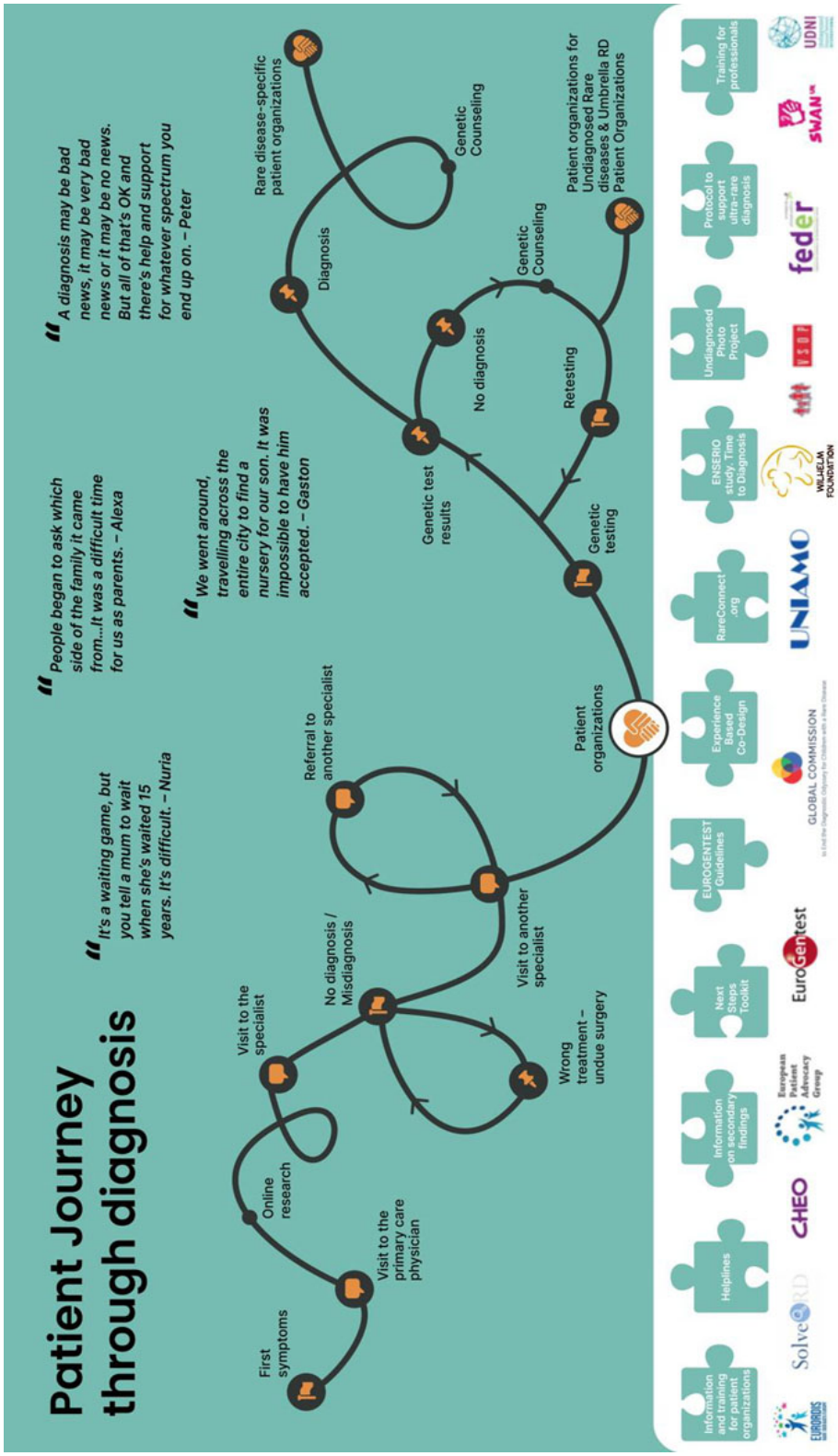
However, these diagrams were not consistent in a number of aspects. Firstly, Eurordis included misdiagnosis and mistreatments in their diagram (Fig. 2.1a), which were not shown in the NORD diagram (Fig. 2.1b). Secondly, the time that certain stages occur varied across the two diagrams both for genetic counselling and support from the rare disease community or patient organisations. These stages also occur multiple times on the Eurordis diagram, but only once in the NORD diagram. Finally, the NORD diagram included stages for the treatment of symptoms and online research of diagnosis, which were not shown in the other diagram.

The differences in the diagrams make it difficult to draw clear conclusions. Furthermore, the rigour of these diagrams may be called into question since there is no description of the methodology behind these diagrams, aside from the statement that the Eurordis diagram was created by a community of patients, scientists and clinicians. Therefore, to establish a robust contextual foundation for the research in this PhD, we compared and evaluated the stages of diagnosis shown in the diagrams to existing literature. Using these findings, we then create our own model to illustrate the experiences typically faced in a diagnostic odyssey.

2.1 Methodology

In this section, we outline the method we used to create our model of the diagnostic odyssey. First, we needed to collate the diagrams into a single list of stages. Then we compared and evaluated the stages of diagnosis shown in the diagrams to existing literature. Following this, we decided which stages should be included in our model, based on their evidence in the literature and relevance to our purpose of understanding more about the clinical experiences of rare disease patients on their journey towards diagnosis.

To collate the stages from both diagrams, we initially created two linear lists to represent the stages in each diagram. When there were multiple paths in either one of the



(a) Diagnostic odyssey diagram by Eurordis [37]

(b) Diagnostic odyssey diagram by NORD [38]

Figure 2.1: Diagnostic odyssey diagrams created by patient organisations

diagrams, the stages in the alternative path were added to the list in order and marked as *or stage* (e.g. or diagnosis), and looped stages were written *repeated stage* (e.g. repeated referrals). We then combined these lists ensuring that, where possible, the stages present in both diagrams were listed at the same point in the combined list. When this was not possible, stages were listed twice to show the different points they occur in each of the diagrams. Stages shown in only one diagram were listed as per that diagram. In cases where one of the diagrams had more detail on a particular stage, a sub-stage was added.

We then identified quantitative studies or reviews which take an overview of the rare disease diagnostic odyssey. We used these reports to assess evidence for the occurrence of each of the stages from the diagrams, and the point in time (relative to the diagram) in which they occur if the diagrams were inconsistent in this fact. We then defined inclusion criteria to assess whether these stages should be included in our model. In particular, stages of diagnosis were included in our model if they were supported by sufficient evidence from the literature and excluded for any of the following reasons: the stage occurs after diagnosis; it does not require an explicit stage as it is a component of or is otherwise represented in other parts of the model; it does not relate to diagnosis (i.e., it is a part of treatment or disease management). Table 2.1 shows the collated stages from each of the diagrams (Fig. 2.1a & Fig. 2.1b), evidence in the literature of these stages, as well as reasons for and decision of inclusion or exclusion from our model.

Table 2.1: Evidence and inclusion decision for the stages of diagnosis depicted in the diagrams

Stages of Diagnosis	Evidence in Surrounding Literature	Inclusion Decision	Reason for Decision
1. Start of symptoms [37,38]	74% of the 66 case studies surveyed in [39] start at symptom onset.	Accept	In both diagrams & evidenced in literature.
2. First primary care physician visit [37,38]	Primary care physicians are the gatekeepers to the rest of the healthcare system [40].	Accept	In both diagrams & evidenced in literature.
2i. Initial testing [37]	Blood test in patient story [7].	Reject	This is a component of the primary care physician visit.

Continuation of Table 2.1

Stages of Diagnosis	Evidence in Surrounding Literature	Inclusion Decision	Reason for Decision
3. Online research [37,38]	Research mentioned in patient story after referral to specialist [7]. Research mentioned in a quote from a parent which does not specify timing [41].	Reject	Not a clinical interaction. Some support from literature, but time of starting online research varies.
4. First visit to specialist [37,38]	Referral in two patient stories [7], referral to many doctors [6,7,13], referral to many specialists [40,41].	Accept	In both diagrams & evidenced in literature.
5. No diagnosis [37,38]	Evidenced by all following stages (if diagnosis was reached these stages wouldn't occur).	Reject	Represented by the continuation of the diagnostic journey (or lack of diagnosis) in our model.
5i. Or suspected diagnosis [37]		Reject	Insufficient evidence in literature.
5ii. Or repeated misdiagnosis [38]	Stated in overview of patient journeys [6,7,13,41,42] in patient story [7]. In 52% of cases (an average of three times) [7].	Accept	In both diagrams & evidenced in literature.
6. Repeated treatment of symptoms [37]	Symptoms treated when no disease specific treatment exist [13].	Reject	Insufficient evidence in literature.
7. Wrong treatment [38]	Wrong treatment [7,13,41,42].	Accept	Evidenced in literature.
7i. undue surgery [38]	Wrong treatment, including undue surgery [13].	Reject	Insufficient evidence in literature.

Continuation of Table 2.1

Stages of Diagnosis	Evidence in Surrounding Literature	Inclusion Decision	Reason for Decision
8. Repeated specialist referrals [37,38]	Stated in overview of patient journeys: many doctors [6,7,13] and many specialists [40,41]. Referral in two patient stories [7].	Accept	In both diagrams & evidenced in literature.
9. Patient organisations [38]		Reject	Not a diagnostic interaction.
10. Repeated genetic testing [37,38]	Stated in overview [43], 2 patient stories [7], 80% of rare diseases are genetic [13].	Accept	In both diagrams & evidenced in literature.
10i. Or correct diagnosis [37]	20% of rare diseases are not genetic [13].	Accept	Trivial, genetic testing is unlikely for non-hereditary conditions.
11. Repeated genetic test results [37,38]	In two patient stories [7].	Reject	This is a component of genetic testing.
12. Repeated no diagnosis [37,38]	In two patient stories [7].	Reject	Represented by continuation of diagnostic journey (or lack of diagnosis) in our model.
13. Repeated genetic counselling [38]		Reject	Not a diagnostic interaction.

Continuation of Table 2.1

Stages of Diagnosis	Evidence in Surrounding Literature	Inclusion Decision	Reason for Decision
14. Patient organisations (undiagnosed/ umbrella groups) [38]		Reject	Not a diagnostic interaction.
15. Repeated retesting [37,38]	Repeated genetic tests in both of the patient stories which included genetic testing [7].	Accept	In both diagrams & evidenced in literature.
16. Correct Diagnosis [37,38]	Diagnostic odyssey ended at definitive diagnosis for 94% of papers surveyed in [39].	Accept	In both diagrams & evidenced in literature.
17. Online research [37]	Research mentioned in patient story after referral to specialist [7], One example [41].	Reject	Post-diagnosis.
18. Genetic counselling [37,38]		Reject	Post-diagnosis.
19. Disease-specific organisations and support [37,38]	Patient organisations and Facebook groups in multiple patient stories [7].	Reject	Post-diagnosis.

End of Table 2.1

2.2 Diagnostic Odyssey

From our analysis above (described in Section 2.1 and presented in Table 2.1), we created a representational artefact for a rare disease diagnostic odyssey. We present this artefact

as a diagram illustrating a typical diagnostic odyssey of a rare disease patient, which can be seen in Figure 2.2. As in the previous diagrams, our representation shows frequent visits to different specialists and repeated genetic testing. In contrast to the winding nature of the diagrams, which may have been used to emphasise the chaos of the diagnostic experience, we represent the stages in a linear format and show repeated stages with looped arrows. We did this to simplify and improve the interpretability of the diagram to facilitate a greater understanding of the diagnostic journey.

Our diagram also presents only the core clinical interactions as a foundation to understand the context of the rare disease diagnostic odyssey and does not include patient support seeking outside of the healthcare system, such as online research and interactions with the rare disease community, which we build upon in future chapters. This was because an overview of clinical interactions is sufficient to characterise this experience from the patient's perspective, whereas a deeper view was needed and these activities should be explored in more detail to understand not only when and whether patients research their health, but also their motivations, time spent, and processes of research. Moreover, this allows our diagram to act as a boundary object by presenting a multi-stakeholder perspective of the diagnostic journey for rare disease patients. Cooney et al. [44] argue that the creation of visual artefacts as boundary objects facilitate effective sense-making in social problem solving. Therefore, this artefact provides a means to aid the problem solving involved in this challenging area. As such, this diagram acts as a launch pad from which opportunities for improving the rare diagnostic odyssey may be identified, thus this acts as a strong basis from which we critically examine the literature.

In contrast, this artefact presents a reductionist view of the diagnostic odyssey, and does not represent the motivations, behaviours and wider experiences of the stakeholders involved. Therefore, in addition to the contextualisation that this diagram provides, it is important to obtain a fuller picture of the stakeholders. As our work takes a patient-focused approach, we must identify the motivations, behaviours and wider experiences of patients. Hence, in Chapter 4 we characterise patient support seeking outside of the healthcare system through a user study to gain more insight into these stages and the support that patients need.

To explore our diagram of the diagnostic odyssey and understand the context in more depth, let us compare it to patient stories of diagnosis. These stories not only provide examples to strengthen our understanding of the context, but they also give insight into

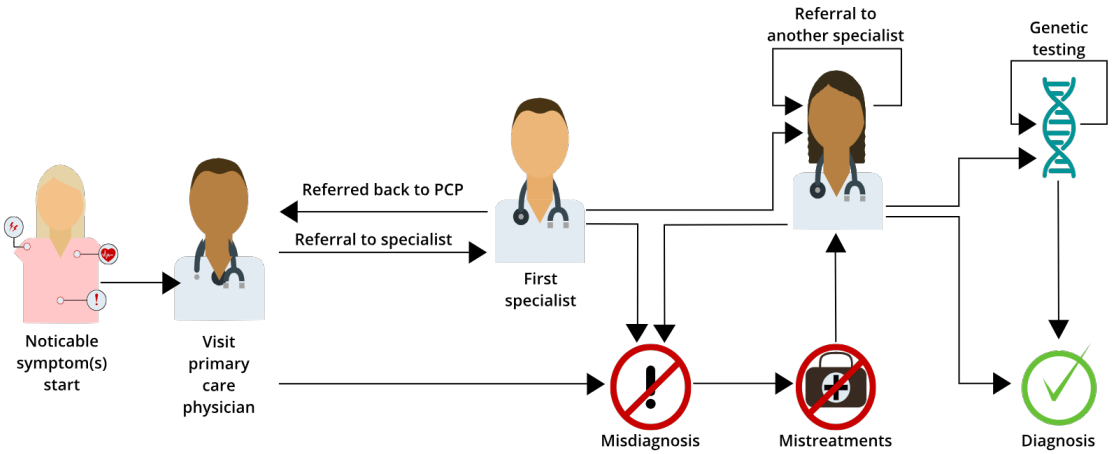


Figure 2.2: The diagnostic odyssey of a typical rare disease patient

the perspectives and impacts that patients have when experiencing a diagnostic odyssey, thus highlighting and extent of the difficulty of obtaining a diagnosis. Using this, we then explore the key challenges within the rare disease diagnostic odyssey and consider the implications that these challenges have on technology.

NORD's website lists many patient stories about their experiences of living with a rare disease. These stories cover a wide range of areas, including testimonies of people refusing to let their condition prevent them from fully embracing life, people overcoming discrimination challenges with their condition, and individuals' accounts of their fight to obtain a diagnosis. Let us explore the two most recent patient stories about diagnosis and compare the experiences described in these stories with the diagram we created.

It was important to assess how representative these stories were of a rare disease patient's diagnostic odyssey. To do this, we identified the last 15 patient stories on NORD that describe the clinical stages of an individual's diagnosis or the diagnosis of a family member. Then for each story, we identified which stages of diagnosis from our model were mentioned. We then calculated the percentage of patient stories which describe each of the stages of diagnosis. This showed that each of the stages was described in the vast majority of patient stories, with the exception of genetic testing which occurred in about one-quarter of the stories. Therefore, we can consider the following two stories to be representative of typical rare disease patient stories.

First, let us consider the story of 10-year-old Sophie, as told by her mother, Jessica. Sophie's story is based in the United States and is about the process and challenges faced

in obtaining her diagnosis of limb girdle muscular dystrophy type 2A/R1, a genetic condition typically characterised by progressive muscular weakness in the legs, upper arms and shoulders, but with various different presentations.[1]

> "*Sophie's medical odyssey began at age 5 after getting very ill* [symptoms start] *with what we were told was pneumonia.* [misdiagnosis 1] {extra explanatory text removed} *When Sophie's pediatrician finally agreed to lab work, after telling me to "relax, it's just a virus," everyone was shocked* [visit primary care physician]. "

As in our diagram, Sophie's story begins with her symptoms starting and a visit to her primary care physician, a pediatrician. This is when Sophie is first misdiagnosed, and her pediatrician trivialises her condition as *just a* virus. Sophie's mother, Jessica, clearly felt that her daughter's condition was more severe than the pediatrician expressed and may have felt unheard or that her concerns were not being valued. The active involvement of patients in their health forms a crucial component of diagnosis when it comes to rare diseases [6,15]. Hence, it is vital that patients and their family members feel that they can contribute to discussions with healthcare providers. This may be prevented if they do not feel listened to, and therefore, patients and family members must be able to advocate for their health.

> "*Her ALT and AST liver enzymes were off the charts, over 300. I was sent to infectious disease* [specialist 1], *then gastroenterology* [specialist 2] *who ordered a liver biopsy after 3 more weeks of high levels found in multiple rounds of bloodwork. At the preop visit, a doctor ran her CK (muscle enzymes) which were extremely elevated – over 10,000 – which can falsely indicate a liver problem, and canceled the surgery. We were sent to cardiology* [specialist 3] *and told she may have cystic fibrosis, so to see a pulmonologist also* [specialist 4]. *She had neither* [misdiagnosis 2&3].
>
> *"We were then told to try a rheumatologist* [specialist 5]. *We saw a wonderful, caring doctor who became my ally in our quest for an answer as Sophie's health continued to decline. An MRI revealed what was thought to be dermatomyositis* [misdiagnosis 4], *later changed to necrotizing myositis* [misdiagnosis 5]. "

[1] https://rarediseases.org/jessica-undiagnosed/

Here, Sophie's story continues to follow Figure 2.2, with five specialist consultations involving several tests. Jessica described one of these specialists as a caring doctor, and as such, considered them an ally in their search for Sophie's diagnosis. The recognition of this specialist as an ally shows the importance of access to actively engaged healthcare professionals, despite the fact that they specialised in rheumatology (and as such were not able to make a diagnosis of this genetic condition), they were still able to offer support which was not found in other consultations. This may be because rheumatologists treat patients with movement issues (a symptom of muscular dystrophy), and as such, they may have had greater sympathy and appreciation for the impact of these issues on Sophie. This shows the impact that a relevant referral can have (even though genetics may have been a more relevant referral) since the support offered by this clinician and their understanding of Sophie's symptoms stood out to Jessica. On the other hand, the gap in support from other clinicians may be due to insufficient time for the 'caring' approach that patients and their families need. This could be improved if the correct referral is made sooner, thus reducing demands on clinicians' time. Therefore, this shows the importance of and challenges regarding the identification of a relevant specialist.

During the specialist consultations, four additional misdiagnoses were made. These diagnostic errors may be because specialists will not typically know of rare conditions outside of their area of expertise. When the underlying condition falls outside of their specialty, clinicians will primarily consider conditions within their expertise or common conditions outside of their expertise. Thus, rare conditions that lie outside of their expertise are unlikely to be taken into consideration until the correct referral is made, resulting in misdiagnoses and frequent specialist referrals. These misdiagnoses can further diminish Jessica's faith in the healthcare system, and inevitably stop the search for alternative conditions and thus delay the identification of the correct diagnosis.

> "*She was immediately admitted to New York Presbyterian Hospital in NYC, where she received high-dose corticosteroids and IVIG treatment every 26 days, inpatient* [mistreatments]. *The treatments took a toll on her. Sophie's whole personality changed. She developed Cushing syndrome and had bouts of very low blood pressure requiring emergency intervention and pauses to treatment. Her numbers would improve, then get worse. I had a feeling we were dealing with something else. I listened to my gut and pushed for answers.* "

Here we see frequent erroneous treatments requiring hospital visits, resulting from Sophie's misdiagnosis of Necrotizing Myositis. This had significant negative impacts beyond the debilitating effects on her health. The treatments were extremely distressing for Sophie, to the point that her personality changed, and she developed a stress-induced condition (Cushing syndrome). Noticing that these treatments did not help Sophie, Jessica questioned the diagnosis. This was a key moment in Sophie's journey towards diagnosis since Jessica started directing and fighting for investigations from this point.

> "*Muscle biopsies were inconclusive, neurologists all said it wasn't of neurological origin. We went to CHOP, Cornell, NYU, Hospital for Special Surgery, Hackensack Children's Hospital, and Morristown Children's Hospital* [specialists 6-12 (in different medical facilities)] *but no one could tell us what exactly it was. I finally got the doctors to agree to genetic testing* [genetic testing]. *Results took months, and in the meantime were spinal taps, CT scans, MRIs, X-rays, endoscopy, colonoscopy. . . they poked and prodded my baby and kept upping her meds.*
>
> "*I found Wisconsin Children's Hospital's Undiagnosed and Rare Disease Program (a NORD Center of Excellence) and emailed Sophie's case* [specialist 13]. *After five months, we were invited out for a week of new tests and another muscle biopsy as well. Within that time, I got a call from genetics in New York to come in and bring someone for support. I was handed a diagnosis of limb girdle muscular dystrophy, a progressive muscle wasting disease, and told she also had myositis.* [misdiagnosis 5 (again)] *In Wisconsin, it was determined she never had myositis, it was only the LGMD 2A, and to get her off all medication as this is incurable. It took 1.5 years to wean Sophie off the steroids.* "

Jessica's advocacy led to numerous specialist consultations and visits to different medical facilities. This advocacy and push for genetic testing led to the identification of the correct diagnosis. However, the geneticist reaffirmed the misdiagnosis of myositis. By emailing Sophie's case to a medical facility in Wisconsin with a rare disease program, this misdiagnosis was ruled out and enabled them to start reducing the erroneous treatments. While Sophie's diagnosis was still challenging and long, it was Jessica's advocacy which led to the correct diagnosis, since she had to push for the tests which identified the genetic cause, and she was the one who reached out to the rare disease facility which ruled out myositis and allowed Sophie to come off the treatments.

> "{extra explanatory text removed} *If we teach medical staff to think outside the box and not stay on a course of their specialty, no matter how fitting they feel a diagnosis is, we can help prevent misdiagnosis and drawn-out diagnosis. Don't let kids get passed around and become a file on a desk.* "

The end of Sophie's story demonstrates the significant impact that this diagnostic journey had on Jessica's view of the healthcare system. She felt like her daughter was put into a box and that she needed to teach healthcare professionals to think differently. This may be a reflection of Jessica's loss of faith or hope in receiving adequate support from the healthcare system. The description of patients as *a file on a desk* suggests that consultations were impersonal and that clinicians were perceived to not care. This shows the significant impact that a diagnostic odyssey has on patients and their family members. Therefore, it is highly important to identify methods to support clinicians, patients and their family members with the challenges faced during diagnosis.

Sophie's story spans the entire diagnostic odyssey diagram, with thirteen specialist visits, six misdiagnoses and years of mistreatments. The identification of a relevant specialist was a significant issue in this story, as evidenced by the fact that each of the referrals were to completely different specialties. This is likely due to Sophie's widespread symptoms affecting multiple organ systems, which is often the case for rare diseases [45]. These suspected or erroneous diagnoses led to years of extremely distressing erroneous treatments. At the end of the story, we the importance of Jessica's advocacy for her daughter. Sophie had a diagnosis and was undergoing treatments, which was later revealed to be in error. While this would usually stop investigations, Jessica felt this diagnosis was incorrect and sought a genetic test which identified the correct diagnosis and reached out to medical facilities for rare diseases, which ruled out the misdiagnosis and started weaning Sophie off of mistreatments.

Now, let us look at Kylee's story of seeking a diagnosis of dopa-responsive dystonia with parkinsonism.[2]

> "*This story is about me and my long journey to a diagnosis. Although I might have had mild symptoms as a child, it was nothing noticeable. My symptoms began in my early 20s with some uncontrolled movement that affected my legs and arms. I*

[2]https://rarediseases.org/kylee-undiagnosed/ (capitalisation and disease names as per patient story). Location could not be included to add context to Kylee's story since the website does not specify location

> *began seeking treatment with my local ER, general practitioner, and neurologists. This diagnostic journey is the worst part of the whole experience and is the reason I am sharing my story.*
>
> *"My diagnosis took about 21 years.*
>
> *"My first onset of symptoms* [symptoms start] *lasted over seven months and got progressively worse as time passed. I went to a neurologist* [specialist 1] *who concluded my increased trouble with walking was due to depression* [misdiagnosis 1]. *He determined that I was depressed because I didn't have a husband, boyfriend, children, or pet. I'd recently moved to the community and was in my early 20s so, to me, that seemed normal. I wasn't distraught about my lack of those relationships. "*

Before Kylee tells her story of diagnosis, she first states the impact it had on her, describing it as the worst part of her experience. Again, this further motivates research into this diagnostic journey to provide better support with the challenges associated with diagnosing rare conditions. Her story starts with the onset of noticeable symptoms, and upon seeking a neurologist, she was misdiagnosed with a mental health condition. Mental health related misdiagnoses can have significantly larger delays in diagnosis than physical health [6]. Therefore, this misdiagnosis would have a significant impact on her journey towards diagnosis. Furthermore, people with rare diseases can feel that their symptoms are trivialised for being *all in their head* when they are misdiagnosed with mental health conditions [6]. This paired with the questioning of her lifestyle, would likely have a significant impact on her view of the healthcare system and willingness to contribute to consultations. Since patient involvement can significantly aid rare disease diagnosis [6, 15], it is highly important to encourage patients to continue to actively engage in consultations when facing adversity.

> *"I next made an appointment with a doctor in a neuroscience office* [specialist 2]*, but I didn't know really who I needed to see. He felt that I had dystonia but wasn't the right type of doctor to diagnose it and encouraged me to see a different neurologist.*
>
> *Next, I went to a neurologist in a bigger city* [specialist 3]. *He had a pretty good reputation, so I was excited to get his input. He told me there was nothing wrong with me and suggested that, evidently, I "needed a lot of attention." At this point I needed a cane to walk. "*

One of the underlying conditions, a form of dystonia, was recognised by a neurologist, but felt they did not have sufficient expertise to diagnose it. This is a significant moment since a specialist with experience in movement disorders would typically be able to diagnose or rule out whether it was dystonia. This again suggests that clinicians lack the resources to make effective referral decisions when it comes to rare diseases.

Instead, a referral is made to a neurologist in a bigger city and with a good reputation. As such, Kylee was excited about this consultation, but instead of getting the answers she was hoping for, the neurologist trivialised her condition and blamed her for the symptoms that she was experiencing and claimed that she was attention-seeking. This shows the fluctuating emotions between hopefulness and disappointment, and we see Kylee's perception of healthcare professionals and the healthcare system deteriorate further. The disregard for a patient's symptoms may also lead to them feeling unheard and thus prevent them from contributing further to discussions with healthcare providers. This again suggests the importance of methods to encourage patient engagement in their journey towards diagnosis.

> "*My General Practitioner* [primary care physician] *asked if I would be willing to go to the Mayo Clinic* [specialist 4], *which I was. I made arrangements for two weeks of testing. I specifically asked about checking into dystonia. At the end of the testing, they told me there was nothing wrong with me in neurology and invited me back for psychiatric testing.*
>
> "*I went home and made appointments with alternative medicine doctors* [specialist 5 (outside of the formal medical system)]. *I found a chiropractor who diagnosed me with a calcium deficiency* [misdiagnosis 2 (outside of the formal medical system)]. *and sold me expensive vitamins* [mistreatment (outside of the formal medical system)]. *I thought they helped, and he must be right because I was feeling better a little at a time.* "

Kylee's journey before this point was met with several negative consultations with healthcare professionals; not only was she blamed for her condition, but it was also trivialised and misdiagnosed. Here she suggests they test for the correct underlying condition. Since this did not result in its identification, she may have felt unable to direct or have a say in the search for a diagnosis. This may have been a key factor in motivating Kylee to seek alternative medicine, which led to misdiagnosis and expensive

mistreatments. This suggests that there was a gap in the support that Kylee could access from the formal healthcare system.

> "*For the next 20 years, I was almost afraid to go to the doctor unless I was desperate. I would have small episodes of not being able to walk well after being active. I would take some calcium and rest and it would go away. I probably should have been seeking out help from other specialists, but I didn't need anyone else to tell me I needed psychiatric help.*
>
> "*In April 2022, I began having symptoms again, but this time they were worse and would not relent. I tried calcium supplements, but they didn't help. I went to an endocrinologist* [specialist 5] *because I had been diagnosed with thyroid issues within those 20 years, so we decided to start there. The endocrinologist spent five minutes with me, ordered a few blood tests, and assured me that whatever was happening couldn't be explained with endocrinology. "*

The impact of Kylee's feelings of disenfranchisement of the healthcare system is shown clearly here. In particular, it resulted in her avoidance of seeking medical attention within the formal healthcare system for twenty years. While she recognised specialist consultations had the potential to help, her experiences of dismissal were too significant for this; she had lost hope for positive consultations and felt she would be misdiagnosed again. It wasn't until she experienced severe and unrelenting symptoms that she started seeking medical attention again. Even then, she sought an endocrinologist, rather than a neurologist, because there was already an established relationship from a different condition. This shows the importance of her trust in healthcare professionals since she spoke to a specialist that she was comfortable with, rather than one that was more relevant to her symptoms. If she had felt heard and hadn't lost hope earlier on in her journey to diagnosis, she may have still had sufficient trust in healthcare professionals to keep pursuing a diagnosis, preventing a twenty-year gap in seeking a diagnosis from the formal healthcare system. Therefore, it is important to explore if and how patients can feel heard and maintain hope during their search for a diagnosis.

> "*Against my better judgment, I considered seeing a neurologist again* [specialist 6]. *I did research this time* [online research (not in Fig. 2.2)] *and learned about the existence of movement disorder specialists. I made it my mission to get an*

> *appointment with one of those doctors as soon as possible. I had to get a referral from neurology, so I found a hospital in another state that had neurologists and movement disorder specialists* [specialist 7].
>
> *"In March of 2023, I had to endure MRI's, EMGs, and test-out medications but finally, I had a diagnosis: dopa-responsive dystonia with parkinsonism. "*

Here, Kylee self-advocates further by researching and identifying suitable specialists herself. This highlights the impact that advocacy can have in obtaining a diagnosis.

With the exception of genetic testing, Kylee's story spans the diagnostic odyssey diagram, but rather than starting by visiting a primary care physician, her story starts with a visit to a specialist. Kylee's odyssey included seven specialist consultations; and two misdiagnoses, one of which resulted in mistreatments. Unlike Sophie's story, Kylee's symptoms clearly related to a given specialty (neurology), but she needed specific expertise within neurology to obtain a diagnosis. Clinicians did not successfully identify facilities and specialists with the required expertise to diagnose Kylee's conditions; instead, she had to research to identify these herself. This not only highlights the importance of self-advocacy and patient involvement, but it also suggests that clinicians may not have the information or support they need to make effective referrals for specific rare conditions. A loss of faith and general disenfranchisement of the healthcare system was evident throughout Kylee's story, likely due to her experience of several negative interactions with clinicians. This may have motivated Kylee to seek alternative medicine, which led to a misdiagnosis and expensive mistreatments. This suggests that there was a gap in the support that Kylee could access from the formal healthcare system.

2.3 Discussion on the Implications on Technology

When contrasted to a diagnosis of a common condition, we can see that the diagnostic odyssey is a far different experience for rare disease patients, not just for their physical health, but also for their well-being and faith in the healthcare system. The stories of Sophie and Kylee show the deterioration of their health from incorrect or delayed treatments; significant impacts on their emotional state in their loss of hope as well as diminished trust and willingness to seek medical attention, having been referred to many specialists without receiving the answers they want and need. In recognising these challenges and identifying the opportunities for improvement, we establish key

opportunities within the research space to support the diagnosis of rare disease patients. Let us now explore the implications that these experiences have on technology.

Sophie's story demonstrated the importance of genetic testing, and while several repeated tests were not required for her, it is often a necessary step in other cases. Therefore, methods to identify and narrow down potential genetic causes may reduce these challenges. In particular, screening and sequencing technologies can identify likely genetic causes far more quickly which may facilitate more efficient testing strategies [46].

The frequent misdiagnoses within the diagnostic odyssey, which was particularly evident in Sophie's story, is a significant component of the rare disease diagnostic odyssey. These misdiagnoses not only resulted in significant emotional and physical harm from mistreatments but also stopped the search for more answers. Even for common conditions, it is important to avoid misdiagnoses, but due to the increased number of misdiagnoses for rare conditions, it is even more important for technology to be robust against misdiagnosis. In particular, rare disease technology must reduce the number of misdiagnoses and in the event that a misdiagnosis is made, recognise them sooner. The importance of this is recognised in Miyachi et al.'s [47] paper which examines the risk of misdiagnosis for diagnostic support system by examining a case with diagnostic errors due to confirmation bias. This poses a difficult challenge for technology which typically considers diagnosis as the *end point*. Instead, a more dynamic and flexible system is needed. In addition, clinicians would need to continue to use technology after a diagnosis is made to identify misdiagnoses, or it would have to continue working in the background. For example, one approach would be to alert clinicians of *red flags* or cases where an alternative (potentially rare) diagnosis is likely.

Specialists may need support with diagnostic decisions for rare conditions within their area of expertise since it is unlikely that they will have prior experience with a given rare disease. This may have significantly reduced the time to diagnosis in Kylee's case, where the correct specialty was identified immediately, but specific expertise within that specialty was needed to obtain a diagnosis. In this case, specialist-specific technology may promote faster diagnosis once patients are referred to a suitable specialist [48,49]. However, since the identification of a relevant specialist is often a significant hurdle to obtaining a rare diagnosis, specialist-specific technologies should be used in addition to technology which follows patients throughout their odyssey, in order to obtain several opportunities to aid diagnosis.

The high number of specialist referrals was a key element of the rare diagnostic odyssey. Both patient stories showed several irrelevant referrals made by clinicians, and Kylee and Jessica had to advocate and actively engage in consultations in order to obtain a suitable specialist referral. This suggests that there is a need for greater support to make more effective referrals, not only for clinicians (particularly primary care physicians who often act as gatekeepers to secondary care) but also to support patients like Kylee who have to research and make referral decisions themselves. Therefore, there is a need for technological interventions which provide support with referral decisions, not just diagnostic decisions, as supported by the surrounding literature [40, 50, 51].

However, several challenges arise when using clinician-facing technology throughout the rare diagnostic odyssey. Firstly, the design must accommodate the varying needs of different clinicians if it is to be used throughout the diagnostic journey. There are three key motives that clinicians may have, the most obvious is to identify the correct diagnosis, but clinicians may also want to make more effective referrals or rule out potential conditions as efficiently as possible (e.g. tests which eliminate a number of likely conditions or the most likely explanations). Moreover, it is unlikely that technology will be adopted across multiple different medical facilities and specialties. Even in the event of the universal adoption of a specific decision support system, clinicians cannot easily share information, even when they have consent from patients [12]. This can result in information loss between consultations thus preventing clinicians from effectively building upon previous consultations. However, patient-facing technology will naturally follow patients throughout their diagnosis and are intimately aware of their symptoms and experiences. Therefore, patient-facing technology lends itself to facilitating support throughout the diagnostic odyssey in ways that the current healthcare system does not allow for clinician-facing technology.

Moreover, patient-facing technology can provide an avenue for people with rare diseases to go if they lose faith in the healthcare system. The use of technology to research health information can prompt people to seek medical attention [52–55] and can provide support when patients feel that they have nowhere else to go [56]. However, while clinician-facing technology in this area has a clear aim, to aid diagnosis, the aim of technology for rare disease patients is far more ambiguous. In particular, as we explored in the patient stories, the experiences of rare disease patients are not limited to those within the healthcare system, and as such the support for patients' needs and desires

can fall beyond the direct implications of diagnosis. Therefore, to create more effective patient-facing technology to support people on their diagnostic journey for a rare disease, we need to ascertain the unmet support needs of rare disease patients. This has been recognised as an important aspect of HCI in health – people who are chronically ill are people first, and patients second, so technology must account for this [16], however, work into the needs of people with undiagnosed rare conditions still remains to be explored.

2.4 Chapter Summary

In this chapter, we identify a comprehensive overview of the clinical interactions that rare disease patients typically experience on their journey towards diagnosis. By understanding this context more clearly, we identify the key factors which delay the diagnosis of rare diseases, which include: frequent referrals to specialists; frequent misdiagnoses; lack of expertise on specific rare diseases within specialties and repeated genetic testing. We then use this overview of the clinical context of a rare diagnosis to explore the design space for interventions to support the diagnosis of rare diseases.

In our exploration of the design space, we argue that clinician-facing technology to support rare disease diagnosis should offer support with clinical decisions which can lead to a correct diagnosis, which is not limited to direct diagnostic decisions. Firstly, in the patient stories described above, we see that clinicians need more support with referral decisions which can be especially challenging for rare diseases which often affect multiple organ systems or require clinicians with very specific expertise, not just clinicians within a given specialty. Furthermore, due to the high misdiagnoses of rare diseases, the diagnosis cannot be the end-point of the support system, and other possible explanations should still be considered beyond this point. Finally, there should be specialist support technologies for once the correct referral is made to support diagnostic decisions that require specific expertise, as well as prioritise testing, particularly within genetics where sequencing or screening technologies can be utilised to identify likely genetic causes.

However, we also identify limitations of clinician-facing technology. In particular, the patient is the only constant throughout their journey towards diagnosis and clinicians cannot easily share information, even with patient consent [12]. Hence, important information may be lost, thus preventing clinicians from fully building upon information from previous consultations. In contrast, patient-facing technology remains with the

patient, and therefore, naturally follows along every stage of the diagnosis. Therefore, we must explore the design space for patient-facing technology further. To fully characterise this space, we must identify the unmet support needs of rare disease patients to ascertain opportunities for intervention, which we explore in Chapter 4.

The exploration of the design space for clinically-based diagnostic technologies for rare disease patients enables us to critically examine approaches to aid diagnosis within the literature. Thus, in the next chapter, we use this context as a lens to identify, in relation to the diagnostic odyssey, how technology contributes to a rare diagnosis. We use this to identify gaps and as a call to action for research within this space to adopt a more context-focused approach.

Chapter 3

Literature Review

Machine learning approaches can utilise data from numerous years and locations. In contrast, the lack of prevalence of rare diseases make it unlikely that a clinician will have experience of a given rare condition before diagnosing a patient. Therefore, motivated by the promise of machine learning within health [28, 29], let us explore the potential of machine learning based pre-diagnostic technology in this area.

Therefore, in order to explore existing approaches and critically examine gaps in the research space, let us systematically review machine learning based pre-diagnostic technology to support the diagnosis of rare diseases. Then, examine where and how a patient's diagnosis is aided across the diagnostic odyssey.

3.1 Systematic Review Approach

3.1.1 Search and Screening Process

The search and screening process shown in Figure 3.1 is as follows. PubMed and the ACM Digital Library were used to identify 1,056 articles from a search string consisting of the concatenation of Machine Learning Terms (Machine Learning, ML, Artificial Intelligence, AI, Algorithm, Clinical Decision Support System) AND Rare Disease Terms (Rare Diseases, Rare Disease, Rare Disorder, Rare Defect, Orphan Disease) AND Diagnosis Terms (Diagnosis, Diagnosing, Diagnostics, Diagnostic, Diagnoses). Each of the individual terms within a category was concatenated with OR. These search strings were constructed to identify papers relating to all three of the following

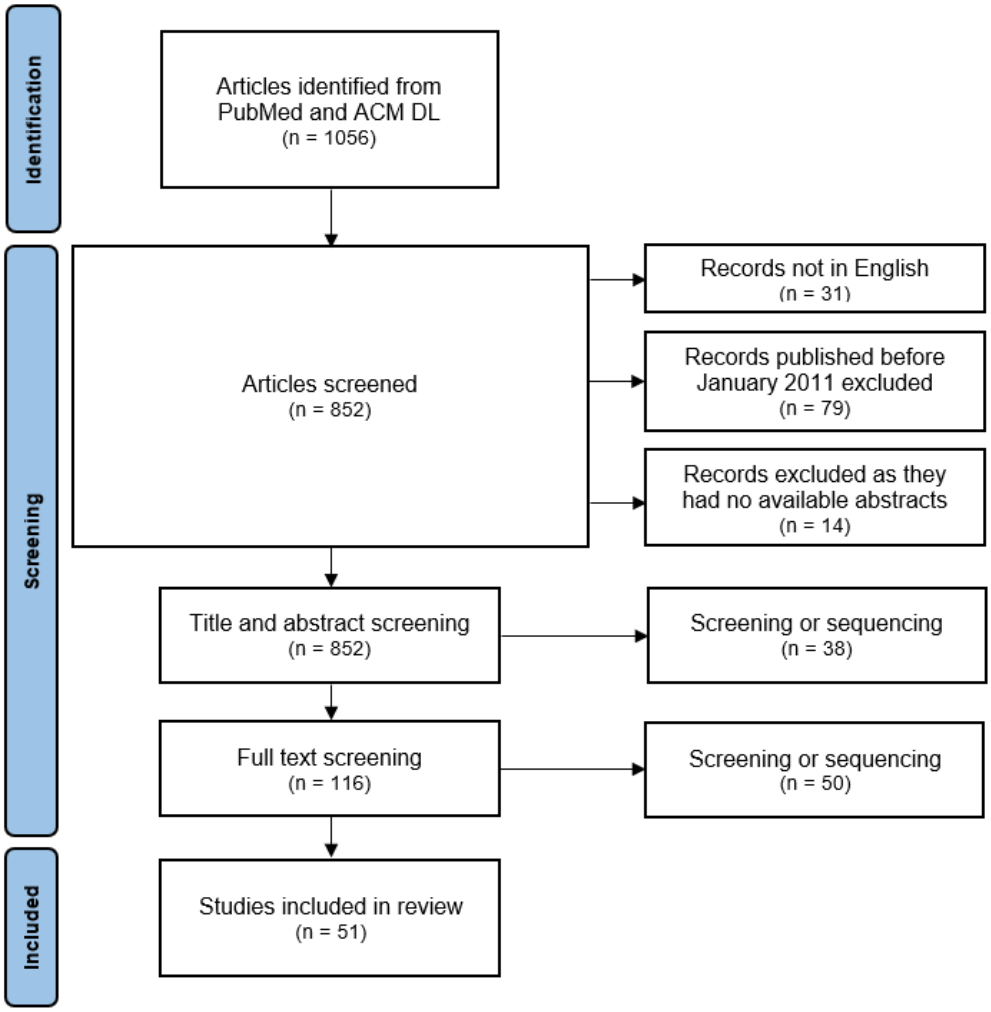


Figure 3.1: PRISMA flow chart of search and screening process based on [57]

concepts: the use of computational methods; the application to rare diseases; and the intention to aid diagnosis.

The first screening process consisted of three stages: firstly, to remove articles that did not have any versions in English; secondly, to remove articles that were published before 1st January 2011[1]; and thirdly, to remove articles with no available abstract. This resulted in 852 remaining articles. Following this procedure, the articles were further screened by assessing the relevance to this study (i.e., whether they met the additional features of

[1]The initial search was performed January 2021, and so captured articles from 1st January 2011 until 1st January 2021 to capture articles from a 10 year period. This was later expanded to include the most recent publications, with the last search performed on the 29th November 2023.

Table 3.1: Eligibility Criteria

Eligibility Criteria
Original peer-reviewed publications in English
Publications released after 1st January 2011
The publication describes a computational-based approach to assist diagnosis
The approach described is directly focused on diagnosing human beings with rare diseases
The publication does not describe testing or screening methods (including single disease diagnosis)

the eligibility criteria outlined in Table 3.1) of the titles and abstracts, resulting in 116 remaining articles. Following this, the full text was screened for relevance, resulting in a total of 51 articles to be included in the review.

3.1.2 Analysis of Papers

Upon the inclusion of each article, basic information was collected including the title, authors, citation, name of publisher (journal/conference name), date of publication and the DOI. The articles were then read to identify the data sets used, algorithms used, targeted end-user, presence of a user interface and any key takeaways from the papers. Following this, collective themes were identified by iteratively reviewing small samples of the literature for noteworthy commonalities. This led to the collection of evaluation approaches and success metrics.

The targeted end-user was collected if they were explicitly named. From this the end-user was categorised into user types, namely: primary healthcare professionals; specialist healthcare professionals and non-experts (people without any medical qualifications). These user types were established by first separating healthcare professionals from non-expert users since this is a significant distinction in the context and experience of users.

Following this, the types of healthcare professionals were considered. Many of the terms used were general and non-specific terms that are used synonymously when describing healthcare professionals (e.g., clinician, physician, doctor). Given that these terms are non-specific, it is likely that they are referring to primary healthcare professionals. The remaining healthcare professionals were categorised as specialist

healthcare professionals as the remaining terms referred to experts within several given specialties. The final category was comprised of patients as well as people without a diagnosis, so we defined this category as non-expert users.

Once we had categorised the specified end-users, the papers which did not explicitly name an end-user (8 papers) were also categorised based on the contextual information of the paper. In particular, we identified specialist healthcare professionals as the intended end-user if the paper aimed to support diagnosis for a specific disease type or required specialist expertise; whereas the intended end-user was deemed to be non-specific healthcare professionals if the paper aimed to support clinicians with a range of diagnoses.

Based on the identification of end-users, we frame the papers' interventions around the context of a rare disease diagnosis which we explored and characterised in Chapter 2. In particular, by comparing the diagnostic odyssey of a rare disease patient to the approaches used, we categorised the papers by the stages in which their approach would intervene in a patient's odyssey. From this, we identified a gap within the provision of continuous support throughout the diagnostic odyssey as well as existing methods of evaluation. Hence, the papers were read again to identify the types of success metrics used, values of these metrics and whether these metrics were continuous or measured over time. Metrics given for a benchmark or feature selection were excluded since these values do not reflect the final model. In cases where multiple metrics were given, we identified a single metric which was representative of the model's performance: the metrics for the best-performing algorithm were taken if multiple algorithms were evaluated; the mean of the metrics was calculated if multiple datasets or parts of a dataset were evaluated, and no overall performance was given. For models which output a rank list, rather than a single disease, accuracy was taken as the presence of the correct disease in the top k rank. Metrics that were only used in one paper were not included in the results.

3.2 Technical Choices

To get an overview of the research, let us now consider the different aspects of the systems described in the papers. There were a range of different objectives in the papers surveyed, including developing a new system/algorithm, applying new methods, improving

Table 3.2: Types of Algorithms Used

Type of Algorithm Used	Frequency	Reference
Neural Networks	19	[48,49,58–74]
Support Vector Machine	9	[66,69,70,75–80]
Bayesian and Statistical Models	9	[62,65,67,69,70,75,77,81,82]
Random Forest and Decision Trees	9	[67,69,70,80,83–85]
Logistic Regression	7	[60,67,69,70,77,80]
Nearest Neighbours	7	[67,69,70,73,75,86,87]
Linear Discriminant Analysis	5	[69,70,80,88]
Natural Language Models	5	[66,74,89–91]
Collaborative Filtering/Similarity Analysis	4	[92–95]
Ranking/Rule Based Algorithms	4	[47,84,96,97]
Other	10	[73,75,98–105]

performance, implementing prototypes and testing, and evaluating systems. We can break the approaches of these systems down into the technical choices, targeted end-user and contextual approach as detailed below. Firstly, let us look at the technical choices, namely, the data and algorithms used in these papers.

3.2.1 Algorithms

The algorithms used were categorised into model types, starting with the most commonly used. Algorithms that were used by only one paper were categorised as *other*. There were also various different algorithms used, as shown in Table 3.2. Neural networks (NN) were the most common type of algorithm used, with 19 papers surveyed in this study using this algorithm. Most of these (9) described an implementation of a new convolutional neural network, others either built upon existing architectures (4), used

Table 3.3: Types of Data Used

Type of Data Used	Frequency Used	Reference
Electronic Health Records	36	[48,49,58–60,62,63,65,66,68,71–75, 77–79,81–85,88–91,94–97,99,101–104]
Online Databases	14	[61,66,67,74,82,86,89,90,92,93,95, 98,99,103]
Ontological or Terminology Data	9	[66,67,78,85,89,91,93–95]
Primary Data	9	[47,64,69,70,76,80,87,92,100]

general adversarial networks (2), recurrent neural networks (1), or did not specify a particular architecture (3). A large number of papers used support vector machine; statistical and Bayesian models (e.g., Naive Bayes, Bayesian Networks); or some form of decision tree (predominantly consisting of an advanced tree algorithm such as random forest or gradient boosting decision trees).

3.2.2 Data Sets

Once the data sets used in the included papers were collected, they were abstracted into categories. This was done by comparing trends within the data sets to establish different types of data used, namely: Electronic Health Records (EHRs); online databases; primary data; and ontological data. Table 3.3 shows the abstracted datasets. There were not many datasets in common across the different papers, with over 70 different data sets used. The few datasets that were used in multiple papers included MIMIC-III, Pubmed, Orphanet, Online Mendelian in Man, Human Phenotype Ontology (HPO) and the Unified Medical Language System (UMLS). Both HPO and UMLS are designed to ensure that the data conforms to a standard vocabulary for disease names, variants and clinical features. These tools were used when multiple data sets were utilised because a standardised vocabulary allows different data sets to be integrated more easily. The wide range of datasets used highlights the several different approaches to aid the diagnosis of rare diseases, likely due to the broad area of conditions, across different specialties, which rare diseases encompass.

Many of the EHRs used in these papers required clinical data from a participating hospital. Access to these datasets requires lengthy ethics processes to ensure the confidentiality and privacy of patients' data. In the papers where EHRs were not utilised, a few different approaches were used to access data: online datasets were sourced; data was self-collected by the researchers; or data was mined from different sources on the internet. Many of the online datasets used consisted of specialist-specific data, and as such, would not suffice for a general diagnostic support system. The only paper which described a patient-facing intervention used primary data, which suggests that there may be a lack of external datasets which represent the patient's perspective.

Therefore, many of the existing datasets are not suitable for pre-diagnostic technology for rare disease patients. Firstly, many datasets are not publicly available, and as such, accessibility is a key issue. Secondly, many datasets do not contain individual patient cases, or only present cases within a given specialty. Thirdly, only one dataset identified in this review contained sufficient temporal information to facilitate a meaningful evaluation – to assess the timeliness of diagnosis. And finally, there are no datasets in this review which represent the patient's perspective in terms of information discovery and language. In addition, datasets based on EHRs contain sensitive information, and as such, it would be unethical to make these datasets freely available.

As a potential method of addressing the lack of accessible data, approaches such as synthetic data have been explored. Synthetic data has for example been used to alleviate issues such as privacy concerns of an individual's health data within healthcare research projects [106]. Other works have generated new data by utilising techniques such as: General Adversarial Networks (GANs) [107–111]; Neural Networks [112–114]; and statistical models [115–117]. As synthetic data, does not provide direct access to real patient data, this can help alleviate the ethical challenges with releasing this data. Ethical challenges relating to re-identification are particularly important for rare diseases where patients are far more likely to be identifiable due to the lack of prevalence of their condition [118].

Four approaches within our review utilised synthetic data, three of which supplemented existing EHRs with synthetic data, with one paper using synthetic data alone. Approaches for synthetic data generation show significant promise in areas where the data is limited, which will inevitably be the case for conditions which are rare. However, synthetic data generation usually requires existing data to build upon, so

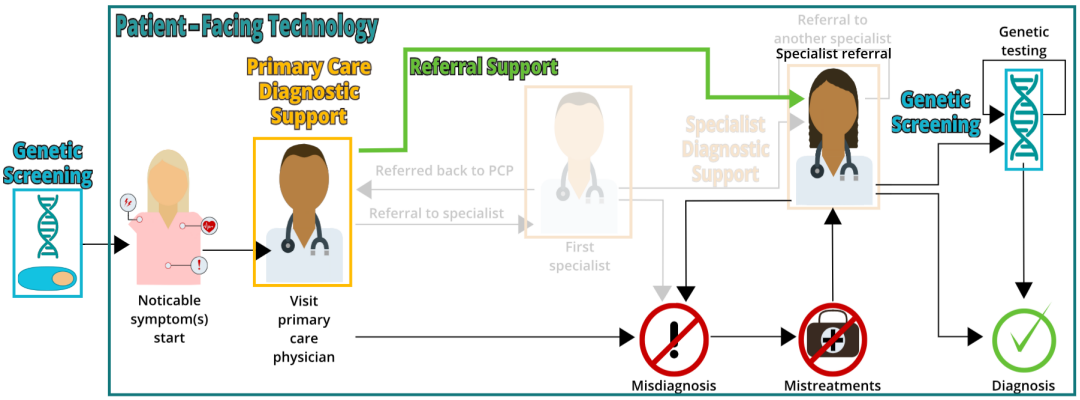


Figure 3.2: Points of intervention within a diagnostic odyssey

standard approaches for generation are not feasible if there is no suitable data to build upon. Moreover, these approaches to generate synthetic data have been constructed as a means to develop clinical decision-making support for healthcare providers and do not represent the patient's input or perspective. As such, these datasets consist of technical, clinical terminology, so they do not represent the perspectives of non-expert individuals.

Therefore, we identify the lack of patient-perspective data as a key gap within this research space. As such, we aim to explore the role that synthetic data can play in the curation of pre-diagnostic technology for rare disease patients where data will naturally be harder to come by. Methods to achieve this have been explored through the use of quantitative persona creation [106] to generate typical examples of individuals representing a population. Hence, a potential approach to bridge this gap may be through the generation of quantitative personas which represent rare disease patients.

3.3 Comparing Design Approaches to the Diagnostic Odyssey

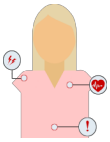
To compare the approaches within this review with the context of a rare disease diagnosis, we first consider at what point each of the papers is aiming to aid the rare disease diagnostic odyssey. This was primarily done by identifying the end-user and context of use (e.g., specialist-specific, primary care, genetic testing). In addition, we noted that sequencing and screening technologies, whilst not included in our review, may promote prenatal diagnosis as well as aid genetic testing at the end of the diagnostic odyssey, so this was also included in our diagram. We then augmented the diagnostic odyssey diagram

from the previous chapter with this context of use, as seen in Figure 3.2. By following along the diagnostic odyssey journey, we can see several opportunities for intervention.

3.3.1 Genetic Screening

Firstly, screening or sequencing technologies can be used to aid genetic testing, not only by screening in pregnancy and increasing the number of pre-natal diagnoses but also by reducing the cycle of genetic testing in the prioritisation of tests which are more likely. Since genetic testing occurs at the start and end of the diagnostic odyssey and does not have many relations to disease classification or referral support, we did not include it in our systematic review. However, through our literature screening, we identified 88 articles describing screening or sequencing technologies. This demonstrates the high attention in this area, likely due to the significant strides made through large-scale research projects, such as the 100,000 Genomes Project[2] or The Human Genome Project[3]. However, not all conditions can be screened during pregnancy, so individuals who do not get diagnosed before birth may still experience a diagnostic odyssey, and whilst these technologies may remove or reduce the cyclic nature of genetic testing, it may take a long time for individuals to reach this stage of their diagnostic odyssey, so let us now consider interventions that may aid in the earlier stages of diagnosis.

3.3.2 Patient-Facing Technology



We can consider the post-natal diagnostic odyssey to begin when an individual first experiences, or first notices, symptoms of their rare condition. At this point, the patient may start interacting with patient-facing technologies. Only one paper by Kühnle et al. [87] described a patient-facing approach. Their work proposed a patient-facing peer-matching website based on the experiences faced by people with rare diseases (collected through questions on the website) to offer patient-facing diagnostic support. As this approach is patient-facing, it can facilitate support throughout an individual's diagnosis. Shen et al. [95] also utilised a patient matching approach, but from a clinical perspective and matched patients based only on their symptoms, rather than other experiences. Two papers by Grigull et al. also involved patients through questionnaires to inform clinical consultations; the first paper [69] aims to identify

[2] www.genomicsengland.co.uk/initiatives/100000-genomes-project
[3] www.genome.gov/human-genome-project

specific rare neuromuscular conditions as a proof of concept, whilst the second paper [80] aimed to identify and highlight the possibility of a rare disease based on behavioural information. As these three clinically-based tools involve patients, they could potentially span a patient's diagnostic journey, but it is not clear if or how they intended different clinicians to use and monitor these.

However, given that only Kühnle et al. utilised a patient-facing approach, it suggests that there may be a lack of focus on pre-diagnostic technology for rare disease patients. In Chapter 2, we discuss the high value of patient-facing technology, since the patient is the only constant throughout their journey towards diagnosis and clinicians cannot easily share information, even with patient consent [12]. Hence, important information may be lost between referrals, thus preventing clinicians from fully building upon information from previous consultations. In contrast, patient-facing technology remains with the patient, and therefore, naturally follows along every stage of the diagnosis. Therefore, patient-facing pre-diagnostic technology to support individuals with rare conditions is a highly promising and under-explored research area.

3.3.3 Primary Care Diagnostic Support

When a patient's symptoms prompt them to consult their primary care physician, clinically-based interventions may start to offer support. 18 of the 51 papers which were included in our review aimed to support primary healthcare professionals with diagnosis decisions [47,60,63,66,72–75,78,82,85,89,90,93,94,103–105]. These works provided novel diagnostic solutions but did not provide support for referral decisions. Primary care physicians are naturally far more familiar with common conditions, so many conditions may be diagnosed without the need for referrals. In contrast, rare conditions are unlikely to be correctly diagnosed at the initial consultation. As such, a significant role of the primary care physician is to make a suitable referral.

However, this presents a significant challenge in a rare disease diagnosis, where due to the genetic nature of 80% of rare conditions [13], symptom manifestations can span multiple organ systems, and unless various symptoms arise and are presented during the initial consultation, the genetic origin of the condition may not be apparent. The challenge of referrals was demonstrated by the patient stories in Chapter 2, and in the cyclic nature of specialist referrals in our diagram. In contrast, if a primary care physician can make more effective referral decisions, then the number of consultations would be

significantly reduced, thus reducing the time spent waiting for referral appointments. In addition, it would reduce unnecessary costs from additional specialist consultations.

3.3.4 Specialist Diagnostic Support

After visiting their primary care physician, patients may consult with one or many different specialists, with the average patient consulting with 5 clinicians [7]. Even within a given specialty, rare conditions may require more specific expertise to diagnose, as was the case in Kylee's story (Chapter 2). The majority of included papers (27 papers [48,49,58,59,61,62,64,65,67,68,70,71,76,79,81,83,84,86,88,91,92,96,98–102]) aimed to support specialists with their diagnostic decisions. This may be particularly helpful for challenging diagnoses within a specialist's expertise, and as we explored in Kylee's story (Chapter 2), this could have significantly positive impacts on rare and hard-to-diagnose conditions.

In addition, two of the papers above [67, 101] aim to utilise clinical data more effectively to facilitate better knowledge sharing across referrals or different patient cases. As such, these may facilitate more productive consultations due to improved knowledge sharing which is otherwise not easily done [12], thus this can also have significant impacts on diagnosis.

However, these approaches do not account for the case where multiple specialist referrals are needed before the correct specialty is identified. If a patient is referred to a clinician whose expertise differs from the specialty in which the patient's diagnosis lies, additional referrals may be necessary. As such, specialist-specific diagnostic support is not sufficient in of itself as there must be an awareness of possible causes outside of the given specialty, and additional support with referral decisions may also be necessary. Without considerations of causes outside of their specific expertise, conditions are likely to be missed, or a patient may be misdiagnosed.

3.3.5 Misdiagnosis

Moreover, rare disease patients are far more likely to be misdiagnosed, with the average patient experiencing three misdiagnoses during their diagnostic odyssey [7]. As such, pre-diagnostic technology to aid diagnosis decisions must be designed to reduce this risk. The risk of misdiagnosis may be evaluated, in part, by the performance of the model, however, when evaluating on specialist data, patients outside

of the specialty may not be included in the data even if they may still be referred to a clinician within this specialty. The evaluations on one paper [47] had a particular focus on misdiagnosis prevention by examining their system on a case where diagnostic errors were made due to previous history. Their system still suggested the misdiagnosis, which was deemed to be due to the confirmation bias, which was reported in the case report, but the correct condition was ranked third. As such, the authors deemed their decision support system would prevent diagnostic errors by physicians. Clearly, more awareness is needed of the implications of biases like these being baked into the algorithm. Thus, diagnostic support intervention for rare conditions requires scrutiny in the evaluation of easily misdiagnosed cases with biases, such as this. Reducing misdiagnoses may play a significant part in reducing the time to reach a diagnosis since misdiagnoses will naturally stop the search for alternate causes.

3.3.6 Cross-Specialty Support

Two remaining papers aimed to support clinical decisions throughout the diagnostic odyssey, with Ronicke et al. [97] aiming to support clinicians with diagnostic decisions through several consultations. Rider et al. [77] aimed to support both referral decisions of primary healthcare professionals and diagnostic decisions of specialists within a given group of rare conditions. Not only does this work recognise referral decisions as a necessary component of diagnostic support for rare conditions, but it also aims to support the diagnostic decisions made by specialists. These approaches aim to span the diagnostic odyssey, and as such offer greater opportunities to reduce the length of the diagnostic odyssey.

3.4 Evaluation Approach

All of these approaches may play a significant part in the diagnosis of rare diseases. This is demonstrated by the high-performance metrics of the algorithms in many of the papers, as shown in Table 3.4. Some papers (33%) evaluated how likely a model is to correctly identify positives by measuring the sensitivity (also known as recall and true positive rate). Sensitivity was often compared with how accurate the predicted positives are (i.e., fewer false positives, which gives a higher true negative) by measuring either

Table 3.4: Success Metrics Used in the Surveyed Papers

	ACC	F1	AUC	Recall	Precision	Specificity
90-100%	11	5	10	6	6	2
80-90%	9	2	3	4	3	3
50-80%	6	3	4	2	4	0
<50%	5	1	0	0	2	0
None	20	40	34	39	35	46

the specificity or precision. This is a trade-off which is commonly considered in the papers as it is not just shown in the aforementioned measures, but it is also represented in the F1-score and the Area Under the (AUC) Receiver Operator Curve. These other measures may account for why only 5 papers measured the specificity of the model.

However, this does not give very much insight into the utility of the diagnostic support system in aiding the diagnosis of a rare disease patient. In particular, the use of these performance metrics to analyse performance on a distinct point is more relevant for diagnosing common conditions which are often diagnosed on the first consultation. In contrast, a positive outcome for patients with rare conditions is the identification of a diagnosis as early in their journey as possible. So, the stage in a patient's journey to diagnosis is key to interpreting what success metrics mean.

The 45 papers (88%) consisting of only primary or only specialist diagnostic support describe one-shot approaches for diagnosis. As such, these metrics are used to evaluate at a single point. In addition, the patient-facing approaches naturally span the diagnostic odyssey, but they are not evaluated over time, and again are treated as one-shot approaches, rather than continuous opportunities for diagnosis. As these approaches do not assess whether a correct diagnosis would be made *earlier*, the potential for each of these approaches to support the diagnostic odyssey of rare disease patients is difficult to interpret.

For example, if a model is evaluated at the point when patients would have obtained their clinical diagnosis, then a high-performing model would not aid their diagnosis, however, if a model performs well at an earlier point than a clinical diagnosis would have been made, then it would aid the diagnostic odyssey. From this example, we can see that a model may appear to be effective at first instance, but it may not suggest the correct disease at an earlier point than other models. Therefore, evaluations for diagnostic

support for rare diseases would be more relevant and indicative if they evaluate whether it reduces the time to diagnosis, and by how much.

Only one paper surveyed in this review evaluated their algorithm for how much it would reduce the time taken to reach a correct diagnosis. Ronicke et al. [97] evaluated whether their system, Ada DX, would suggest the correct disease before the time of diagnosis by evaluating their system for each consultation that a patient has. This was done by firstly extracting the dates of each visit and all pieces of clinical evidence and secondly removing any evidence which would not be available at the date of the relative visit. Then the performance was evaluated for all visits following the first documented visit regarding the corresponding condition until clinical diagnosis is attained. This identifies the first point at which Ada DX would suggest the correct disease, which can then be compared to the time of diagnosis. Ronicke et al. found that Ada DX top 5 fit disease list suggested the correct disease at the first documented visit for only 33.3% of cases. Therefore, if Ada DX was only evaluated at the first documented visit, it would undermine the usefulness of the system since although this may appear low, Ada DX suggested the correct disease before clinical diagnosis for 53.8% of cases. This shows a non-trivial difference in the evaluation of this system since reducing the time to diagnosis for over half of the cases is highly significant in this context, however, to only show a single-point accuracy of 33.3% does not accurately portray the effectiveness of this system. Therefore, technology to support the diagnosis of rare diseases must be evaluated at multiple points to get an accurate impression of its effectiveness.

If a temporally aware evaluation approach is taken in the future, it would promote the design of pre-diagnostic technology which integrates well with a patient's journey towards diagnosis. This could open a new avenue of research to apply different algorithms to this area which could prove more effective for continued use. Considering pre-diagnostic technology for rare diseases which work continuously increases its potential to support more challenging diagnoses. For example, unusual presentations of diseases and ultra-rare diseases are by their nature less prevalent and therefore, may not be well represented in the data, resulting in poor performance for these conditions. If a long-term approach is taken for rare disease diagnostic support, they may be more likely to suggest the correct condition as more information comes to light, making more common conditions less likely.

While the evaluation approach presented by Ronicke et al. satisfies this measure, it may not be feasible in several research projects since the breakdown of cases into each of the clinical visits may not be possible (i.e., this information may not be present in the data, or it may be too time-consuming). Therefore, there exists a need for approaches to evaluate pre-diagnostic technologies for rare diseases which do not require significant editing of data for each evaluation. Moreover, this approach aims to support evaluations of clinically-based technologies, so other methods may be required for patient-facing technologies.

3.5 User Interface

If any kind of user interface (UI) was mentioned in a paper at any point or could be seen from the figures, it was said to have a user interface. Otherwise, it was said to not have described a user interface. In this study, 39% of the papers surveyed described the design of the user interface for their decision support system, while 61% only specified creating the back-end of the system. This suggests that existing works in this space may be in the early stages of development with little focus on how the user would interact with the system. Moreover, many of the papers that created a user interface described it in little detail or did not discuss its applicability in a real setting. This implies that there is a perceived lack of importance or interest in this area. Therefore, there is a need for further research into human-centred approaches for pre-diagnostic technology for rare disease patients. Hence, in addition to our review of approaches within machine learning, we must also draw from HCI to further explore the research space for patient-focused approaches.

3.6 Future Directions for Rare Disease Pre-Diagnostic Technology

In this section, we have highlighted several gaps within the approaches to designing pre-diagnostic technology. By relating existing works to the context of the diagnosis of rare diseases, we identify unique challenges and opportunities for intervention within this space.

Firstly, for clinically-based technologies, we highlight the need for support with referral decisions, especially for primary care physicians who are typically the first clinical interaction in the diagnostic odyssey. As patients consult with numerous specialists during their diagnostic odyssey [6,7], greater support with these challenging referral decisions may shorten the diagnostic odyssey by reducing the number of required clinical consultations. In addition, given that multiple specialists may be consulted, we argue that specialist support may involve aiding referral or diagnostic decisions for referred patients with conditions which lie outside of their specialty.

Secondly, we argue that a key gap within this area is that existing methods of evaluation are not sufficient for assessing the impact of rare disease pre-diagnostic technology. In particular, evaluating the accuracy, specificity, sensitivity, etc. at a single distinct point offers little interpretability for whether it would reduce the time taken for a patient to be diagnosed. One approach by Ronicke et al. [97] did evaluate whether their system suggested the correct diagnosis at each consultation before the clinical diagnosis was made. However, the data required a significant amount of editing to identify the state of the patient for each consultation. This edited dataset could not be made publicly available as it would violate the privacy of the patients. However, recreating this process each time is not always feasible since the time in which symptoms were first presented to clinicians may not be available in other datasets and researchers may not have the time to edit their dataset in this manner. In addition, this approach aims to evaluate clinically-based technologies and different methods may be required to evaluate patient-facing technology.

Thirdly, we identify the lack of data which is representative of the patient's perspective since most of the data sources used consist of clinical terminology and may not show the symptom progression, only that which is presented during a given consultation. As such, we suggest that the use of quantitative persona generation may have potential in this area in providing a data source. Moreover, quantitative patient personas have been utilised within HCI to perform evaluations on recommendation systems [119], so they may have potential to facilitate evaluations of patient-facing pre-diagnostic technology. Furthermore, vignettes, consisting of patient personas from a clinical perspective, have been used to assess the knowledge and clinical reasoning skills of clinical experts [120] and pre-diagnostic clinically-based technology [24, 121]. It follows that the generation

of quantitative personas which are representative of the patient's perspective may also facilitate evaluations for patient-facing pre-diagnostic technology.

Finally, another key gap we identify is the lack of research for patient-facing interventions. The only paper reviewed which described a patient-facing intervention was by Kühnle et al. [87]. Whilst this intervention showed promising results, they did not adopt a human-centred approach. Participatory approaches for design have been long established within HCI [122], and Vargas et al. [123] argue that community participation is vital for effective research into the design and the evaluation of patient-facing technology. Therefore, we suggest that a human-centred approach for the design and evaluation of pre-diagnostic technology for rare disease patients is a key gap which we aim to address in our thesis.

Therefore, this thesis addresses the three key gaps which relate to patient-facing technology. In particular, we provide an evaluation approach which assesses pre-diagnostic patient-facing technologies for their information-seeking capabilities (i.e., the relevancy and timeliness of information support) as well as the supportive capabilities (i.e., the provision of the unmet support needs of rare disease patients). Secondly, we curate a synthetic dataset which represents the patient's perspective, and thus facilitates more data-driven approaches for patient facing technology. Finally, we utilise a human-centred approach to the development of pre-diagnostic technology for rare disease patients, to support the informational and empathetic needs of people seeking a rare diagnosis.

3.7 Patient-Facing Technology within HCI

A human-centred approach for the design and evaluation of pre-diagnostic technology for rare disease patients is a key gap which we aim to address in our thesis. As such, let us explore the design space for patient-facing technology further. In order to do this, let us explore approaches outside of rare disease interventions in order to obtain more grounding in the literature for patient-facing approaches within HCI.

It is common practice for patients to use information or social platforms in order to discover new insights about their health, with services such as Google, Facebook and ChatGPT being popular choices [20–24]. Health queries can relate to information, from the treatment of a common cold, to new management techniques of a chronic condition and treatment of a new injury. Online health information seeking may happen

throughout the diagnostic odyssey, from potentially before contact with a clinician, to long after an initial consultation has taken place. The information-seeking behaviours, in addition to the utilisation of technologies which are not specifically designed for health, suggest that patients require further support in addition to that which is available during clinical consultations [55, 124].

Existing works to provide patients with further support in the domain of HCI and consumer health have focused on the promotion of healthy behaviours [125–127] and facilitating better monitoring of health for specific conditions and disabilities, such as epilepsy [128], respiratory conditions [129] and wheelchair users [130].

Min et al. [128] and Li et al. [130] both present co-design approaches to characterise the support needs of the intended end-users to establish design implications. In contrast, the three papers on the promotion of healthy behaviours utilise a co-production approach to evaluate user satisfaction.

In addition to assessing user satisfaction, Molina et al. [127] evaluate the perceived impact on physical activity whereas Harmon et al. [126] also incorporates a numerical evaluation for learning by measuring participants' performance on carbohydrate estimation of images. Ahmed et al. [129] evaluate only the performance of their breathing rate measurements from Samsung headphones and not user satisfaction.

These works highlight the use of human-centred approaches in supporting patient's needs and increasing user satisfaction. However, these approaches do not support patients who are seeking a diagnosis for their condition and previous surveys of the research landscape indicate that a great level of attention is placed on targeted condition management, but the area of supporting patients seeking diagnosis receives less attention [131, 132]. Therefore, let us explore works that focus on pre-diagnostic patient-held technology.

3.7.1 Pre-diagnostic Patient-Facing Technology

Many works in pre-diagnostic patient-facing technology focus on self-diagnosis. For example, Zhao et al. [52] aim to facilitate early diagnosis by creating a patient-facing self-diagnosis application which presents a patient with two displays: either the symptoms alone or symptoms with associated body parts. They then perform disease matching which utilises a knowledge graph based on electronic health records. The data used in this approach is not explicitly described, but it utilises clinical electronic health records,

and as such likely includes clinical terminology which is not understandable to the patient. Hammoud et al. [121] also aimed to facilitate patient self-diagnosis through a symptom checker which utilises a probabilistic graph model and used gold-standard vignettes to evaluate its performance.

Muneeswari et al. [53] aimed to provide a privacy-preserving 'self-diagnosis' platform. Their paper does not describe any diagnostic classification but instead empowers patients with collated data recorded from several different sensors in a timely, privacy-preserving manner. These three approaches show significant potential in the identification of the underlying cause of patients' symptoms as well as information support from sensors. However, none of these approaches used a human-centred approach. This is likely due to their focus on performance, rather than on user satisfaction or wider support needs.

In contrast, You et al. [133] evaluate user satisfaction as well as unmet needs from existing chatbots. This identified different conversational styles which were used to inform the design of a new chatbot which was then subsequently evaluated using a within-subjects approach to assess participants' trust, satisfaction and perceptions of the chatbot's efficiency, likeability, effectiveness, transparency, human likeness and empathy. Whilst Li et al. [134] investigate the psychological aspects which impact recognition and trust of self-diagnosis systems and then evaluate an existing commercial self-diagnosis app to assess participants' confidence, perceived usefulness, the sufficiency of information support, understandability of diagnosis and reasoning, willingness to adopt AI self-diagnosis tools. Both of these approaches aim to facilitate greater support and user satisfaction; however, they do not evaluate performance.

As we can see from the above literature, many of the existing approaches focus only on the performance or on the supportive capabilities of technology, not on both. However, both of these aspects are important to understand the positive outcomes that technology may have on patients. As such, both the supportive capabilities and the informational support are necessary components of technology in this area. Therefore, we should design technology not only around the informational needs of patients but also around their unmet support needs. This would encourage the curation of interventions which not only support rare disease patients with their clinical challenges but also their wider challenges as people.

3.7.2 Technology for Rare Disease Patients

Approaches for patient-facing technology for rare diseases have recognised the importance of meeting the wider support needs of patients. For example, Gundersen [56] discusses the emotional need for information for those parenting a child with a rare disorder. In addition, MacLeod et al. [135] found that people who are diagnosed with rare conditions face several challenges with social relationships and clinical interactions. They later build upon these findings to identify support networks from other rare disease patients [118].

Therefore, existing research recognises the wider needs of rare disease patients as a significant factor in technology usage, however, these approaches discuss patients who have already obtained a diagnosis and as such do not support those seeking a diagnosis. The quest to obtain a diagnosis is the first, and often the most significant, challenge that rare disease patients face [6]. Therefore, there is a need for pre-diagnostic technology which facilitates information discovery and aids the unmet support needs of patients.

Clearly, pre-diagnostic technology for rare disease patients remains an emerging research area. Therefore, let us draw from the lived experiences of rare disease patients; namely, let us examine the information seeking or support seeking behaviours of rare disease patients. Whilst, minimal studies have been conducted on the information seeking behaviours of rare disease patients, it has become evident throughout this research that many people with rare diseases use social media platforms, such as Facebook, to interact with one another and participate in both empathetic support giving and receiving as well as socially-orientated information seeking and sharing. This is supported by a study of the behaviours of people with Ehlers-Danlos Syndrome (EDS), which showed that 89% participated in peer support groups on social media and that many participants also used academic publications, organisational websites and Google for health information seeking [136].

However, condition specific peer support groups can only be identified following a confirmed or suspected diagnosis. Therefore, it follows that if peer support technologies can match patients to people who share their experiences prior to diagnosis, they would be well received by undiagnosed rare disease patients.

3.8 Chapter Summary

In this systematic review, we identified several gaps within the literature. In particular, we identify the need for supportive pre-diagnostic technology for rare disease patients. Existing approaches either aim to support diagnosed patients or aim to diagnose patients. However, as we identified in Chapter 2, rare disease patients face a long and arduous journey to obtain a diagnosis. As such, we suggest that pre-diagnostic technology may be able to support people with rare diseases in their needs both as patients and as people. In particular, supporting the informational needs of patients may facilitate greater conversations during clinical consultations and empower patients to play a more active role in their diagnosis, whilst supporting their wider needs as people may reduce the emotional and social burdens faced by rare disease patients throughout this journey. Hence, we explore this space in Chapter 4 by utilising a human-centred approach to identify the unmet support needs of rare disease patients and develop prototype pre-diagnostic technologies to aid these support needs.

In addition, we identified a lack of data which is representative of the patient's perspective. Inspired by the use of quantitative personas within HCI, we suggest that the use of quantitative personas which represent a patient's perspective may provide an opportunity to address this gap and thus facilitate additional patient-facing approaches within this emerging area. As such, we present a persona generation approach to create our Patient Perspective Dataset in Chapter 6 of this thesis.

We also identified the evaluation approaches to be a key gap within the literature. In particular, we argue that patient evaluations should utilise empirical evaluations to assess whether rare disease patients' support needs are met in addition to performance evaluation on the information support capabilities. Moreover, we discuss that performance evaluations of technology, both clinically-based and patient-facing, must consider diagnosis time as a key aspect to interpret the value it can have on rare disease patients. Therefore, we present and conduct a new evaluation approach in Chapter 5 and Chapter 7, which utilises a two-phase mixed-methods approach to evaluate patient-facing pre-diagnostic technology.

Therefore, by critically comparing existing research to the context of a rare disease diagnostic odyssey, we have identified three key gaps which we explore in this thesis. Firstly, there exists a need for human-centred pre-diagnostic technology for rare disease

patients to establish unmet support needs and opportunities for improvement. Secondly, there exists a need for data which represents the patient's perspective. Finally, there exists a need for evaluation approaches which assess pre-diagnostic patient-facing technologies for their information-seeking capabilities (i.e., the relevancy and timeliness of information support) as well as the supportive capabilities (i.e., the provision of the unmet support needs of rare disease patients). In the next chapter, we aim to explore the design space by establishing the unmet support needs of rare disease patients through a human-centred approach.

Chapter 4

Exploring Support Needs of Rare Disease Patients and Technological Implications

This chapter seeks to firstly discover opportunities within the rare disease space to support the diagnostic journeys of patients, and secondly, explore the technological implications of these opportunities to present a design space for digital intervention on patient-facing technology for rare disease patients. Building on our understanding of the clinical context of rare disease diagnosis outlined in Chapter 2, we explore the wider impacts of a diagnostic odyssey on people with rare diseases. From this, we identify and characterise the unmet support needs of people with rare diseases when seeking a diagnosis and then explore potential avenues for patient-facing technology to fill this gap. This explores the design space for technology and verifies it by evaluating a mobile phone application, Puzzle, as a potential intervention within this space.

To explore the design space, we conducted a four-part user study with rare disease patients, consisting of two workshops that were preceded and followed by questionnaires. Firstly, this explored the obstacles or challenges that were most significant to patients on their journeys towards diagnosis. Secondly, by building upon the identified challenges and unmet needs, this established a design space for patient-facing technological interventions for rare diseases. This patient-centred approach explores the wider needs of patients and, combined with the implications identified from clinical interactions

in Chapter 2, it provides a clear design space for digital interventions within the rare disease context.

As seen in the previous chapter, several studies have been conducted to provide or explore the potential of diagnostic support tools for rare diseases which are primarily focused on the clinician's perspective on the issue. However, patients live with their conditions and know their symptoms intimately, in such detail that brief consultations cannot do them justice. It follows that patient involvement plays a crucial role in obtaining a diagnosis [6]. Furthermore, the UK Strategy for Rare Diseases [15] outlines a number of policies to be implemented to support diagnosis and early intervention. One of these strategies is recognising the role patients can play in helping clinicians with referral and diagnosis decisions when they have access to suitable resources.

Conversely, only a small number of studies have been conducted to consider tools for rare disease patients to use themselves as part of the diagnostic journey, with works such as Kühnle et al. [87]. Furthermore, these studies did not take a human-centred approach to involve the patients throughout the design process. This has been identified as a crucial factor in the adoption of technologies, particularly in the domain of health technologies [137].

While many user studies have been conducted to design technology for patients with common diseases, their diagnostic experiences can vary significantly from rare disease patients, and therefore, we cannot assume that their needs from technology would be the same. On the contrary, we hypothesise that their needs significantly differ since, as shown in Chapter 2, rare disease patients experience several obstacles on their journey to seek a diagnosis, including frequent consultations and referrals, several misdiagnoses and significant delays. The average rare disease patient will consult five clinicians, receive three misdiagnoses and wait four years before they receive a diagnosis [7]. These experiences led to a sense of disenfranchisement from the healthcare system, diminished hope in receiving answers and greater advocacy from patients or their family members. This advocacy was shown to be a significant factor to enable diagnosis in the patient stories examined in Chapter 2, however we infer that patients may not have sufficient support from the formal healthcare system since one of the patients used alternative medicine after several consultations.

Therefore, patients must have access to the resources they need in order to have greater advocacy and to play an active role during their journey to obtain a diagnosis.

Hence, this chapter presents the first methodologically robust and patient-driven design of pre-diagnostic technologies for rare diseases. Namely, rare disease patients are placed at the centre of the design process of a prototype technology designed to support rare disease patients by involving them at each stage to ensure this research is centred around their needs and desires. Firstly, a workshop to understand their needs during their journey towards diagnosis; secondly, a workshop to explore what technologies can better equip them and inform design choices of technology; and finally, a questionnaire to evaluate what aspects of this technology are useful and explore the potential for improvements. This showed that rare disease patients desire technology that goes beyond typical information retrieval tasks, but instead prefer technology to support them with clinician-patient communication as well as with finding peers who understand their experiences. As such, we identify an area of the design space which had previously not been explored, highlighting the need for a clear exploration of the design space.

4.1 Methodology

A four-stage study was conducted: a preliminary questionnaire followed by two workshops and a prototype evaluation in the form of a questionnaire. The preliminary questionnaire built upon the diagnostic odyssey diagram in Chapter 2 and established the participants' experiences across the diagnostic odyssey. The first workshop aimed to identify common challenges, barriers and significant moments on the journey to diagnosis as well as areas that could be improved in the future, thus highlighting their unmet support needs. The second workshop aimed to determine how technology can provide support for the previously unmet support needs of rare disease patients during their journey toward diagnosis as well as which devices, information capture, insight extraction, interfaces and functionalities might be useful.

This approach was taken to involve rare disease patients at each of the key stages of the design. Firstly, defining the problem statement - for the research to be wholly patient-centred, patients need to be involved in this process since they understand best the most significant problems that they face. Secondly, exploring opportunities for technological interventions - involving patients directly in the design of technology is important to include their desired features and to avoid superfluous features [137]. And finally,

evaluating the prototype - after creating the prototype, it is important to re-evaluate the design to establish improvements and refinements, a key principle in HCI [138].

As part of the studies, we collected two questionnaires and written notes which were all collated by the lead researcher. Sessions were also recorded, and transcripts were written up. These transcripts were analysed thematically according to the theoretical method described by Braun and Clarke [139] by the lead author to see general themes in the experiences of rare disease diagnosis. More specifically, each of the six-phases in Braun and Clarke's inductive method was followed. The first phase was to increase familiarity with the data, so the recordings and transcripts were listened to and read several times. This provided an initial idea of the trends in the data, which helped to inform the remaining stages. The second phase involved the generation of initial codes. First, open coding was used for every segment of text that related to patients' experiences of diagnosis or their attitudes towards health technologies. Once this was finished, codes were evaluated and modified where relevant. The third phase was to generate initial themes. This was done by analysing the codes generated for shared meanings and overarching narratives which in turn formed the themes and sub-themes. The fourth phase consisted of reviewing the themes identified in the previous step. In particular, the transcript relating to each of the codes for each theme was compared to the theme. Both the codes and the themes were assessed for coherence, distinctiveness and significance. Once this phase was completed, the fifth phase was conducted – defining and naming the themes. This phase involved critically assessing the essence of each theme, to explore what the theme is saying, and how each of the themes and sub-themes relate to one another. Following this, the themes were named to ensure conciseness, informativeness, and memorability. The final phase of writing-up was then conducted, using notes from the processes above as framework to build upon.

4.1.1 Participants

The preliminary questionnaire and two workshops were designed to have the same participants throughout each stage in order to build upon the previous stages. A total of five participants were identified via an email sent out to university staff and students in addition to an existing pool of previously identified eligible participants and snowballing methods. This group of participants consisted of 2 male and 3 female participants,

with ages ranging from 20-40 years old[1]. Participants were included in the study if they self-identified to meet the following criteria: they were at least 18 years old; were diagnosed with a rare disease in the UK and were able to give their informed consent. They were excluded if their time to diagnosis was less than one year, in order to eliminate people who did not experience the hurdles typical in a rare diagnosis and as such would have less to offer in discussions.

Recruitment of rare disease patients presents challenges due to their low visibility and lack of social, collective identification, resulting in a small sample size. Therefore, to increase the consistency of the findings, it was necessary to restrict the location of diagnosis to the UK due to the variability of healthcare systems. In particular, general practitioners act as gatekeepers for any secondary care in the UK's National Health Service (NHS), as such patients in the UK do not typically seek a given specialist of their own accord but require a referral. Also, NHS consultations have no direct cost, and expensive medical fees may prevent people from seeking medical attention. These factors can differ in other countries, such as the US [140].

For the prototype evaluation questionnaire, 21 participants were recruited via rare disease groups on social media. The workshop participants were also invited to participate in the evaluation. Participants were included in the study if they met the following criteria: they were at least 18 years old either with a rare disease or a Syndrome Without A Name (SWAN) and were able to give their informed consent. Restrictions on location and time to diagnosis were removed due to the increased scale of this part of the study. Participants were also included if they had a SWAN, defined as *'a genetic condition so rare that it is often impossible to diagnose'* [141]. This would include people who have not yet received a diagnosis, making them more suited to evaluate the app.

4.2 Preliminary Questionnaire

The five participants were asked to fill out a preliminary questionnaire before the first workshop, but due to time constraints, one participant was not able to do this. The purpose of this questionnaire was to identify participants' experiences of seeking health information, their interactions with patient support groups and any other key moments they experienced during their diagnostic journey. This had two purposes, firstly to

[1]Individual demographics and disease names are not provided to ensure the anonymity of participants as the risks of re-identification are greater for individuals with rare diseases [118]

characterise the support-seeking activities of patients, in relation to our diagnostic odyssey diagram; and secondly to refine the workshops and lines of questioning based on their experiences.

4.2.1 Results

Figure 4.1 shows the diagnostic odyssey diagram with the patient-led interactions of the four participants who answered the questionnaire.

All participants engaged in health information-seeking activities for several hours each week, ranging from 1-2 hours per week (1 participant) to over 4 hours per week (2 participants). This began after the first symptoms for half of the participants and after the first specialist visit for the remaining participants. Health information seeking was prompted by specific experiences for all participants; this experience varied among participants but included the suggestion of suspected diagnoses from clinicians, severity of symptoms, and ineffectiveness of medications. Three participants used search engines and one participant used a specific diagnostic health application as well as trusted websites.

Three of the four participants interacted with patient groups, which we define as any interaction with patients, groups of patients, or other people with similar experiences of diagnosis, including patient organisations and support groups. This consisted of disease-specific patient groups which were identified through recommendations from healthcare professionals (2 participants), family and friends (2 participants), or online research (1 participant). Patient group interactions began after the first specialist visit (1 participant) or additional specialist visits (2 participants).

Two participants identified other key moments which aided their diagnosis. One participant reached out to extended family after experiencing a misdiagnosis, which led to the identification of a diagnosed family member. The other participant received support from their supervisor following intervals of increased sick leave, which prompted them to seek a diagnosis. This increased their appreciation for peer support and health advocacy but also led to frustration regarding the lack of treatment options. Both of these key moments relate to increased advocacy, either by questioning a misdiagnosis and looking for information outside of the healthcare system, or due to increased awareness and support. Since the key moments for these two participants were specific to their cases and not necessarily generalisable, they were not added to the diagram.

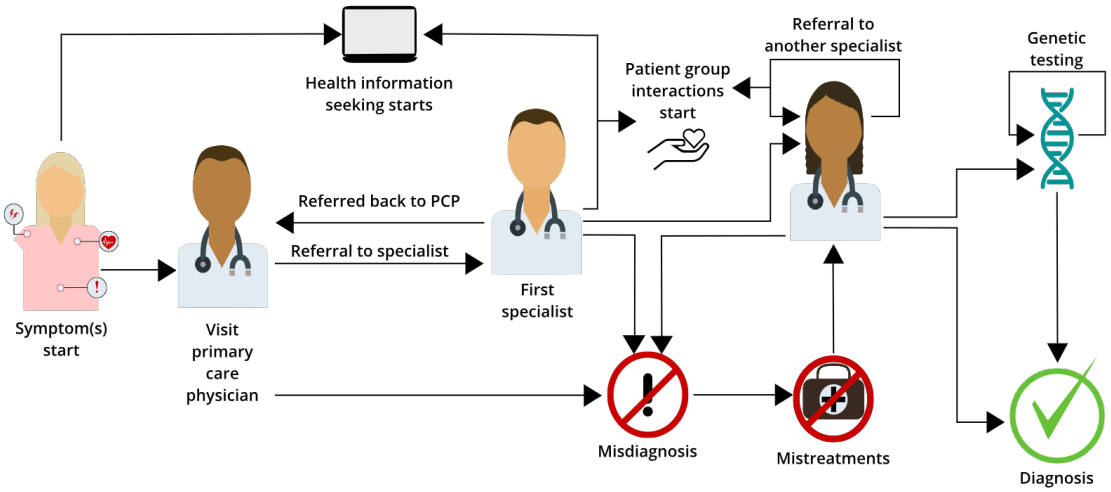


Figure 4.1: Patient interactions during the diagnostic odyssey

4.3 What are the support needs of rare disease patients?

The five participants for workshop one were invited to participate in a one-hour workshop conducted online via Zoom[2]. The preliminary questionnaire informed the workshop to include more questions on information-seeking behaviour due to the large amount of time spent seeking health information. There was also a focus on engagement with patient groups and organisations as, although there were high levels of engagement reported in the questionnaire, participants were only introduced to them after visiting specialists. Therefore, the workshop could explore the types of support gained from these activities as well as additional support that could be beneficial. As a welcome to the workshop, a brief icebreaker activity was held to allow participants to become more engaged and comfortable with one another. Participants were asked to discuss experiences of diagnosis anonymously via a Mentimeter board, for it to be revealed shortly to the group. Hence, this exemplified the shared experiences among the group. The majority of the session then followed a semi-structured focus group approach to investigate participants' own experiences of their diagnostic journey and their barriers to obtaining a diagnosis.

[2]As stated in Chapter 1, this study was conducted virtually due to Covid-19

4.4 Results: Workshop One

The initial activity to list words which best-described participants' journeys towards diagnosis revealed it was a *"difficult"* and emotionally straining time (described as *"depressing"*, *"frustrating"* and *"demoralising"*). A sense of disenfranchisement from the healthcare system was also shown in the responses to this task. In particular, participants felt that they were *"misunderstood"* in their experiences of seeking a diagnosis which was *"lacking a personal approach"*, and they had feelings of *"dismissal"*. The analysis of the responses to questions and discussions provided by the workshop's participants yielded three major themes: a perceived lack of care from healthcare professionals; insufficient empathy from peers; and diminished view of existing technologies to support information discovery during their diagnostic journey.

4.4.1 Perceived lack of care from healthcare professionals

The most predominant theme from the first workshop was a perceived lack of care from healthcare professionals. Participants identified several issues such as: a lack of faith in their doctor's abilities; feelings of dismissal and the need to convince their doctor of their state of health; being passed on from one doctor to the next; poor social interactions with doctors; difficulties and delays with getting referrals. This theme presented an issue with the doctor-patient relationship resulting in diminished trust and communication issues. For example, one participant said *"there was no one was taking ultimate responsibility of me"*, while another stated *"no one's listening to me"*, and another *"I've been told that what was happening to me wasn't real and I ended up in A&E"*. These hindered relationships are likely due to a combination of issues faced by rare disease patients, including the high number of misdiagnoses [6,7]; patients being disregarded and 'accused' of being neurotic or a hypochondriac [7] as well as significant delays and wait times for correct diagnosis [6,7]. One of the participants also reported experiencing 'brain fog' with their condition, which may contribute to challenges with clinician communication, as they may not remember to mention symptoms during their consultation.

4.4.2 Insufficient empathy from peers

Another significant theme that arose from the first workshop was participants having difficulty explaining their condition to their friends and family, with a number of them

feeling blamed for their condition. Others lacked a support network, found their peers had unrealistic expectations of them and were treated as a burden. This was likely due to non-rare individuals being unable to relate to the challenges faced by rare disease patients, with one participant stating *"most of my family still do not understand"* ten years after their diagnosis as well as peers struggling to face the severity of the situation.

4.4.3 Issues with technology

There were also several challenges with utilising existing technologies including: difficulty finding and understanding information; difficulty accessing online health information; misinformation, unreliability and a lack of trust in forums. For example, one participant said *"there are so many conditions with those [the participant's] symptoms, so it was a bit difficult to actually come across [my condition]"*. Therefore, patients are not satisfied with available technology; they need technology which will give them a voice and support them in a meaningful way.

4.4.4 Intervention Opportunities

Our analysis and feedback from participants clearly show that there are issues surrounding information-gathering needs, peer support and interactions with healthcare professionals. We can consider then that individuals with a rare disease have multiple different types of needs, not only as patients but also as people. While many health technologies support patients' needs to access health information, rare disease patients may benefit from technologies which support them with clinical interactions as well as emotional and social needs.

4.5 How Can Technology Support Patients Experiencing These Problems?

We invited the participants from workshop one to return for a second online workshop on Zoom[3]. Due to a medical emergency, one of the five participants was not able to attend. Following an icebreaker activity, three low-fidelity visual probes were presented to participants. Following this, a summary of the findings from the previous workshop

[3]As stated in Chapter 1, this study was conducted virtually due to Covid-19

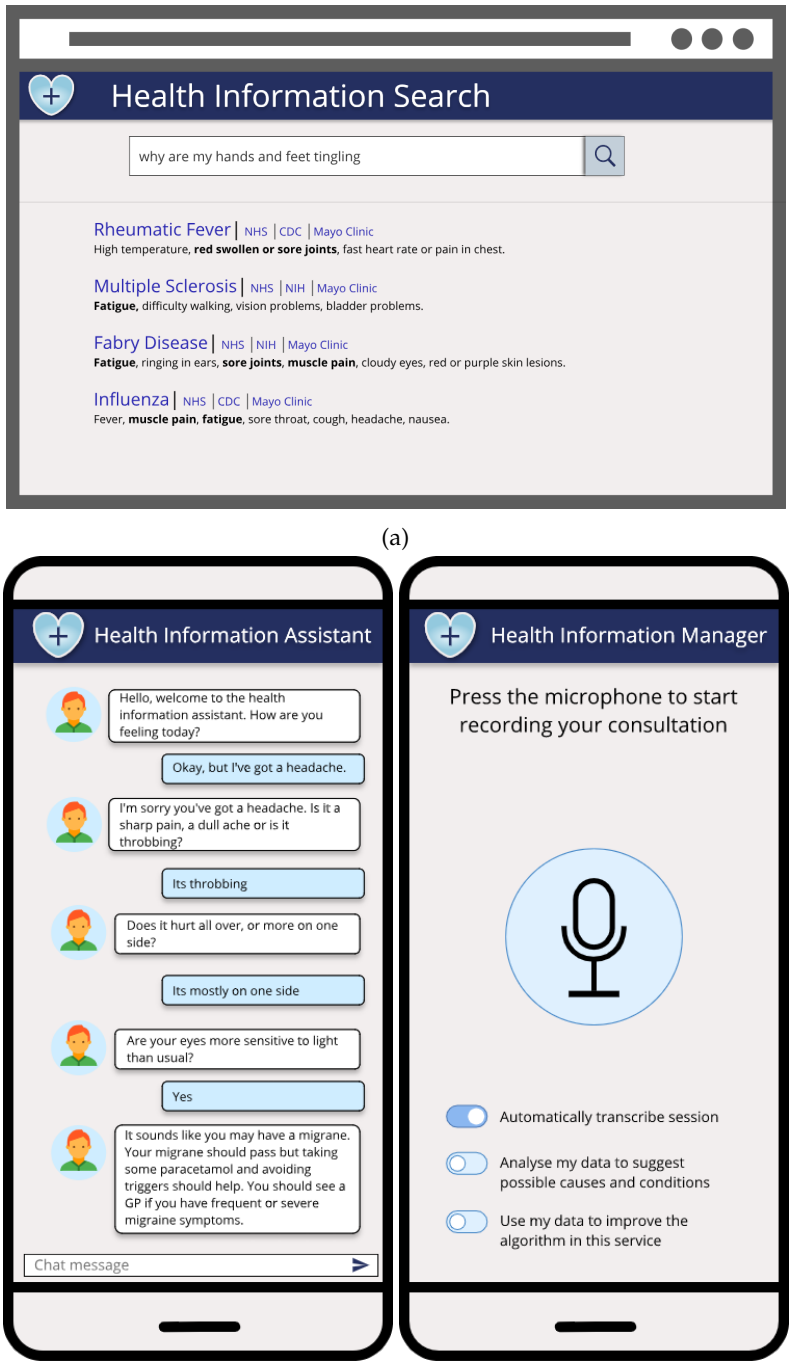


Figure 4.2: The three basic design probes for Workshop 2. (a) Probe 1: a search engine with results relating to previously searched symptoms, (b) Probe 2: a chatbot to inquire about symptoms, (c) Probe 3: a recorder to automatically transcribe consultations and identify potential causes.

was reiterated to prepare participants to explore and build on the challenges identified in the previous workshop. After the workshop, participants were given the opportunity to share any additional thoughts.

Probes have been used throughout HCI, but their definition and purpose vary significantly within this research space [142]. Boehner et al. [143] define cultural probes or design probes as *"designed objects, physical packets containing open-ended, provocative and oblique tasks to support early participant engagement with the design process"*, but highlight the varying use of probes as research tools. This study uses probes as material grounding to facilitate conversations within critical design. In particular, these probes served to provoke thoughts and ideas of technology that would be helpful in the rare disease diagnostic journey. They were designed to satisfy three main attributes: to expand participants' thinking beyond basic or existing technologies; to narrow participants' thinking from unrealistic technology and finally to leave room to encourage participants to come up with ideas.

The three basic design probes consisted of: a search engine which uses search history as a symptom list to refine future searches (Fig. 4.2a); a chatbot for messaging when experiencing symptoms which would gather more information on symptom manifestation and analyse symptoms for potential diagnoses (Fig. 4.2b); and an automatic transcription device for doctor consultations with optional additional inputted notes from the patient (Fig. 4.2c).

The design probes were initially devised based on our examination of the diagnostic odyssey. Following this, they were then presented to the patient advocacy team at Amicus Therapeutics and compared to our findings in the first workshop to refine and improve the suitability for the context and lived experiences of patients. In particular, Figure 4.2a aimed to provide better support within search engines for the lengthy time of diagnosis by identifying search results related to previous symptoms in addition to the current search. Figure 4.2b aimed to alleviate difficulty remembering symptoms due to brain fog[4], by providing access to chat histories of symptoms. Figure 4.2c aimed to provide support with difficulty understanding or remembering conversations with healthcare providers [6] as well as with feelings of dismissal or need to convince healthcare professionals by allowing them to refer back to previous consultations.

[4]brain fog is present in several rare diseases associated with cognitive impairment including Fabry disease [144] and ME [145]

Since search engines and chatbots are often already used by patients to gather health information [20–24], Figure 4.2a and Figure 4.2b portrayed methods of tailoring them for rare diseases, whereas Figure 4.2c expanded beyond existing approaches. This was to ensure that participants were aware of adaptation of existing technologies if familiarity was important to them, but more importantly, that they were not limited to existing approaches.

4.6 Results: Workshop Two

Participants' regard of the ubiquity of technology, which is *"more easily available"* than information sought from healthcare providers was evident from the icebreaker activity at the start of the workshop. Despite this factor, participants generally had negative views on the use of technology to support them during diagnosis as they described online health information as *"opinionated"* and *"confusing"*. This suggests that the ubiquity of technology may be a significant draw for rare disease patients, but their needs are not met by existing technology, which they found to be *"non personal"* and *"obscure"*. This may be because many technologies used for health prioritise popularity, and as such, users may find it difficult to identify information relating to their rare disease, which inherently will not be a *popular* result since most people will find the content to be irrelevant.

During discussions, the value of the ubiquity of technology was emphasised further. In particular, participants felt that they would want to be able to access technology at any time, both during and after diagnosis. *"[If] a new symptom or something comes up as well, then you can easily see if that's related ... [otherwise you have to] remember to do the research when you get back"*. Due to this, they felt that a mobile application would be best due to its portability which facilitates the searching of symptoms as and when they arise. Participants also wanted support from technology to continue following diagnosis *"not just the support through the, the actual diagnosis, but to support you through, well for as long as you need that support for"*. This further highlights the need for additional support since, despite the negative experiences of technology demonstrated during the icebreaker, there still exists interest in obtaining support from technology both during and beyond their diagnosis.

4.6.1 Clinician-Patient Mediation

The idea to use technology to provide clinician-patient mediation arose from exploring the potential of technology to alleviate the issues raised in the previous workshop, namely a perceived lack of care from healthcare professionals. Specifically, participants were interested in using a symptom logger to record and monitor their symptoms regularly on an app, which they could then share with their clinicians: *"it would be really handy to have something like a log or something like, you know, diaries of symptoms that the doctor can look at ... And they could see everything I've logged and be like, okay, this is interesting."* Furthermore, participants felt that a symptom logger would help combat issues with forgetting important symptoms during the short consultation window. *"They have all the information right there without me forgetting ... details and ... how you were feeling at that time."*

In addition, participants found that existing technology used for health information seeking elevated health anxiety *"never Google your symptoms because you'll think you're gonna die."* As such, they did not wish to receive patient-facing diagnostic suggestions, with one participant expressing at the start of the workshop that they were *"sceptical of technology for this kind of thing"* and had a dislike of pre-diagnostic patient-facing technology in general due to the *"high risk of self-misdiagnosis, which from my own experience can cause more problems than it can help ... [and] worse than that people [are] falling through the system"*. As such, it was deemed that any diagnostic suggestions should be presented directly to the clinician, and not to the patient. The participant who was previously against the use of pre-diagnostic technology expressed a change of mind and felt that it *"would be beneficial"* to have a symptom logger which is shared with one's clinician.

In order to share their symptoms, participants wanted to be able to notify their doctor of certain symptoms using the app and felt that it would save them time as symptoms could be quickly assessed as benign from a quick view rather than from a time-consuming appointment *"the GP'll get this notification ... and after a while they, they do see you and it's like, 'oh no, actually you're completely fine' ... Although it's not, you know, upfront, you've got that GP scrutiny somewhere down the line"*. The desire for a clinical presence on the platform may be due to the negative experiences of using technology in the past, and concerns for misinformation and self-diagnosis.

In addition to sharing with their clinician, some participants wanted the option to share their symptom data anonymously to databases *"it could help sort of add to a database, I guess, with other people. So even like anonymously, you could choose to submit that*

information to help other people who are trying to find their diagnosis" or non-anonymously to family members *"sharing information with certain family members ... so that they know what you're going through"*. However, participants also expressed concerns about sharing this sensitive data *"sharing information becomes a little bit risky with all these data protection ... I'm not sure I'd be comfortable with [it] other than me being a statistic."* Also, some symptoms may be more personal than others *"I think it should be an option ... [so] You can share some symptoms rather than share completely everything"* and participants also feel that they should clearly know *"who can see potentially your profile"*. It is obvious that any health information must be kept private and secure, but these concerns may be greater in this case due to experiences of a lack of empathetic support and any stigma associated with symptoms or the overall condition.

Participants also felt that a reminder to log symptoms would be important to get the most out of the symptom logger *"all you need then is the app, for example, to ping you ... 'don't forget to fill this in today' "* and without these reminders *"the risk is you don't, you don't use it effectively"*. In addition, participants wanted to be able to log symptom severity *"on a scale of one to 10, how bad is this symptom sort of thing ... I think that would be really useful. I really like that idea."*

4.6.2 Peer Support

Participants expressed an interest in emotional and mental health support specific to their condition and experiences during their journey towards diagnosis, suggesting that technology should *"point you in the right direction of ... therapies or ... support groups"*. This led to the suggestion of similarity analysis to be performed on symptoms logged to connect people with their peers going through similar experiences. Profiles could be matched one-to-one or could be clustered into social networking groups to facilitate peer support and symptom management advice. Participants expressed keen interest in this idea, stating that *"it is really good to have someone to talk ... [to who] goes through pretty much the exact same things that I do."*

Most participants wanted to communicate through public platforms *"a small network of people with similar symptoms, similar conditions that I think there's benefit"* and messaging and *"I quite like that ... I also do like the idea of having, having like the option [for users] to ... go on to mine and message me"*. Some participants would prefer a message board

to one-to-one messages *"you could have like, a messaging board or, or something like that where you could message other people."*

Two of the four participants felt that symptom management suggestions would be a very helpful aspect of communicating with peers *"symptom management is ... a big part of my condition"*. One participant described the benefits of discussing symptom management during a retreat for their condition. Another participant felt that *"It would be good to know what other people do to help ... their symptoms as well."* stating that *"you're only going to find those [symptom management strategies] out if you've got a place where you can talk to people about what you are going through. ... I would definitely be on board with that"*.

> *"I heard a lot of different, sort of, ways to manage symptoms like home remedies. It's like a big thing is, I need to drink a lot of water, which can be hard sometimes. So, people are like recommending certain fruits that contain a lot of water and that sort of thing. So, ... just like simple things like that, that just made it a bit easier to manage, I think would be really nice."*

However, several safety concerns arise from sharing and discussing symptom management. In particular, participants were wary of hearing about other people's experiences without a clinician's advice, *"I would consult people who have ... got similar symptoms to see what they have done and what their GP has suggested to them. But I'm not gonna do anything without consulting my GP"*. Moreover, one participant was particularly concerned about speaking to other undiagnosed people and stated *"I believe in science and knowledge. So, the only person, the most reliable people ... are GP[s] or specialist[s]."*, but also reflected that seeing people with similar experiences *"helps you mentally because you just see this is not only you"*. Therefore, participants felt that providing an option to interact: only in peer support, only in symptom management advice, or in both would help to alleviate this. In addition, it would be necessary to add a disclaimer stating to not make any medical decisions without consulting your doctor.

In addition to concerns around the need for clinical input, there exist other safety concerns around the platform. Firstly, while this was not brought up by participants, it is clear that if this app were to be implemented, strategies must be put in place to protect users' privacy. As the design is purely conceptual, this is not an issue at this stage. Secondly, participants raised concerns about bad actors and felt it was important to be able to flag inappropriate behaviour. *"It'd be probably useful to flag it for inappropriateness*

as well, or this is not as useful". Participants also expressed a need for regulation on the platform and a desire to prevent users from over-posting. For example, one participant proposed that *"you can only have access to the app if your GP or your specialists thinks it would be beneficial for you"* due to their concerns about *"somebody who isn't necessarily relevant"* using this platform and that it only *"exists for people like you."*

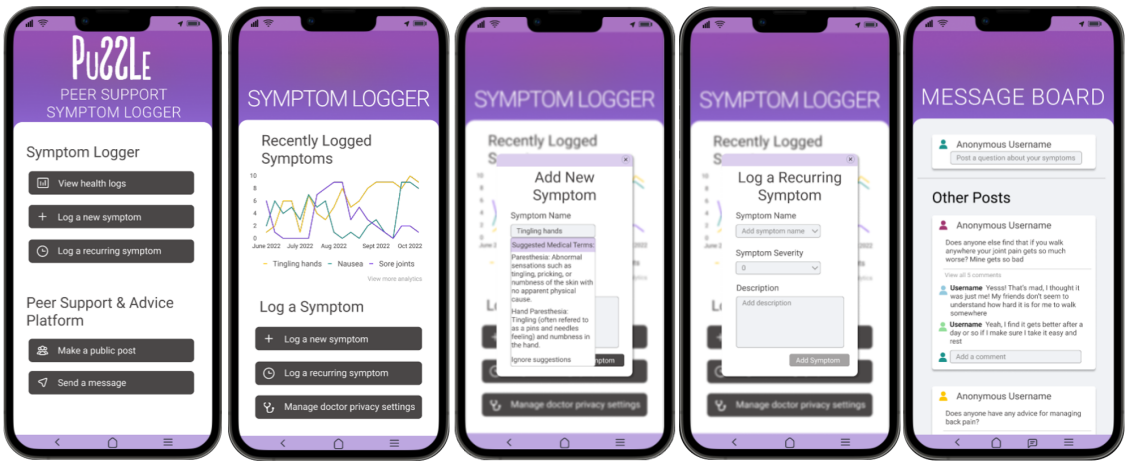
In contrast, other participants stated *"I personally, I don't think like a doctor's recommendation is the best thing, because I personally have had a lot of difficulty even just getting in contact with doctors sometimes."* and *"I don't think a doctor would recommend it. I don't think that would encourage people who need support to get it."*. Following this, the participant retracted their suggestion. Therefore, participants wanted the app to be freely available to the public with no requirements or recommendations needed from a doctor but maintained concerns about users who do not need this support. One solution that was deemed more suitable was that the app only allows group interaction after a week or so of logging to prevent people from joining and interacting impulsively.

Participants recognised the importance of specificity and personalisation to make the most of technology of this kind, as shown by the several concerns or suggestions pertaining to this. Firstly, participants wanted to be able to flag posts for relevance *"if you wanna post something ... you could like have a little flag at the top that says ... I'm looking for symptom management [or] I'm looking for support"*. In contrast to previous experiences with one-size-fits-all technology, the use of personalisation may improve the relevance of the content so that individuals are still able to view less popular content if it is more relevant to them.

Secondly, in addition to filters for peer support and symptom management advice, participants expressed interest in having other filters. *"I can't think about [it off] the top of my head, but there's probably more levels than that than just the peer [support] and health [advice] ... I think there's definitely a benefit in having some kind of tailorship around that."* This again demonstrates the importance of personalisation.

Finally, participants felt it was important that relevance was prioritised over the size of groups *"if it's too broad, I think you, you get the risk of it being just like a, a forum and it can become a little bit overwhelming ... [and] quite annoying"*. Therefore, smaller groups were deemed preferable to create more specific and relevant groups. These suggestions demonstrate the need for high relevance to their individual needs in terms of content and matches, likely due to unmet needs from popularity-focused technologies.

4.7 Prototype Design and Evaluation



(a) (b) (c) (d) (e)

Figure 4.3: Prototype application (Puzzle) screenshots: (a) Home screen, (b) Symptom logger main page, (c) Logging new symptoms, (d) Logging recurring symptoms, and (e) Patient matching message board.

4.7.1 Prototype Design

Our results from the second workshop provided specific design implications for technology. Namely, participants wanted a mobile phone application consisting of a symptom logger which could be shared with their healthcare provider and matched users to other individuals or groups with similar symptoms. Using these findings, a prototype was designed on Figma to demonstrate these key functionalities. We named the prototype Puzzle due to the puzzling nature of a rare diagnosis and this app is intended to act as a companion to patients during the challenging period of diagnosis.

Figure 4.3 shows the main screens of Puzzle. Puzzle consists of two main parts: firstly, a symptom logger where users can record symptom occurrences, the severity of the symptoms and a description of this occurrence (Fig. 4.3b and Fig. 4.3c) and secondly, a peer support platform which enables users to interact with other patients (Fig. 4.3e).

The symptom logger enables users to log new and reoccurring symptoms to track how their symptoms change over time. To further support this, users can view graphs which show the logged severity of multiple symptoms over time. To facilitate clinician-patient mediation, these symptom logs may be shared directly with users' clinicians.

The symptom logger would also translate patient terminology into clinical terminology (Fig. 4.3c), allowing the patient to select the word and definition from a list of suggested medical terms which match their symptom. This aimed to help users research their health in parallel with the app but would also improve peer-to-peer matching by ensuring consistent terminology.

The peer support platform provides similar functionalities to typical social media platforms, namely: message boards which allow people to post to groups; private messaging and group messaging. It would, however, differ from these platforms by creating group matches using symptom logs to match users to other people who have similar experiences to them. This aims to mitigate the issue raised that the peers of rare disease patients 'just don't understand'.

4.7.2 Prototype Evaluation

This prototype was evaluated in a follow-up questionnaire to assess whether the resulting app design would be useful for individuals with rare conditions. 21 participants were recruited via patient groups on social media (i.e., *'a genetic condition so rare that it is often impossible to diagnose'* [141]). The workshop participants were also invited to participate in the evaluation. Participants were asked to watch a video which demonstrates the functionalities of the prototype [5]. Then they were asked to answer the following Likert-scale questions (ranging between strongly disagree and strongly agree):

- Q1 The app presented in the video would have been helpful during my diagnostic journey.
- Q2 If it was available, I would have used the app presented in the video during my diagnostic journey.
- Q3 I would have benefited from using the symptom logger to manage my health information before I was diagnosed.
- Q4 I would want to share these symptom logs with my clinician.
- Q5 I would benefit from being matched to groups with similar symptom patterns.

These questions were then followed with open questions for constructive feedback, namely: *'What would you most like to change about this app?'* and *'What functions would*

[5] Video of Puzzle shown to participants can be found here: https://youtu.be/HGlCVgDxFE4

Table 4.1: Participant responses to a video demonstration of Puzzle

I would have...	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
Found Puzzle helpful (Q1)	33.3%	33.3%	9.5%	14.3%	9.5%
Used Puzzle (Q2)	38.1%	33.3%	4.8%	9.5%	14.3%
Benefited from the symptom logger (Q3)	47.6%	33.3%	4.8%	4.8%	9.5%
Shared logs with my clinician (Q4)	52.4%	38.1%	0.0%	0.0%	9.5%
Benefited from the peer matching (Q5)	28.6%	33.3%	28.6%	4.8%	4.8%

you most like to add to this app?'. This helped to understand why certain features were favoured over others.

4.7.2.1 Results

Table 4.1 shows that participants generally responded positively to the app with the majority (67%) of participants reporting that the app would have been helpful during their diagnostic journey and half of them stating they strongly agree with the statement. Furthermore, 71.4% of participants said that they would use the app during their diagnostic journey (38.1% strongly agree, 33.3% agree).

The best regarded feature was the ability to share symptom logs with one's clinician with 90.5% wanting to use this feature with the remaining 9.5% strongly disagreeing. This is likely due to the difficulty of communication with clinicians, as expressed by participants in the further comments question.

However, if doctor-patient trust is severely diminished, this feature may be seen as unhelpful. For example, one participant stated *"The app may depend too much on clinicians listening to patient feedback on diagnosis. Clinicians would need [to] buy-in on this to be useful"*. This poses a difficult challenge for this research area, and future work should investigate if and how clinicians can be encouraged to further involve patients and give them a greater voice in their diagnostic journey.

The vast majority (81%) of participants stated that they would have benefited from using the symptom logger during their diagnostic journey. This suggests that

participants would benefit from information management support *"This is a superb idea. In the pas[t] I have tried writing symptoms down daily and endeavoured to explain these to a doctor"*. This implies that information management plays an important part in effective communication with one's clinician.

The most contentious feature was to match patients to groups with similar symptom patterns. While the majority of people (61.9%) still felt that they would benefit from this feature, 28% of participants gave a neutral response and 9.5% of participants said they would not benefit from this feature. This may be due to concerns for moderation, misinformation and privacy which arose when participants were asked what they would most like to change about the app with a number of participants preferring the social aspect to be opt-in or opt-out.

4.8 Chapter Summary

In this chapter, we explore the design space for patient-facing technology to support people seeking a rare diagnosis. Building on the understanding of the clinical context of diagnosis, we identify the impact and wider issues of this context to identify the unmet support needs of rare disease patients. As such, we identify a large portion of the design space that is typically unexplored. In particular, we present a human-centred study with rare disease patients to design and create technologies that can support these individuals during the diagnosis process. Conversely, previous work has largely been focused on the clinician perspective within this process [78–80, 146–149]. Work on pre-diagnostic technology that does focus on patient perspectives often draws from existing literature to create features such as patient matching to promote informational support (i.e., discover possible diagnoses, coping strategies, and treatments) through rare disease patient groups [87] and do not focus on the wider needs of rare disease patients.

Our approach used exploratory and human-centred design workshops with rare disease patients to create and evaluate our design prototype, called Puzzle. Puzzle builds on the concept of Kühnle et al. [87] by not only facilitating group matching, but also including a symptom logger which could be shared with the user's clinician to give patients a greater voice in their diagnostic journey. Future work should investigate if and how *clinicians* can be encouraged to further involve patients in their health. Additionally, in our approach, users would be matched based on symptoms logged,

rather than the 53 questions required for RarePairs, reducing additional mental and temporal stressors on patients.

Many existing patient-facing health technologies [150–153] provide informational support (i.e., researching, discovering, and exploring more information about one's health). In contrast, Puzzle focuses on supporting information management (symptom logging), providing access to empathetic social support (peer matching) and support for clinician communication (clinician sharing) with the intention to be used in parallel with informational support technologies. By working with patients to understand the main challenges they face, before considering how technology could mitigate some of these issues, we understand that the needs of rare disease patients go beyond information seeking. This demonstrates the need for our work in exploring the design space to promote the development of patient-facing technology which considers the needs of people with rare diseases beyond their clinical needs.

The evaluation of the prototype created in this chapter demonstrates the impact on patient-facing technology curation of establishing a comprehensive design space which accounts for both the human side of the diagnostic odyssey in addition to the clinical side. Furthermore, it verifies the findings of our exploration of the design space, and as such, provides a starting point to investigate patient support tools which go beyond the information-seeking needs during the diagnostic journey. In order to build upon this starting point, we need to implement prototypes to perform evaluations on first-hand use. This would provide far greater insights to assess design potential than remote evaluation using a video. However, Chapter 3 demonstrates that existing evaluation approaches are not sufficient to evaluate technology. Therefore, before we discuss the implementation of our prototypes, we need to first identify suitable methods to evaluate technology. This is explored in the next chapter, where we propose a new method to evaluate patient-facing technology for rare disease patients. Due to the clear direction for the next stages of this research, these multiple avenues covered in Chapter 5 and Chapter 6 were explored simultaneously. However, to maintain a logical order for clarity, we present these chapters linearly and separately.

Chapter 5

Methodology of Simulation Task to Evaluate Pre-Diagnostic Patient-Facing Prototypes

Our findings from Chapter 2, 3, and 4 form a comprehensive exploration of the design space of pre-diagnostic technologies for rare disease patients. In particular, we have identified clear implications for both patient-facing and clinically-based technologies. Now we have presented a clear design space for rare disease pre-diagnostic technologies, we can explore what are the positive outcomes of rare disease pre-diagnostic technology and identify methods for evaluation. Then, in the following chapter, we describe our curation of the materials required for this study: a website and time-series patient personas to facilitate our study design in addition to three patient support prototypes which we then evaluate using the methodological approach described in this chapter.

In this chapter, we draw from the context of the diagnostic odyssey and our exploration of the design space to provide a comprehensive definition of the positive outcomes of pre-diagnostic technology for rare diseases. Following this, we explore the challenges associated with evaluating the additional aspects in our definition of the positive outcomes of pre-diagnostic technology for rare diseases and discuss the limitations of other approaches which are either insufficient or infeasible for this purpose. Following this, we investigate potential avenues for evaluation and, drawing from existing methodological approaches, we identify a new approach which involves

recreating the information-seeking tasks and contexts of rare disease patients. The approach we present in this chapter aims to facilitate preliminary evaluations which provide a comprehensive assessment to examine the aspects that constitute a positive outcome for pre-diagnostic technology in the context of rare diseases.

5.1 Defining Positive Outcomes for Pre-Diagnostic Technology for Rare Diseases

As we explored in Chapter 4, several aspects constitute a positive outcome in pre-diagnostic technology for rare disease patients. For clinician-facing technologies, we can consider positive outcomes to simply pertain to diagnostic efficacy, whereas the aspects of positive outcomes for patient-facing technologies can be considered as two main categories: information-seeking capabilities and provision of support. Let us consider the different definitions of positive outcomes, first in terms of the diagnostic efficacy of clinical technology and then in terms of information-seeking capabilities of patient-facing technologies by using our definition of clinical positive outcomes as a basis.

5.1.1 Diagnostic Efficacy of Clinical Technology

As we discussed in Chapter 3, clinically-based diagnostic tools should not simply measure single-point success metrics but should also account for the temporal factor of the rare diagnosis. In particular, we can consider each of the many clinical visits throughout the lengthy diagnostic odyssey to be an opportunity to obtain a correct diagnosis. As such, pre-diagnostic technology provides a positive outcome if the correct underlying condition is suggested at the *earliest* possible point. A delayed diagnosis may not only have severe health consequences from erroneous or delayed treatments, but it also incurs unnecessary expenses from additional diagnostic consultations or tests. Therefore, the temporal factor of pre-diagnostic technologies is a significant aspect which cannot be overlooked when evaluating positive outcomes in this context.

In addition to this, rare conditions are often missed, and diagnostic errors occur far more frequently in a rare diagnosis, which is demonstrated by the fact that the average rare disease patient is misdiagnosed three times. Therefore, diagnosis typically results in a high number of false negatives for rare conditions, so the consistency of an accurate diagnostic suggestion is also an important factor to consider. In addition,

there may be a trade-off between early diagnosis of those with a condition versus potential misdiagnosis of those without (i.e., false positives) which may increase the chances of misdiagnoses. Hence, it is of even more importance to consider the risk of misdiagnosis in a rare diagnosis, so consistency is a vital component in the measure of positive outcomes. Therefore, incorporating accuracy, consistency and temporal factors, our definition of positive outcomes is as follows: *clinically-based technologies provide a positive outcome if they consistently present the clinician with a patient's correct underlying condition at the earliest possible stage of diagnosis.*

5.1.2 Information Seeking Capabilities of Patient Technology

This definition of positive outcomes differs for patient-facing technology since only clinicians have the required expertise to make a formal diagnosis. Instead, patients may bring their health research findings to direct consultations. For patient-facing technology, the risk of false positives is reduced as the outcomes would be suggestions to physicians who would then be able to use directed tests to filter out, however it may impact mental health as well as physical health and resources[1].

As with clinical technologies, we can consider positive outcomes in terms of what is presented to the clinician, but patients will also evaluate these suggestions for their perceived relevance. As such, they will only bring findings to consultations that are deemed to be relevant, so an important aspect of positive outcomes is whether patients consider conditions to be potential underlying causes when performing health searches since this would determine what findings are brought to the clinician. Therefore, our definition of positive outcomes is as follows: *patient-facing technologies provide positive information-seeking outcomes if they facilitate the patient's consistent and correct identification of their underlying condition at the earliest possible stage of diagnosis.*

It is worth noting that this definition assumes that patients would bring their findings to their healthcare providers. Therefore, given that we now have a definition of positive outcomes in terms of information-seeking capabilities, let us now consider our definition of positive outcomes in terms of meeting the support needs of patients, including self-advocacy and the willingness of patients to report findings.

[1]Research into the impacts of online health information and false positives can be found in Emily Nielsen's MSc work (not examinable for the PhD thesis) https://www.swansea.ac.uk/computational-foundry/epsrc-centre-for-doctoral-training/msc-projects/emily-nielsen/

5.1.3 Supportive Capabilities of Patient Technology

To ascertain the supportive capabilities of technology, we need to identify whether patient support needs are met. While we identified the commonly shared support needs of rare disease patients in Chapter 4, we cannot limit our definition to these needs due to the personal nature of support. In particular, the support needs of an individual may vary on several factors (e.g., an individual's disposition or access to existing support). So, by maintaining a broad definition, we can consider, for each individual, both the identified support requirements and the degree to which these support needs are fulfilled.

Therefore, in this aspect of design, we consider positive outcomes to be dependent on whether and how effectively technology alleviates all possible support needs. As such, *we consider patient-facing technologies to have a positive support outcome if they are able to satisfy, to the greatest extent, as many patient support needs as possible*

5.1.4 Final Definitions of Positive Outcomes

Therefore, our final definitions of positive outcomes for pre-diagnostic technology for rare disease patients are as follows. We consider clinical technology for rare disease diagnosis to provide a positive outcome if:

> *It consistently presents the clinician with a patient's correct underlying condition at the earliest possible stage of diagnosis.*

We consider pre-diagnostic technology for individuals with a rare disease to provide a positive outcome if:

> *Firstly, it facilitates the patient's consistent and correct identification of their underlying condition at the earliest possible stage of diagnosis. And secondly, it satisfies, to the greatest extent, as many patient support needs as possible.*

It is worth noting that whether a clinician is able to perform a diagnosis based on the suggestions presented to them (either by technology or by a patient) is another factor to be considered. However, as this would significantly increase the complexity of evaluations, we consider this to be outside of the scope of a preliminary evaluation. As such, we leave this factor for evaluations of prototypes which are in the later stages of development.

5.2 Challenges of Evaluation Methods

Now that we have established a clear definition of positive outcomes, the next logical course is to establish a method to measure or evaluate pre-diagnostic technology for rare disease patients based on this. Current approaches used to evaluate rare disease pre-diagnostic technology do not consider all the aspects of positive outcomes that our definition covers. Thus, we need to identify suitable approaches to evaluate rare disease pre-diagnostic technology which examines all the aspects that we have identified to be factors of positive outcomes. However, evaluating technologies for undiagnosed rare disease patients presents several challenges due to the prolonged time to diagnosis, recruitment challenges and ethical considerations.

Firstly, due to the lengthy diagnosis of a rare disease patient, potentially lasting 30 years [6], it is not feasible to perform long-term deployments within the timescale of a PhD project. Moreover, even beyond the scope of a PhD, many other academic research projects do not have the temporal and financial resources required to conduct evaluations on deployments of this length.

To address the lengthy time of diagnosis, one approach would be to conduct several short-term deployments utilised at different points in the diagnostic odyssey. This may provide insight into information-seeking and support needs at different stages of diagnosis which may account for whether technology aids in the identification of the correct condition and support needs for a given stage of diagnosis, however, it will not evaluate whether the identification is at the earliest stage possible, which we define as a crucial element of a positive outcome.

Secondly, individuals with rare diseases represent a limited pool of participants in which to contact and participate in studies and design workshops. This is not only due to the lack of prevalence of individual conditions, but also a lack of social collective identification, since many patients do not know that their condition is classed as rare. One approach which does not have this issue, which was used by many evaluations of pre-diagnostic technologies in the literature, was the use of numerical or performance-based evaluations based purely on the information output of prototypes may be utilised. However, since patients evaluate health findings and only bring findings to consultations that are deemed to be relevant, we need to ascertain the findings that patients take on

board, not just the ones that are presented to them. Therefore, an empirical study is necessary to assess the level of positive outcomes of patient-facing technology.

Alternatively, there exist approaches to recruit participants from hard-to-reach groups, as demonstrated by Xu et al. [154], which may alleviate this challenge. However, the undiagnosed rare disease population presents additional challenges in terms of identification, since they include all manner of conditions and are defined only by their low prevalence. As such, the manifestations of these conditions vary significantly and similarities in experience are only a result of their rarity. As a result, we cannot utilise recruitment strategies around clinical presentations.

Moreover, as non-experts, we cannot expect to successfully identify patients before clinical experts are able to. Therefore, this leaves us with only the shared experiences resulting from the rarity of conditions as a basis from which to recruit. That is if their experiences of diagnosis follow the pattern of the rare disease diagnostic odyssey illustrated in Figure 2.2. However, this would only identify patients who have already experienced a diagnostic odyssey, and as such would not capture information from earlier stages of diagnosis.

Finally, it is unethical to deploy prototypes which could potentially hinder the diagnostic journey as this may lead to severe consequences on patients' health. For example, a misdiagnosis may be made if an incorrect condition is suspected based on patient findings from using the prototype. Therefore, sufficient prior evaluations to evaluate the safety of prototypes must be conducted to ensure prototypes are safe before any real-world deployment takes place.

5.3 Exploring Potential Avenues for Evaluation

Because of the reasons described above, in situ, long-term deployments in this context are not merely difficult; they are far beyond the boundaries of a reasonable or practical solution, rendering them effectively impossible. To address this gap, we aim to identify an alternative approach to measure all of the factors of positive outcomes in this context, without the need to identify, recruit and deploy interventions at the start of the diagnostic odyssey.

Due to the vastly different natures of the information-seeking and support capabilities of technology, let us first consider a potential approach to evaluate the information-seeking capabilities of technology. In particular, we aim to design a laboratory study that evaluates our definition of positive outcomes but mitigates these challenges by reducing the search for diagnosis into a controlled setting which mirrors the real-world context. As part of our design process, let us explore and draw from research methods which offer solutions and insight for similar challenges.

5.3.1 Drawing from Existing Evaluation Approaches

As resource constraints often present issues for a wide range of academic research projects, there exist several methods to evaluate technology which do not require in situ, long-term deployments. Firstly, Nielsen's work on discount methods of evaluation [155] (i.e., user and task observation, scenarios, simplified thinking aloud, and heuristic evaluation) have been widely accepted as an effective approach to evaluate usability where the *'best'* approach is not obtainable due to resource constraints. The use of discount methods, which are considered good methods but not necessarily the best methods, to provide feedback early on in the design process was shown to facilitate evaluations in situations where the methods were not feasible.

This motivation aligns with our own, however, our object at this stage is to evaluate design features for their utility, not their usability. Therefore, inspired by this work, we aim to identify a new approach to evaluate the utility of the design features in the context of pre-diagnostic technology to aid information-seeking and wider support needs of rare disease patients. To do this, we draw from two main methodological approaches, game-based and simulation-based laboratory studies, which mirror the tasks and contexts of an in situ evaluation.

5.3.1.1 Gamification and Task-Based Approaches

The use of gamification for evaluating technological interventions apply game mechanics, such as leaderboards and scoring systems, to facilitate evaluations which typically revolve around an end-user's task. In particular, by recreating the real-world tasks of intended end-users and setting the object of the game to be the end-users' desired outcomes of those tasks, these approaches can evaluate interventions for their performance on the typical tasks of end-users.

Game-based approaches have been utilised to facilitate evaluations of a number of studies which evaluate technology for information-seeking capabilities. Firstly, Mackenzie et al. [156] evaluated what properties of emails allow re-finding, by giving participants 5 seconds to read an email, asking them 5 yes or no questions and then giving them 120 seconds to find the email. Two studies compared participants' abilities to perform information retrieval tasks in order to design and validate a user-interaction model which predicts the difficulty of a search task. He et al. [157] asked participants to find 10 relevant documents for a given prompt and scored participants by the number of clicks remaining, whereas Ageev et al. [158] scored participants by the number of correct answers out of ten difficult search tasks. All three of these approaches highlight the potential of game-based approaches in evaluating information discovery. Moreover, we see further potential of game-based or task-based design in the work of Riegler et al. [159] which gradually reveals blocks of images and asks participants to guess the subject of the image as quickly as possible in order to create data for computer vision algorithms. This mirrors health information seeking across the diagnostic journey in that it reveals information (i.e. symptoms and clinical findings) over time and participants are tasked with identifying what this information alludes to (i.e. the underlying condition) as quickly as possible. Therefore, the utilisation of tasks within game-based design approaches shows high potential for evaluation of patient facing pre-diagnostic technologies. However, patients do not know if their health information searches are correct or not, so a game-based approach, which would require the reveal of the correct answer, would not be a reasonable evaluation method.

Clearly, the recreation of common information-seeking tasks to evaluate search engines provides a direct parallel to health information-seeking tasks. However, the use of gamification is not suitable for our context since patients do not know if their findings are correct when seeking health information. In addition, it is inappropriate to utilise game mechanics for studies relating to health due to the sensitive nature of this topic. Instead, we may utilise the task-based aspects of this approach to mirror health information-seeking tasks.

In contrast to general information-seeking, health information-seeking tasks are far more context-dependent. In particular, patient health information-seeking tasks are intrinsically linked to experienced symptoms and other health information. Therefore, we are only able to create a realistic task if we can recreate the rare disease context in

terms of the discovery of observable patient traits (i.e., symptoms and findings from clinical investigations), known as phenotypes.

5.3.1.2 Simulation-Based Approaches

Simulation-based studies recreate the real-world *contexts* of the intended end-users, typically through visual or tangible experiences to evaluate the suitability of technology within its intended context. Simulation-based design has been widely adopted across other disciplines for many years, but recent work argues that simulation has unmet potential in the research space within the field of HCI [160].

Existing approaches within HCI have primarily utilised virtual or augmented reality, but some papers utilise a simulation-based evaluation approach to assess mobile phone applications for information retrieval and usability [161–163]. In addition, Guerrier et al. [164] simulated health contexts to evaluate Cerebral Palsy interventions, that is, they simulate involuntary movements as a specific targeted symptom. Again, each of these approaches draw parallels to our context, as we aim to evaluate the information seeking capabilities of pre-diagnostic technology for rare disease patients. In our case, we cannot simulate the full experience of symptoms and the clinical journey that rare disease patients go on. Even if it were feasible in practice, it would be unethical to subject participants to potentially challenging or painful experiences. Therefore, we propose to simulate only the data-orientated aspect of a rare disease journey towards diagnosis. In particular, we want to present participants with the phenotypes of rare disease patients in the order in which patients discover these traits. This draws a clear parallel to the work by Zhang and Balog [165] which simulates users to evaluate a recommender system. In our case, we want to simulate the informational journey of rare disease patients to evaluate a recommender system for patient matching. This provides the contextual knowledge base from which patients perform health information-seeking tasks.

5.3.2 Simulation Task-Based Approach

Therefore, drawing from these two approaches, we can utilise an information-based simulation to provide the context of the health information-seeking tasks that rare disease patients undertake. This allows participants to perform health information-seeking tasks which mirror those of rare disease patients. We envisage a simulation where the participant is asked to pretend they are a patient with an unknown condition, they are

successively given information about emerging symptoms, which would take many years to emerge in real life. At each stage, they are asked to use a given tool to investigate the symptoms of their assigned profile to make a potential diagnosis.

Now, let us consider and refine the characteristics of the simulation task required to evaluate the information-seeking capabilities of pre-diagnostic technology for rare disease patients. Then, in the next chapter, we will describe our curation of the materials required to provide the simulation of the data-orientated context of rare disease patients. In particular, we create time-series patient personas which reveal the phenotypes of a given (synthetic) patient over a series of stages in the simulation task.

5.4 Recreating Information Seeking Tasks of Rare Disease Patients

Now that we have an idea of the general approach to take, we need to ascertain several aspects of the information-seeking task. Firstly, let us broadly define the real-world objective of rare disease patients based on our definition of positive outcomes: patients are able to identify their underlying disorder at the earliest stage in their diagnosis. Thus, abstracting from this, we task participants with the identification of the persona's underlying disorder at the earliest point in the simulation as possible.

Before we can define more concretely and quantify this objective for evaluation, we need to characterise the task in more detail. In particular, several variables need to be considered to both mirror as closely the real-world task and maintain an appropriate level of difficulty for the study. Namely, we must identify appropriate values for the number of guesses per round, the quantity of rounds, and the duration of rounds must be identified.

5.4.1 Defining the Number of Guesses

To identify a suitable number of guesses, let us draw from the information-seeking behaviours of rare disease patients. Patients will evaluate health findings for perceived relevance and compare with their experience of living with their condition, but do not have the required expertise to perform clinical evaluations. As such, they may bring multiple findings during clinical consultations if they are initially deemed to be relevant to obtain the discernment that only clinicians are qualified to make. As clinical

consultations typically last less than 10 minutes[2], it is unlikely that patients will have the opportunity to communicate many findings during this time window. In addition, even if they do have time, they may not present all findings if they do not consider them to be sufficiently relevant. Therefore, in our recreation of the task, we allow participants to make up to three guesses to identify the underlying condition to mirror the limitation of patient input for each clinical stage whilst maintaining the ability to abstain from putting forward conditions that are deemed to be irrelevant.

5.4.2 Defining the Duration and Quantity of Rounds

Setting a time limit for each round and quantifying the number of rounds per simulation task would affect the difficulty of the study task. To get the most useful data from our study, we needed to evaluate this difficulty. If the task is too difficult, participants will score poorly no matter what tool they are assigned, but if the task is too easy, they will perform well no matter what tool they are assigned. We need to easily discriminate between tools for their information-seeking capabilities, hence we perform an effect maximisation (i.e., we identify the level of difficulty which maximises the differentiation between the tools). As such, we make the task reasonably challenging, but still attainable for many participants. Therefore, we conduct two pilot studies to evaluate the suitability of these variables for our study. Before we do this, let us establish initial variables by drawing from the typical clinical experiences of rare disease patients.

The number of rounds in our simulation should relate to the number of clinical consultations since each consultation can be considered as an opportunity for diagnosis. The average rare disease patient consulted with five clinicians before diagnosis [7], so we might assume that there should be five rounds. However, while some patients began researching their health at the onset of their symptoms[3], others started researching following their first specialist referral or visit[4]. This means that patients do not necessarily start bringing their health findings to all five consultations, instead, they may only bring findings from the second or third consultation. Hence, they will bring findings for two to five consultations, so we should have two to five rounds in our simulation.

[2]https://www.bmj.com/content/359/bmj.j5172

[3]Reported by two participants in Chapter 4

[4]Reported by two participants in Chapter 4; in both diagnostic odyssey diagrams (as well as one paper in Table 2.1) in our characterisation of the diagnostic odyssey in Chapter 2

Since the conditions must remain consistent among participants to increase the reliability of results, we need to identify a single value within this range. Therefore, now that we have a basis to work from based on the real-world context, let us refine this further based on the difficulty of the simulation task. We consider two rounds to be insufficient to evaluate the temporal aspect of information discovery, so taking an initial value of three stages in the simulation, let us evaluate the difficulty of the task through our pilot studies.

Setting a time limit for rounds is another important factor for consistent results. There are no quantifiable time limits in a real-world context, only time constraints pertaining to referral waiting times and the amount of available time patients have. This may vary depending on the referral, however, to obtain meaningful data, we must maintain a consistent time limit across all rounds. As such, we do not base the time limit on the real-world context, but instead on the utility of the study and the difficulty of the task. We initially adopted a time limit of three minutes as we deemed this to likely be an appropriate level of difficulty.

5.4.2.1 Pilot Studies to Evaluate Difficulty

In our first pilot, we evaluated the difficulty for three rounds, each lasting three minutes. In this pilot, one participant[5] (an associate professor in computer science at Swansea University) played through three simulations and was observed by three other researchers in computer science at Swansea University (a PhD student, a lecturer and a professor). They were informed to evaluate the suitability of the task to obtain good-quality data from the study and discussed out loud any challenges they found both during and after the simulation. It was noted, both by the participant and the observers, that the first round was particularly challenging, and as such, may not get the best data.

The participant felt that they didn't have enough time to feel fully informed and it was deemed that additional time should be given to participants to allow for initial ideation. In contrast, the full-time was not needed for the last two rounds, with the participant finishing between 45 and 60 seconds before the time ran out. Therefore, we need to increase the time for the first round of information discovery and reduce the time from the final two information rounds.

[5]The number of participants, whilst low, sufficed for the purposes of the pilots (to get a reasonable estimate to avoid wasting resources in the main experiment)

Responses are assessed based on how early in the simulation is the round in which the correct disorder is identified, so these rounds must remain uniform to improve interpretability. Therefore, we propose four timed rounds lasting two minutes, but with only three information discovery rounds. In particular, the first two rounds present the same set of initial symptoms with no additional phenotypes revealed in the second round, but with new phenotypes revealed in the final two rounds. This allows more time for initial ideation, reduces the time for later rounds in the simulation and ensures the duration of the rounds remains consistent.

Following this, we conducted another pilot with three participants[5] (a rare disease clinical expert, a lecturer in computer science and a professor of computer science). They each played the same simulation task using Google and could discuss amongst each other any challenges they found or thoughts which came to mind. The participants found the task challenging, and usually required most of the time, but did not always require the full time. As such, this level of difficulty was deemed appropriate for the study.

5.4.3 Defining the Objective and Quantifying Success

Based on our definition of what makes a positive outcome for pre-diagnostic technology for rare disease patients, we can consider the objective of rare disease patients during health information seeking to be the identification of their underlying disorder as early in their diagnosis as possible. Hence, we task participants with the identification of the persona's underlying disorder at the earliest point in the simulation as possible. More specifically, let us define a metric to quantify the different aspects of our definition of positive outcomes: timeliness, accuracy and consistency.

First, we need to determine which guesses are accurate. For this, we will need to clean participant responses for spelling and UTF-8 encoding issues; and then compare responses to a dictionary of synonymous terms for the underlying condition of the assigned persona. Any guesses which were included in the dictionary were identified as correct. In addition, the raw inputs were manually checked for automation errors.

Given the time limitation of clinical consultations, a patient's most confident finding would be presented first and prioritised over others, so it follows that the order of guesses is key. So, the score, s, of a given round, r, was based on the order of the guesses, as follows:

$$s_r = \begin{cases} 3, & \text{if the correct answer is the top guess} \\ 2, & \text{if the correct answer is the second guess} \\ 1, & \text{if the correct answer is the bottom guess} \\ 0, & \text{if there are no correct answers} \end{cases}$$

Now that we have a measure for accuracy, we need to consider the timeliness of participant responses. To do this, we could potentially measure the time taken to respond. However, the diagnostic odyssey has distinct consultations which provide opportunities for clinical diagnosis. As such, we measure temporal factors discretely, not continuously, so, we measure responses timeliness based on which round they obtain the correct answer, rather than the specific time taken to obtain the correct answer.

The score given to a completed task could naturally be considered the sum of the score of all rounds. However, as discussed earlier, consistency is a key factor of positive outcomes, so we must reduce the influence of lucky and inconsistent guesses. Therefore, we only scored rounds from the point where the correct answer remained in at least one of the three guesses for all the following rounds, which we denote the point of correctness, p. Hence, we define the total score S of a completed task as:

$$S = \sum_{r=p}^{4} (s_r).$$

That is, we add scores for all rounds from the point where the correct answer is in at least one of the three guesses for all following rounds. This measure assesses not only how fast the correct underlying condition can be identified as a potential cause, but also how consistent the guesses are.

5.4.4 Methodological Procedure

Now we have established a measurable task which allows us to empirically evaluate pre-diagnostic technology for rare disease patients. Let us now summarise the task and discuss the details of the procedural approach.

Firstly, we adopt a within-subjects design, so participants repeat this task for different prototypes. This can reduce the impacts of individual differences on data collection. This is important here given that health literacy and technological familiarity may vary

significantly among participants. However, to reduce the learning effects on our data collection, we permute the conditions among participants so that all six permutations of the three patient prototypes to be evaluated are distributed equally among participants.

Secondly, given that our task-based study simulates the context of rare disease patients, we no longer require rare disease patients at the stage of this informational evaluation. Of course, we need to ascertain their perspectives of technology, but we can utilise a qualitative accompanying approach to ascertain the impacts of patient-facing technology on rare disease patients. Recruiting from the general public enables the collection of statistically significant data, which would otherwise be highly demanding on resources.

The final resulting study procedure is as follows. First, participants are assigned one of the prototypes and a patient persona consisting of symptoms and clinical findings. They must only use the assigned prototype but are permitted pen and paper if they want to take notes. The persona's phenotypes are gradually revealed in a series of timed stages or rounds to imitate the information discovery of rare disease patients. During these rounds, participants are tasked with the identification of the persona's underlying disorder at the earliest stage in the simulation as possible. Participants then repeat this process, each time with a new persona and prototype, until they have completed the simulation tasks for all three of the prototypes. The response collection, timing of the rounds, and the prototype and persona assignment were all conducted through a website, illustrated in Figure 6.4.

In addition to the scoring of participants' guesses, participants' experiences will be recorded through the following questions at the end of all three tasks:

Q1 The tool was suitable for this task (strongly agree-strongly disagree).
Q2 The tool was easy to understand and use (strongly agree-strongly disagree).
Q3 Out of the tools I used, the tool I would prefer is (each of the three tools)
Q4 I would prefer to use another tool (yes: free text/no).

This is significant in providing more insight into why the interventions evaluated perform well or poorly. In addition, it provides comparison within and beyond the study of preferred tools.

5.5 Chapter Summary

In this chapter, we draw from the context of the diagnostic odyssey and our exploration of the design space to provide a comprehensive definition of the positive outcomes of pre-diagnostic technology for rare diseases. Following this, we explore the challenges associated with evaluating the additional aspects in our definition of the positive outcomes of pre-diagnostic technology for rare diseases and discuss the limitations of other approaches which are either insufficient or infeasible for this purpose. Following this, we investigate potential avenues for evaluation and, drawing from existing methodological approaches, we identify a new approach which involves recreating the information-seeking tasks and contexts of rare disease patients. The approach we present in this chapter aims to facilitate preliminary evaluations which provide a comprehensive assessment to examine the aspects that constitute a positive outcome for pre-diagnostic technology in the context of rare diseases.

Chapter 6

Implementation of Pre-Diagnostic Patient-Facing Prototypes and Simulation Data

Now that we have established an evaluation approach, we need to create the materials required for this evaluation. Firstly, we describe the curation of the Time-Series Persona Dataset to provide the crucial contextual information in which participants can perform the health information-seeking task described in our evaluation approach in Chapter 5. Secondly, we describe our implementation of a website to present the personas to participants at each stage of the simulation task. Finally, we implement pre-diagnostic patient support tools for rare diseases which will be evaluated in our study.

For the final part of the study, we build upon our findings in Chapter 4 which identified three potential avenues for design: information management; clinician-patient mediation and support; and providing access to empathetic support. Puzzle facilitated this through a symptom logger; clinician-patient communication support; and patient matching (connecting users to people experiencing the same symptoms as them). In this chapter, we build upon this work by implementing a patient support tool to facilitate further evaluation of the patient-matching aspect of Puzzle in comparison with other tools. In particular, we describe the curation of the following three support tools: Peer Matching - a recommender system to match patients to people with similar experiences; MaladyHelp - a rare disease specific search engine with support groups; Google Custom

- Google search with a specified dataset. All of the work regarding the development of the retrieval corpora data and implementation of the search engines was conducted by Souradeep Ghosh, primarily for his MSc project work, and supervised by Emily Nielsen.

6.1 Data Curation Process Overview

We have three main types of data used in this chapter. First, the retrieval corpora data, which consists of a range of documents used to fine-tune MaladyHelp and provides the retrieval data for MaladyHelp and Google. Second, is the Static User Profile Dataset, which consists of several patient profiles with realistic phenotypes associated with a given disease. Third, is the Time-Series Persona Dataset which forms the patient personas in the simulation study. Figure 6.1 provides an overview of the processes to curate each of these three datasets. By augmenting Orphanet's data with patient perspective information (layman terminology and patient information discovery), we curate the Patient Perspective Dataset. This dataset provides a basis from which we generate the Time-Series Persona Dataset and the Static User Profile Dataset. We describe this process in more detail below.

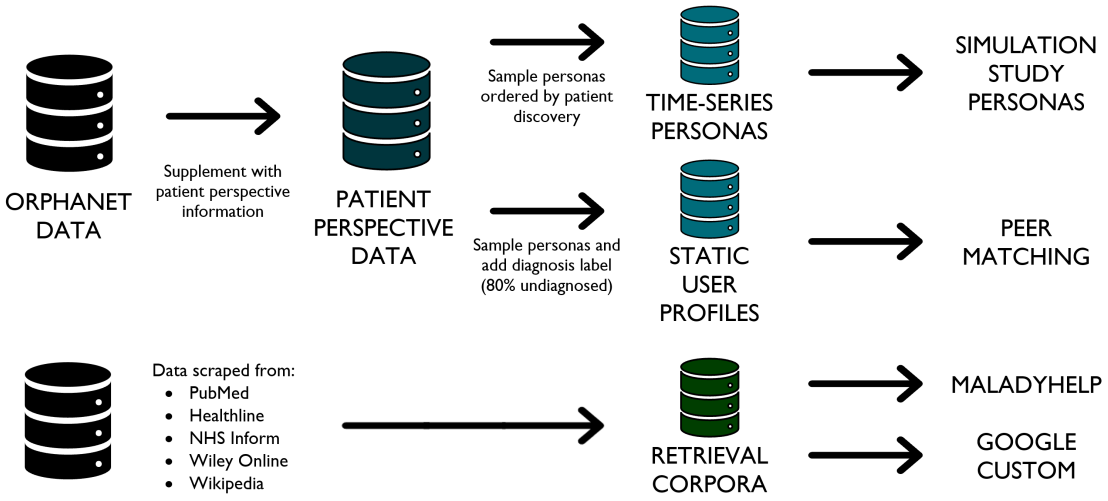


Figure 6.1: Overview of the curation process for each of the three datasets

6.1.1 Identifying Disorders for Evaluation

To establish a challenging but tractable basis on which to evaluate the potential of the intervention tools, we adopt a subset of three main rare diseases: Fabry Disease[1], Gaucher Disease[2], and hypermobile Ehlers-Danlos Syndrome (hEDS)[3] as the basis of our empirical work. These were chosen because they are well documented, have varying diagnostic difficulties, and are easily misdiagnosed [166–170]. Thus, they provide a suitably challenging information-seeking task but also have sufficient documentation to gather the resources for the databases for each of the prototypes to be evaluated.

For the purposes of the evaluation approach, we need additional conditions in the databases of the three prototypes for evaluation in order to sufficiently assess their performance. The inclusion criteria were chosen to create a real-life distribution of the experimental context. That is, the three main rare diseases provided the basis for the Time-Series Persona Dataset, but the prototypes also included similar conditions (i.e., conditions that are most likely to be mistaken as the underlying condition).

The inclusion of misdiagnoses was guided by the relevant literature on the disease in question, such as 39% and 15% of patients with Fabry who were misdiagnosed with rheumatic fever and arthritis, respectively [171]. Similarly, in another research study, patients with Gaucher disease revealed a consistent pattern of previous misdiagnoses that included leukemia, hepatic cirrhosis, and idiopathic avascular necrosis [172]. When selecting the additional disease, we ensured that only specific named conditions were included and excluded psychological conditions. In all, we identified an additional 16 rare (rheumatic fever, dermatomyositis, erythromelalgia, myelofibrosis, five rare forms of leukemia, two rare types of avascular necrosis, two rare forms of rheumatoid arthritis, vEDS, cEDS, cardiac-valvular EDS) and 8 common disorders (arthritis, fibromyalgia, myelofibrosis, leukemia, idiopathic avascular necrosis, hepatic cirrhosis, chronic fatigue syndrome, rheumatoid arthritis) which individuals with the three misdiagnoses are commonly misdiagnosed with [170, 173–178]. Therefore, including the three main disorders,

[1] Fabry Disease:
my.clevelandclinic.org/health/diseases/16235-fabry-disease
rarediseases.org/rare-diseases/fabry-disease/
ninds.nih.gov/health-information/disorders/fabry-disease
medlineplus.gov/genetics/condition/fabry-disease/

[2] Gaucher Disease:
mayoclinic.org/diseases-conditions/gauchers-disease
wikipedia.org/wiki/Gaucher's_disease/
ncbi.nlm.nih.gov/books/NBK1269/
hopkinsmedicine.org/health/conditions-and-diseases/gaucher-disease/

[3] hEDS:
nhs.uk/conditions/ehlers-danlos-syndromes/
ehlers-danlos.com/what-is-eds/hypermobile-ehlers-danlos-syndrome-heds/
ncbi.nlm.nih.gov/books/NBK1279/
rarediseases.info.nih.gov/diseases/2081/hypermobile-ehlers-danlos-syndrome/

we have a total of 19 rare disorders (including rare sub-types of disorders), and 16 disorders (excluding rare sub-types of disorders which are included in common conditions).

The Static User Profile Database for Peer Matching should be representative of the user base in a real-world context. Peer Matching is designed for rare disease patients and aims to cater for those facing a diagnostic odyssey, so people with common conditions would not have the need for this tool. Therefore, for the purpose of our evaluation, the Static User Profiles Database for Peer Matching included the 19 rare conditions. The retrieval corpora data for the search engines included both the rare and common disorders, but excluded the rare sub-types of disorders, resulting in a total of 16 disorders.

6.2 Curating the Patient Perspective Dataset

Both the Static User Profile Database for Peer Matching and the Time-Series Persona Dataset for the simulation study require realistic patient data which represents the patient's perspective. However, many sources of patient data are obtained from Electronic Health Records (EHR) which consist of technical, clinical terminology [179, 180]. These datasets show patients, their condition, and their phenotypic information, either stored as codes or in raw text. However, they are often difficult to access and do not represent a patient's perspective, since this would naturally consist of non-expert language. Some datasets exist which reflect the patient perspective, for example, a number of companies, such as Apple or Google, collect health data (e.g. symptom logging and sensor readings) from smart devices, but these datasets are considered proprietary information and as such are not publicly available. We therefore propose an approach to generate synthetic patient data consisting of a given disorder, as well as non-expert terms for their phenotypes.

We need to create unique and realistic profiles to ensure both our Time-Series Persona Dataset and Static User Profile Dataset is as close to real data as possible, so we need to use some form of real patient data as a basis. Since data reflecting the patient perspective is not easily accessible, we must use clinical patient data as a basis to create patient perspective profiles. Therefore, several aspects of the clinical patient data will need to be edited to make it representative of the patient's perspective.

This requires a significant amount of work to alter individual distinct patient records, conversely, a knowledge base which serves as a foundation to generate specific patient profiles is more efficient to edit. In particular, we can alter the knowledge base to

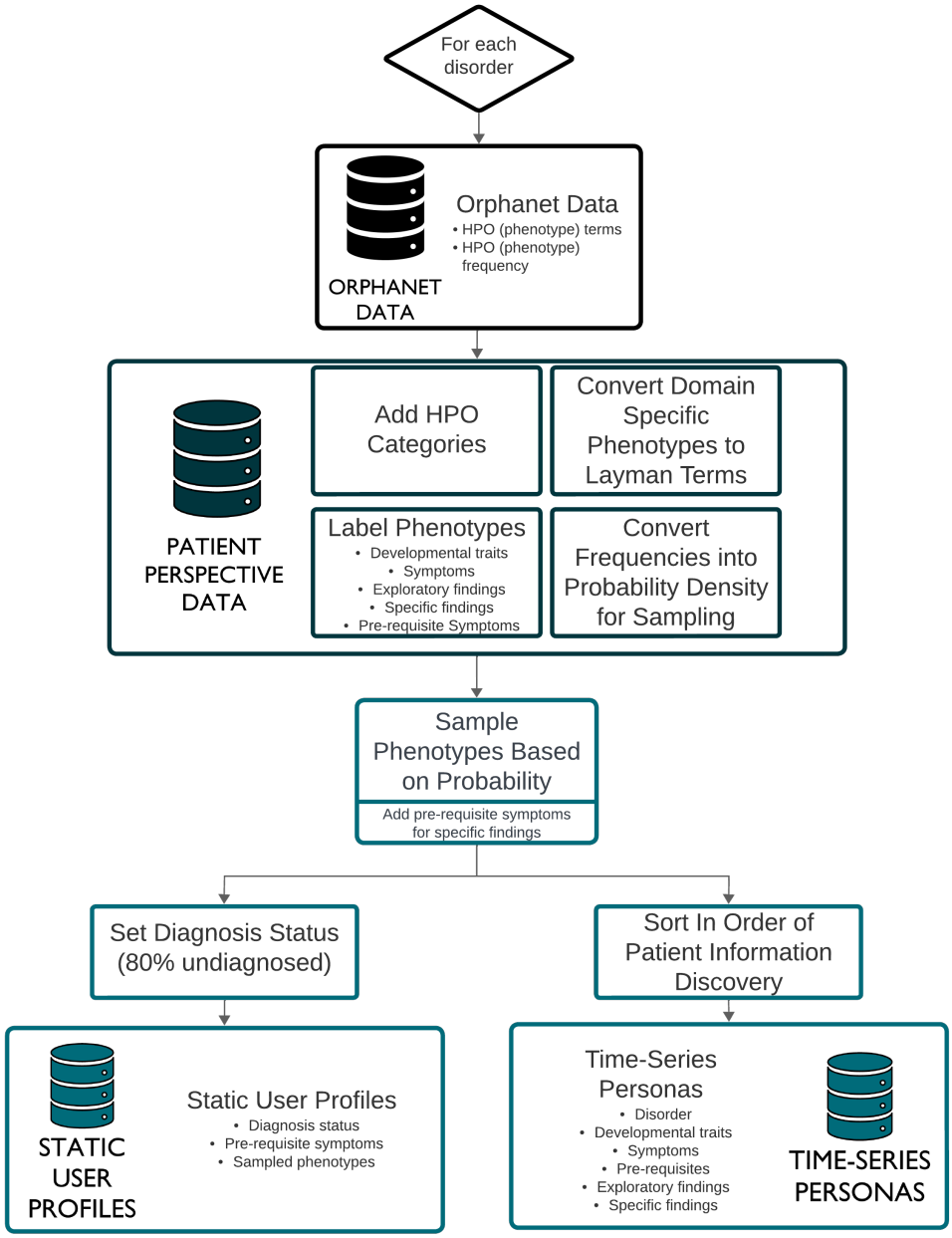


Figure 6.2: Data generation processes for the Static User Profile Database and the Time-Series Persona Dataset

create the Patient Perspective Dataset to act as a base from which we can generate as many patient profiles we choose without additional work, so it is important to identify a knowledge base that enables this.

Orphanet has endeavoured to gather and improve knowledge on rare diseases since 1997, and as such has created several knowledge bases which provide insight into many different aspects of rare diseases. One knowledge base that Orphanet curated is their rare diseases and phenotypes knowledge base [181]. This contains phenotypes as well as the corresponding frequency that each phenotype that occurs within each given disorder, i.e. the percentage of patients with a given phenotype for each disorder. In addition, it consists of standardised clinical terminology, is widely used for rare disease research, is frequently updated, multi-lingual, open source, and spans thousands of rare disorders. Moreover, it enables reproducibility and ensures that, if used in future development, it remains up to date and this process can be expanded to curate large comprehensive datasets. Therefore, Orphanet's knowledge base provides a suitable starting point to generate the Patient Perspective Dataset which provides the base from which we generate both the Time-Series Persona Dataset and the Static User Profile Database to populate Peer Matching. Below we first describe the process used to create the Patient Perspective Dataset, then we present the remaining processes for the Static User Profile Database as well as the Time-Series Persona Dataset.

Now that we have chosen the clinical data to base our patient profiles on, we need to alter this data to represent the perspective of rare disease patients. This process is illustrated in Figure 6.2, which shows that there are two main steps to create the Static User Profile Database and the Time-Series Persona Dataset. Firstly, we create the Patient Perspective Dataset - we enrich the Orphanet data by augmenting the phenotype information for each disease with: HPO categories, layman terms, phenotype discovery group (i.e., development traits, symptoms, exploratory clinical findings, specific clinical findings), and probability of occurrence. Secondly, we dynamically generate a range of varied patient profiles from the Patient Perspective Data, using the probability of occurrence and some perturbation noise, each patient profile samples a proportional amount of phenotypes. After this, two separate processes follow. For the Static User Profile Database, a diagnosis status and a name are assigned to the patient profile to create a user. For the Time-Series Persona Dataset, the phenotypes are sequenced based on their discovery group and divided into the three informational rounds of the

simulation. We describe these steps in more detail in the following sections. To aid in the understanding of each stage within this process, an example is provided for the phenotype *gastroesophageal reflux*, which is associated with hEDS throughout this section.

6.2.1 Identifying Patient Terminology for Phenotypes

For each phenotype associated with the 19 rare disorders, the knowledge base used standardised clinical terminology, namely Human Phenotype Ontology (HPO) terms [182]. HPO not only provides a standardised list of clinical terminology, but it also has synonyms and definitions and classifies phenotypes into categories of the human body (i.e., organ systems and parts of the body). All of this information is accessible on their website[4].

Using HPO's synonyms and definitions, we created initial patient terms. In particular, where synonyms were deemed to be patient- or non-expert-friendly, these were chosen. If there were no synonyms or only expert synonyms, the definitions were checked for short phrases which could be considered synonymous terms in themselves. Phenotypes that did not have clear patient-friendly terms from HPO were researched and discussed among two to four non-experts (people who were not healthcare professionals) until a term was unanimously agreed upon.

Once all terms were finalised by the non-experts, an experienced healthcare professional was shown the original HPO terms as well as the non-expert terms that we created for the three main disorders. They then checked that the patient phenotypes matched the original HPO term. The non-expert terms were then updated according to the suggestions made by the healthcare professional.

Example of Gastroesophageal Reflux:

The HPO website lists several synonyms (*acid reflux*, *acid reflux disease*, *GERD*, *gastroesophageal reflux*, *gastro-oesophageal reflux*, *gastroesophageal reflux disease*, *heartburn*) for *gastroesophageal reflux*. *Heartburn* was deemed to be a clear non-expert-friendly term by the author, so this term was chosen as the initial patient term for *gastroesophageal reflux*. Upon the healthcare professional's review of the initial terms, the term *heartburn* did not require revision. We will now refer to *gastroesophageal reflux* as *heartburn* for all further examples.

[4] hpo.jax.org/

6.2.2 Labelling Phenotypes with their Discovery Group and HPO Category

Now that we have finalised our layman phenotype terms, let us consider the order of discovery of the different types of patient data we have and group them accordingly. This provides the crucial sequencing information which we later use to create the Time-Series Persona Dataset. We can consider each phenotype to either be symptoms (i.e., patient observable) or clinical findings (i.e., not observable by the patient). Since patients would clearly identify phenotypes they can observe themselves before those which require a clinical investigation, these phenotypes should come first. Some symptoms, such as *feeding difficulties in infancy*, would be observable from birth or a very young age. Therefore, it is important to distinguish these phenotypes, so the first two discovery groups are: symptoms and developmental traits.

In addition, we can consider clinical findings to have two main types: exploratory clinical findings and specific clinical findings. We define exploratory clinical findings to be traits that are likely to be identified from routine investigations (e.g., blood tests or a standard physical examination). As such, many of these findings will be identified during or shortly after a patient visits their primary care physician.

We define specific clinical findings as traits which are likely to only be identified from specific investigations. It would be unrealistic for specific investigations to be conducted without the presence of symptoms to prompt these investigations. For example, the probability of discovering an *abnormal myocardium morphology* (abnormal heart wall muscle) is increased with the presence of cardiovascular symptoms, such as chest pain. Therefore, for each specific finding, we must ensure that the pre-requisite symptoms that are needed to prompt the necessary investigations are also sampled along with the specific finding.

First, we need to establish which symptoms should be considered pre-requisites for a given specific finding. To do this, we augment each phenotype in our data with its HPO category (i.e., organ systems and other physiological categories) as gathered from HPO's website. Then, for each sampled specific clinical finding, we additionally sample pre-requisite symptoms with the same HPO category as the finding. We generate one to two pre-requisite symptoms for each specific clinical finding using a random number generator to add perturbation noise. For example, if *abnormal myocardium morphology* was sampled, cardiovascular symptoms such as *heart palpitations* would also be sampled.

In addition, some phenotypes, such as anaemia, may be considered conditions in of themselves. As such, there would naturally be symptoms associated with them, however, this was not present in the data. Since a clinician will know these conditions well, by denoting the presence of the condition, the associated symptoms may be implied. However, a patient will not necessarily recognise these associated symptoms. Therefore, we also add key symptoms associated with conditions that are present in the dataset. The underlying symptoms of phenotypes of this nature were identified by a healthcare professional and as such these symptoms were added and categorised into a new discovery group, pre-requisites. This ensured that they were only sampled if they are a pre-requisite of a phenotype that has been sampled, and otherwise cannot be sampled.
Example of Gastroesophageal Reflux (Heartburn):
Clearly, heartburn is a patient-observable phenotype and is not necessarily present at birth, so the discovery group for heartburn would be symptoms (and as such does not require pre-requisite phenotypes). In addition, the hEDS page on HPO's website[5] categorises heartburn as *Digestive System*. So the discovery group would be *symptoms* and the HPO category would be *Digestive System*.

Upon sampling of a specific clinical finding relating to the digestive system, such as *Gastrointestinal dysmotility* (abnormal contractions in the intestine), pre-requisite phenotypes relating to the digestive system are needed. So, *heartburn* is one of many phenotypes relating to the digestive system that may be sampled as a pre-requisite for *Gastrointestinal dysmotility*.

6.2.3 Generating Phenotypes Based on their Frequency

To generate a realistic range of phenotypes for each patient profile for a given disease we transformed unstructured frequency values from Orphanet's ontology into a probability density. The Orphanet ontology consisted of categorical values stored as strings (namely, always present: 100 %, very frequent: 99%-80% frequent: 79%-30%, occasional: 29%-5%, rare: 4%-1%, excluded: 0%). For values in a range, we took the mid-point of the percentage values and converted these to numbers between 0 and 1 (see Table 6.1).

To generate unique patient profiles, we take a sample of phenotypes for a given disorder by randomly sampling from their defined frequency distribution using Numpy's

[5]https://hpo.jax.org/app/browse/disease/ORPHA:285

Table 6.1: Probability values for the categorical percentages

Categorical Value	Frequency Range	Probability
Always Present	100%	1
Very Frequent	99%–80%	0.895
Frequent	79%–30%	0.545
Occasional	29%–5%	0.170
Rare	4%–1%	0.025
Excluded	0%	0

random choice function[6]. Since the phenotypes in each discovery group are generated separately, we need to normalise our frequencies for each discovery group (this is a requirement of the sampling function). Since we sample within each of the discovery groups, we lose the distribution of each group, so we set the number of phenotypes sampled to be representative of the weight of the discovery group's frequency. In addition, the proportion of phenotypes for each discovery group will vary from patient to patient, so we add some perturbation noise to the number of phenotypes sampled within the discovery group.

Now, let us formalise this as an equation for clarity. Given a disorder there is a set D of phenotypes for that disorder and a collection G_1, G_2, G_3, G_4, G_5 of the discovery groups for the disorder (subsets of phenotypes consisting of *developmental traits*, *symptoms*, *clinical findings*, *specific findings* and *pre-requisite symptoms*) that have the properties:

1. $G_k \subset D$ – a discovery group for a disorder only includes phenotypes for that disorder
2. $D = \bigcup(G_k)$ – every phenotype for the disorder is in a discovery group
3. $G_k \cap G_l = \emptyset$ for $k \neq l$ – no phenotype in more than one discovery group

So, we define the weight, w, of the discovery group as:

$$w = \frac{\sum\limits_{i \in G_k} F_i}{\sum\limits_{i \in D} F_i}$$

where F_i is the frequency of phenotype i. Therefore, given the total number of desired phenotypes per patient profile T, the size of phenotypes sampled within a

[6] numpy.org/doc/stable/reference/random/generated/numpy.random.choice.html

specific discovery group is $T * w$ with some perturbation noise and then rounded to the nearest whole number.

Example of Sampling Phenotypes:

Due to the complexity of the data, let us consider a simplified example for this process, rather than relating to the precise data. Consider an example where we generate data with 10 phenotypes per patient profile for a disorder with a total number of 7 phenotypes, the probabilities of which are [0.3, 0.1, 0.2, 0.6, 0.3, 0.6, 0.8]. Now, let us calculate the number of phenotypes from the *symptoms* discovery group, where the probabilities of each symptom are [0.3, 0.1, 0.2]. Using the formula above, we can calculate that the number of symptoms to be sampled is:

$$10 * \frac{0.3 + 0.1 + 0.2}{0.3 + 0.1 + 0.2 + 0.6 + 0.3 + 0.6 + 0.8} = 2.$$

Therefore, we have a method of creating a set of phenotypes for a rare disease patient, which provides the basis for our patient profiles.

6.2.4 Creating the Static User Profile Database for Peer Matching

Using the method described above, we sample patient personas from the Patient Perspective Dataset. Next, we augment the data to form the Static User Profile Database for Peer Matching. First, we set a diagnostic status for users. Since this is intended as a pre-diagnostic application, we'll assume the majority of users are undiagnosed, however, this system is not limited to pre-diagnosis, so some people with diagnoses will be on the system. Therefore, we populate 80% of the users in the Static User Profile Database to be undiagnosed, the remaining users' diagnostic status displays the sampled disorder name. In a real-world context, the database for Peer Matching would consist of the profiles of the users of the system. Therefore, in order to imply that the profiles shown would be based on those using the system, we include randomly generated first names for each of the profiles. Therefore, the Static User Profile Database portrays a patient profile with a name, diagnosis status and a list of phenotypes.

6.2.5 Creating the Time-Series Persona Dataset

In this section, we outline our process to create the Time-Series Persona Dataset. In particular, we utilise the first round of the data generation process above as a basis, we then curate our Time-Series Persona Dataset by dynamically generating a range of varied patient personas which we augment with sequencing information. In particular, using the probability occurrence and some perturbation noise, we create a static patient profile from the Patient Perspective Dataset by sampling a proportional amount of phenotypes. We then use the discovery groups to sequence the phenotypes based on the patient information discovery of these different types of phenotypes. These steps are described in more detail in Section 6.2.5.1.

Now, let us use the Patient Perspective Dataset that we generated above to provide a basis from which we can generate the Time-Series Persona Dataset. Given that the Patient Perspective Dataset was based on a rigorous knowledge base, it provides a realistic static basis for the Time-Series Persona Dataset. However, to simulate the informational journey of a rare disease patient, we need to augment the data with critical sequencing information which is representative of a patient's discovery of phenotypes.

6.2.5.1 Adding Sequencing Information

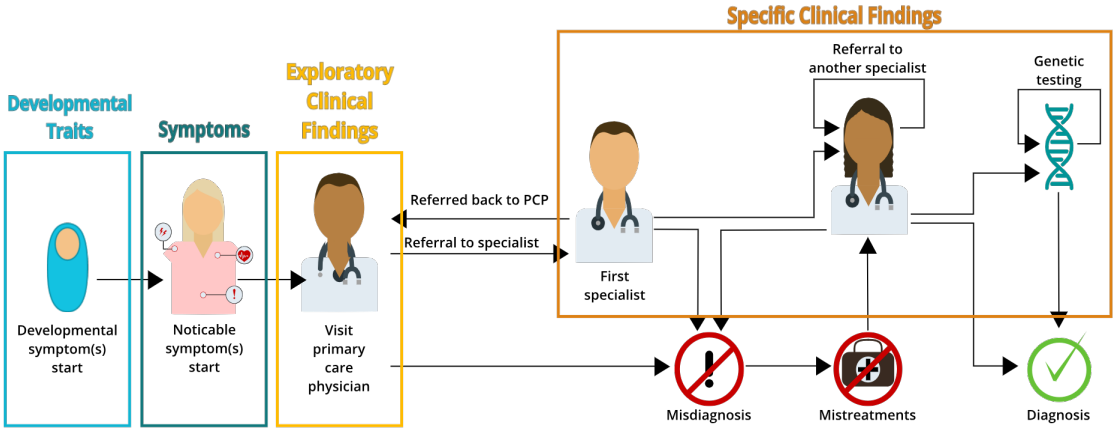


Figure 6.3: Information discovery during the diagnostic odyssey of a rare disease patient

First, we need to recreate the temporal aspect of information discovery from a patient's perspective. In order to do this, let us relate the phenotypes from our data to the diagnostic odyssey diagram that we created in Chapter 2. In particular, as shown

in Figure 6.3 we can augment the patient journey diagram with the discovery groups, showing the order in which clinical discoveries are made and observed by the patient. The different types of data which we divided into the following discovery groups: developmental traits; symptoms; exploratory clinical findings; specific clinical findings; and clinician-identified pre-requisite symptoms. This grouping can be utilised to provide the synthetic data with critical sequencing information to facilitate the revelation of phenotypes as each discovery group is perceived by the patient for each of the three information discovery rounds of the simulation.

Patients discover the different types of phenotypes at different rounds of their diagnostic journey. Firstly, phenotypes which are observable by patients (i.e., symptoms, developmental traits and pre-requisite symptoms) would clearly be discovered first. Since developmental traits are present from an early age, it follows that these phenotypes should occur first, followed by other non-developmental symptoms (the clinician-identified pre-requisite symptoms should also be shown at the same time as the other symptoms).

Secondly, once the patient seeks medical help, clinicians will start providing them with more information from tests or physical examinations. Routine investigations are often made on the first few visits, so exploratory findings will be identified first. Specific findings will occur towards the end of the diagnostic odyssey, since these findings require specific investigations based on the phenotypes that have been observed thus far. Therefore, to ensure that the Time-Series Persona Dataset is representative of a rare disease odyssey, we order the synthetic data so that developmental traits come first, followed by symptoms (including pre-requisites), exploratory findings, and finally specific findings.

Now that we have a basic order, we need to divide the phenotypes among the three information discovery rounds. To do this, we distribute the sorted phenotypes equally into three parts, each of which provides the list of phenotypes for the information discovery round. This ensured that the simulation rounds were consistent in the amount of information that was revealed.

Table 6.2 shows an example of the final generated Time-Series Persona Dataset for each of the three conditions included in the laboratory study. Now that we have our data, we must verify its suitability for our simulation task. In particular, we need to assess how realistic the Time-Series Persona Dataset is in comparison to real patient cases. This verifies whether the dataset is sufficient for the simulation part of our laboratory study.

Table 6.2: Time-Series Persona Dataset example showing each of the three conditions evaluated in the study

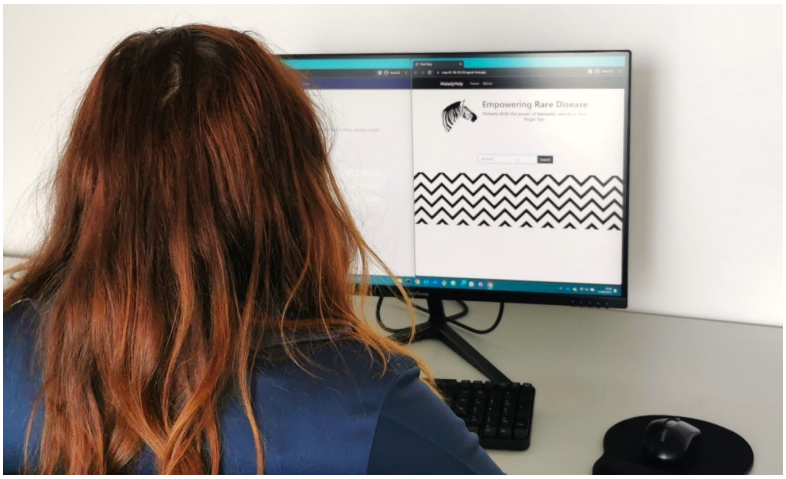
Condition	Round 1 & 2	Round 3	Round 4
Hypermobile Ehlers Danlos Syndrome (hEDS)	Sleep disturbance, Joint dislocation, Stretchy skin, Elbow dislocation, Muscle pain	Fatigue, Heartburn, Thin skin, Depressivity, Vertigo	Nausea and vomiting, Soft skin, Constipation, Gastrointestinal dysmotility, Extra bones in the cranium
Fabry Disease	Small dark-red spot, Poor appetite and weight loss, Lack of sweating, Nausea and vomiting	Joint pain, Blood in urine, Thickened skin, Vision loss	Kidney damage/Kidney disease, Cataract, Optic atrophy, Corneal dystrophy
Gaucher Disease	Squint, Corneal opacity, Joint pain, Abdominal pain, Recurrent fractures	Tremor, Falling, Delayed skeletal maturation, Recurrent fractures	Osteopenia: Decreased bone density, Enlarged liver, Cranial nerve paralysis, Enlarged spleen

6.2.5.2 Assessing Realism of the Time-Series Persona Dataset

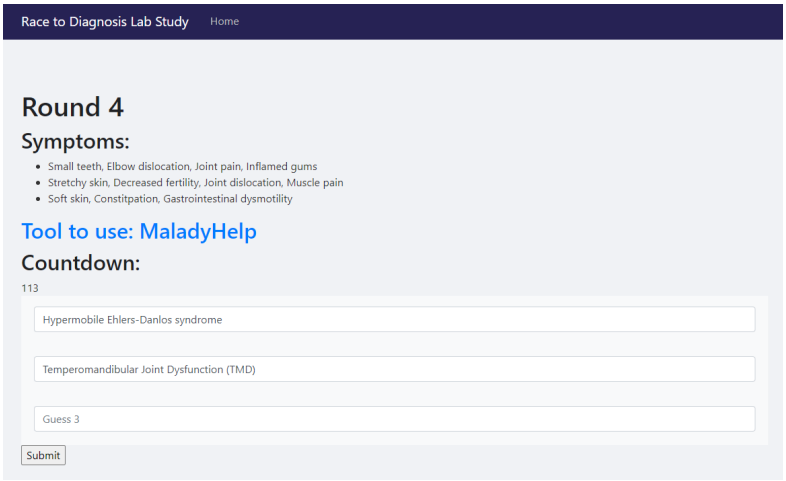
A healthcare practitioner with experience in primary care and rare diseases tested the Time-Series Persona Dataset by participating in a simulation task using Google and the underlying persona was revealed at the end. They were asked about the realism of the task based on their clinical experience. They expressed that the patient persona made sense and was similar to their clinical experiences.

However, they discussed that when making diagnostic decisions, they would typically inquire about the family history of the patient; in addition, they would inquire more about the duration of symptoms (i.e., whether they are chronic or acute). Given that patients do not typically consider their family history until a genetic condition is suspected [183,184][7], we did not consider this data feature to be necessary for the study. Moreover, we also deemed it could potentially lead participants to actively pursue genetic causes who would not have otherwise considered a genetic condition.

[7]These books show i) doctors have to explain or justify questions about family history and ii) most patients know the importance of healthy lifestyle, but not necessarily family history.



(a) Participant performing study tasks



(b) Screenshot of Simulation Study Persona

Figure 6.4: Simulation study and information support tools

The duration of symptoms, however, would be a relevant addition to the simulation, but as we did not have this information in the dataset, we could not add this aspect. For future studies, we could explore methods to curate additional data features, such as the duration of symptoms. However, we deemed the current data to be sufficient for a preliminary evaluation study.

6.3 Lab Study Website Implementation

Now we have a suitable process to generate the Time-Series Persona Dataset, we need to find a suitable means to present the data to participants so that they can complete the simulation task. In addition, we need to collect participant responses for each round, stored with the corresponding prototype used and phenotypes shown. To facilitate this, we build a website to display unique personas from the Time-Series Persona Dataset. In particular, we need the website to: i) assign participants a unique persona and prototype to use for each of the three simulation tasks; ii) display the time remaining for each round; iii) automatically move to the next round if time runs out; and iv) collect participant responses for each simulation and each round and save as a CSV file. This section describes the implementation of this website.

Souradeep Ghosh, an MSc student supervised by Emily Nielsen, created an initial prototype of the website which presented phenotypes for a given participant by displaying the patient persona data stored in a CSV file and facilitated a timer countdown as well as redirecting the participant to the next page. However, this prototype did not allow multiple participants to use it simultaneously, it was not able to store participants' guesses and only presented symptoms of the current round (not previous rounds). Emily Nielsen then made the changes, described below, to facilitate these remaining functionalities.

The facilitation of simultaneous use by multiple participants requires the use of session variables stored in SQL to track the participant, number of simulation tasks completed, number of rounds completed, and responses. These variables are initiated when the participant first opens the website and are updated at each of the information discovery rounds and the end of each simulation. This ensures that for the current simulation, the corresponding assigned tool is linked on the page, as shown in Figure 6.4b, and that only the phenotypes for the current and previous rounds (within the current simulation task) are shown. Since this study utilises a within-groups design, we wanted to minimise the learning effects of the simulation task on the data obtained from this study, so we permuted the order that the tools were assigned to participants so that all six possible permutations of tool assignments were distributed equally among participants. Upon completion of the final simulation task, participants are redirected to a questionnaire to gather data on their experience of using each of the three prototypes.

As the prototypes assigned to the participants are also websites, they will need to switch between the assigned tool and the lab study website. As such, multiple monitors or a larger screen present an advantage to allow easier switching between screens. Therefore, to ensure equal footing in terms of screen size and environment, we conduct this study in a computer room where the office equipment, in particular the monitor, is uniform across participants. In future, to improve usability, the prototypes could be integrated within the lab study website.

The website was then tested, firstly for its ability to handle multiple simultaneous users (i.e., session variables remained distinct for the assigned patient profile, current assigned tool, their round of progress for simulation tasks and their rounds). The website was also penetration tested by Emily Nielsen and Connor Clarkson (another PhD researcher with experience in security testing) to check for security issues (e.g., SQL injection, cross-site scripting) and it was deemed secure for the study.

6.4 Prototype Implementation

Chapter 4 identified three potential avenues for design: information management; clinician-patient mediation and support; and providing access to empathetic support. Puzzle facilitated this through a symptom logger; clinician-patient communication support; and patient matching (connecting users to people experiencing the same symptoms as them). In this chapter, we build upon this work by implementing a patient support tool to facilitate further evaluation and comparison with other tools.

Many publicly available applications offer symptom tracking/logging functionalities[8] or clinician-patient technologies[9]. In contrast to the numerous works in symptom loggers and clinician-patient technologies [185–194], there exists little research which explores patient matching. As we discussed in Chapter 3, many of these works propose patient matching to aid clinicians with the diagnosis of rare conditions [195–198]. The only paper included in our literature review which utilises a patient-facing approach [87] aimed to facilitate information discovery and did not focus on facilitating matches to aid empathetic support. Therefore, in this chapter we explore the potential for supportive

[8]Symptom tracking/logging:
flaredown.com/ www.sympleapp.com/ careclinic.io/ bearable.app/ www.skinvision.com/

[9]Clinician-patient communication:
theclinician.com/remote-monitoring www.treatmyocd.com/ www.allyhealth.net curoflow.se/ www.nhsapp.service.nhs.uk

technologies to provide access to empathetic support through peer matching, thus we present Peer Matching, our prototype to pair users with similar symptoms.

We also require other prototypes to compare Peer Matching against to allow for later evaluation. In particular, since Google is commonly used to look for health information [199], we want to compare Peer Matching against Google. In addition, we designed MaladyHelp, a rare disease-specific search engine, which in addition to promoting information discovery, also provides links to patient support groups. As such, it aims to support rare disease patients' needs both for information-seeking and for empathetic social support. Therefore, this chapter describes our process to create three patient-facing supportive technologies for undiagnosed rare disease patients which will be evaluated in Chapter 7.

6.4.1 Symptom-Based Peer Matching Tool

To populate the recommender system, we used the Static User Profile Database consisting of synthetic patient profiles generated in Section 6.2, totalling 665 profiles (35 per disorder) and spanning the 19 different rare disorders. Now we have our data, we need to develop an algorithm to match patients based on their shared symptoms.

6.4.1.1 Peer Matching Algorithm Development

The development of our algorithm was inspired by the work of Shen et al. [200], who found that the Tanimoto (also known as Jaccard) similarity score performed best for a clinician-facing collaborative filtering task to group patients based on their phenotypes. Namely, for each pair of patients i and j, the Jaccard similarity between them is:

$$J(i,j) = \frac{|Phe_i \cap Phe_j|}{|Phe_i| + |Phe_j| - |Phe_i \cap Phe_j|},$$

where Phe_i is the total set of phenotypes for a patient i. To identify phenotypes in common, we use fuzzy string matching to identify the closest string match to our dictionary of patient terms (i.e., the terms identified in Section 6.2.1). Our implementation of this part of the approach described by Shen et al. facilitated the identification of a user's matches showing profiles sorted from the most to the least shared symptoms.

Shen et al. then used this similarity score to find the K Nearest Neighbours (KNN) and Threshold Patient Neighbours (TPN). However, during our testing of the one-to-one matches, we noticed that even the top matches would typically only have a few

symptoms in common. As such, we deemed that our patient profiles were too varied and the population too small to accommodate reliable group matches. While we could overcome this issue by generating more patient profiles, we wanted to ensure that the Static User Profile Database remained small since rare diseases are not prevalent enough to provide a large user base.

Furthermore, group matches may not facilitate the flexibility needed to provide the most empathetic support for conditions which are especially rare. In particular, since some conditions may only affect a couple of individuals globally, matching them to a group which does not relate to them may result in them matching (poorly) to a group of people with a specific condition. This may make them feel that they do not belong if their symptoms are different to the group. In contrast, by matching users to other patients with the most shared symptoms, their matches may still be able to support them with that aspect of their condition. This is evidenced by the fact that patient support groups for people with hypermobile Ehlers-Danlos Syndrome often also include people with hypermobility spectrum disorder (previously known as joint hypermobility syndrome)[10], as their symptoms are similar, and as such they can offer support. Therefore, we use the Jaccard similarity score to sort a list of profiles, so that the most similar matches appear first.

6.4.1.2 Peer Matching Interface and Functionalities

Now that we have our algorithm for Peer Matching which identifies the closest pairings of individuals based on their symptoms, we need to create an interface for it and decide how the user interacts with the system. We created a Python-based web application using Flask to facilitate users' interactions with the system. This application needed to be simple enough for participants to use, however, optimising user-friendly features was not a key goal at this stage, since we first needed to evaluate whether the algorithm and functionalities were suitable. The opening page on Peer Matching (Fig. 6.5a) intended to loosely imitate a login page to present the idea that the profiles were made by users of the system.

[10]Facebook groups for individuals affected by hEDS and hypermobility spectrum disorder:
Hypermobile EDS / Hypermobility Spectrum Disorder
Hypermobile Ehlers Danlos Syndrome & Hypermobility Spectrum Disorder
Parents of children with Ehlers-Danlos and Hypermobility Syndrome
Ehlers-Danlos syndrome & hypermobility syndrome looking for friends

Upon submitting their name, symptoms and the number of matches they want to find, users were presented with the next page, as illustrated in Figure 6.5b. This presents the user with a list of patient profiles (as described in Section 6.2) consisting of first names, their state of diagnosis (i.e., undiagnosed or the name of their diagnosis), and symptoms. In the list of patient profiles presented, the symptoms in common with the user were shown in bold to improve readability. From this page, users can also add additional symptoms and update their list of matches accordingly. By default, the best matches (i.e., the people with the most shared symptoms) will be returned, additionally, the user can choose to only see people who have a diagnosis (this option is shown on both pages).

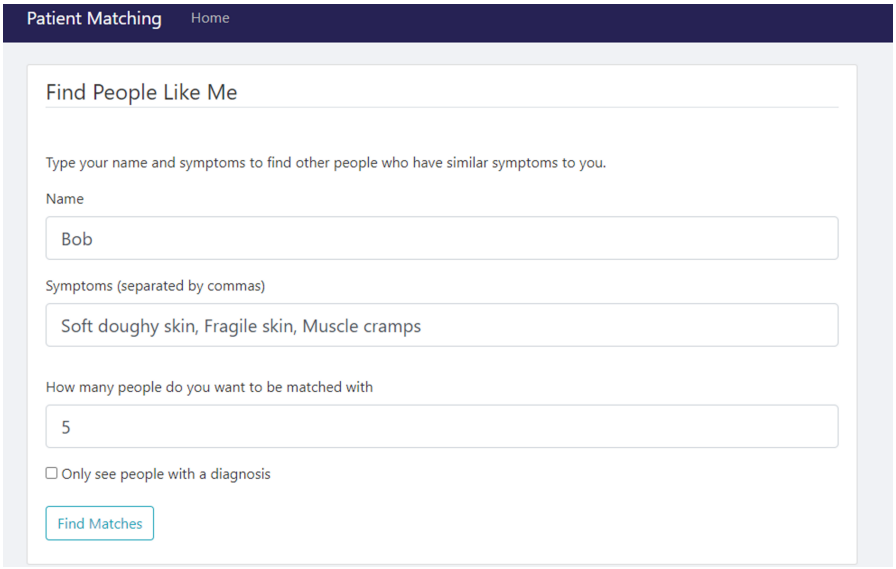
6.4.2 Search Engines

Next, let us describe our implementation of the search engines. Since search engines, such as Google, are commonly used by patients, we consider Google to be our baseline for evaluating new prototypes. In addition, inspired by the work of Find Zebra [201] which highlights the potential of a rare disease-specific search engine for clinician-facing technology, we consider its potential within patient-facing approaches and create MaladyHelp, a patient-facing rare disease-specific search engine. To aid comparison with MaladyHelp, Google Custom is used to restrict the database (see Section 6.4.2.3). Both MaladyHelp and Google Custom require a corpus of data targeted to disease diagnosis.

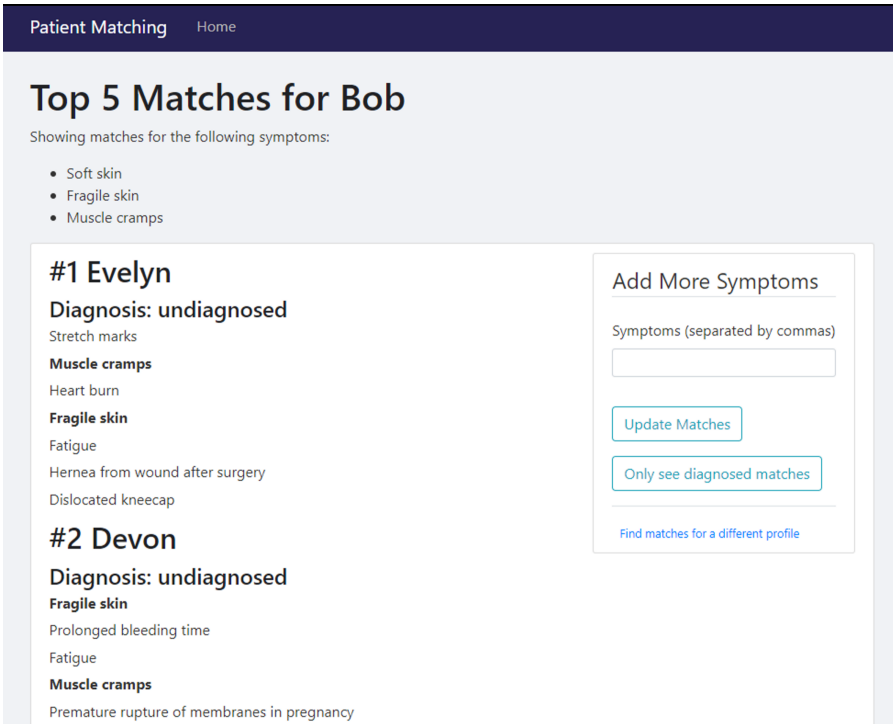
6.4.2.1 Retrieval Corpora Data

It was essential to have a dataset that reflected a real-life corpus relative to the prevalence of a disease for the conditions included. Therefore, we curated a corpus comprising 150 long documents to fine-tune the pre-trained model for MaladyHelp and act as the search corpora for the lab studies. These documents were curated by scraping abstracts or essential elements from the following publicly available websites PubMed[11], Healthline[12], NHS Inform[13], Wiley Online[14], and Wikipedia[15]. The first four websites were chosen due to their reliability as sources of information. Whilst Wikipedia articles can vary in

[11] pubmed.ncbi.nlm.nih.gov/
[12] www.healthline.com/
[13] www.nhsinform.scot/
[14] onlinelibrary.wiley.com/
[15] www.wikipedia.org/



(a) User input of Peer Matching



(b) Resulting patient matches from Peer Matching

Figure 6.5: Peer Matching website screenshots

reliability, it contained articles on rare conditions where there were limited resources elsewhere, so Wikipedia was also included in the websites that we scraped from.

The number of documents retrieved for each of the additional rare and common diseases described above was dependent on the number of Google search results. We calculated the weighting of a given disease, i, by dividing the number of its results on Google, n, by the sum of results for all diseases, D.

$$\frac{n_i}{\sum_{j=0}^{D} n_j}$$

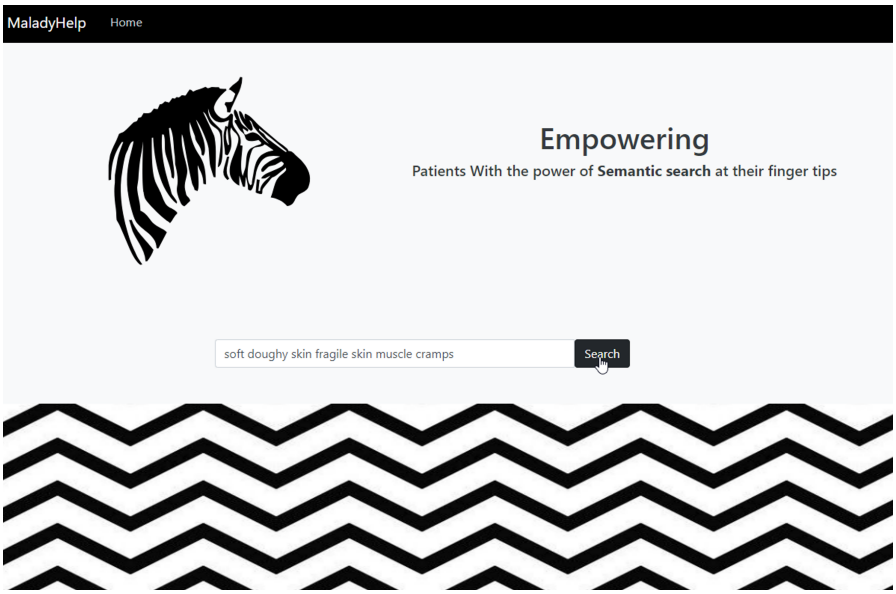
However, this resulted in a highly disproportionate data set with Arthritis carrying 54% of weight. To provide a reflective but greater balance to the dataset, for a weight W, the number of documents retrieved was defined by $5 + 200W$ except those exceeding 20 which were capped. Table 6.3 shows the distribution of the documents and our calculation of the disorder's weight in the dataset. This provided between 5 and 20 documents on each condition, with the higher numbers of documents for common conditions and the lowest number for rare conditions.

6.4.2.2 MaladyHelp Rare Disease Search Engine

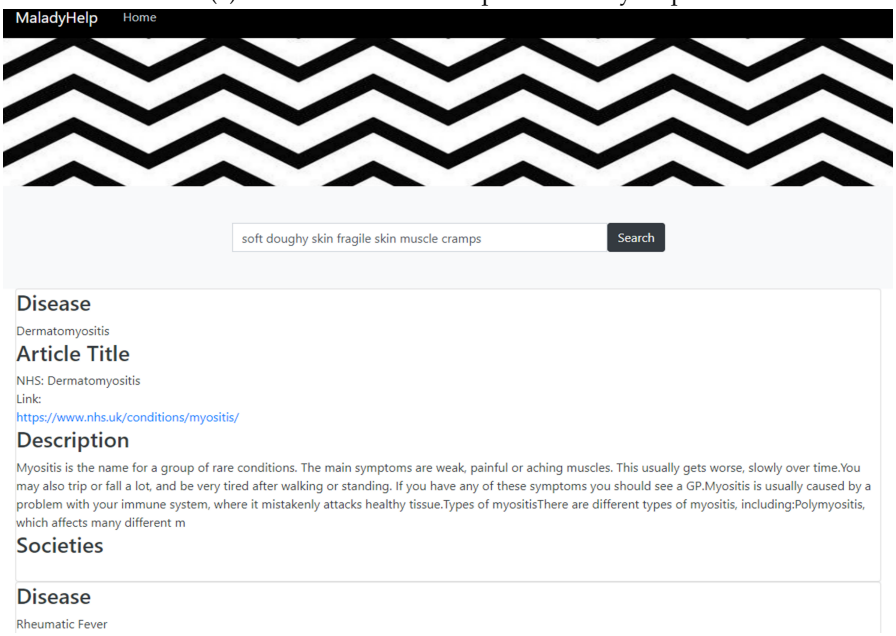
We developed MaladyHelp, a search engine with a focus on rare disease information retrieval. While Google performs searches over a large general data set, systems such as Find Zebra provide tailored data sets focused on rare diseases to enable the querying and information-seeking tasks for clinicians exploring rare diseases. MaladyHelp has been designed to adopt a similar approach to provide a patient-facing search engine.

By using our own specialist search engine, rather than Find Zebra, we could control the corpus of data across our study prototypes. We build on Find Zebra's work, utilising modern approaches to information retrieval tasks. In particular, Find Zebra utilises a frequency-based similarity metric TF-IDF (Term Frequency - Inverse Document Frequency) whereas MaladyHelp makes use of a pre-trained BERT transformer model which is fine-tuned on the rare disease corpora data described above (Section 6.4.2.1).

The interface for MaladyHelp was also a web application that was implemented using Flask. The opening page (Fig. 6.6a) showed a zebra and a search bar much like any other search engine. The image of a zebra has been adopted as a symbol of rare diseases in protest of the mentality *'if you hear hoofbeats think horse, not zebra'*, that is to think of common conditions before rare ones. Upon the input of search terms, the search results, as



(a) Home screen search input of MaladyHelp



(b) Search results from MaladyHelp

Figure 6.6: MaladyHelp screenshots

Table 6.3: Example of Document Distribution

Disease	Google Results (in millions)	Weight	Documents Retrieved
hEDS	1	<0.01	5
Gaucher	17	0.01	6
Fabry	19	0.01	6
Rheumatic fever	30	0.01	7
Arthritis/ rheumatoid arthritis	1,420	0.54	20
Fibromyalgia	139	0.05	16
Dermatomyositis	6	<0.01	5
Erythromelalgia	0.3	<0.01	5
Myelofibrosis	4	<0.01	5
Leukemia	251	0.10	20
Idiopathic avascular necrosis	0.3	<0.01	5
Hepatic cirrhosis	296	0.11	20
Chronic fatigue syndrome	307	0.12	20
Joint Hypermobility Syndrome	4	<0.01	5
Other EDS subtypes (vEDS & cEDS)	5	<0.01	5

shown in Figure 6.6b, are listed in terms of their disease, the article title, and a description which is scraped from abstracts or summaries. In addition to this, societies, or patient groups for the disease in the results are listed. This could facilitate users' communication with people who have the condition that they suspect they have. Thus, this could provide a source of empathetic support and social methods of information discovery.

6.4.2.3 Google Search Engine

Since we are only evaluating a small subset of data (three target rare diseases plus the additional 16 rare and 8 common disorders and all their associated symptoms), we use Google's Programmable Search Engine[16] to keep the data retrieval task consistent across all three arms of the experiment (Google, MaladyHelp and Peer Symptom Matching

[16] programmablesearchengine.google.com/about/

Tool). Limiting the data Google searches to match that in the MaladyHelp database provides meaningful information retrieval comparison to the other tools used in this study. Without this limitation, Google's large dataset would provide a significant amount of noise for the specific information retrieval task in our study. Therefore, this would prevent a fair evaluation against the other prototypes to be evaluated in our study which have smaller datasets for the purposes of this initial evaluation.

6.5 Chapter Summary

In this chapter, we presented Peer Matching, a novel intervention which matches patients to people who share their symptoms to address the isolation associated with rare diagnosis. To provide a user base for Peer Matching and the patient profiles for the simulation study, we also presented a data generation method for our Patient Perspective Dataset by augmenting a clinical dataset with layman terms and patient information discovery groups. This provides the basis for the Time-Series Persona Dataset (the simulation study personas to mirror the informational journey of a rare disease patient) and the Static User Profile Database (the user base for Peer Matching). In addition, we created two search engines to evaluate Peer Matching against: Google (as a baseline) and MaladyHelp (a search engine designed specifically for rare diseases). Now we want to assess whether these prototypes fulfil the support needs identified in Chapter 2 and Chapter 4. In addition, we examine the suitability of our methodological approach by examining whether it provides sufficient discrimination between the three prototypes.

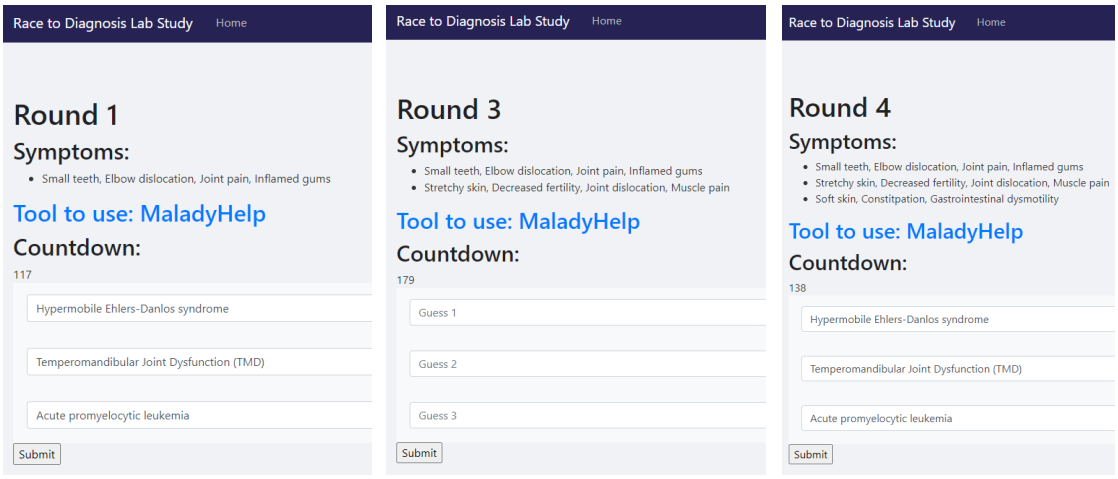
Chapter 7

Empirical Evaluation of Patient-Facing Prototypes

Now that our methodological approach has been established and the required materials have been curated, we can now evaluate our three patient-facing pre-diagnostic prototypes. Using this method, we can evaluate Peer Matching for its positive outcomes on rare disease patients. As such, we can assess the suitability of our methodological approach in terms of the utility of our findings. Therefore, in this chapter, we conduct a mixed-methods empirical evaluation consisting of two parts. The quantitative part, as described in Chapter 5, evaluates the information-seeking capabilities of patient-facing pre-diagnostic prototypes and the qualitative part evaluates their supportive capabilities. For the qualitative part we first ask rare disease patients about their experiences prior to diagnosis and then, utilising an adaptation of the quantitative simulation task so that participants become familiar with the patient-facing pre-diagnostic prototypes, we ask whether the prototype would affect the experiences or hurdles with diagnosis that they describe. Utilising both of these approaches, we are able to perform a preliminary evaluation which comprehensively assesses the positive outcomes of patient-facing pre-diagnostic technology.

7.1 Quantitative Evaluation of Information Seeking Capabilities

To evaluate the information-seeking capabilities of our three prototypes, we utilise the methodological approach that we propose in Chapter 5. In particular, participants were asked to open a website where they were assigned a patient persona and tasked with finding the underlying condition, as early in the simulation as possible. This task was designed to evaluate how well users perform when using the different prototypes for seeking health information to empirically evaluate the utility of each of the prototypes for non-expert health information seeking.



(a) (b) (c)

Figure 7.1: Phenotypes given for each stage of an example simulation task (a) phenotypes for round 1 and 2, (b) phenotypes for round 3, (c) phenotypes for round 4.

Each task consisted of four rounds where participants could input into the study website up to three answers to guess the underlying condition of their assigned persona. For the first two rounds, the same set of initial symptoms was presented, and no additional phenotypes were revealed in the second round to allow time for initial ideation. New phenotypes were then given for subsequent rounds. The website view for the three information discovery rounds (rounds which present new phenotypes) can be seen in Figure 7.1.

Participants are asked to complete three iterations of the simulation task (i.e., three tasks with four rounds to be completed within each task). Each task iteration presented

a new persona that displays phenotypes for another condition with a different prototype to be used (either Google, Peer Matching or MaladyHelp). The order in which the prototypes were assigned was permuted between participants.

7.1.1 Participants

Participants were recruited via email on Swansea University mailing lists and self-identified to be eligible if they didn't consider themselves to be health experts; they were over the age of 18; and were able to provide their own consent. In total, 49 students from the faculty of science and engineering attended the study, which was held in a computer room on university grounds and were awarded Amazon vouchers upon completion of the study. Clearly, these participants represent a highly educated demographic from STEM backgrounds. This approach facilitated recruitment of large numbers of participants, but further studies could use a broader recruitment strategy to improve the generalisability of results.

We conducted this study in a computer room where the office equipment, in particular the monitor, is uniform across participants. This was because participants will need to view the lab study website and use their assigned prototypes throughout the study. As such, they will need to switch between the assigned prototype and the lab study website, so multiple monitors or a larger screen present an advantage to allow splitting or easier switching between screens. Thus, the fixed location was necessary to ensure equal footing in terms of screen size and environment.

7.2 Qualitative Evaluation of Support

Our qualitative evaluation aimed to examine whether each of the prototypes provides the support that rare disease patients need, as characterised in Chapter 4. Clearly, participants must use a prototype to evaluate its potential to meet their wider support needs. However, it is highly unethical for participants to use technology which could potentially hinder their diagnosis and, as a result, their health. Therefore, to provide an activity so that participants could become familiar with and explore their assigned prototype without using their own symptoms, we build upon the simulation task approach in Chapter 5.

The use of the simulation task allows participants to use the prototypes without performing their own health searches, thus preventing impacts on diagnosis. In addition,

given that the simulation task imitates the information-seeking behaviours of rare disease patients, it can promote reflection and comparison of using the assigned prototype with their own experiences. However, the existing task is used to measure participants' performance, so we must adapt the task so that participants become sufficiently familiar with the prototypes for evaluation.

To gather qualitative information on the supportive capabilities, we conduct a three-part interview. However, we cannot ask questions revolving around the design space as it is important that we do not lead participants in our questioning. As such, when interviewing participants about their experience of using the assigned prototype, we can only ask questions about the impact of the prototype on each participant's personal support needs. Therefore, we adopt a three-phase study for the qualitative evaluation of the prototypes. This consists of a prototype interaction simulation task which was preceded and followed by semi-structured interviews.

The pre-interaction interview aimed to firstly highlight the challenges that each specific participant had faced during their diagnosis and any aids and support they employed; and secondly sought to encourage reflection on their diagnostic journey in preparation for the interaction task. This provided a basis for identifying the support needs of each participant, which facilitated the post-interaction interview (without leading the participant) to enquire which of these needs and experiences would be affected by the use of the prototype. By asking about participants' wider experiences first, outside the context of technology, we can evaluate the wider aspects of the design space (i.e., empathetic support needs) that may typically be otherwise ignored within this context.

The prototype interaction task remained similar to the quantitative part of the lab study but used a between-subjects design to allow more time for the interviews. In contrast to the quantitative evaluation which evaluated the accuracy or performance of the prototypes, the main object for this task was for participants to become familiar enough with the prototype to evaluate it effectively. As such, additional time was necessary to facilitate greater familiarity and their guesses were not collected.

Finally, participants were interviewed about their experience of using the prototype and how it would have affected them during their journey towards diagnosis. This part of the study refers to the previous interview to identify whether the prototype would affect each of the support needs and challenges discussed in the first interview.

To gather the qualitative data, the simulation task primarily remained the same as in the quantitative part of the study. The key change in the task was that they were given an increased time of six minutes per round to allow for more familiarity with the prototypes. As a result of this, the additional round for initial ideation was no longer necessary since the time had sufficiently increased, so participants only played three rounds, each of which provided new symptoms and findings as in the informational rounds in the lab study. Their responses were not gathered and scored as they were in the quantitative part of the lab study, but the task remained the same.

Participants then completed post-interaction interviews that reflected on the prototypes and their experiences discussed during the pre-interaction interview. We aimed to understand the suitability of the prototypes throughout patients' experience of diagnosis and gathered the participants' evaluation of the prototypes for the specific needs they had, rather than limiting the reflection to previous experiences with technology. We aimed to do this by encouraging reflection on their positive and negative experiences, thus identifying key hurdles and support needs. By identifying the support needs of a specific individual, we can then ask questions about these needs to evaluate how effectively the prototypes meet each of their personal support needs.

7.2.1 Analysing Responses

The audio from each session was recorded and fully transcribed. These transcripts were analysed thematically using the Braun and Clark method [139] to identify the challenges, desired support and support needs from participants' experiences of diagnosis. In particular, the transcripts were anonymised, and then open coding was performed by two independent coders (Emily Nielsen and Rory Clark, a PhD student) on every segment of text that related to patients' experiences of diagnosis or their attitudes towards health technologies. Once this was finished, codes were discussed, evaluated and modified where relevant. Following this, initial themes were identified and agreed upon by the two coders.

7.2.2 Participants

Sixteen participants were recruited through three different means: an existing pool of participants[1] and advertisements on rare disease social media groups (with admin permission). Participants self-identified to be eligible for one of two groups: the rare disease group if they are diagnosed with or being treated for a disease that is considered rare (according to https://rarediseases.org/rare-diseases/); or the delayed diagnosis group if they have a condition which took longer than one year to diagnose. They also had to be old enough to remember and be involved in their journey towards diagnosis; over the age of 18; and be able to provide their own consent.

7.3 Quantitative Results

A total of 49 participants completed the quantitative part of the study, each completing the simulation task three times, consisting of four rounds and three guesses per round. Therefore, this gives us 1,764 data points for which we calculate the participants' scores using the method defined in Section 5.4.3.

Table 7.1: Statistics on the scores of each of the three prototypes compared with random chance (higher values show greater participant performance)

	Google Custom	MaladyHelp	Peer Matching
Mean	2.20	0.86	7.02
Variance	9.88	4.86	23.61
Coefficient of Variation	1.43	2.57	0.69

We then calculated the mean, variance and coefficient of variation of the participants' scores for each of the three prototypes. Due to the challenging nature of the task, there were numerous participants with a score of 0, so the variance will inevitably increase for better-performing prototypes. A more suitable measure, in this case, is the coefficient of variation since it measures the variance in relation to the mean. Table 7.1 shows the mean, variance, and coefficient of variation for each of the three prototypes. We can see that Peer Matching has the highest performance among users with a mean score of 7.02, whereas the mean score of Google Custom is 2.20 and MaladyHelp is

[1]One participant was recruited after they participated in the Workshops in Chapter 4 and two people were recruited after they participated in the quantitative part of this simulation-based study

0.86. However, the mean random chance score is even lower than that of 0.29, which demonstrates the difficulty of the task.

7.3.1 Non-Parametric Empirical Tests

Based on Table 7.1, it appears that Peer Matching does reduce time to diagnostic insights. To test this, we establish whether (i) the prototypes behave any better than random chance, and (ii) whether the prototype choice makes a difference. As is evident from Figure 7.2 and Figure 7.4, the metric used does not follow a normal distribution, hence we use non-parametric empirical tests.

7.3.1.1 Comparison to Random Chance

First, we calculate a p-value for the null hypothesis that each of the three prototypes does not change the performance when compared to random chance. To do this, we first generated round scores based on random chance. Since there were 23 disorders in the data sets, we generated guesses which had a 1 in 23 chance of being correct for four rounds each consisting of three guesses. We then calculated the score using the same method as the participants. Then we sorted the random chance scores and calculated the mean for every 49 samples. This gave us a sorted list of random chance means, so given n, the number of means where $\mu_{\text{Random Chance}} > \mu_{\text{prototype}}$, we calculated the p-value for each prototype as $p = \frac{n}{\text{total random chance means}}$. Therefore, the p-value for Google Custom is $p = \frac{1041}{20{,}000} = 0.05205$, the p-value for MaladyHelp is $p = \frac{2493}{20{,}000} = 0.12465$ and the p-value for Peer Matching is $p = \frac{9}{20{,}000} = 0.00045$.

7.3.1.2 Permutation Testing

We also performed an empirical analysis using a permutation test. The labels of the assigned prototypes were permuted 20,000 times while retaining the scores collected in our initial results. For each permutation, we then calculated a test statistic. The ANOVA metric is a commonly used test statistic, but given that our data is not normally distributed, this metric is not suitable. Instead, we select a mod-based metric, which provides a more stable metric for non-normal data. Hence, we define our test statistic to be the average difference between the means for all three prototypes

$$T = \frac{|\mu_M - \mu_G| + |\mu_P - \mu_G| + |\mu_P - \mu_M|}{3}.$$

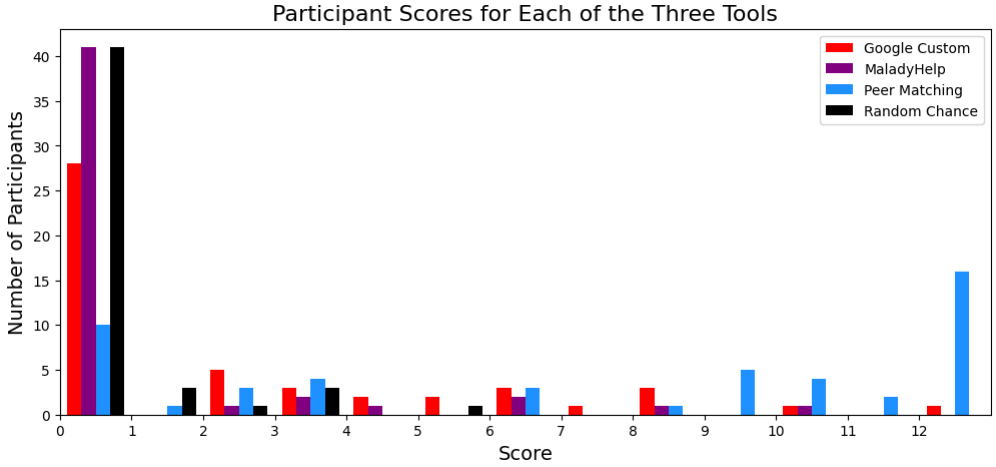


Figure 7.2: Participants scores compared to random chance

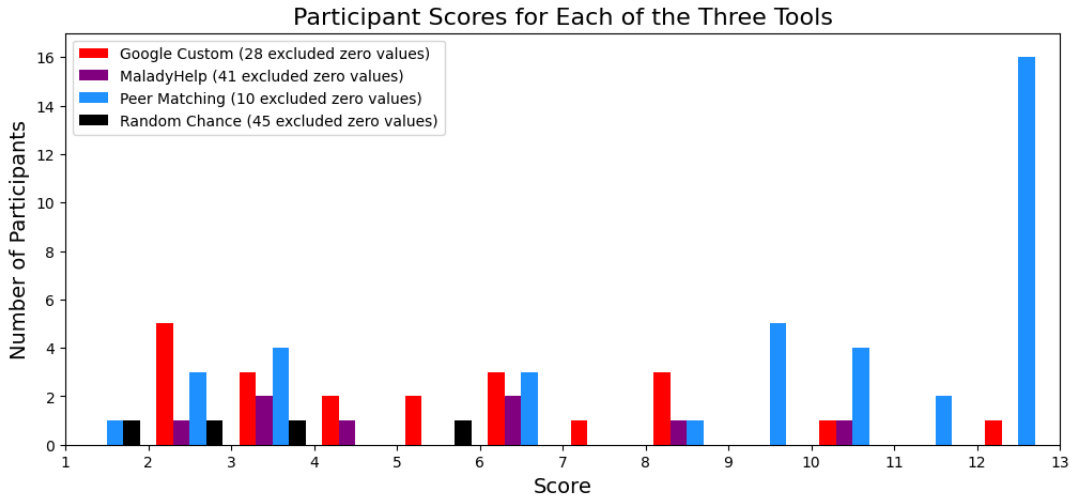


Figure 7.3: Non-zero participant scores

We can calculate the p-value by comparing the test statistic for each permutation to the test statistic for the observed data. Since T_{observed} remained the highest test statistic, we have $p < \frac{1}{20{,}000} = 0.00005$. So, there is a statistically significant difference between the three prototypes.

7.3.2 Graphical Distributions of Participant Scores

Figure 7.2 shows the participant scores for each of the three prototypes. As we can see, the majority of the participants never got the answer correct for the tasks that

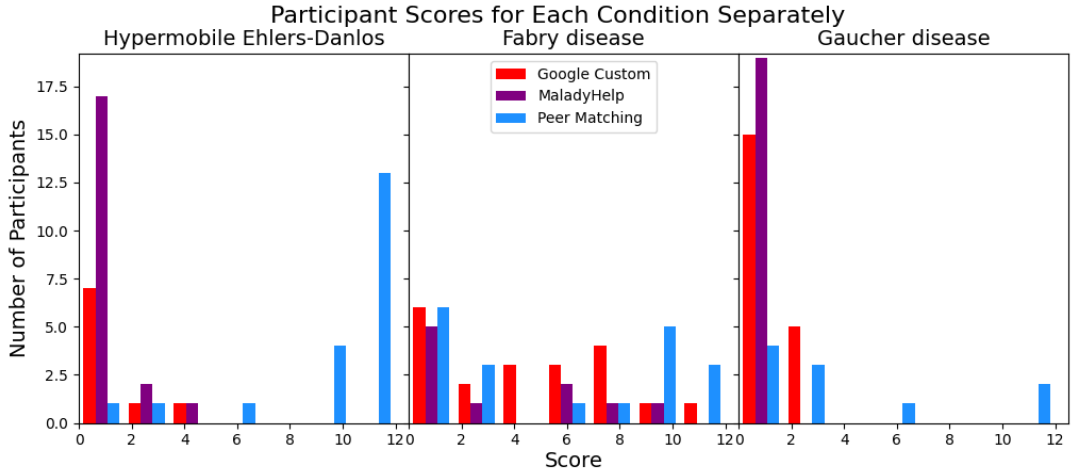


Figure 7.4: Scores for each individual condition

they were assigned MaladyHelp and Google Custom. The very large number of zero scores makes it hard to see the details in Figure 7.2, so Figure 7.3 shows the distribution of participants' non-zero scores. This clearly shows a high number of participants obtaining the maximum score of twelve when using Peer Matching, with a fairly even distribution of scores for Google Custom, whereas MaladyHelp only has four participants scoring higher than three.

So, we can initially see a greater performance from Peer Matching overall. Let us now consider the disease-specific score distributions to explore the performance of each prototype in more detail. Figure 7.4 shows this distribution, and we can see that both MaladyHelp and Google Custom performed significantly better for Fabry disease than for the other two conditions. Gaucher disease was particularly challenging; most participants scored zero and only Peer Matching had participants score more than two points.

Peer Matching also performed significantly better for hEDS than for Fabry disease or Gaucher disease. As such, we must check whether Peer Matching still achieves a statistically significant improvement without hEDS. Therefore, we use the same methods described above to calculate the mean, coefficient of variation and statistical significance. This showed that Peer Matching still performed best with a mean of 4.9, whilst Google Custom only got 2.53 and MaladyHelp 1.10. The coefficient of variation increased to 0.94 but was still less than Google Custom (1.31) and MaladyHelp (2.41) Moreover,

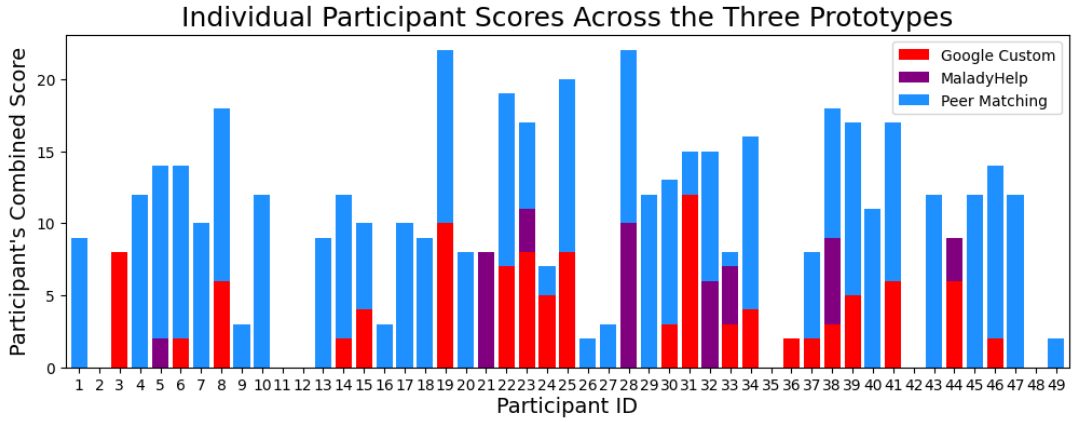


Figure 7.5: Individual participant scores across the three simulation tasks

it remained statistically significant with a p-value of 0.00655 for random chance, and 0.000075 for the permutation test.

Now, let us consider the different scores among individual participants to compare performance individually. As we can see in Figure 7.5, six participants did not score at all, whilst nine individuals scored more than 15 points across the three simulation tasks. This may be due to varying levels of health literacy and familiarity with technology across participants. However, only three individuals obtained a non-zero score for all three prototypes, which again highlights the difference in performance of the three prototypes. In addition, 21 participants only obtained a score for one of the prototypes - 18 of which were obtained when using Peer Matching. This suggests that Peer Matching may be able to identify rare conditions where information retrieval systems are not able to.

7.3.3 Questionnaire Responses

Table 7.2 shows that both Peer Matching and MaladyHelp were considered to be suitable (64% MaladyHelp, 84% Peer Matching) and easy to use (88% MaladyHelp, 88% Peer Matching). In contrast, Google Custom was not deemed to be suitable or easy to use by many participants. This may be because Google is designed to be able to handle much larger databases and as such does not perform well on a smaller database, one participant stated that *"Google Customs didn't return any results"*. In addition, Google Custom shows several adverts which may be distracting and make the task harder

Table 7.2: Participant responses to a video demonstration of Puzzle

For the task...	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
Google was suitable	6%	22%	20%	29%	22%
MaladyHelp was suitable	27%	37%	22%	10%	4%
Peer Matching was suitable	49%	35%	8%	6%	2%
Google was easy to understand and use	6%	22%	29%	16%	27%
MaladyHelp was easy to understand and use	45%	43%	4%	6%	2%
Peer Matching was easy to understand and use	41%	47%	8%	0%	4%

to complete; *"google was very confusing as there were so many adverts, diversions and just a lot of irrelevant information"*.

69% preferred to use one of the tools from this study than readily available technologies used for health information (e.g., ChatGPT). Within those who preferred one of the tools from this study, 59% preferred Peer Matching, 38% preferred MaladyHelp and 3% preferred Google Custom. This suggests that Peer Matching may be well received, but there is still a need for health-related search engines. The most common preferences of the remaining 31% were clinical expertise (either abstaining from technology outside of clinical consultations or using technology involving doctors), the NHS website, or large language models (i.e., ChatGPT).

The NHS website and clinical expertise both highlight the importance of trusted medical guidance when seeking health information. In contrast, ChatGPT's 'hallucination' problem is well known and suggests that the validity of the information found on them cannot be trusted. Participants may prefer ChatGPT because they find it easier to direct health information searches with a conversational interface as it can work more dynamically than the prototypes considered in this study.

7.4 Qualitative Results

We define themes from our thematic analysis in terms of the implications that our prototypes have on participants' experiences. Namely, three distinct themes were identified: *Restoring Faith in the Healthcare System, Building Support Networks,* and *Facilitating Information Seeking*, each of which contains in some way a form of knowing more about their rare disease. This demonstrates the importance of information and understanding for rare disease patients.

We compare these themes which highlight participants' challenges and desired support from their experiences of diagnosis to the findings of each of the interventions to ascertain their potential to support and alleviate challenges faced by people experiencing a rare or challenging diagnosis. In particular, we will draw from both the quantitative analysis of the interventions and the participants' qualitative responses to the interventions.

7.4.1 Restoring Faith in the Healthcare System

Participants faced several challenges in their experiences of the healthcare system. As such, they had an increased appreciation for a confirmed diagnosis, to a far greater capacity than would typically be expected, which was apparent throughout the interviews. The importance of diagnosis was not simply to open up treatment options, but the alleviation of obtaining closure and knowledge of specific answers *"Even if it doesn't cure the whole situation completely, ... I'll be relieved"*. As such, it was seen as a significant goal that participants were striving for and therefore shows the significance of high-performing prototypes from the simulation-based lab study. This analysis showed that Peer Matching performed significantly better than MaladyHelp or Google at identifying the underlying cause of the three conditions in this study. This shows significant promise for identifying potential avenues to explore within clinical investigations, however, additional investigations would be necessary to establish whether Peer Matching continues to perform well in a real-world setting.

Whilst the high performance of Peer Matching is significant, it is not the role of these interventions to provide patients with a diagnosis, only healthcare professionals have the required expertise for this responsibility. Instead, the diagnostic impact of this technology is to facilitate greater consultations with healthcare professionals. Participants generally recognised this fact saying that diagnosis *"has to be [made by] a doctor"* and

considered technology as an avenue to direct and input into their consultations and clinical investigations *"from things I've checked on Google, I would love to talk to ... a medical professional [who] can add their knowledge"*. Therefore, we now need to ascertain if and how these three interventions could input into consultations.

These two aspects reveal that, in the perspectives of those who experienced a rare or challenging diagnosis, receiving a formal clinical diagnosis is seen as one of the most important goals to search for. Once participants have a name to call their illness, they not only feel relieved and reassured in their experiences but can begin to move on and seek help in the form of both treatment and emotional support.

7.4.1.1 Patient Findings: Self-Advocacy During Consultations

Participants expressed interest in sharing findings from Peer Matching with their healthcare professionals and believed that it would aid diagnosis by directing clinical investigations to check for specific conditions, for example, one participant said *"you can be matched with people who have already been through what you're going through. And potentially, if it is the same exact, like diagnosis, you can kind of give that to a doctor I'm guessing, and they'll be able to check for those things and diagnose you faster"*.

On the other hand, one participant felt that healthcare professionals would say " *'just because you've got those symptoms doesn't mean you've got that condition' "* and that they needed to see evidence themselves to direct clinical investigations; *"unless ... healthcare professionals ... could do it as well"*. Another participant also felt the addition of healthcare professionals in Peer Matching would aid the identification of relevant specialists. If Peer Matching facilitated *"finding people in the field and linking them to certain illnesses on this app"*, they felt *"it would have helped me a lot"* to reduce the cycle of referrals. Therefore, exploring ways to identify and contact relevant healthcare professionals in Peer Matching could potentially improve patients' satisfaction.

Participants using Google *"couldn't find any results"* that were relevant, and one participant blamed themselves for the lack of relevant results *"[if I was] familiar with the usage, I might ... draw more information"*. So, participants had little findings to bring back at all. Whereas participants using MaladyHelp felt that *"... discuss[ing findings] with your doctor ... [is] a good starting point"* to explore potential answers, but believed they might be dismissed and not necessarily be able to direct clinical investigations *"you could go in with ... ideas and then they'll [healthcare professionals] go, 'well, no' "*. Therefore,

Peer Matching shows the most potential not only to allow patients to input into their consultations but also to bring more relevant content.

Many participants expressed their frustration at not having a way to accurately describe what they were feeling and the associated disenfranchisement that resulted from being frequently dismissed in the past: *"It's just this, or it's just that, or its nothing... you didn't need to come in"*. Patients expressed frequent distress and frustration that, until they were formally diagnosed, they would often have trouble convincing clinicians and peers that they had a health issue. It also demonstrates that this view of diagnosis being the *Ultimate Goal* can have resounding effects on the participant - frequent dismissal of symptoms can lead patients to feel that they would not get answers; *They'd [healthcare professionals] always be like, "oh its nothing, go on, off you go"*. Participants frequently discussed how healthcare professionals would trend towards a common diagnosis that did not align with their symptoms, and this resulted in patient doubt towards both their illness and healthcare providers; "*Just lack of trust in the professionals to do anything... I've kind of given up*".

Participants recognised that addressing these issues was outside the scope of technology, stating that the lack of interest from healthcare professionals was *"down to the people in general, like the doctors ... rather than something technology can fix"*. However, our findings suggest the interventions may reduce the burdens and negative impacts associated with several problems, including healthcare professionals dismissing health issues; and avoiding rare conditions; resulting in a loss of motivation and self-advocacy.

7.4.1.2 Reducing Feelings of Dismissal and Doubt of Health Issues

Participants felt that finding people with shared symptoms on Peer Matching provided validation for their symptoms *"Like, the fact that somebody else is experiencing that as well ... could mean that it is part of the bigger picture"*. This may reduce the impact of dismissed experiences by providing evidence which supports the genuity of their symptoms. On the other hand, participants felt that symptoms not shown in the results of MaladyHelp and Google caused additional doubts regarding symptoms *"[the search results] might not fit all the symptoms ... it might make you guess again ... am I experiencing this?"*. While Peer Matching is very unlikely to find a perfect match of an individual's combination of symptoms, participants did not express any issues with this. This may be because they are viewing individual cases, so the absence of symptoms does not necessarily

mean that other people don't experience them, whereas results from search engines may be perceived as an exhaustive list of symptoms. Therefore, Peer Matching shows potential to provide validation and relief for individuals seeking a challenging diagnosis, whereas MaladyHelp and Google aggravate this issue.

7.4.1.3 Dare to Think Rare

One participant stated that Peer Matching findings could prompt their doctor to look for new or rare causes and speculates that their doctor would say *" 'oh yeah, maybe this is something I did miss. And I know that to check for this specific thing, you'll have to have this' ... It'll make it way easier, and I think more stress-free as well"*, however, there were no findings to suggest that MaladyHelp or Google would aid patients to direct clinical investigations to be more specific and targeted to rare or missed conditions. This suggests that Peer Matching may support participants with their experiences of healthcare professionals fitting them into a box and not considering alternatives outside of their existing knowledge and scope. Therefore, our findings show that Peer Matching may increase patients' self-advocacy and encourage their healthcare providers to *dare to think rare*.

7.4.1.4 Increasing Motivation and Self-Advocacy

Participants expressed that Peer Matching would encourage them to play a more active role and increase their self-advocacy due to the identification of new avenues for investigation. For example, one participant stated *"I'd be more inclined to keep booking doctor's appointments."* and that it was hard to justify booking appointments without bringing new ideas to the table, but that findings from Peer Matching *"give me a reason to keep going and booking [consultations]"*. Participants felt that MaladyHelp also provided *"a list of all the different things it could be ... [which might] keep you motivated"* and therefore, may also facilitate greater self-advocacy. The identification of new avenues for investigation would arm patients with rare or diagnostically challenging conditions with the information-seeking opportunities that are available for people with common conditions, for whom services like Google cater. Therefore, Peer Matching and MaladyHelp were seen to facilitate this, resulting in a stronger voice and a more active role in their health.

Peer Matching not only had the highest performance in the quantitative study, but it was seen to facilitate and encourage greater patient input and self-advocacy in

consultations as well as reducing some of the negative impacts from the disenfranchisement of the healthcare system; whereas MaladyHelp showed more promise than Google Custom in the participants' perception of it but had the lowest performance in the quantitative study. However, the incorporation of clinicians and specialists, either through direct use or communication channels, could be a potential avenue for improvement. Hence, Peer Matching shows the most promise in supporting individuals with their challenging diagnosis.

7.4.2 Facilitating Information Seeking

Another constant theme throughout these interviews was the presence and seeming ubiquity of online health information, which primarily was sought through search engines. Participants would frequently talk about utilising Google Search to explore more about their symptoms and provide them with potential ways to address them that they could then later discuss with their clinician: *"Google was my only backup"*. Two key characteristics were attributed to the use of technology, primarily search engines, for information seeking: their accessibility, and their inappropriateness. These two characteristics would often be discussed in tandem with the caveat that utilising said search engines was not an effective way to gain answers, and highlighted their desires and requirements of technology, both in terms of the positive parts of search engines: accessibility regarding resources; perceived privacy and lack of judgement; ubiquity and constant availability; and the negative aspects: difficulty finding relevant content, fear-provoking.

First, we explore the positive aspects for the use of technology for information seeking. Accessibility was present across all three prototypes; however, Peer Matching does not offer as much perceived privacy or ubiquity as the search engines due to its involvement with other people. However, participants felt it created a safe space where they won't be judged *"I feel I'll be able to relate with someone ... I won't be judged for what I'm going through"*. As Peer Matching was only used in a controlled environment and there was no real patient data collected, there were no privacy and security issues for this stage of testing, however, careful consideration and safeguards would need to be in place to ensure that Peer Matching provides patients with a safe space. We explore these positive aspects in the next three sections, the following two sections then explore the negative aspects.

7.4.2.1 Financially Accessible

The costs for transport, money lost due to time taken away from work, and direct expenses of healthcare created a significant barrier to attending consultations, whereas all of the prototypes were freely accessible online. This provided an easier more accessible path to knowledge discovery *"it seems as cost-effective for me ... hospitals usually cost money and inconveniences. ... So, using the Peer Matching ... [will] help me to save cost[s]"* from the comfort of the participant's home. While this is beneficial, it is not suitable to replace consultations with technology use. This may contribute to participants' motivation to seek information themselves and bring these findings to their healthcare provider. Therefore, it is worth exploring whether incorporating clinician interactions within technology is feasible without impacting its accessibility.

7.4.2.2 Privacy Preserving

As search engines, MaladyHelp and Google provide privacy *"[a] search engine has, has that advantage that ... it's not necessarily going to pin it down to you"*. Participants also expressed appreciation for the lack of judgement received from technology *"there will be no form of judgement on my health"*.

As Peer Matching connected people, it could be expected that privacy issues would arise, but only one participant raised this as a concern and said *"my privacy ... it's a big deal for me"* but later said that *"I feel I'll be able to relate with someone ... I won't be judged for what I'm going through"*. So, while Peer Matching will not provide the same level of anonymity as a search engine, the community formed through it may provide a safe space to allow patients to be open about the challenges they face. As Peer Matching was only used in a controlled environment and there was no real patient data collected, there were no privacy and security issues for this stage of testing, however, careful consideration and safeguards would need to be in place to ensure that Peer Matching provides patients with a safe space.

7.4.2.3 Ubiquitous

Technology was described as the only thing that was constantly there for participants *"at some point, Google was my only backup"*. Again, this would intrinsically be present with MaladyHelp and Google Custom. On the other hand, support from Peer Matching may

not always be present given that people will not always be online. When matches are offline, while their profile and list of symptoms would be visible, users wouldn't be able to receive instant responses from communications. Therefore, the support from communications and social interactions would not be constant within Peer Matching, however, the information and support that arise from the identification of each match's symptoms and conditions would still be present, whereas Google and MaladyHelp are more constant.

7.4.2.4 Information Relevancy

The specificity of MaladyHelp and Peer Matching were both perceived to improve the relevance of results; however, the quantitative results suggest that Peer Matching offers stronger performance in this regard. This suggests that Peer Matching can better fulfil patients' needs, in contrast to past experiences of technology which typically cater for the average person, making it difficult to find things that relate to them.

Participants recognised that search engines generally cater to the average person. This was discussed as a particular problem with Google showing *"broader thing[s] that might be irrelevant"*, thus creating *"a pin in a haystack"* problem, making it difficult to find relevant results for their diagnostically challenging condition. However, some participants preferred a broader outlook from search engines as a starting point to then refine within specialist content *"I want it to be broad and to think of everything that it could be"*. Participants expressed relief in MaladyHelp's specificity, stating *"not having to first, click through 5000 things"* to find a relevant result reduces the challenge in information-seeking tasks. Participants felt that Peer Matching would *"help me narrow down my, my search"* and that using *"it makes the process much more easier ... [as] you're checking for specific things"*. Therefore, both MaladyHelp and Peer Matching show promise in finding more relevant content from the perspective of participants, however, the quantitative results show that Google performs better than MaladyHelp.

7.4.2.5 Fear Provoking

Participants found Google *"comes up with like the most serious cases"* which they found to be scary. Participants also felt that when using MaladyHelp they would *"get a little bit afraid ... [and they] don't want to ... [see] negative outcomes"*. On the other hand, participants felt that Peer Matching *"is probably closer to what it would be than ... the*

most extreme version", but they recognised that if their condition was life-threatening, then it would still evoke similar fears.

Therefore, all three prototypes were financially accessible. MaladyHelp and Google provided greater privacy and ubiquity than Peer Matching, although participants felt it provided a safe space with no judgement. Participants expressed difficulty finding relevant content with Google, whereas they found Peer Matching and MaladyHelp had more specific content.

As a whole, Peer Matching shows the most promise in providing support with diagnosis through more impactful and relevant consultations by bringing patient findings to clinicians so that they *dare to think rare*. The quantitative study showed that Peer Matching performed best with the identification of relevant information and potential causes, while the interviews showed that Peer Matching and MaladyHelp resulted in greater self-advocacy, and interest in working with their healthcare provider when compared with Google. Some participants felt that there would be greater recognition from healthcare professionals if they were directly incorporated into Peer Matching.

Participants also felt that Peer Matching would reduce their disbelief of symptoms, whilst MaladyHelp and Google would cause additional doubts. Peer Matching also showed significant promise in providing a community to act as advocates for one another as well as providing emotional and empathetic support. It is clear that Peer Matching cannot fulfil all the social needs of patients, since in-person contact can provide more rich interactions which technological communications do not offer, and matches cannot fill the roles of familial relationships. However, participants felt that such a community would reduce feelings of isolation and provide relief that they are not alone.

The theme of search engines brought up several desires and issues with technology. The financial accessibility was present across all three prototypes; however, Peer Matching did not offer as much privacy or ubiquity as the search engines due to its involvement with other people, but participants felt it created a safe space where they wouldn't be judged.

Participants also found that search engines typically cater for the average person, making it difficult to find things that relate to them. The specificity of MaladyHelp and Peer Matching were both seen to improve the relevance of results; however, the quantitative results suggest that Peer Matching is better. Overall, Peer Matching shows significant promise to support patients facing a diagnostic odyssey in a range of ways

including: information discovery; greater advocacy; and social support - all of which affect the well-being and outlook of patients.

7.4.3 Facilitating Empathetic Support Networks

Participants stressed the importance of support networks, usually consisting of family members. Support networks acted as advocates for their diagnosis, provided instrumental support (i.e., giving lifts to consultations), or simply just listened and provided emotional support. It is clear that technology cannot provide familial support, however, our findings suggest that Peer Matching shows the potential to alleviate the feelings of isolation associated with those without a support network as well as increased empathy from connections on Peer Matching.

The only form of social support in the design of MaladyHelp was linked to support groups relating to conditions in the search results; however, none of the participants discussed this. Google did not have any social aspects in the design. As such, participants expressed no significant effects on support groups from the use of these prototypes (aside from the effects of diagnosis being obtained faster), since they did not easily facilitate connections to others, so in this section, we will primarily address the potential of Peer Matching for the types of support covered in this theme.

7.4.3.1 Advocacy

Having an advocate was a significant component of the support received from diagnosis. This role was usually taken on by close relationships, often a parent. Peer Matching would not facilitate the level of support which can be attained from family, however, participants felt that they could advocate for one another and *"figure it out together"* by working as a team and collaborating with their clinical investigations *"I was checked for this, you might want to get checked for that"*. As said earlier, our findings suggest that both MaladyHelp and Peer Matching facilitate self-advocacy through information discovery, but Peer Matching may also provide an external source of advocacy.

7.4.3.2 In Person vs Online Support

Peer Matching would not facilitate instrumental support since this generally requires a physical presence. One participant stated that online connections they had made through

their own journey of diagnosis were *"... not so personal as like really meeting someone else like face to face'* and that the *"connection is a little bit missing"*. However, another participant felt they could not interact with patient groups in person *"I didn't really feel welcomed in that space"* because *"there was no one with my illness ... And then with Peer Matching, I think it's a lot easier for me because I found someone'*. Therefore, Peer Matching cannot give the rich connections you can get in real life, however, it may help people who cannot easily access support groups in person to access it in a more comfortable setting.

7.4.3.3 Emotional and Empathetic Support

Participants who didn't have support networks described feelings of isolation and craved connections with people who could relate to them *"I felt left out ... [I] wish[ed] I could find someone that could share in what I go through"*. Moreover, people who did have support networks in place still felt connections to people going through the same experiences was important and that *"just having a community"* can help to know *"you're not alone ... you're not making everything up"*.

Participants felt that Peer Matching would provide this much-needed empathetic and emotional support even just from matching people and displaying symptoms *"I'm finding empathy here. There's someone who feels like I do. I'm not alone ... I just have that awareness that there's someone going through that with me"*. However, to *"maximise the benefits"*, they would need to be able to communicate with the matches *"if I could contact them, ... [it could facilitate] sharing knowledge and like emotional support"*.

Therefore, Peer Matching shows the potential to alleviate feelings of isolation experienced during the diagnostic odyssey by identifying individuals who can offer empathetic support. Moreover, participants felt that this support would be much greater if they were able to communicate with these individuals. In addition, these individuals could also advocate for one another, and utilise social means of information seeking.

7.5 Chapter Summary

Peer Matching presents the most promise both in terms of the quantitative and qualitative parts of our evaluation. These simulation-based laboratory studies showed that Peer Matching facilitated faster and more accurate information discovery in comparison to MaladyHelp or Google. Participants also felt that the knowledge discovery on Peer

Matching, both from viewing other profiles and from interactions would get them much closer to attaining diagnosis; and as such would be empowered to play a more active role in their health, which is recognised as an important factor in obtaining a rare diagnosis [6, 15].

Furthermore, Peer Matching shows significant potential in providing patients with relief and support with the wider issues they face during their diagnostic odyssey. In particular, participants felt that using Peer Matching would reduce their feelings of isolation and build a community of peers to work together and find answers. This need for a support network significantly impacted participants' experiences of diagnosis, reducing the vast amounts of frustration that they experienced.

In this chapter, we demonstrate that Peer Matching shows potential to aid clinical experiences and clinician-patient relationships; facilitate empathetic support networks; and provide better facilitation of information-seeking. Therefore, Peer Matching shows the most promise to support the previously unmet needs identified in our exploration of the design space in Chapter 2 and Chapter 4.

Chapter 8

Discussion and Conclusion

8.1 Implications on Methodological Approach

As discussed in Chapter 5, we developed a new methodological approach due to several unique challenges that arise when conducting studies with rare disease patients:

- The lengthy diagnosis of a rare disease patient, potentially lasting 30 years [6] makes for an infeasibly long study
- Individuals with rare diseases represent a limited pool of participants in which to contact and recruit for human-centred studies
- As non-experts, we cannot expect to successfully identify patients before clinical experts are able to (i.e. before they are diagnosed)
- It is unethical to deploy prototypes which could potentially hinder the diagnostic journey

As a result of this, in situ, long-term deployments in this context are far beyond the boundaries of a reasonable or practical solution, rendering them effectively impossible. Our methodological approach facilitates an evaluation which assesses the positive outcomes of patient-facing prototypes without the need to identify, recruit and deploy interventions at the start of the diagnostic odyssey.

The majority of evaluation approaches were comprised of purely numerical evaluations (i.e. success metrics such as accuracy, sensitivity, specificity, etc.) and did not account for the wider and more personal impacts on rare disease patients. However, we argue that numerical evaluations are not sufficient to assess positive outcomes of pre-diagnostic technology for rare disease patients; instead, an empirical study is

necessary for two key reasons. Firstly, patients evaluate health information and only bring findings to consultations that are deemed to be relevant, so evaluating raw results makes it difficult to interpret what findings would be brought to a clinician. Secondly, given the challenging emotional and social aspects of seeking a rare diagnosis, it is important to assess the wider impacts on rare disease patients which cannot be captured through numerical evaluations.

Our evaluation provided insight into both of these aspects. Moreover, by utilising a simulation-based design, we were able to obtain results in a significantly more resource-efficient way, without the need for a lengthy and expensive clinical trial. Our approach not only assessed the informational capabilities of prototypes (as seen by previous evaluations of pre-diagnostic prototypes) but also assessed empathetic support which was otherwise overlooked in evaluations.

In addition, our approach evaluates the timeliness of diagnosis, a critical aspect of a positive outcome for rare diseases. Considering both the informational capabilities and the empathetic support capabilities, we were able to discriminate between prototypes for their potential to provide positive outcomes for rare disease patients. As such, this provided feedback from end-users during the early stages of development and thus revealed initial avenues for improvements in the design of these interventions. Moreover, our evaluations eliminated prototypes which were not suited for purpose, thus preventing additional resources wasted on interventions which do not cater effectively for the needs of rare disease patients. Therefore, this approach provides a low-resource avenue to perform evaluations on technology which would otherwise require significantly high resources.

The addition of resource-efficient evaluation approaches would open up this area to more experimentation and creative freedom since the significant resource cost (or complete lack) of an effective evaluation may have previously deterred researchers from risking a potentially negative outcome. Therefore, this could significantly improve the rate of innovation in this field. In addition, it facilitates iterative prototype developments, which is important for effective design. opens doors within this research area.

There exists a significant body of work on self-care technologies for common chronic conditions [128, 129, 202]. However, people with rare conditions experience unique challenges in the self-management of their health and well-being, including a lack of empathy and understanding from peers, perceived lack of care and disenfranchisement from the healthcare system, and a general lack of knowledge around their condition.

Works by MacLeod et al. [118, 135, 203] contrast and compare the challenges faced by people with chronic conditions with those faced by people with rare diseases. However, these works look at supporting people with a diagnosis. In contrast, our work explores the potential of pre-diagnostic technology to support rare disease patients with patient-provider collaborations and self-management. As we demonstrate in Chapter 3, patient-facing support for undiagnosed rare disease patients currently receives very little attention from the research community. We speculate that this is due to a lack of appropriate methods of evaluation that account for the wider experiences of rare disease patients. Therefore, we mitigate a significant hurdle for researchers in this area by providing a low-resource, ethical approach to perform evaluations.

8.1.1 Limitations and Risk-Factors

It is important to recognise the key limitations of our approach. Firstly, the number of conditions included in the study was sufficient to facilitate early discrimination between our prototypes, however, further evaluations should include a greater number of conditions to provide a greater assessment of the likelihood of false positives. This may be achieved through automatically generating more patient cases by training a Large Language Model (LLM) on the data curated in this thesis. Secondly, this approach requires participants to retrospectively reflect on their experiences. Therefore, some aspects may be forgotten, or participants may feel more distant from past experiences. However, as stated above, the identification of people with rare diseases, before it is known that they have a rare disease is not feasible. In addition, distancing between past experiences may reduce the potential of participant distress in discussing their diagnostic experience. Another limitation of this approach is that it does not assess the long-term impacts of pre-diagnostic technology. However, as we discuss in Chapter 5, it would be unethical to conduct an in-situ longitudinal study since deploying prototypes which could potentially hinder the diagnostic journey as this may lead to severe consequences on patients' health. In particular, pre-diagnostic patient matching systems may result in a misdiagnosis may be made if an incorrect condition is suspected based on patient findings from using the prototype; there exists a risk of misconduct or anti-social behaviour on a pre-diagnostic patient matching system; in addition, sensitive data would be required for a pre-diagnostic patient matching system, so it is important to mitigate any risks for data breaches. Our evaluation approach, in addition to further evaluations on data

security, could be used to perform several evaluations to assess the safety of prototypes until the potential risks of harm are outweighed by the potential benefits of an in-situ evaluation. Any in-situ evaluations following this should also be carefully monitored to prevent anti-social behaviour and misconduct. Overall, our methodological approach overcomes significant issues with existing approaches, which outweigh the limitations of this approach. Thus, this methodology provides a suitable approach to perform early evaluations of pre-diagnostic prototypes for rare disease patients.

8.2 Thesis Overview

This thesis presents four key contributions: the exploration of the design space for pre-diagnostic rare disease technology; the identification of the lived experiences of rare disease patients; the development of a novel methodological approach for evaluation; and the generation and publication of open-source data [1] and code[2]. Within this chapter, we explore the overall contribution of the work presented in this thesis, collating the contributions from each chapter and discussing the implications for the research space. First, we revisit our overview of the challenges of a rare disease diagnosis. Then the research questions from Chapter 1 are discussed and related to the contributions presented in each chapter. Finally, we reflect on the overall contributions of this thesis and discuss implications for future work within this research space.

8.2.1 Revisiting the Problem Statement

Clinicians are taught *"when you hear hoofbeats think horse, not zebra"*[3]as a metaphor to search for common conditions over rare ones. But zebras exist and rare disease patients exist. As such, rare disease patients may often be overlooked, demonstrated by the insufficient knowledge of rare diseases as self-confessed by 94.6% of clinicians [19]. But as a rough estimate, the average full-time general practitioner will see 10 rare disease patients per week[4]. This may be one of the contributing factors for why patients face an incredibly challenging diagnostic experience. Patients are frequently passed from one doctor to the next, misdiagnosed multiple times and are constantly seeking but

[1] github.com/902549/patient_perspective_data
[2] github.com/902549/peer-matching
[3] Quote by Dr Theodore Woodward
[4] General Practitioners see on average 37 patients per day[1], 1 in 17 people have a rare disease [32]

not obtaining the answers they want and need [6,7]. Beyond this, patients experience additional strains on their social interactions, feeling that their friends and family simply do not understand what they are going through[5].

Many people use popular sites to support their health searches, including ChatGPT, Google and Facebook [20–24]. But, in their prioritisation of popularity, they do not cater for rare disease patients. In addition, there are several hurdles in curating and evaluating patient-facing pre-diagnostic technology for rare disease patients, including the lack of data which represents the patient's perspective as well as insufficient methods of evaluation. Therefore, the research for this PhD aims to open up this research space by: establishing a design space for the needs of people with rare diseases, not only as patients but also as people; designing a novel, low-resource evaluation approach which is suited to the context; and providing our Patient Perspective Dataset.

8.3 Research Questions

Here we discuss the contribution of each chapter in relation to the research questions. For convenience, the full list of research questions is provided again below. Figure 8.1 shows the specific outputs and contributions for each of the research questions explored in this thesis as well as the corresponding chapters. This shows that Chapter 2, Chapter 3 and Chapter 4 each contribute to answer **RQ1** and **RQ2**. **RQ1** explores the challenges faced by undiagnosed rare disease patients to identify opportunities for improvement. The exploration of **RQ1** did not refine this exploration to technological implications to prevent limitations of scope due to perceptions of the role of technology. Following this, **RQ2** explored the implications for technology by building upon the identified challenges and opportunities for improvement through a human-centred design approach and comparing to existing approaches within the literature. By separating these research aims, we identified the need for greater empathetic support, a previously unexplored area of pre-diagnostic technology for rare disease patients. Thus, our approach formed a comprehensive exploration of the design space for pre-diagnostic technology for rare disease patients. In addition, we demonstrate in Chapter 3 that there is a lack of data and suitable evaluation methods for patient-facing pre-diagnostic technology for rare

[5]See Chapter 4

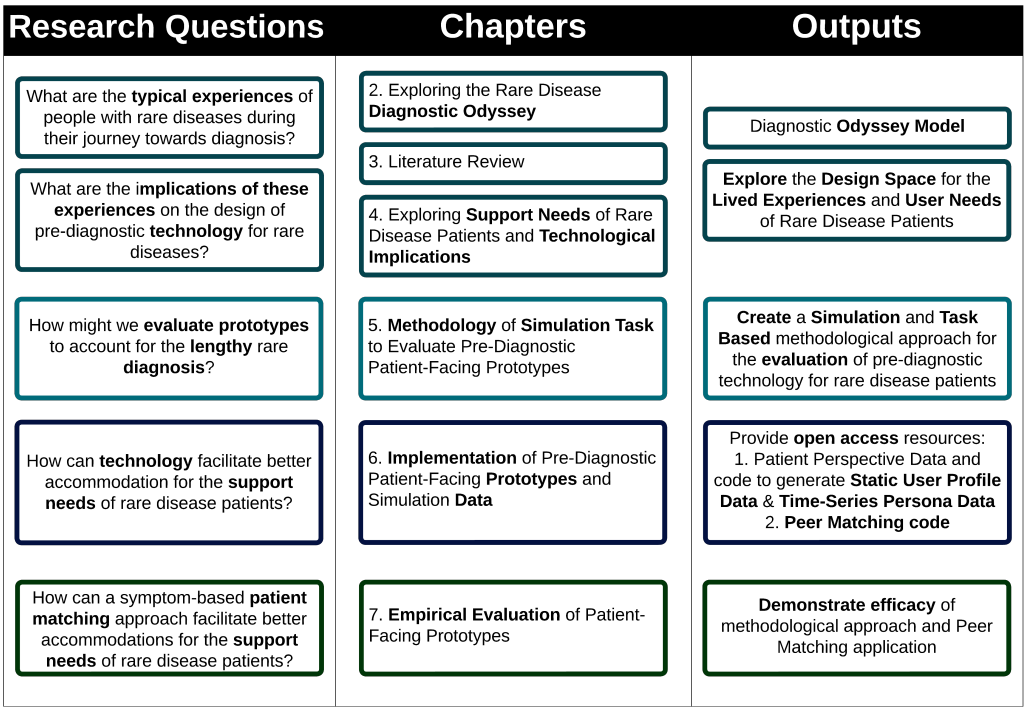


Figure 8.1: Diagram linking research questions to each chapter and the contributions

disease patients and argue that this may be why there is little focus on this area within the research community. This leads us to our next three research questions.

RQ3 explores the avenues in which we could evaluate pre-diagnostic technology for rare disease patients and identifies a new methodological approach which accounts for the lengthy diagnostic time. Namely, we propose a simulation- and task-based study to assess how quickly can participants consistently identify the correct underlying disorder of a given patient profile. This requires realistic time-series personas to recreate the informational journey of a rare disease patient, the curation of which is explored in **RQ4** to create the Time-Series Persona Dataset. In addition, **RQ4** explores the curation of prototype implementations to be evaluated in our simulation-based study, thus creating Peer Matching and the Static User Profile Dataset (which provides the user base for Peer Matching). Following this, **RQ5** carries out this Simulation-Task approach to explore whether Peer Matching can facilitate greater accommodations for the support needs of rare disease patients. In our exploration of this research question, we demonstrate that Peer Matching significantly outperforms the MaladyHelp search engine and Google Custom, by facilitating greater empathetic support and information discovery.

RQ1. **What are the typical experiences of people with rare diseases during their journey towards diagnosis?**

a) What are the typical clinical interactions that people with rare diseases face when seeking a diagnosis?
b) How are people with rare diseases affected by typical clinical interactions when seeking a diagnosis?
c) Where are there opportunities for improvement within the rare disease diagnostic odyssey?

RQ2. **What are the implications of these experiences on the design of pre-diagnostic technology for rare diseases?**

a) What are the implications of clinical interactions on the design of technology to support diagnostic decisions?
b) To what extent is the design space of pre-diagnostic technology for rare diseases being met?
c) What are the implications of the opportunities for improvement identified in the design of technology?

RQ3. **How might we evaluate prototypes to account for the lengthy rare diagnosis?**

a) How can we evaluate whether prototypes promote faster knowledge discovery for rare diseases?
b) How can we evaluate whether technology facilitates better accommodations for the support needs of rare disease patients?

RQ4. **How can technology facilitate better accommodation for the support needs of rare disease patients?**

a) What are the data requirements for patient-facing technology for people with rare diseases?
b) How can we generate data which is representative of terminology and information discovery from the patient's perspective?
c) How can machine learning interventions support the information-seeking needs of people with rare diseases?
d) How can machine learning interventions identify pairings between people with rare diseases which offer empathetic support?

RQ5. **How can a symptom-based patient matching approach facilitate better accommodations for the support needs of rare disease patients?**

a) How can Peer Matching facilitate greater support for the information-seeking needs of people with rare diseases?
b) How can Peer Matching facilitate empathetic support through pairings between people with rare diseases?

RQ1. What are the typical experiences of people with rare diseases during their journey towards diagnosis?

Our first research question explores the context of the diagnostic odyssey of people with rare diseases. To explore this, we first explore the clinical interactions that rare disease patients face and then consider the wider impacts on patients when experiencing a diagnostic odyssey. Chapter 2 provides the answer for the first part of research question 1:

a) What are the typical clinical interactions that people with rare diseases face when seeking a diagnosis?

In particular, Chapter 2 presents a thorough examination of the clinical interactions that individuals with rare diseases typically encounter during their diagnostic journey. Key factors contributing to delays in rare disease diagnoses are identified, including frequent referrals to specialists, misdiagnoses, a lack of expertise in specific rare diseases (even within specialties), and repetitive genetic testing.

In addition to the exploration of the clinical interactions of rare disease patients, we need to explore the wider impacts of this clinical context and identify the opportunities for improvement. Therefore, we ask the following two sub-questions which are covered in Chapter 4:

b) How are people with rare diseases affected by typical clinical interactions when seeking a diagnosis?
c) Where are there opportunities for improvement within the rare disease diagnostic odyssey?

Chapter 4 utilised human-centred workshops with rare disease patients to explore these research questions. In particular, it first explores the impacts of typical clinical interactions of rare disease patients and identifies the unmet support needs of rare disease patients, a largely unexplored aspect of the design space. Then, we explore opportunities to improve the experience of diagnosis for rare disease patients based on these unmet support needs. In particular, we identify three key issues that rare disease patients face. Firstly, participants *lacked empathetic connections and understanding from peers*. In particular, participants lacked a support network and felt that their need for empathetic connections was not realised in their existing circles who blamed them for their condition; had

unrealistic expectations of them; and treated them like a burden. Secondly, participants perceived a *lack of care from healthcare professionals*; participants expressed that healthcare providers: lacked sufficient knowledge or experience; dismissed or did not listen to them; and did not take responsibility for their care. In addition to this, participants experienced poor social interactions with doctors as well as difficulties and delays with getting referrals. Thirdly, participants had *unmet needs from existing technology* consisting of: a lack of access; difficulty in finding and understanding information; and concerns over misinformation and unreliability, presenting a general lack of trust.

Therefore, we have identified key unmet support needs for rare disease patients. Thus, the findings in this chapter, combined with Chapter 2, fully characterise the challenges that rare disease patients face and opportunities for improvement within these challenges. Having established the problem space, the next step was to explore the implications that it has for patient-facing technology supporting individuals seeking a rare disease diagnosis. This leads us to our second research question:

RQ2. What are the implications of these experiences on the design of pre-diagnostic technology for rare diseases?

For this research question, we draw from Chapters 2 to 4 in order to form a comprehensive exploration of the design space. In particular, Chapter 2 and Chapter 3 explore the first two sub-questions:

a) What are the implications of clinical interactions on the design of technology to support diagnostic decisions?
b) To what extent is this design space being met?

Chapter 2 discusses the implications of our understanding of the clinical context on pre-diagnostic interventions in supporting rare disease diagnosis. Within this design space, we argue that clinician-facing technology should extend support beyond direct diagnostic decisions. The need for assistance with challenging referral decisions, ongoing considerations of errors following (mis)diagnosis, specialist support technologies and prioritised genetic testing. Despite the potential of clinician-facing technology, we discuss several limitations of our examination from the context, particularly regarding information sharing among clinicians. Patient-facing technology is highlighted as a

more consistent companion throughout the diagnostic journey, ensuring continuity of information as well as opportunity for support.

Our identification of the implications of clinical interactions on the design of technology provides a lens to facilitate critical examinations of the literature in Chapter 3. As such, we explore the second sub-question, and in doing so highlight significant gaps in rare disease literature concerning the diagnostic journey. For clinically-based technologies, we emphasise the need for better support with referral decisions, particularly for primary care physicians and argue that streamlining the referral process would significantly reduce diagnosis time. In addition, given that multiple specialists may be consulted during the diagnostic odyssey, we argue that specialist support may involve aiding further referral decisions for referred patients with conditions which lie outside of their specialty.

A key gap which we identify is in the evaluation of the utility of rare disease pre-diagnostic technology, emphasising the limitations of existing methods and the need for more nuanced approaches that consider patient-facing technologies. In particular, in contrast to common conditions which evaluate performance at a single point, we argue that success for rare disease patients lies in the reduction of diagnosis time. As such, single-instance evaluations are hard to interpret in this context since we cannot compare this evaluation to the time of clinical diagnosis - assessing a diagnostic decision support system at the point in which the correct clinical diagnosis would have been made is very different to assessing it at the point when a patient first presents symptoms. As such, we answer both of the above sub-questions further and re-evaluate the definition of success in pre-diagnostic technology for rare diseases.

For patient-facing technology within rare diseases, we identify four critical gaps: the absence of human-centred approaches in the design of pre-diagnostic technology; the lack of approaches which account for the lived experiences of rare disease patients; the need for data which is representative of the patient's perspective; and the necessity for evaluation approaches encompassing both information support capabilities and provision of unmet support needs. There exists a general lack of pre-diagnostic technology tailored to the unique challenges faced by people with rare conditions, as existing research predominantly focuses on clinician-facing approaches or post-diagnosis support. Only one paper in our systematic review presented patient-facing technology, but it did not use a human-centred approach. Recognising the importance of human-centred design

approaches, we argue for the integration of human-centred approaches in both the design and evaluation of pre-diagnostic technology for rare diseases. In addition, as with clinically-based technology for rare disease diagnosis, we need to revisit what defines a positive outcome and identify measures of evaluation which assess the timeliness of information support capabilities as well as the provision of support needs.

c) What are the implications of these experiences of rare disease patients on the design of technology?

The findings from the first workshop in Chapter 4 provided us with the support needs of rare disease patients. To answer the above question, the next workshop explored how rare disease patients envision technology meeting these needs. In particular, utilising a human-centred approach, the design features of a mobile phone application are established. Based on this, a low-fidelity prototype for a mobile phone application called Puzzle was developed and then evaluated again by patients. In contrast to existing pre-diagnostic technology, Puzzle explores the potential to facilitate information management, empathetic social support, and clinician communication aids, aligning with the broader needs of rare disease patients beyond information seeking. The evaluation of Puzzle demonstrates the significance of considering both human and clinical aspects in the design space as the majority of participants felt that it would be helpful (67%) and that they would want to have used it during their diagnosis (71%). This affirmed the need for patient-facing technology to address the needs of individuals with rare diseases as people, not just as patients.

Therefore, all of the contributions in relation to the first two research questions combine to provide a comprehensive exploration of the design space for patient-facing pre-diagnostic technology for rare disease patients. Existing works predominantly focus on clinician-facing technologies for rare disease patients, and the one paper [87] included in our systematic review which did utilise patient-facing technologies does not use a human-centred approach. Vargas et al. [123] argue that community participation is vital for effective research into the design and evaluation of patient-facing technology. Our human-centred approach explored the design space and identified directions for future research (information management, empathetic social support, and clinician communication aids), thus providing a human-centred basis to direct future research.

Moreover, several design implications have also been identified for clinically-based technologies. Our examination of the clinical interactions of rare disease patients facilitated our examination of current approaches and opportunities for intervention by understanding the impact of clinically-based technologies on patients. Thus, by systematically reviewing existing approaches within the literature, we highlight key gaps and opportunities for new context-driven approaches for both clinically-based and patient-based pre-diagnostic technology for rare diseases. In particular, as patients consult with numerous specialists during their diagnostic odyssey [6,7], we argue that clinically-based technologies would have a greater opportunity to shorten the diagnostic odyssey if they provide support with challenging referral decisions throughout a patient's diagnosis in addition to diagnostic decisions, by reducing the number of required clinical consultations. Only two papers from our systematic review aimed to support clinical decisions throughout the diagnostic odyssey [77,97]. Ronicke et al. [97] did not aim to support referral decisions as well as diagnostic decisions, so only Rider et al. [77] aimed to support both diagnosis and referral decisions throughout the diagnostic odyssey. Therefore, highlighting the need for referral and diagnostic support systems throughout the diagnostic odyssey may serve to direct future research to shorten the diagnostic odyssey of rare disease patients.

By adopting a patient-focused and context-specific approach, we explored several avenues within our research that we would not have considered at the start of the project. In particular, there was a shift in our research approach at this point. Our examination of the context and experiences of rare disease patients aided the identification of the wider support needs of rare disease patients, as well as the lack of focus on these needs in the research area. This may be due to the limitations in terms of evaluation methods and available data which is representative of the patient's perspective. As such, the remaining research questions aim to not only create and evaluate our own intervention but to also facilitate future research projects by providing data and code resources in addition to a novel evaluation approach for patient-facing technology.

In answering these two research questions, we present the first reported robust human-centred study with rare disease patients to design and create technologies that can support these individuals during the diagnosis process. Previous work has largely been focused on the clinician perspective, with only one paper in our systematic review taking a patient-focused approach [87]. This paper utilised existing literature

to create features such as patient matching in order to promote informational support (i.e. discover possible diagnoses, coping strategies, and treatments) through rare disease patient groups. Moreover, many of the existing patient-facing pre-diagnostic technologies which do not focus on rare diseases [52, 53, 121] provide informational support (i.e. researching, discovering, exploring more information about one's health). However, our design workshops highlight the importance of empathetic support. By adopting a human-centred approach and working with patients to understand the challenges they face, before considering how technology could mitigate these issues, we understand that the needs of rare disease patients go beyond information seeking, and identify an unexplored area of the design space.

RQ3. How might we evaluate prototypes to account for the lengthy rare diagnosis?

Our third research question explored our evaluation approach which we present in Chapter 5 and carry out in Chapter 7. The first part of the lab study, presented in Chapter 5 explores the research question:

a) How can we evaluate whether prototypes promote faster knowledge discovery for rare diseases?

In answering this sub-question, we draw from the context of the diagnostic odyssey and our exploration of the design space to provide a definition of the positive outcomes of pre-diagnostic technology for rare diseases. This defines what we must measure to evaluate technology, and thus provides the basis from which we develop our methodological approach. Following this, we assessed the feasibility of existing approaches to incorporate the additional aspects from our definition of the positive outcomes of pre-diagnostic technology for rare diseases and identified these approaches to be insufficient or infeasible for this purpose. Thus, drawing from the task aspects of game-based design as well as the recreation of contexts in simulation-based studies, we identify a new evaluation methodology which involves the recreation of information-seeking tasks and contexts of rare disease patients. The approach we present in this chapter aims to facilitate preliminary evaluations which provide a more meaningful assessment to examine the aspects that constitute a positive outcome for pre-diagnostic technology in the context of rare diseases.

The next sub-question was explored in Chapter 7:

b) How can we evaluate whether technology facilitates better accommodations for the support needs of rare disease patients?

To provide the qualitative component of our evaluation, we built upon our simulation-task approach for the quantitative evaluation of information discovery. This time, the task aimed to familiarise participants with the prototypes for effective evaluation by facilitating first-hand use without impacting their health and diagnosis decisions. This task was preceded and followed by interviews. The pre-interaction interview highlighted challenges faced during diagnosis, thus identifying support needs to facilitate non-leading questions for the post-interaction interview. By aligning the evaluation with participants' unique experiences, the study sought to assess how effectively the prototypes addressed the support needs of rare disease patients during the diagnostic process.

Only one paper surveyed in our systematic review evaluated their algorithm for how much it would reduce the time taken to reach a correct diagnosis. Ronicke et al. [97] evaluated whether their system, Ada DX, would suggest the correct disease before the time of diagnosis by evaluating their system for each consultation that a patient has. Ronicke et al. found that Ada DX top 5 fit disease list suggested the correct disease at the first documented visit for only 33.3% of cases. Therefore, if Ada DX was only evaluated at the first documented visit, it would undermine the usefulness of the system since although this may appear low, Ada DX suggested the correct disease before clinical diagnosis for 53.8% of cases. This shows a non-trivial difference in the evaluation of this system since reducing the time to diagnosis for over half of the cases is highly significant in this context, however, to only show a single-point accuracy of 33.3% does not accurately portray the effectiveness of this system. Therefore, technology to support the diagnosis of rare diseases must be evaluated at multiple points to get an accurate impression of its effectiveness.

However, none of the works on patient-facing technology utilise a temporal evaluation of efficacy. Even outside of the area of rare diseases, many evaluation approaches for pre-diagnostic patient-facing technology either assess numerical performance [52, 53, 121] or user satisfaction [133, 134], but do not typically assess both. In contrast, our methodological approach provides an evaluation which not only assesses the performance temporally but also assesses user satisfaction (provision of support for the wider support

needs) of rare disease patients. As such, our approach facilitates evaluation which accounts for the lengthy time taken to obtain a diagnosis as well as the wider support needs of rare disease patients, thus providing a more meaningful and contextually appropriate evaluation approach for rare disease technology.

In addition, even for clinically-based technology, Ronicke's approach may not be feasible in several research projects since the breakdown of cases into each of the clinical visits is not always possible (i.e., this information may not be present in the data, or it may be too time-consuming). However, our simulation task study may be adapted to support evaluations on the performance of clinically-based technology to aid rare disease diagnosis with only a few minor adaptations (i.e., patient cases are sampled based on the original HPO terms, and the ordering of phenotypes is adjusted to the clinician's perspective - symptoms and developmental traits are presented at the first consultation).

Clearly, our evaluation approach facilitated sufficient discrimination between different pre-diagnostic patient-facing interventions. By evaluating this technology with a simulation-based design, we were able to obtain results in a significantly more resource-efficient way; we were able to perform early evaluations without a lengthy and expensive clinical trial. This not only revealed initial avenues for improvements in the design of these interventions, but it also eliminated prototypes which were not suited for purpose. Therefore, this approach can be used to perform low-cost evaluations on technology which would otherwise require significantly high resources. This encourages experimentation and creative freedom in this field as well as iterative developments of prototypes, which is important for effective design, and thus opens doors within this research area.

RQ4. How can technology facilitate better accommodation for the support needs of rare disease patients?

Chapter 6 explores the fourth research question, which explores the development of the Patient Perspective Dataset and patient-facing pre-diagnostic prototypes: Peer Matching; MaladyHelp and Google Custom. Our first two sub-questions explore the data requirements for our prototype as well as the personas required for our methodological approach:

a) What are the data requirements for patient-facing technology for people with rare diseases?
b) How can we generate data which is representative of terminology and information discovery from the patient's perspective?

We highlight the lack of data which represents the patient's perspective - existing data uses clinical terminology and often lacks the temporal component of information discovery (i.e., symptoms progress or clinical findings are identified). As such, we present a method for generating synthetic patient profiles from a clinical knowledge base, providing the Patient Perspective Data which serves as the base data for our Time-Series Persona Dataset and the the Static User Profile Dataset (which provides the user-base for Peer Matching). Firstly, we enrich Orphanet's dataset by augmenting the phenotype information for each disease with: HPO categories, layman terms, phenotype discovery group (i.e., development traits, symptoms, exploratory clinical findings, specific clinical findings), and probability of occurrence. Secondly, we dynamically generate a range of varied patient profiles from the above base data, using the probability occurrence and some perturbation noise, each patient profile samples from the base data a proportional amount of phenotypes. After the curation of the base data, two separate processes follow. For the Peer Matching user base, a diagnosis status and a name are assigned to the patient profile to create a user. For the time-series personas, the phenotypes are sequenced based on their discovery group and divided into the three informational rounds of the simulation. This provides data for both the curation and evaluation of pre-diagnostic technology for rare disease patients.

c) How can machine learning interventions support the information-seeking needs of people with rare diseases?
d) How can machine learning interventions identify pairings between people with rare diseases which offer empathetic support?

In Chapter 6, we describe our curation of Peer Matching, connecting patients with similar symptoms in order to address the isolation associated with a rare disease diagnosis. As a small proportion of the (synthetic) user base in Peer Matching are assigned to be diagnosed, we also suggest that it may support information discovery. This then facilitates a more data-driven approach to supporting people seeking a diagnosis of a

rare condition. The patient matching system developed by Shen et al. [200] also utilises a data driven approach, but aims to support the clinician and not the patients. In contrast, the only patient-facing approach identified in our systematic review [87] utilised a low data approach to develop a patient matching system and required the input of a long questionnaire before patients can find potential matches. Our approach thus facilitates lower demand on patients time by curating the data required for patient-facing pre-diagnostic technology.

RQ5. How can a symptom-based patient matching approach facilitate better accommodations for the support needs of rare disease patients?

We explore the final research question in Chapter 7. In particular, we demonstrate that not only does Peer Matching facilitate empathetic support through pairings of rare disease patients, but it also facilitates faster information seeking.

a) How can Peer Matching facilitate greater support for the information-seeking needs of people with rare diseases?
b) How can Peer Matching facilitate empathetic support through pairings between people with rare diseases?

Peer Matching emerges as the most promising solution in both the quantitative and qualitative phases of our evaluation. Simulation-based laboratory studies reveal that Peer Matching ($\mu = 7.02$) enables faster and more accurate information discovery compared to MaladyHelp ($\mu = 0.86$) or Google ($\mu = 2.20$), with a statistical significance of $p < 0.0005$ for two non-parametric empirical tests. In addition, Peer Matching was also deemed most suitable for the information-seeking task (49% of participants strongly agreed, total agreement 84%) in comparison to Google (6% strongly, 22% total) and MaladyHelp (27% strongly, 64% total).

Participants express confidence that Peer Matching's knowledge discovery features would significantly contribute to their journey toward diagnosis, empowering them to play a more active role in their health—an essential aspect in rare disease diagnosis. Moreover, Peer Matching demonstrates substantial potential in addressing the broader challenges rare disease patients face during their diagnostic odyssey. Participants anticipate reduced feelings of isolation and the establishment of a supportive peer community, easing the frustration associated with their diagnostic experiences. This

highlights Peer Matching as a comprehensive solution that addresses clinical, emotional, and informational support needs, offering significant promise for previously unmet needs identified in our exploration of the design space.

Kühnle et al. [87] present a patient matching system, but does not use a human-centred design approach. In contrast, we utilised a human-centred approach to the design of Puzzle which was informed the implementation of Peer Matching. As such, users would be matched based on symptoms logged, rather than the 53 questions required for RarePairs, reducing mental and temporal stressors on patients. Moreover, Kühnle et al. focus on information seeking, whilst, Peer Matching aims to support the wider support needs of rare disease patients, including the empathetic support needs. MacLeod et al. [118] focus on supporting rare disease patients with their lived experiences by matching people with the same rare condition. However, MacLeod at al.'s approach aims to support those already diagnosed. In contrast our approach facilitates peer matching between patients experiencing similar symptoms, and as such, may provide support during the arduous journey of seeking a diagnosis.

8.4 Future Work

In our exploration of the above research questions, we present four core contributions relating to the research space of designing and evaluating pre-diagnostic technologies for rare disease patients. In particular, we aim to pave the way for future research by firstly exploring the design space for both patient-facing and clinically-based pre-diagnostic technologies for rare disease patients; secondly by establishing the lived experiences of rare disease patients; thirdly by providing a methodological approach which facilitates the evaluation of patient-facing pre-diagnostic technologies; and finally in our curation of open-source materials for patient-facing pre-diagnostic technologies. All of these contributions facilitate future work within this area by: providing direction for future research by characterising the design space; facilitating meaningful evaluations in a practical time frame; and providing required resources for prototype curation. Let us discuss in this section more specific examples for which future work could be utilised.

8.4.1 Clinically-Based Interventions

We present several design implications to improve the contextual suitability of clinically-based interventions for rare disease diagnosis within our literature review (Chapter 3). The most significant implication that we do not explore within the thesis is in providing support for the referral decisions for rare disease patients made by clinicians. This may have significant impacts on reducing the diagnostic odyssey, by identifying suitable expertise earlier, thus reducing additional referrals and saving time for both clinicians and patients.

In addition, we highlight the need for evaluation approaches which assess technology based on the time taken to diagnosis, rather than as a single-point accuracy. Ronicke et al. [97] performed a time-based evaluation, however, it required significant manual edits to separate data based on the information that would be available at specific clinical visits. Researchers may not have sufficient time to perform these manual edits. Therefore, future work may explore whether the simulation task that we present in this thesis may be adapted to facilitate evaluations for clinician-facing technology. Moreover, future research may examine whether a similar simulation-task approach could support low-resource early evaluation of patient-facing technology for common conditions.

8.4.2 Simulation-Task and Data

Our Time-Series Persona Dataset facilitated evaluations for a small number of conditions. A larger dataset of patient phenotypes would be necessary to allow for evaluations on a greater scale. Large Language Models (LLMs) have shown significant promise in generating textual data based on a given input. Pre-trained transformer models, such as BART may be fine-tuned on the manually curated Patient Perspective Dataset, in addition to the text on HPO's website to provide a model which will translate clinical terminology to patient terminology. In addition, due to the variability of outputs from models like ChatGPT, this may also provide multiple synonymous terms, thus adding to the realism of the patient data.

8.4.3 Patient-Facing Interventions

The curation of an LLM based on our Patient Perspective Dataset may also facilitate translations from patient terminology into standardised terms, which would be a key

component in facilitating effective symptom-based matching in a real-world context. As such, this may be incorporated into Peer Matching to make it viable in a real-world context where many synonymous terms may be used.

Another avenue for future work is to explore the integration of Peer Matching with a symptom logger and the ability to share data with clinicians, as in the design of the Puzzle prototype. With logged symptoms, data such as frequency of symptoms as well as severity may facilitate a more complex matching algorithm by weighting the severity or frequency of symptoms.

8.5 Concluding Remarks

Whilst this thesis topically spans a range of areas, from human-centred design approaches to generating data and presenting a new methodological approach, the core of this thesis revolves around rare disease patients. By putting patients at the centre of this thesis, we **explore their needs and desires,** thus opening the research space to include the unmet support needs of rare disease patients. By putting patients at the centre of this thesis, we **identify limitations with existing measures of success** in technology and highlight the importance of a reduction in the time of diagnosis. This provides the basis for which we **develop the methodological approach** used in this thesis; our understanding of the information-seeking behaviours of rare disease patients was a critical component in shaping our simulation task to mirror the information-seeking tasks that rare disease patients undertake. By putting patients at the centre of this thesis, we **curate applications that are designed around their needs**; and **data that represents their perspective**. All of these contributions stem from a critical examination of the experiences that rare disease patients go through on their journeys towards diagnosis.

Bibliography

[1] Genomics England. (2016) Rare disease genomics. [Online]. https://www.genomicsengland.co.uk/understanding-genomics/rare-disease-genomics/

[2] International Federation of Pharmaceutical Manufacturers & Associations. (2023, Feb.) Rare diseases: shaping a future with no-one left behind. International Federation of Pharmaceutical Manufacturers & Associations. [Online]. https://www.ifpma.org/publications/rare-diseases-shaping-a-future-with-no-one-left-behind/

[3] G. M. Abozaid, K. Kerr, A. McKnight, and H. A. Al-Omar, "Criteria to define rare diseases and orphan drugs: a systematic review protocol," *BMJ Open*, vol. 12, no. 7, p. e062126, Jul. 2022. [Online]. https://bmjopen.bmj.com/content/12/7/e062126

[4] European Commission. (2020) Rare diseases. European Commission. [Online]. https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/rare-diseases_en

[5] M Haendel et al., "How many rare diseases are there?" *Nature Reviews Drug Discovery*, vol. 19, no. 2, pp. 77–78, 2020. [Online]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7771654/

[6] F. Faurisson. (2004) Survey of the delay in diagnosis for 8 rare diseases in europe: Eurordiscare2. [Online]. https://www.eurordis.org/wp-content/uploads/2009/12/EURORDISCARE_FULLBOOKr.pdf

[7] E. Muir. (2016) The rare reality – an insight into the patient and family experience of rare disease. Rare Disease UK. [Online]. https://www.raredisease.org.uk/media/1588/

the-rare-reality-an-insight-into-the-patient-and-family-experience-of-rare-disease.pdf

[8] V. D'Alessio. (2022) The long journey to a rare disease diagnosis. [Online]. https://ec.europa.eu/research-and-innovation/en/horizon-magazine/long-journey-rare-disease-diagnosis

[9] N. K. Vanan. (2018) The rare disease diagnostic odyssey: A medical student's perspective. [Online]. https://www.findacure.org.uk/2018/07/06/the-rare-disease-diagnostic-odyssey-a-medical-students-perspective/

[10] J. E. Mellerio, "The challenges of clinical trials in rare diseases," *British Journal of Dermatology*, vol. 187, no. 4, pp. 453–454, 2022.

[11] P. Kaufmann, A. R. Pariser, and C. Austin, "From scientific discovery to treatments for rare diseases - the view from the national center for advancing translational sciences - office of rare diseases research," *Orphanet Journal of Rare Diseases*, vol. 13, no. 1, p. 196, Nov. 2018. [Online]. https://doi.org/10.1186/s13023-018-0936-x

[12] Shire Human Genetic Technologies. (2013, Apr.) Rare disease impact report: Insights from patients and the medical community. International Federation of Pharmaceutical Manufacturers & Associations. [Online]. https://www.undiagnosed.org.uk/

[13] EURORDIS. (2005) Rare diseases: understanding this public health priority. [Online]. https://www.eurordis.org/publications/rare-diseases-understanding-this-public-health-priority/

[14] E. E. Nielsen, T. Owen, M. Roach, and A. Dix, "A patient centred approach to rare disease technology," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023.

[15] Department of Health UK. (2013) The uk strategy for rare diseases. [Online]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf

[16] K. Mack, E. J. McDonnell, L. Findlater, and H. D. Evans, "Chronically underaddressed: Considerations for hci accessibility practice with chronically ill people," in *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '22. New York, NY, USA: Association for Computing Machinery, Oct. 2022. [Online]. https://doi.org/10.1145/3517428.3544803

[17] A. Halloy, E. Simon, and F. Hejoaka, "Defining patient's experiential knowledge: Who, what and how patients know. a narrative critical review." *Sociology of health & illness*, vol. 45, pp. 405–422, Feb. 2023.

[18] A. T. McCray and K. LeBlanc, "Patients as partners in rare disease diagnosis and research." *The Yale journal of biology and medicine*, vol. 94, pp. 687–692, Dec. 2021.

[19] D. Walkowiak and J. Domaradzki, "Are rare diseases overlooked by medical education? awareness of rare diseases among physicians in poland: an explanatory study." *Orphanet journal of rare diseases*, vol. 16, p. 400, Sep. 2021.

[20] C. McNab, "What social media offers to health professionals and citizens." *Bulletin of the World Health Organization*, vol. 87, p. 566, Aug. 2009.

[21] M. De Choudhury, M. R. Morris, and R. W. White, "Seeking and sharing health information online: Comparing search engines and social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, Apr. 2014, p. 1365–1376.

[22] H. Tang and J. H. K. Ng, "Googling for a diagnosis—use of google as a diagnostic aid: internet based study," *Bmj*, vol. 333, no. 7579, pp. 1143–1145, 2006. [Online]. https://www.bmj.com/content/333/7579/1143

[23] Y. Shahsavar and A. Choudhury, "User intentions to use chatgpt for self-diagnosis and health-related purposes: Cross-sectional survey study," *JMIR Hum Factors*, vol. 10, p. e47564, May 2023. [Online]. http://www.ncbi.nlm.nih.gov/pubmed/37195756

[24] L. Mehnen, S. Gruarin, M. Vasileva, and B. Knapp, "Chatgpt as a medical doctor? a diagnostic accuracy study on common and rare diseases," *medRxiv*, Jan. 2023. [Online]. http://medrxiv.org/content/early/2023/04/27/2023.04.20.23288859.abstract

[25] M. K. Depping, N. Uhlenbusch, Y. von Kodolitsch, H. F. E. Klose, V.-F. Mautner, and B. Löwe, "Supportive care needs of patients with rare chronic diseases: multimethod, cross-sectional study." *Orphanet journal of rare diseases*, vol. 16, p. 44, Jan. 2021.

[26] J. C. Long, S. Best, B. Nic Giolla Easpaig, S. Hatem, Z. Fehlberg, J. Christodoulou, and J. Braithwaite, "Needs of people with rare diseases that can be supported by electronic resources: a scoping review." *BMJ open*, vol. 12, p. e060394, Sep. 2022.

[27] J. McMullan, A. L. Crowe, C. Bailie, K. Moore, L. S. McMullan, N. Shamandi, H. McAneney, and A. J. McKnight, "Improvements needed to support people living and working with a rare disease in northern ireland: current rare disease support perceived as inadequate," *Orphanet Journal of Rare Diseases*, vol. 15, no. 1, p. 315, Nov. 2020.

[28] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare." *Future healthcare journal*, vol. 6, pp. 94–98, Jun. 2019.

[29] H. Habehh and S. Gohel, "Machine learning in healthcare." *Current genomics*, vol. 22, pp. 291–300, Dec. 2021.

[30] C. Avgerou, "The significance of context in information systems and organizational change," *Information Systems Journal*, vol. 11, no. 1, pp. 43–63, Jul. 2001. [Online]. https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2575.2001.00095.x

[31] P. Nielsen and S. Sahay, "A critical review of the role of technology and context in digital health research." *Digital health*, vol. 8, p. 20552076221109554, Jun. 2022.

[32] Rare Disease UK. (2021) What is a rare disease? Genetic Alliance. [Online]. https://www.raredisease.org.uk/what-is-a-rare-disease/

[33] Rare Disease Day. (2019) Frequently asked questions. NORD. [Online]. https://rarediseases.org/wp-content/uploads/2019/01/RDD-FAQ-2019.pdf

[34] S. Nguengang Wakap, D. M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, D. Murphy, Y. Le Cam, and A. Rath, "Estimating cumulative point prevalence of rare diseases: analysis of the orphanet database." *European journal of human genetics: EJHG*, vol. 28, pp. 165–173, Feb. 2020.

[35] C. C. Y. Chung, Hong Kong Genome Project, A. T. W. Chu, and B. H. Y. Chung, "Rare disease emerging as a global public health priority," *Frontiers in Public Health*, vol. 10, Oct. 2022. [Online]. https://www.frontiersin.org/articles/10.3389/fpubh.2022.1028545

[36] R. Valdez, L. Ouyang, and J. Bolen, "Public health and rare diseases: Oxymoron no more." *Preventing chronic disease*, vol. 13, p. E05, Jan. 2016.

[37] V. DonFrancesco. (2021) New patient journey infographic gives a glimpse into the diagnostic odyssey. [Online]. https://rarediseases.org/new-patient-journey-infographic-gives-a-glimpse-into-the-diagnostic-odyssey/

[38] EURORDIS. (2020) Solve-rd infographic on the patient journey to diagnosis. [Online]. https://www.eurordis.org/publication/diagnosisjourneyinfographic

[39] N. Black, F. P. Martineau, and T. Manacorda. (2015) Diagnostic odyssey for rare diseases: exploration of potential indicators. Policy Innovation Research Unit (PIRU), London School of Hygiene & Tropical Medicine. [Online]. https://api.semanticscholar.org/CorpusID:73745507

[40] E. de Vries, L. Fransen, M. van den Aker, and B. R. Meijboom, "Preventing gatekeeping delays in the diagnosis of rare diseases," *The British journal of general practice: the journal of the Royal College of General Practitioners*, vol. 68, no. 668, p. 145–146, Mar. 2018.

[41] J. S. Hausmann, K. G. Lomax, A. Shapiro, and K. Durrant, "The patient journey to diagnosis and treatment of autoinflammatory diseases," *Orphanet Journal of Rare Diseases*, vol. 13, no. 1, p. 156, Sep. 2018.

[42] W. R. Evans and I. Rafi, "Rare diseases in general practice: recognising the zebras among the horses," *British Journal of General Practice*, vol. 66, no. 652, pp. 550–551, 2016. [Online]. https://bjgp.org/content/66/652/550

[43] E. Mengel, J. Gaedeke, H. Gothe, S. Krupka, A. Lachmann, J. Reinke, and C. Ohlmeier, "The patient journey of patients with fabry disease, gaucher disease and mucopolysaccharidosis type ii: A german-wide telephone survey," *PLOS ONE*, vol. 15, no. 12, pp. 1–13, Dec. 2021.

[44] R. Cooney, N. Stewart, T. Ivanka, and N. Haslem, "Representational artefacts in social problem solving: A study from occupational rehabilitation," *Design Studies*, vol. 56, pp. 149–168, May 2018. [Online]. https://www.sciencedirect.com/science/article/pii/S0142694X17300881

[45] D. of Health and S. Care. (2021, Jan.) The uk rare diseases framework. Department of Health and Social Care. [Online]. https://www.gov.uk/government/publications/uk-rare-diseases-framework

[46] M.-J. Kim, M.-S. Yum, G. H. Seo, Y. Lee, H. N. Jang, T.-S. Ko, and B. H. Lee, "Clinical application of whole exome sequencing to identify rare but remediable neurologic disorders." *Journal of clinical medicine*, vol. 9, Nov. 2020.

[47] Y. Miyachi, O. Ishii, and K. Torigoe, "Design, implementation, and evaluation of the computer-aided clinical decision support system based on learning-to-rank: collaboration between physicians and machine learning in the differential diagnosis process," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 26, 2023.

[48] J. Haubold, K. Zeng, S. Farhand, S. Stalke, H. Steinberg, D. Bos, M. Meetschen, A. Kureishi, S. Zensen, T. Goeser, S. Maier, M. Forsting, and F. Nensa, "Ai co-pilot: content-based image retrieval for the reading of rare diseases in chest ct," *Scientific Reports*, vol. 13, no. 1, p. 4336, 2023.

[49] Z. Liu, Y. Chen, Y. Zhang, S. Ran, C. Cheng, and G. Yang, "Diagnosis of arrhythmias with few abnormal ecg samples using metric-based meta learning." *Computers in biology and medicine*, vol. 153, p. 106465, Feb. 2023.

[50] W. R. Evans, "Dare to think rare: diagnostic delay and rare diseases," *The British journal of general practice: the journal of the Royal College of General Practitioners*, vol. 68, no. 670, p. 224–225, May 2018.

[51] T. Willmen, L. Willmen, A. Pankow, S. Ronicke, H. Gabriel, and A. D. Wagner, "Rare diseases: why is a rapid referral to an expert center so important?" *BMC health services research*, vol. 23, p. 904, Aug. 2023.

[52] F. Zhao, J. Xu, and Y. Lin, "Multi-view neural network integrating knowledge for patient self-diagnosis," in *Proceedings of the 3rd International Conference on Machine*

Learning and Soft Computing, ser. ICMLSC '19. New York, NY, USA: Association for Computing Machinery, Jan. 2019, p. 33–38.

[53] G. Muneeswari, S. Sajithra Varun, R. Hegde, S. Sharon Priya, P. Josephin Shermila, and A. Prasanth, "Self-diagnosis platform via iot-based privacy preserving medical data," *Measurement: Sensors*, vol. 25, p. 100636, Feb. 2023. [Online]. https://www.sciencedirect.com/science/article/pii/S2665917422002707

[54] N. Hallowell, S. Badger, F. McKay, A. Kerasidou, and C. Nellåker, "Democratising or disrupting diagnosis? ethical issues raised by the use of ai tools for rare disease diagnosis," *SSM - Qualitative Research in Health*, vol. 3, p. 100240, 2023. [Online]. https://www.sciencedirect.com/science/article/pii/S2667321523000240

[55] N. Radionova, E. ög, A.-J. Wetzel, M. A. Rieger, and C. Preiser, "Impacts of symptom checkers for laypersons' self-diagnosis on physicians in primary care: Scoping review," *J Med Internet Res*, vol. 25, p. e39219, May 2023. [Online]. http://www.ncbi.nlm.nih.gov/pubmed/37247214

[56] T. Gundersen, "'one wants to know what a chromosome is': the internet as a coping resource when adjusting to life parenting a child with a rare genetic disorder," *Sociology of Health & Illness*, vol. 33, no. 1, pp. 81–95, Jan. 2011. [Online]. https://doi.org/10.1111/j.1467-9566.2010.01277.x

[57] M. J. Page et al., "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. 71, 2021. [Online]. https://www.bmj.com/content/372/bmj.n71

[58] Y. A. Veturi et al., "Syntheye: Investigating the impact of synthetic data on artificial intelligence-assisted gene diagnosis of inherited retinal disease," *Ophthalmology Science*, vol. 3, no. 2, p. 100258, 2023. [Online]. https://www.sciencedirect.com/science/article/pii/S2666914522001476

[59] S. Pati et al., "Federated learning enables big data for rare cancer boundary detection," *Nature Communications*, vol. 13, no. 1, p. 7346, Dec. 2022.

[60] I.-E. Nogues, J. Wen, Y. Lin, M. Liu, S. K. Tedeschi, A. Geva, T. Cai, and C. Hong, "Weakly semi-supervised phenotyping using electronic health records." *Journal of biomedical informatics*, vol. 134, p. 104175, Oct. 2022.

[61] Y. Kabeya et al., "Deep convolutional neural network-based algorithm for muscle biopsy diagnosis." *Laboratory investigation; a journal of technical methods and pathology*, vol. 102, pp. 220–226, Mar. 2022.

[62] J. D. Rudie, A. M. Rauschecker, L. Xie, J. Wang, M. T. Duong, E. J. Botzolakis, A. Kovalovich, J. M. Egan, T. Cook, R. N. Bryan, I. M. Nasrallah, S. Mohan, and J. C. Gee, "Subspecialty-level deep gray matter differential diagnoses with deep learning and bayesian networks on clinical brain mri: A pilot study." *Radiology. Artificial intelligence*, vol. 2, p. e190146, Sep. 2020.

[63] I. Sánchez Fernández, E. Yang, P. Calvachi, M. Amengual-Gual, J. Y. Wu, D. Krueger, H. Northrup, M. E. Bebin, M. Sahin, K.-H. Yu, and J. M. Peters, "Deep learning in rare disease. Detection of tubers in tuberous sclerosis complex," *PLoS ONE*, vol. 15, no. 4, p. e0232376, Apr. 2020. [Online]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7190137/

[64] B. Pode-Shakked, Y. Finezilber, Y. Levi, S. Liber, N. Fleischer, L. Greenbaum, and A. Raas-Rothschild, "Shared facial phenotype of patients with mucolipidosis type iv: A clinical observation reaffirmed by next generation phenotyping." *European journal of medical genetics*, vol. 63, p. 103927, Jul. 2020.

[65] A. M. Rauschecker, J. D. Rudie, L. Xie, J. Wang, M. T. Duong, E. J. Botzolakis, A. M. Kovalovich, J. Egan, T. C. Cook, R. N. Bryan, I. M. Nasrallah, S. Mohan, and J. C. Gee, "Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain mri." *Radiology*, vol. 295, pp. 626–637, Jun. 2020.

[66] X. Li, Y. Wang, D. Wang, W. Yuan, D. Peng, and Q. Mei, "Improving rare disease classification using imperfect knowledge graph," *BMC Medical Informatics and Decision Making*, vol. 19, no. 5, p. 238, Dec. 2019.

[67] J. Jia, R. Wang, Z. An, Y. Guo, X. Ni, and T. Shi, "RDAD: A Machine Learning System to Support Phenotype-Based Rare Disease Diagnosis," *Frontiers in Genetics*, vol. 9, Dec. 2018.

[68] C. Lam, C. Yu, L. Huang, and D. Rubin, "Retinal Lesion Detection With Deep Learning Using Image Patches," *Investigative Ophthalmology & Visual Science*, vol. 59, no. 1, pp. 590–596, Jan. 2018.

[69] L. Grigull, W. Lechner, S. Petri, K. Kollewe, R. Dengler, S. Mehmecke, U. Schumacher, T. Lücke, C. Schneider-Gold, C. Köhler, A.-K. Güttsches, X. Kortum, and F. Klawonn, "Diagnostic support for selected neuromuscular diseases using answer-pattern recognition and data mining techniques: a proof of concept multicenter prospective trial," *BMC medical informatics and decision making*, vol. 16, p. 31, Mar. 2016.

[70] A.-K. Rother, N. Schwerk, F. Brinkmann, F. Klawonn, W. Lechner, and L. Grigull, "Diagnostic support for selected paediatric pulmonary diseases using answer-pattern recognition in questionnaires based on combined data mining applications–a monocentric observational pilot study." *PloS one*, vol. 10, p. e0135180, 2015.

[71] K. Sidiropoulos, D. Glotsos, S. Kostopoulos, P. Ravazoula, I. Kalatzis, D. Cavouras, and J. Stonham, "Real time decision support system for diagnosis of rare cancers, trained in parallel, on a graphics processing unit," *Computers in Biology and Medicine*, vol. 42, no. 4, pp. 376–386, Apr. 2012. [Online]. https://www.sciencedirect.com/science/article/pii/S0010482511002381

[72] H. Ren, J. Wang, and W. X. Zhao, "Generative adversarial networks enhanced pre-training for insufficient electronic health records modeling," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 3810–3818.

[73] L. Qiu, S. Gorantla, V. Rajan, and B. C. Y. Tan, "Multi-disease predictive analytics: A clinical knowledge-aware approach," *ACM Trans. Manage. Inf. Syst.*, vol. 12, no. 3, Jun. 2021.

[74] M. McDermott, B. Nestor, E. Kim, W. Zhang, A. Goldenberg, P. Szolovits, and M. Ghassemi, "A comprehensive ehr timeseries pre-training benchmark," in *Proceedings of the Conference on Health, Inference, and Learning*, ser. CHIL '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 257–278.

[75] R. Noll, M. Minor, A. Berger, L. Naab, M. Bay, H. Storf, and J. Schaaf, "Conception, development and validation of classification methods for coding support of rare diseases using artificial intelligence." *Studies in health technology and informatics*, vol. 295, pp. 422–425, Jun. 2022.

[76] A. R. Porras, M. Summar, and M. G. Linguraru, "Objective differential diagnosis of noonan and williams-beuren syndromes in diverse populations using quantitative facial phenotyping." *Molecular Genetics & Genomic Medicine*, vol. 9, p. e1636, May 2021.

[77] N. L. Rider, G. Cahill, T. Motazedi, L. Wei, A. Kurian, L. M. Noroski, F. O. Seeborg, I. K. Chinn, and K. Roberts, "Pi prob: A risk prediction and clinical guidance system for evaluating patients with recurrent infections." *PloS one*, vol. 16, p. e0237285, 2021.

[78] M. M. Segal, R. George, P. Waltman, A. W. El-Hattab, K. N. James, V. Stanley, and J. Gleeson, "Clinician-centric diagnosis of rare genetic diseases: performance of a gene pertinence metric in decision support for clinicians," *Orphanet Journal of Rare Diseases*, vol. 15, no. 1, p. 191, Jul. 2020.

[79] A. M. Cohen, S. Chamberlin, T. Deloughery, M. Nguyen, S. Bedrick, S. Meninger, J. J. Ko, J. J. Amin, A. J. Wei, and W. Hersh, "Detecting rare diseases in electronic health records using machine learning and knowledge engineering: Case study of acute hepatic porphyria," *PloS One*, vol. 15, no. 7, p. e0235574, 2020.

[80] L. Grigull, S. Mehmecke, A.-K. Rother, S. Blöß, C. Klemann, U. Schumacher, U. Mücke, X. Kortum, W. Lechner, and F. Klawonn, "Common pre-diagnostic features in individuals with different rare diseases represent a key for diagnostic support with computerized pattern recognition?" *PloS One*, vol. 14, no. 10, p. e0222637, 2019.

[81] B. Ehsani-Moghaddam, J. A. Queenan, J. MacKenzie, and R. V. Birtwhistle, "Mucopolysaccharidosis type II detection by Naïve Bayes Classifier: An example of patient classification for a rare disease using electronic medical records from the Canadian Primary Care Sentinel Surveillance Network," *PloS One*, vol. 13, no. 12, p. e0209018, Dec. 2018.

[82] A. B. Holmes, A. Hawson, F. Liu, C. Friedman, H. Khiabanian, and R. Rabadan, "Discovering disease associations by integrating electronic clinical data and medical literature." *PloS one*, vol. 6, p. e21132, 2011.

[83] D. Araújo-Vilar et al., "Lipoddx: a mobile application for identification of rare lipodystrophy syndromes." *Orphanet journal of rare diseases*, vol. 15, p. 81, Apr. 2020.

[84] L. Jamian, L. Wheless, L. J. Crofford, and A. Barnado, "Rule-based and machine learning algorithms identify patients with systemic sclerosis accurately in the electronic health record," *Arthritis Research & Therapy*, vol. 21, no. 1, p. 305, Dec. 2019.

[85] Q. Li, K. Zhao, C. D. Bustamante, X. Ma, and W. H. Wong, "Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis," *Genetics in Medicine*, vol. 21, no. 9, pp. 2126–2134, Sep. 2019. [Online]. https://www.nature.com/articles/s41436-019-0439-8

[86] T.-C. Hsieh et al, "Gestaltmatcher facilitates rare disease matching using facial phenotype descriptors," *Nature Genetics*, vol. 54, no. 3, pp. 349–357, Feb. 2022.

[87] L. Kühnle, U. Mücke, W. M. Lechner, F. Klawonn, and L. Grigull, "Development of a social network for people without a diagnosis (rarepairs): Evaluation study," *Journal of Medical Internet Research*, vol. 22, no. 9, p. e21849, Sep. 2020. [Online]. http://www.jmir.org/2020/9/e21849/

[88] G. Mangeat, R. Ouellette, M. Wabartha, B. De Leener, M. Plattén, V. Danylaité Karrenbauer, M. Warntjes, N. Stikov, C. Mainero, J. Cohen-Adad, and T. Granberg, "Machine learning and multiparametric brain mri to differentiate hereditary diffuse leukodystrophy with spheroids from multiple sclerosis." *Journal of neuroimaging : official journal of the American Society of Neuroimaging*, vol. 30, pp. 674–682, Sep. 2020.

[89] H. Dong, V. Suárez-Paniagua, H. Zhang, M. Wang, A. Casey, E. Davidson, J. Chen, B. Alex, W. Whiteley, and H. Wu, "Ontology-driven and weakly supervised rare disease identification from clinical notes," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 86, 2023.

[90] H. Dong, V. Suarez-Paniagua, H. Zhang, M. Wang, E. Whitfield, and H. Wu, "Rare disease identification from clinical notes with ontologies and weak supervision." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2021, pp. 2294–2298, Nov. 2021.

[91] A. Berglund, M. Olsen, M. Andersen, E. H. Nielsen, U. Feldt-Rasmussen, C. Kistorp, C. H. Gravholt, and K. Stochhholm, "Evaluation of icd-10 algorithms to identify hypopituitary patients in the danish national patient registry." *Clinical epidemiology*, vol. 9, pp. 75–82, 2017.

[92] L. M. Echeverry-Quiceno, E. Candelo, E. Gómez, P. Solís, D. Ramírez, D. Ortiz, A. González, X. Sevillano, J. C. Cuéllar, H. Pachajoa, and N. Martínez-Abadías, "Population-specific facial traits and diagnosis accuracy of genetic and rare diseases in an admixed colombian population," *Scientific Reports*, vol. 13, no. 1, p. 6869, 2023.

[93] T. Fujiwara, J.-M. Shin, and A. Yamaguchi, "Advances in the development of pubcasefinder, including the new application programming interface and matching algorithm," *Human Mutation*, vol. 43, no. 6, pp. 734–742, Jun. 2022.

[94] X. Chen, C. Faviez, M. Vincent, N. Garcelon, S. Saunier, and A. Burgun, "Identification of similar patients through medical concept embedding from electronic health records: A feasibility study for rare disease diagnosis." *Studies in health technology and informatics*, vol. 281, pp. 600–604, May 2021.

[95] F. Shen, S. Liu, Y. Wang, A. Wen, L. Wang, and H. Liu, "Utilization of Electronic Medical Records and Biomedical Literature to Support the Diagnosis of Rare Diseases Using Data Fusion and Collaborative Filtering Approaches," *JMIR Medical Informatics*, vol. 6, no. 4, p. e11301, Oct. 2018. [Online]. https://medinform.jmir.org/2018/4/e11301

[96] L. Kraus, O. Kremmyda, T. Bremova-Ertl, S. Barceló, K. Feil, and M. Strupp, "An algorithm as a diagnostic tool for central ocular motor disorders, also to diagnose rare disorders." *Orphanet journal of rare diseases*, vol. 14, p. 193, Aug. 2019.

[97] S. Ronicke, M. C. Hirsch, E. Türk, K. Larionov, D. Tientcheu, and A. D. Wagner, "Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study," *Orphanet Journal of Rare Diseases*, vol. 14, no. 1, p. 69, Mar. 2019.

[98] L. Lyu, I. E. Toubal, and K. Palaniappan, "Multi-expert deep networks for multi-disease detection in retinal fundus images," in *2022 44th Annual International*

Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow, Scotland, United Kingdom: IEEE, Jul. 2022, pp. 1818–1822.

[99] R. Zhang, E. Haihong, L. Yuan, J. He, H. Zhang, S. Zhang, Y. Wang, M. Song, and L. Wang, "Mbnm: Multi-branch network based on memory features for long-tailed medical image recognition." *Computer methods and programs in biomedicine*, vol. 212, p. 106448, Nov. 2021.

[100] Y. Gurovich, "The path to and impact of disease recognition with ai." *IEEE pulse*, vol. 11, pp. 13–16, Jan. 2020.

[101] R. Colbaugh, K. Glass, C. Rudolf, and M. Tremblay Volv Global Lausanne Switzerland, "Learning to identify rare disease patients from electronic health records." *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2018, pp. 340–347, 2018.

[102] C. Yang, Z. Li, H. Fan, Y. Chen, N. Lin, and N. Wei, "Serg: A sequence-to-sequence model for chinese ecg report generation," in *Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence*, ser. RICAI '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 708–712.

[103] Y. Tan, C. Yang, X. Wei, C. Chen, W. Liu, L. Li, J. Zhou, and X. Zheng, "Metacare++: Meta-learning with hierarchical subtyping for cold-start diagnosis prediction in healthcare data," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 449–459.

[104] W. He, X. Mao, C. Ma, Y. Huang, J. M. Hernàndez-Lobato, and T. Chen, "Bsoda: A bipartite scalable framework for online disease diagnosis," in *Proceedings of the ACM Web Conference 2022*, ser. WWW '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 2511–2521.

[105] T. Okumura and M. Tagawa, "Hierarchical representation of differential diagnosis lists for clinical decision support systems," in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '14. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014, p. 231–234.

[106] A. Yale, S. Dash, R. Dutta, I. Guyon, A. Pavao, and K. P. Bennett, "Generation and evaluation of privacy preserving synthetic health data," *Neurocomputing*, vol. 416, pp. 244–255, Nov. 2020. [Online]. https://www.sciencedirect.com/science/article/pii/S0925231220305117

[107] F. Akbari, K. Sartipi, and N. Archer, "Synthetic behavior sequence generation using generative adversarial networks," *ACM Trans. Comput. Healthcare*, vol. 4, no. 1, Feb. 2023.

[108] N. R. Veeraragavan and J. F. Nygård, "Securing federated gans: Enabling synthetic data generation for health registry consortiums," in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, ser. ARES '23. New York, NY, USA: Association for Computing Machinery, Aug. 2023.

[109] H. Zhang, Z. Huang, and Z. Lv, "Medical image synthetic data augmentation using gan," in *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, ser. CSAE '20. New York, NY, USA: Association for Computing Machinery, Oct. 2020.

[110] M. Guillaudeux, O. Rousseau, J. Petot, Z. Bennis, C.-A. Dein, T. Goronflot, N. Vince, S. Limou, M. Karakachoff, M. Wargny, and P.-A. Gourraud, "Patient-centric synthetic data generation, no reason to risk re-identification in biomedical data analysis." *NPJ digital medicine*, vol. 6, p. 37, Mar. 2023.

[111] I. H. Rather and S. Kumar, "Generative adversarial network based synthetic data training model for lightweight convolutional neural networks," *Multimedia tools and applications*, vol. 82, no. 20, pp. 1–23, May 2023.

[112] K. Ryu, Y. Nam, S.-M. Gho, J. Jang, H.-J. Lee, J. Cha, H. J. Baek, J. Park, and D.-H. Kim, "Data-driven synthetic mri flair artifact correction via deep neural network." *Journal of magnetic resonance imaging : JMRI*, vol. 50, pp. 1413–1423, Nov. 2019.

[113] R. Ganguli, R. Lad, A. Lin, and X. Yu, "Novel generative recurrent neural network framework to produce accurate, applicable, and deidentified synthetic medical data for patients with metastatic cancer." *JCO clinical cancer informatics*, vol. 7, p. e2200125, May 2023.

[114] M. Nikbakht, A. H. Gazi, J. Zia, S. An, D. J. Lin, O. T. Inan, and R. Kamaleswaran, "Synthetic seismocardiogram generation using a transformer-based neural network." *Journal of the American Medical Informatics Association : JAMIA*, vol. 30, pp. 1266–1273, Jun. 2023.

[115] J. Dahmen and D. Cook, "Synsys: A synthetic data generation system for healthcare applications," *Sensors*, vol. 19, no. 5, p. 1181, Mar. 2019.

[116] D. Kaur, M. Sobiesk, S. Patil, J. Liu, P. Bhagat, A. Gupta, and N. Markuzon, "Application of bayesian networks to generate synthetic health data." *Journal of the American Medical Informatics Association : JAMIA*, vol. 28, pp. 801–811, Mar. 2021.

[117] G. Gogoshin, S. Branciamore, and A. S. Rodin, "Synthetic data generation with probabilistic bayesian networks." *Mathematical biosciences and engineering : MBE*, vol. 18, pp. 8603–8621, Oct. 2021.

[118] H. MacLeod, G. Bastin, L. S. Liu, K. Siek, and K. Connelly, "'"be grateful you don't have a real disease": Understanding rare disease relationships," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, May 2017, p. 1660–1673. [Online]. https://doi.org/10.1145/3025453.3025796

[119] J. Misztal-Radecka and B. Indurkhya, "Persona prototypes for improving the qualitative evaluation of recommendation systems," in *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, ser. UMAP '20 Adjunct. New York, NY, USA: Association for Computing Machinery, Jul. 2020, p. 206–212. [Online]. https://doi.org/10.1145/3386392.3399297

[120] K. S. E. Payton and J. B. Gould, "Vignette research methodology: An essential tool for quality improvement collaboratives." *Healthcare (Basel, Switzerland)*, vol. 11, Dec. 2022.

[121] M. Hammoud, S. Douglas, M. Darmach, S. Alawneh, S. Sanyal, and Y. Kanbour, "Avey: An accurate ai algorithm for self-diagnosis," *medRxiv*, p. 2022.03.08.22272076, Jan. 2022. [Online]. http://medrxiv.org/content/early/2022/03/11/2022.03.08.22272076.abstract

[122] R. Hartson, P. Pyla, R. Hartson, and P. Pyla, "Chapter 19 - background: Design," in *The UX Book (Second Edition)*. Boston: Morgan Kaufmann, Jan. 2019, pp. 397–401. [Online]. https://www.sciencedirect.com/science/article/pii/B9780128053423000199

[123] C. Vargas, J. Whelan, J. Brimblecombe, and S. Allender, "Co-creation, co-design, co-production for public health - a perspective on definitions and distinctions," *Public Health Research & Practice*, vol. 32, p. e3222211, 2022.

[124] J. Zhang and S. P. Curley, "Exploring explanation effects on consumers' trust in online recommender agents," *International Journal of Human–Computer Interaction*, vol. 34, no. 5, pp. 421–432, 2018.

[125] R. Gouveia and D. A. Epstein, "This watchface fits with my tattoos: Investigating customisation needs and preferences in personal tracking," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023.

[126] S. M. Harmon, E. M. Heitkemper, L. Mamykina, and M. L. Hwang, "Are we healthier together? two strategies for supporting macronutrient assessment skills and how the crowd can help (or not)," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, Nov. 2022.

[127] M. D. Molina, E. S. Zhan, D. Agnihotri, S. Abdullah, and P. Deka, "Motivation to use fitness application for improving physical activity among hispanic users: The pivotal role of interactivity and relatedness," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023.

[128] A. Min, W. R. Miller, L. M. Rocha, K. Börner, R. Brattig Correia, and P. C. Shih, "Understanding contexts and challenges of information management for epilepsy care," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023.

[129] T. Ahmed, M. M. Rahman, E. Nemati, M. Y. Ahmed, J. Kuang, and A. J. Gao, "Remote breathing rate tracking in stationary position using the motion and acoustic

sensors of earables," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023.

[130] Y. Li, F. M. Li, and P. Carrington, "Breaking the "inescapable" cycle of pain: Supporting wheelchair users' upper extremity health awareness and management with tracking technologies," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023.

[131] C. Free, G. Phillips, L. Galli, L. Watson, L. Felix, P. Edwards, V. Patel, and A. Haines, "The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review." *PLoS medicine*, vol. 10, p. e1001362, Jan. 2013.

[132] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Patel, and A. Haines, "The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis." *PLoS medicine*, vol. 10, p. e1001363, Jan. 2013.

[133] Y. You, C.-H. Tsai, Y. Li, F. Ma, C. Heron, and X. Gui, "Beyond self-diagnosis: How a chatbot-based symptom checker should respond," *ACM Trans. Comput.-Hum. Interact.*, Mar. 2023.

[134] W. Li, X. Fan, H. Zhu, J. Wu, and D. Teng, "Research on the influencing factors of user trust based on artificial intelligence self diagnosis system," in *Proceedings of the ACM Turing Celebration Conference - China*, ser. ACM TURC '20. New York, NY, USA: Association for Computing Machinery, Oct. 2020, p. 197–202.

[135] H. MacLeod, K. Oakes, D. Geisler, K. Connelly, and K. Siek, "Rare world: Towards technology for rare diseases," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, Apr. 2015, p. 1145–1154. [Online]. https://doi.org/10.1145/2702123.2702494

[136] S. Ashtari and A. Taylor, "Patients with rare diseases and the power of online support groups: Implications for the medical community." *JMIR formative research*, vol. 7, p. e41610, Sep. 2023.

[137] A. Kushniruk and C. Nøhr, "Participatory design, user involvement and health it evaluation," *Stud Health Technol Inform*, vol. 222, pp. 139–151, Jan. 2016.

[138] J. D. Gould and C. Lewis, "Designing for usability: Key principles and what designers think," *Commun. ACM*, vol. 28, no. 3, p. 300–311, Mar. 1985.

[139] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[140] C. Ham, "Money can't buy you satisfaction." *BMJ (Clinical research ed.)*, vol. 330, pp. 597–9, Mar. 2005.

[141] Swan UK. (2021) Home. Genetic Alliance. [Online]. https://www.undiagnosed.org.uk/

[142] S. Çerçi, M. E. Cecchinato, and J. Vines, "How design researchers interpret probes: Understanding the critical intentions of a designerly approach to research," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021. [Online]. https://doi.org/10.1145/3411764.3445328

[143] K. Boehner, J. Vertesi, P. Sengers, and P. Dourish, "How hci interprets the probes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. New York, NY, USA: Association for Computing Machinery, Apr. 2007, p. 1077–1086. [Online]. https://doi.org/10.1145/1240624.1240789

[144] L. Sigmundsdottir, M. C. Tchan, A. A. Knopman, G. C. Menzies, J. Batchelor, and D. O. Sillence, "Cognitive and psychological functioning in fabry disease." *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, vol. 29, pp. 642–50, Nov. 2014.

[145] T. Kitami et al, "Deep phenotyping of myalgic encephalomyelitis/chronic fatigue syndrome in japanese population." *Scientific reports*, vol. 10, p. 19933, Nov. 2020.

[146] L. S. Canas et al., "Prion disease diagnosis using subject-specific imaging biomarkers within a multi-kernel Gaussian process," *NeuroImage. Clinical*, vol. 24, p. 102051, Oct. 2019.

[147] K. Chennen, T. Weber, X. Lornage, A. Kress, J. Böhm, J. Thompson, J. Laporte, and O. Poch, "MISTIC: A prediction tool to reveal disease-relevant deleterious missense variants," *PLOS ONE*, vol. 15, no. 7, p. e0236962, Jul. 2020. [Online]. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0236962

[148] T. Fujiwara, Y. Yamamoto, J.-D. Kim, O. Buske, and T. Takagi, "PubCaseFinder: A Case-Report-Based, Phenotype-Driven Differential-Diagnosis System for Rare Diseases," *American Journal of Human Genetics*, vol. 103, no. 3, pp. 389–399, Sep. 2018.

[149] X. Li, L. Yu, Y. Jin, C.-W. Fu, L. Xing, and P.-A. Heng, "Difficulty-aware meta-learning for rare disease diagnosis," in *Proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds., vol. 12261. Cham: Springer International Publishing, Jul. 2020, pp. 357–366.

[150] D. Trivedi and V. Gopalakrishnan, "Medicrawl - a web search engine for diseases," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 2021, pp. 0148–0157.

[151] J. W. Keeling, A. M. Turner, E. E. Allen, S. A. Rowe, J. A. Merrill, E. D. Liddy, and H. R. Turtle, "Development and evaluation of a prototype search engine to meet public health information needs." *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2011, pp. 693–700, 2011. [Online]. https://pubmed.ncbi.nlm.nih.gov/22195125/

[152] W. Honekamp and H. Ostermann, "Evaluation of a prototype health information system using the fitt framework." *Informatics in primary care*, vol. 19, pp. 47–9, Jan. 2011.

[153] Y. Sun and S. S. Sundar, "Exploring the effects of interactive dialogue in improving user control for explainable online symptom checkers," in *Extended Abstracts of the*

2022 CHI Conference on Human Factors in Computing Systems, ser. CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022.

[154] X. Xu, X. Yan, and T. R. Dillahunt, "Reaching hard-to-reach populations: An analysis of survey recruitment methods," in *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, p. 428–432.

[155] J. Nielsen, *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. [Online]. https://dl.acm.org/doi/abs/10.5555/529793

[156] J. Mackenzie, K. Gupta, F. Qiao, A. H. Awadallah, and M. Shokouhi, "Exploring user behavior in email re-finding tasks," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, May 2019, p. 1245–1255.

[157] J. He, M. Bron, L. Azzopardi, and A. de Vries, "Studying user browsing behavior through gamified search tasks," in *Proceedings of the First International Workshop on Gamification for Information Retrieval*, ser. GamifIR '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 49–52.

[158] M. Ageev, Q. Guo, D. Lagun, and E. Agichtein, "Find it if you can: A game for modeling different types of web search success using interaction data," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '11. New York, NY, USA: Association for Computing Machinery, Jul. 2011, p. 345–354.

[159] M. Riegler, R. Eg, L. Calvet, M. Lux, P. Halvorsen, and C. Griwodz, "Playing around the eye tracker - a serious game based dataset," in *In proceedings of the GamifIR'15 Workshop*. Vienna, Austria: CEUR Workshop Proceedings, Mar. 2015. [Online]. https://ceur-ws.org/Vol-1345/gamifir15_5.pdf

[160] R. Murray-Smith, A. Oulasvirta, A. Howes, J. Müller, A. Ikkala, M. Bachinski, A. Fleig, F. Fischer, and M. Klar, "What simulation can do for hci research," *Interactions*, vol. 29, no. 6, p. 48–53, Nov. 2022.

[161] K. Leichtenstern, E. André, and M. Rehm, "Using the hybrid simulation for early user evaluations of pervasive interactions," in *Proceedings of the 6th Nordic*

Conference on Human-Computer Interaction: Extending Boundaries, ser. NordiCHI '10. New York, NY, USA: Association for Computing Machinery, Oct. 2010, p. 315–324. [Online]. https://doi.org/10.1145/1868914.1868952

[162] R. W. White, I. Ruthven, J. M. Jose, and C. J. V. Rijsbergen, "Evaluating implicit feedback models using searcher simulations," *ACM Transactions on Information Systems*, vol. 23, no. 3, p. 325–361, Jul. 2005. [Online]. https://doi.org/10.1145/1080343.1080347

[163] O. Bouidghaghen, L. Tamine-Lechani, G. Pasi, G. Cabanac, M. Boughanem, and C. da Costa Pereira, "Prioritized aggregation of multiple context dimensions in mobile ir," in *Proceedings of the 7th Asia Conference on Information Retrieval Technology*, ser. AIRS'11. Berlin, Heidelberg: Springer-Verlag, Dec. 2011, p. 169–180. [Online]. https://doi.org/10.1007/978-3-642-25631-8_16

[164] Y. Guerrier, J. Naveteur, C. Kolski, and F. Anceaux, "Discount evaluation of preliminary versions of systems dedicated to users with cerebral palsy: Simulation of involuntary movements in non-disabled participants," in *Universal Access in Human-Computer Interaction. Design Methods and User Experience*, M. Antona and C. Stephanidis, Eds. Cham: Springer International Publishing, Jul. 2021, pp. 71–88.

[165] S. Zhang and K. Balog, "Evaluating conversational recommender systems via user simulation," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1512–1520. [Online]. https://doi.org/10.1145/3394486.3403202

[166] P. Colomba, C. Zizzo, R. Alessandro, G. Cammarata, S. Scalia, A. Giordano, M. Pieroni, L. Sicurella, L. Amico, A. Burlina, and G. Duro, "Fabry disease and multiple sclerosis misdiagnosis: the role of family history and neurological signs." *Oncotarget*, vol. 9, pp. 7758–7762, Jan. 2018.

[167] E. C. H. Feldman, K. J. Homan, S. E. Williams, T. V. Ting, K. R. Goldschneider, and S. Kashikar-Zuck, "A narrative review of the literature on illness uncertainty in hypermobile ehlers-danlos syndrome: implications for research and clinical practice." *Pediatric rheumatology online journal*, vol. 21, p. 121, Oct. 2023.

[168] C. M. E. Halverson, S. Cao, S. M. Perkins, and C. A. Francomano, "Comorbidity, misdiagnoses, and the diagnostic odyssey in patients with hypermobile ehlers-danlos syndrome," *Genetics in Medicine Open*, vol. 1, no. 1, p. 100812, Apr. 2023. [Online]. https://www.sciencedirect.com/science/article/pii/S294977442300821X

[169] C. L. Marchesoni, N. Roa, A. M. Pardal, P. Neumann, G. Cáceres, P. Martínez, I. Kisinovsky, S. Bianchi, A. L. Tarabuso, and R. C. Reisin, "Misdiagnosis in fabry disease." *The Journal of pediatrics*, vol. 156, pp. 828–31, May 2010.

[170] P. K. Mistry, M. D. Cappellini, E. Lukina, H. Özsan, S. Mach Pascual, H. Rosenbaum, M. Helena Solano, Z. Spigelman, J. Villarrubia, N. P. Watman, and G. Massenkeil, "A reappraisal of gaucher disease–diagnosis and disease management algorithms," *Am. J. Hematol.*, vol. 86, no. 1, pp. 110–115, Jan. 2011.

[171] A. Mehta, R. Ricci, U. Widmer, F. Dehout, A. Garcia de Lorenzo, C. Kampmann, A. Linhart, G. Sunder-Plassmann, M. Ries, and M. Beck, "Fabry disease defined: baseline clinical manifestations of 366 patients in the fabry outcome survey," *European journal of clinical investigation*, vol. 34, no. 3, pp. 236–242, 2004.

[172] P. K. Mistry, M. D. Cappellini, E. Lukina, H. Özsan, S. M. Pascual, H. Rosenbaum, M. H. Solano, Z. Spigelman, J. Villarrubia, N. P. Watman *et al.*, "Consensus conference: A reappraisal of gaucher disease-diagnosis and disease management algorithms," *American journal of hematology*, vol. 86, no. 1, p. 110, Jan. 2011.

[173] A. Mehta, R. Ricci, U. Widmer, F. Dehout, A. Garcia de Lorenzo, C. Kampmann, A. Linhart, G. Sunder-Plassmann, M. Ries, and M. Beck, "Fabry disease defined: baseline clinical manifestations of 366 patients in the fabry outcome survey," *European Journal of Clinical Investigation*, vol. 34, no. 3, pp. 236–242, Mar. 2004.

[174] W. Häuser, S. Perrot, C. Sommer, Y. Shir, and M.-A. Fitzcharles, "Diagnostic confounders of chronic widespread pain: not always fibromyalgia." *Pain reports*, vol. 2, p. e598, May 2017.

[175] M. Castori, C. Celletti, F. Camerota, and P. Grammatico, "Chronic fatigue syndrome is commonly diagnosed in patients with ehlers-danlos syndrome hypermobility

type/joint hypermobility syndrome." *Clinical and experimental rheumatology*, vol. 29, pp. 597–8, May 2011. [Online]. https://pubmed.ncbi.nlm.nih.gov/21722506/

[176] B. Berglund, G. Nordström, and K. Lützén, "Living a restricted life with ehlers-danlos syndrome (eds)," *International Journal of Nursing Studies*, vol. 37, no. 2, pp. 111–118, Apr. 2000. [Online]. https://www.sciencedirect.com/science/article/pii/S002074899900067X

[177] K. S. Yew, K. A. Kamps-Schmitt, and R. Borge, "Hypermobile ehlers-danlos syndrome and hypermobility spectrum disorders." *American family physician*, vol. 103, pp. 481–492, Apr. 2021. [Online]. https://pubmed.ncbi.nlm.nih.gov/33856167/

[178] S. A. Hershenfeld, S. Wasim, V. McNiven, M. Parikh, P. Majewski, H. Faghfoury, and J. So, "Psychiatric disorders in ehlers-danlos syndrome are frequent, diverse and strongly associated with pain," *Rheumatology International*, vol. 36, no. 3, pp. 341–348, Oct. 2016.

[179] Q. Wen, Z. Ouyang, J. Zhang, Y. Qian, Y. Ye, and C. Zhang, "Disentangled dynamic heterogeneous graph learning for opioid overdose prediction," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 2009–2019.

[180] C. S. Kruse, B. Smith, H. Vanderlinden, and A. Nealand, "Security techniques for the electronic health records." *Journal of medical systems*, vol. 41, p. 127, Aug. 2017.

[181] Orphanet. (1997) Phenotypes - orphadata. INSERM. [Online]. https://www.orphadata.com/phenotypes/

[182] S. Köhler et al., "The human phenotype ontology in 2021," *Nucleic Acids Res*, vol. 49, no. D1, pp. D1207–D1217, Jan. 2021.

[183] Genetic Alliance and The New York Mid-Atlantic Consortium for Genetic and Newborn Screening Services, *Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals*. Washington (DC): Genetic Alliance, Jul. 2009. [Online]. https://pubmed.ncbi.nlm.nih.gov/23304754/

[184] H. K. Walker, W. D. Hall, and J. W. Hurst, *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed. Boston: Butterworth-Heinemann Ltd, Apr. 1990, chapter 215 The Family History. [Online]. https://www.ncbi.nlm.nih.gov/books/NBK201/

[185] B. E. Dixon, A. M. Jabour, E. O'Kelly Phillips, and D. G. Marrero, "Challenges when engaging diabetic patients and their clinicians in using e-health technologies to improve clinical outcomes," in *Proceedings of the 8th International Conference on Universal Access in Human-Computer Interaction. Aging and Assistive Environments - Volume 8515*. Berlin, Heidelberg: Springer-Verlag, Jun. 2014, p. 238–247.

[186] S. Houben, M. Frost, and J. E. Bardram, "Hypr device: Mobile support for hybrid patient records," in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ser. ITS '14. New York, NY, USA: Association for Computing Machinery, Nov. 2014, p. 249–258.

[187] S. Houben, M. Frost, and J. E. Bardram, "Collaborative affordances of hybrid patient record technologies in medical work," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: Association for Computing Machinery, Feb. 2015, p. 785–797.

[188] S. Houben, M. Schmidt, M. Frost, and J. E. Bardram, "Hybrid patient record: supporting hybrid interaction in clinical wards," in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '15. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), May 2015, p. 270–271. [Online]. https://dl.acm.org/doi/10.5555/2826165.2826215

[189] S. R. Mishra, A. D. Miller, S. Haldar, M. Khelifi, J. Eschler, R. G. Elera, A. H. Pollack, and W. Pratt, "Supporting collaborative health tracking in the hospital: Patients' perspectives," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, Apr. 2018, p. 1–14.

[190] Y. G. Rajwan and G. R. Kim, "Medical information visualization conceptual model for patient-physician health communication," in *Proceedings of the 1st ACM*

International Health Informatics Symposium, ser. IHI '10. New York, NY, USA: Association for Computing Machinery, Nov. 2010, p. 512–516.

[191] F.-Y. Rao, J. Cao, E. Bertino, and M. Kantarcioglu, "Hybrid private record linkage: Separating differentially private synopses from matching records," *ACM Trans. Priv. Secur.*, vol. 22, no. 3, Apr. 2019.

[192] Y. Wang, Y. Sun, Z. Ma, L. Gao, and Y. Xu, "A hybrid model for named entity recognition on chinese electronic medical records," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 2, Apr. 2021.

[193] Z. Wu, C. Xiao, and J. Sun, "Medlink: De-identified patient health record linkage," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '23. New York, NY, USA: Association for Computing Machinery, Aug. 2023, p. 2672–2682.

[194] D. Wyatt, S. Lampon, and C. McKevitt, "Delivering healthcare's 'triple aim': electronic health records and the health research participant in the uk national health service," *Sociol Health Illn*, vol. 42, no. 6, pp. 1312–1327, Jul. 2020.

[195] Q. Bi, L. Shen, R. Evans, Z. Zhang, S. Wang, W. Dai, and C. Liu, "Determining the topic evolution and sentiment polarity for albinism in a chinese online health community: Machine learning and social network analysis." *Journal of Medical Internet Research Medical Informatics*, vol. 8, no. 5, p. e17813, May 2020.

[196] M. M. Gottlieb, D. J. Arenillas, S. Maithripala, Z. D. Maurer, M. Tarailo Graovac, L. Armstrong, M. Patel, C. van Karnebeek, and W. W. Wasserman, "Geneyenta: a phenotype-based rare disease case matching tool based on online dating algorithms for the acceleration of exome interpretation." *Human mutation*, vol. 36, no. 4, pp. 432–438, Apr. 2015.

[197] C. Rasi, D. Nilsson, M. Magnusson, N. Lesko, K. Lagerstedt-Robinson, A. Wedell, A. Lindstrand, V. Wirta, and H. Stranneheim, "Patientmatcher: A customizable python-based open-source tool for matching undiagnosed rare disease patients via the matchmaker exchange network," *Human Mutation*, vol. 43, no. 6, pp. 708–716, Jun. 2022.

[198] J. Yates, A. Gutiérrez-Sacristán, V. Jouhet, K. LeBlanc, C. Esteves, T. N. DeSain, N. Benik, J. Stedman, N. Palmer, G. Mellon, I. Kohane, and P. Avillach, "Finding commonalities in rare diseases through the undiagnosed diseases network." *Journal of the American Medical Informatics Association*, vol. 28, no. 8, pp. 1694–1702, Jul. 2021.

[199] R. Sriram. (2019, Aug.) New data reveals how many of us are misdiagnosing ourselves using google. The Kolabtree Blog. [Online]. https://www.kolabtree.com/blog/new-data-reveals-how-many-of-us-are-misdiagnosing-ourselves-using-google/

[200] F. Shen, S. Liu, Y. Wang, L. Wang, N. Afzal, and H. Liu, "Leveraging collaborative filtering to accelerate rare disease diagnosis." *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2017, pp. 1554–1563, Apr. 2017. [Online]. https://pubmed.ncbi.nlm.nih.gov/29854225/

[201] R. Dragusin, P. Petcu, C. Lioma, B. Larsen, H. L. Jørgensen, I. J. Cox, L. K. Hansen, P. Ingwersen, and O. Winther, "Findzebra: A search engine for rare diseases," *International Journal of Medical Informatics*, vol. 82, no. 6, pp. 528–538, Jun. 2013. [Online]. https://www.sciencedirect.com/science/article/pii/S1386505613000166

[202] F. Nunes and G. Fitzpatrick, "Self-care technologies and collaboration," *International Journal of Human–Computer Interaction*, vol. 31, no. 12, pp. 869–881, Nov. 2015.

[203] H. MacLeod, S. Yang, K. Oakes, K. Connelly, and S. Natarajan, "Identifying rare diseases from behavioural data: A machine learning approach," in *Proceedings of 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. Washington, DC, US: IEEE, Jun. 2016, pp. 130–139.