Improving GPs' knowledge of the benefits and harms of treatments to support decision making in multimorbidity: qualitative research and co-design of a novel electronic information resource

Julian Stephen Treadwell

Thesis submitted for the degree of

Doctor of Philosophy in Primary Health Care

Green Templeton College

Trinity 2023

Abstract

Background

General practitioners (GPs) regularly prescribe prolonged treatments for long-term conditions. However, GPs may benefit from further understanding of the absolute benefits and harms of these treatments, enhancing their ability to engage in shared decision-making and manage multimorbidity and polypharmacy.

Aim

To produce and evaluate a website to provide information on the benefits and harms of treatments for long-term conditions in a way that can be understood by GPs and potentially integrated into their practice.

Methods

The study consisted of three parts. First, a qualitative interview study and framework analysis with GPs exploring their attitudes to and understanding of the quantitative benefits and harms of treatments. Second, a participatory co-design process to design the website, coupled with a pragmatic approach to evidence collation to provide clinical content. Finally, an exploratory evaluation study of the website using online focus groups.

Results

The interview study reported findings on GPs' understanding of quantitative information on the benefits and harms of treatments which informed the co-design research. The co-design

research resulted in the creation of a website, *www.gpevidence.org* , which presents complex scientific information on treatment effect sizes and the nature and quality of the relevant clinical evidence.

The evaluation study showed that participating GPs were able to understand the clinical information on *GP Evidence,* and that in hypothetical scenarios this might change their prescribing practice. Some participants found some information confusing. There was limited evidence that this new information could be integrated into complex decision-making for multimorbidity and polypharmacy.

Conclusion

The aim of producing a website able to deliver information on the benefits and harms of treatments for long-term conditions to GPs was achieved. Further research is needed to evaluate the effect of *GP Evidence* in real-world practice.

Acknowledgements

Firstly, I would like to thank my team of academic supervisors: Prof. Trisha Greenhalgh, Prof. Kamal Mahtani, Prof. Gina Neff, Dr Jo Crocker and Dr Sietse Wieringa. It has been such a privilege to be supported by these wise and experience people who have brought such a range of perspectives to this project.

I am grateful to the National Institute for Health and Care Research who funded this research with a Doctoral Research Fellowship.

The members of the expert-patient steering committee gave invaluable support and guidance and were generous in time and spirit: Jon Brassey, Andy Hutchinson, Emily Lam, Kamal Mahtani, Mandy Payne, Jackie Walumbe.

Special thanks to Mandy Payne, whose editorial expertise and insights into healthcare experiences were so valuable in the final stages of production of *GP Evidence*.

Many thanks to colleagues in the Nuffield Department of Primary Care Health Sciences who were generous enough with their time to assist me with aspects of this work. Laura Armitage, Kiana Collins, Laura Heath, Olu Onyimadu and Aziza Wahedi undertook data reviews on quantitative content for *GP Evidence*. Rebekah Burrow and Amelia Talbot undertook user-testing with three GP participants.

A huge number of people have advised, inspired and assisted me over the last 10 years since the idea for this project arose. Each step on this long journey would probably not have happened without them: Patrick Craig-Macfeely, David Haslam, Sally Hayes, Iona Heath,

Richard Lehman, Margaret McCartney, Ray Moynihan, all those in the Overdiagnosis group at the RCGP and Imran Rafi to name but a few.

My partner Rebekah has been along with me on every step of this journey. Thank you for your love and companionship, encouragement, wisdom and support without which I may not have made it through to the end!

To my beautiful, brilliant boys Reuben and Gabriel, who came into this world during the course of this project: I would love to thank you for your help but frankly, you have been a little distracting.

Table of Contents

List of Tables

List of Figures

List of abbreviations

ACE-inhibitor	Angiotensin converting enzyme inhibitor
ARB	Angiotensin receptor-blocker
CDL	Content Design London
EBM	Evidence-based medicine
GP	General Practitioner
HRT	Hormone replacement therapy
IPDAS	International Patient Decision Aid Standards (IPDAS)
JAD	Joint application design (workshop)
MRCGP	Membership of the Royal College of General Practitioners
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
PDA	Patient decision aid
QIRx	Quantitative information on the benefits and harms of treatments for long-term conditions
QOF	Quality and outcomes framework

RCT	Randomised controlled trial
RtD	Research through design
SDM	Shared decision-making
UX	User experience

1 Part 1. Chapter 1: Introduction

This thesis describes the development through research of the website gpevidence.org. Aimed at General Practitioners, *GP Evidence* provides information on the benefits and harms of treatments for common long term conditions alongside summaries of the underlying evidence. The need for such an information resource became apparent to me from my own practice as a GP and as I later discovered, was being articulated in the literature. What had initially seemed like a simple idea turned out to be much more complex. The creation of something usable and useful drew on a wide range of previous research and debate. My own research further defined the problem to be addressed and explored this solution, paving the way for an iterative process of research through design. This was interwoven with a tailored, pragmatic strategy for the collation and curation of clinical evidence – the information content of the website.

I will begin this introduction with a personal narrative. Firstly, to explain my motivation to undertake this project, but also because my experience and understanding reflect to some extent the evolution of evidence-based medicine (EBM) and the practice of primary care in the UK – the fields to which this research makes a contribution.

1.1 A Personal Narrative

I graduated from medical school in 1994 and completed GP training in 1998. Evidence-based medicine was recently established as a leading paradigm. The idea that GPs should ensure their populations were treated according to the best available evidence was beginning to evolve. Clinical audit was the new "big idea", indeed a compulsory part of my GP Training was

to complete an audit cycle in the practice. I looked at how many of our patients with a previous history of cardiovascular disease were being prescribed aspirin – only about 70%.

I remember discussions with my GP trainer about the value of preventive treatments. The drug reps were telling us about bisphosphonates for the prevention of osteoporotic fractures – was a number-needed-to-treat (NNT) of 30 really good medicine? There was laughter at a GP trainee teaching session when a lipidologist raised the possibility that one day we might prescribe statins without lengthy attempts at dietary restriction to lower cholesterol. How would we persuade patients of the value of new preventive therapies? It seemed astonishing that one diagnosis (a myocardial infarction) would result in multiple lifelong medications (aspirin, beta-blocker, statin, ACE inhibitor).

Over this period I received good training in the principles of EBM and critical appraisal of research. The examination for the Membership of the Royal College of General Practitioners included a critical appraisal of research papers. The idea was that we would apply these skills, finding and using evidence to inform our clinical practice.

By the early 2000s I was a partner in a GP practice, including a role as lead for cardiovascular prevention. The concept of clinician-level appraisal of evidence had given way to guideline-led practice, whereby as GPs we were encouraged to offer standardised treatments to patients with the understanding that the assessment of the clinical evidence had been done for us. This change was accelerated with the introduction of a new GP contract in 2004 including the Quality and Outcomes Framework (QOF), conferring financial incentives to deliver systematic testing and treatment for a number of long-term conditions.

Looking back, this shift in thinking and practice appeared to happen naturally, perhaps as a consequence of the inherently unrealistic nature of the idea of clinicians seeking, appraising and acting on the results of original research. I recall feeling enthusiastic about implementing guideline-based care for our practice population, assuming that it represented high value medicine for the individual patients.

Over the next decade, my perspective began to shift. The number of clinical guidelines to follow expanded, as had the number of patients with long term conditions and the number of interventions on offer. The QOF expanded too, increasing pressure on GPs to prescribe more interventions for more patients. Multiple medicines became the norm from middle age. I began to feel a sense of unease – perceiving harms from over-medicalisation and wondering just how valuable much of what we were doing actually was. However, for years the busyness of clinical practice prevented me from exploring this deeply. The enquiring mind of the GP registrar was overshadowed by the pragmatic, task oriented mind of the practising doctor.

In 2012, an edition of the British Medical Journal landed on the doormat. On the cover was a cartoon of a hapless patient lying on the ground being assaulted by a tribe of Lilliputian medics wielding scanners and syringes. It contained a feature article, “Preventing Overdiagnosis: how to stop harming the healthy”(1) which eloquently described the mechanisms of expanding medicalisation and the harmful consequences arising alongside sometimes small benefits. This spoke immediately to my sense of unease described above and revealed a world of research and academic activity which had hitherto been unknown to me. Simultaneously it announced an international conference to explore the subject. My decision to get involved and contribute to this was the starting point for what would become a ten year journey.

Whilst exploring the varied topics around overdiagnosis and “Too Much Medicine”, one issue grabbed my attention, relating strongly to my clinical practice and re-awakening a line of thinking from my earlier career. It was that the size (or likelihood) of the benefits of treatments for long term conditions were, from the perspective of an individual patient, highly variable. In addition, the strength of evidence to support guideline recommendations was variable. The clinical guidelines I had been following for the last decade gave the impression of uniformly high value treatments, with high certainty evidence. Though information about evidence quality, the chance of treatment benefits and (to a lesser extent) harms was available in the full guideline documents, exploring these took hours or days and much of the complex information was over my head.

There followed a revelation: that I had often been unwittingly prescribing treatments with an inflated idea of their benefits; the limited benefits of tight glycaemic control in Type 2 diabetes compared to moderate control provide a good example (2). Conversely, I had been sluggish in initiating higher-value treatments such as anticoagulation in atrial fibrillation because my understanding of the benefit-harm ratio was flawed.

It is worth emphasising at this point, that although my initial interest was stimulated by a perception of overdiagnosis and overtreatment, I came to appreciate that under-diagnosis and under-treatment are as much of a problem. These two problems have been described as “conjoined twins” (3). My project addresses both.

From this revelation came a simple idea: “Wouldn’t it be great if information on the benefits and harms of treatments was easily available to GPs and their patients? The evidence is there, and could easily be put online.”

It struck me that this might profoundly change the way we practice. Rather than uncritically following guidelines, GPs could offer patients real choice and tailor complex treatment regimens, thereby coming closer to maximising benefit and minimising harms in line with their patients' preferences.

This thesis describes how the idea became a reality, through research and design.

1.2 Overall aim and research questions

The overall aim of the doctoral project was to develop and evaluate a website to improve GPs' understanding of the benefits and harms of treatments for long-term conditions. This aim will be further described at the end of chapter 2 (section 2.10). To achieve this aim, the research addressed three main questions:

1. What are GPs' needs and preferences for accessing evidence-based quantitative information on the benefits and harms of medications when managing patients with multi-morbidity and polypharmacy?
2. How can we draw on design principles and clinician and patient input to co-design an interactive information resource to support informed conversations and shared decision-making?
3. To what extent might such a resource be usable in practice?

1.3 Overview

The project had three main phases of work to address these research questions.

1. A qualitative interview study with GPs with a framework analysis.
2. Design and creation of the *GP Evidence* website combining a variety of participatory co-design methods and a structured, pragmatic approach to clinical evidence collation.
3. An exploratory evaluation study of the finished website with GPs, using online focus groups with an embedded online survey.

1.4 Outline of thesis chapters

The thesis is structured in three parts, reflecting the research questions and phases of work.

Part 1

Chapter 1: Introduction and context

Chapter 1 describes my personal background, outline of the thesis and my epistemological approach. I include a brief summary of training and exploratory work undertaken prior to commencing the doctoral research which directly informed it. With reference to the literature, I discuss the context in which the project takes place from a perspective of clinical practice, evidence-based medicine, clinical guidelines, multimorbidity and polypharmacy.

Chapter 2: Background and aims

Chapter 2 summarises literature relevant to clinicians' understanding of the benefits and harms of treatments, including an online survey of British GPs I undertook prior to the doctoral research. I summarise research on GPs' relationship with clinical guidelines, outline work over the last two decades in the field of shared-decision making, and summarise what

is known about patients' preferences in this area. I describe patient and public involvement in the research and conclude by summarising the case for the development of *GP Evidence* and the aims and scope of the project.

Chapter 3: GP interviews

Chapter 3 describes a series of qualitative interviews with GPs with two aims. Firstly, to explore their attitudes to, and understanding of the quantitative benefits and harms of treatments for long term conditions as a foundation for the next phase of research. Secondly, to undertake user-testing of existing websites with similar goals to *GP Evidence*, to explore the GPs' needs and preferences regarding the presentation of information. I begin by justifying and explaining my chosen methods of semi-structured interviews including the "think-aloud" technique and a framework method of analysis. The main results of the study with discussion and conclusions are included in this chapter. A sub-set of the methods and results are reported in chapters 4 and 6 respectively where they fit most appropriately in the design research phase.

Part 2

Chapter 4: Methods part 1. Design research - methods and principles

Chapter 4 summarises my exploration of the principles of design and design research, with examples from inside and outside healthcare. I explain my understanding of an overall approach to design which was applied throughout this phase of research. I then describe a series of discrete design research methods which were used to develop *GP Evidence*.

Chapter 5: Methods part 2. Evidence collation - methods, approach and process

Chapter 5 describes my approach to the selection, collation and presentation of clinical evidence. This aspect of the project required a strategy that was simultaneously rigorous and trustworthy whilst being pragmatic enough to be feasible. I describe the development of an approach in partnership with a patient and expert group, who went on to form a steering committee for the project.

Chapter 6: Results part 1. The design of *GP Evidence*

Chapter 6 describes the iterative design research process which took place over three years, employing the methods described in Chapter 4. Using examples of content from the website, I illustrate how answers were found regarding how to present complex content in a way which GPs could understand, would find appealing and which would be usable in practice.

Chapter 7: Results part 2. Selection, collation and presentation of clinical evidence

Chapter 7 shows how the approach to evidence collation described in chapter 5 was used to develop the clinical content for *GP Evidence*. Drawing from the finished website, I illustrate and discuss a range of examples. Some of these show a relatively straightforward process of evidence sourcing and development of clinical content. Others reveal challenges which arose requiring careful consideration, compromise and innovative methods of presentation in order to produce clinical content which was simultaneously understandable and useful whilst remaining robust and trustworthy.

Part 3

Chapter 8: Exploratory evaluation study

Chapter 8 describes an online focus group study with GPs to provide an evaluation of the completed *GP Evidence* website. The aim was to explore whether the design and content successfully enabled GPs to access and understand information in a way which might be useful and usable in practice. I begin by justifying and explaining my chosen methods: online focus groups with an embedded online survey and a descriptive thematic analysis. The results of the study with discussion and conclusions about the design and content of *GP Evidence* are included in this chapter.

Chapter 9: Summary and Discussion

I review the findings of the project and discuss the strengths and limitations of the research. I consider what this has added to the field and suggest ideas for further research.

1.5 Theoretical perspective – pragmatism

In this thesis, I take a pragmatic perspective in accordance with the overall aim of developing a novel information resource for use in the real world. The main consideration is practical: will this resource be useful and helpful?

Pragmatism as a scientific philosophy judges the value of knowledge by its usefulness in a specific context, to answer practical questions or inform specific actions (4, 5). This is in contrast to a positivist approach to scientific enquiry which seeks to provide generalisable truths from empirical evidence (6). Pragmatic experimentation centres on the idea of active

enquiry with attention to real-world conditions where the success of the research process is measured by its (predicted or observed) consequences (4, 5). This approach to experimentation looks forward, in contrast to traditional empiricism, which observes antecedent phenomena to draw conclusions (5). Pragmatism employs abductive reasoning, where imagination is used to provide a provisional best explanation to account for observed phenomena and/or subsequently propose solutions or theory (7). This approach is well suited to this project, in which I assimilate findings from a variety of methods of enquiry to produce provisional results (as designs) which are subsequently tested for their usefulness.

It is important to acknowledge that the summaries of scientific evidence which will populate the *GP Evidence* website are themselves derived from a positivist scientific method of enquiry (randomised controlled trials and meta-analysis)(8). Here, more-or-less generalisable estimates of the effect of treatments are derived from experimental enquiry. Throughout, I accept this as an appropriate approach to answer these types of questions (albeit with consideration of the applicability of findings to all patients). I adopt a pragmatic approach to the research to find new *ways of communicating* this scientific evidence.

1.6 Research through design

Research through design (RtD) is an enquiry focused on producing a contribution of knowledge (9). This knowledge may be theoretical, but most commonly is manifested in the form of "*new, conceptually rich artefacts*" which are themselves the results of the research (10). These artefacts may be physical objects, systems, environments, services or in this case, a novel website.

The results or artefacts do not claim to represent a “truth”, but rather a new finding which solves a particular problem for the current situation. One view of the artefacts is that they can be seen as a proposition or placeholder which can perform a novel function in the here and now, but which also open up new space for further design (11). RtD is generative, it is concerned with the future: what a new product might look like, what its effect might be, can a new artefact transform the world from its current state to a preferred state? (9, 10).

RtD involves three steps: (1) a grounding in understanding of people and contexts through investigation, (2) a process of ideation to generate potential solutions and (3) iteration of proposed solutions to refine the final artefact (9). This approach underpins the design of this doctoral research.

The theoretical perspective of pragmatism aligns well with RtD. Stompff *et al* suggest pragmatism provides “solid ground” for design research (12). They describe the way in which designers justify their choices made by whether an artefact “works or not” and relate this to the pragmatist philosopher Dewey’s definition of knowledge as *“knowing what to do in an evolving situation in order to attain a goal”*. Pragmatism’s emphasis on abductive reasoning supports the designerly activity of creative problem solving. Likewise, its forward-looking view where research findings are judged by their contextual consequences, aligns with the designer’s goal of producing an artefact which will have an effect in the world.

1.7 Exploratory and developmental work and training 2013-2018

This section summarises some pieces of work and activities in the years prior to my doctoral research, which contributed to my understanding of the field.

In 2013 I designed and co-led a workshop at the first international Preventing Overdiagnosis conference (13) exploring the potential effect of clinical guidelines with patients with multimorbity. Using a fictional case study of an elderly person with polypharmacy, we considered how their treatment regime might look if multiple guidelines were rigidly followed, compared to a scenario where an understanding of absolute benefits and harms of treatments were brought to bear. We asked the audience of academics, clinicians and public participants to consider what features an imaginary ideal clinical guideline would have to support better personalised care. This produced a set of suggestions which informed the design of *GP Evidence* and are summarised in Table 1.1.

Building on ideas from this the workshop, in 2014 I designed a pilot version of the *GP Evidence* website and had the opportunity to present this to senior board members at the National Institute for Health and Care Excellence (NICE). I subsequently worked with a NICE clinical fellow to populate it with clinical content from some sample guidelines. This was a useful-test-of-principle exercise and I went on to present the work at conferences and workshops. However, I learned that the task was much more challenging than I had naively imagined. I realised I needed to find a way to develop the idea with proper resource and support, and acquire more advanced academic skills.

This interaction with NICE led to a number of roles with the organisation. I was a GP member for a Standing Committee for Guideline Updates over 2016-2017 and gained valuable insight

into the guideline development process. I was invited to join NICE's "Shared decision making collaborative" which launched in 2015 (14). This was a developmental exercise which ran for three years to inform NICE's strategy in this area. I participated in a number of stakeholder workshops and was exposed to high level expertise and policy discussion. Since 2017, I have been a co-moderator of the NICE GP reference panel, which is an online group of approximately 150 GPs who are invited to give feedback on clinical guidelines in development. I have gained much insight into GPs' needs and how the guideline development process can respond to these.

Around the same time as I started the doctoral fellowship, I was invited by NICE to chair a committee entitled "Connect diabetes pathway committee". It was tasked with exploring how NICE might make its guideline content more accessible and usable, taking diabetes as an example condition. Comprising topic experts, clinicians and patient representatives, it met on four occasions, one of which was a one-day workshop where participants were asked to outline their various information needs and priorities for a diabetes guideline and imagine how this might be best presented. I was able to draw on these findings to extract a set of suggestions which might be applied to *GP Evidence* – these too are summarised in Table 1.1.

I was a GP participant in the Academy of Medical Sciences enquiry and report, "Enhancing the use of scientific evidence to judge the potential benefits and harms of medicines" (15) and the Academy of Royal Medical Colleges "Choosing wisely" project (16)**.** I was exposed to high level debate on aspects of medical evidence and its communication to professionals and public, and contributed to recommendations on both projects.

Over 2013-2016 I delivered more than 20 teaching sessions to fellow GPs on aspects of overdiagnosis, polypharmacy and evidence-based medicine. I found my colleagues shared a similar lack of understanding of treatment benefits and harms, were frustrated by the limitations of guideline-driven practice and were enthusiastic about the idea of an information resource such as *GP Evidence*.

In 2018 I held a workshop with a group of approximately 20 GPs as part of a conference organised by the Royal College of General Practitioners' overdiagnosis group. We discussed the challenges we faced around understanding the benefits and harms of treatments for long-term conditions and developed a list of ideas and requirements from an imaginary new information resource which might support us in this area. Findings from this exercise were recorded and contributed to the design of *GP Evidence*, included in Table 1.1.

Table 1.1 Summary of ideas and recommendations from three events

Workshop. Preventing Overdiagnosis Conference. **Dartmouth College Sept 2013**	**Workshop. RCGP Overdiagnosis Group Conference.** **University of Birmingham 2018**	**NICE Connect Diabetes Pathway committee** **London 2019**
Participants: Approximately 70: Researchers, clinicians, public.	**Participants:** Approximately 40: General Practitioners	**Participants:** Approximately 20: Research experts, clinicians, patients, guideline producers.

A resource which provides supportive evidence rather than prescriptive guidance Development should be led by generalists rather than specialists to avoid potentially narrow perspective. Development should be free from conflicts of interest. It should be freely accessible Produced by a trustworthy organisation Should have significant patient input into development Quantitative evidence should be presented in formats which support ease of understanding. Should include a time-frame and be presented in relation to baseline risks where possible. Present data about meaningful clinical endpoints first. If presenting surrogate outcomes be clear that these are less important. Include information on trial populations to support transparency and consideration of applicability to individual patients. Include information on any limitations to the evidence. Be open about controversies and uncertainties in the evidence. Include information on special groups where appropriate and possible, e.g. the elderly or particular ethnic groups. Include information on non-drug/lifestyle measures. Present these with equal weight/priority.	Provide evidence upon which guideline recommendations are made. Make explicit that guidelines are not rules, to help support clinicians to not rigidly follow guidelines when not appropriate. Visual presentation of benefits and harms: use infographics. Present simple data: risk reductions, numbers-needed-to-treat. Clearly identify the aims and objective of therapy. Be clear about clinical outcomes. Note any gaps and uncertainty in the evidence. Empower patients to make their own change by presenting evidence on lifestyle measures. Could a resource provide or support medico-legal consequences of decision-making? Could it have a function as a tool for continuing professional development?	"Tell us why" (with regard to a guideline recommendation). Provide a simple statement for the reason for a guideline recommendation. Provide quantitative information on treatment benefits and harms to support shared decision-making. Provide visual summaries of evidence including choices and comparisons where possible. Use mapping and layering of information to deal with complexity. Could a ranking of treatment choices be made where appropriate? Provide a summary of clinical evidence and a statement about its strength. Link recommendations to relevant co-morbidities. Include warning flags for special conditions such as frailty or safety alerts. Include lifestyle advice Provide patient-facing information or information for clinicians to help patients. Include practical advice. For example how to take medicines. Include advice about when *not* to prescribe or when to stop medications.

I will refer to the findings in this table in chapter 6, where they contribute to the development of a prototype for *GP Evidence*. The three exercises produced sets of responses with much in common, despite quite different groups of participants. They were united by a desire to have accessible, understandable information about clinical evidence. The more academic audience at the 2013 workshop had raised points about the details of clinical evidence which may bias perceptions, for example the use of surrogate end-points and the importance of considering the time-frame over which benefits or harms may occur. The GP participants in 2018 were particularly interested in the ease of understanding clinical evidence and how it would fit into their practice, for example clarifying that guidelines are not "rules" and wanting support regarding medico-legal issues. The NICE workshop in 2019 focussed additionally on practical aspects and potential patient-facing content, reflecting the committee composition with a significant patient membership.

It is important to acknowledge that the first two workshops were attended by professionals with a particular interest in avoiding the harms of over-medicalisation – a perspective which could potentially bias their responses. I was careful to interpret their suggestions with this in mind and later, during the course of developing *GP Evidence* maintained a principle of presenting clinical evidence in a neutral manner.

1.7.1 NIHR In-Practice Fellowship

I completed an MSc in Evidence-based health care comprising six taught modules:

- *The practice of evidence-based health care.*
- *Introduction to study design and research methods.*
- *Systematic reviews.*

- *Realist reviews and realist evaluation.*
- *Knowledge into action.*
- *Qualitative research methods*.

This equipped me with a grounding in quantitative and qualitative research methods, critical appraisal, reporting standards and evidence synthesis as well as implementation science, evidence-into-practice and evaluation in complex settings. My dissertation research was a online survey of GPs' knowledge on the benefits and harms of treatments which informs this doctoral research.

1.8 Context

In the remainder of this chapter, I discuss the context in which the project takes place from a perspective of clinical practice, evidence-based medicine, clinical guidelines, multimorbidity and polypharmacy.

1.9 Evidence-based medicine

In 1996, the British Medical Journal published what was to become a seminal paper by Sackett and colleagues: "Evidence based medicine: what it is and what it isn't"(17). The authors made a defence of the relatively young discipline of evidence-based medicine (EBM) which had been criticised for promoting "cookbook" medicine and presenting a risk that medical practice would be "hijacked" by managers at the expense of clinical freedom. Their now famous definition of EBM,

> *"…the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients"*

was supported by a definition of clinical expertise,

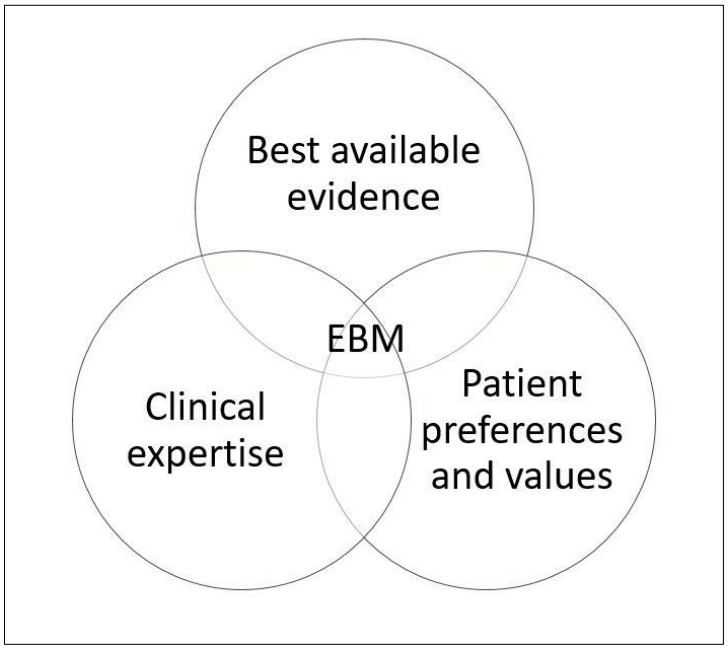
> *"…the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice."*

and an emphasis on the importance of patients' preferences,

> *"…thoughtful identification and compassionate use of individual patients' predicaments, rights, and preferences in making clinical decisions about their care."*

Their definition of what constitutes "best available evidence" was broad, but prioritised the randomised controlled trial (RCT) as the preferred source of evidence about treatment interventions. This triad is often presented as a Venn diagram, with EBM at its centre (fig 1.1).

Figure 1.1 The evidence-based medicine Venn diagram



Nearly thirty years later, their definition and description still offer a positive ideal for evidence-based practice, which has done so much to enhance the quality of medical care around the world. However, as I will describe below, some of the fears of early critics of the

movement have certainly come to pass, despite the sound intentions of the founding thinkers. The work of this thesis can be viewed as part of an ongoing effort to encourage and support a practice of EBM which sits in the centre of this Venn diagram.

1.9.1 Early visions of EBM meet reality in primary care

A 1992 paper from the Evidence-based Medicine Working Group described what clinical practice might look like in the new EBM paradigm(18). They describe a physician wishing to find out the prognosis of a man following his first tonic-clonic seizure. She was to visit the hospital library, perform a computerised literature search, select one paper from 25 (having pre-defined the criteria it should meet), critically appraise it and then apply it to this patient. The authors estimated this would take 30 minutes.

The principles described in this paper were developed and studied over the next two decades, forming the basis for the teaching and practice of EBM. As late as 2005, they informed the Sicily statement on evidence-based practice, which placed them at the centre of curriculum design and professional competencies (Fig 1.2)(19).

Figure 1.2 The 5-step process of evidence-based practice. Adapted from Evidence-based medicine working group 1992

The 5-step process of evidence-based practice

- Translation of uncertainty to an answerable clinical question
- Systematic retrieval of best available evidence
- Critical appraisal of evidence for validity, clinical relevance and applicability
- Application of results in practice
- Evaluation of performance

The 5 steps were included in general practice curricula in several countries, and critical appraisal was part of the examination for membership of the Royal College of General

Practitioners by the late 1990s (20, 21). Though theoretically excellent and supported by trials of teaching methods and effectiveness (19), this vision of EBM required a well-trained workforce with the time and motivation to practice it. This was unrealistic. A 1998 survey of English GPs found not just a lack of resources, confidence, skills and time to practice in this way, but even among those who had received specific training, a preference to use simpler guidelines or summaries (20).Over the next decade, research on primary care clinicians repeated these themes in their findings, with authors often suggesting further education, resources or barrier removal as solutions, rather than question the ambition of the original EBM vision (22-25).

By 2010, even early champions of EBM were writing about the overwhelming volume of scientific evidence: *"Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?"* and proposed accessible systematic reviews as a key component of the solution (26). By this time however, the desire of clinicians for usable evidence summaries (in the face of the impossibility of finding and appraising individual studies), combined with other agendas in healthcare systems (discussed below), led to the development of clinical guidelines and their widespread adoption in practice.

1.10 Clinical guidelines

A frequent starting point in the modern literature for the discussion of clinical guidelines (hereafter referred to as "guidelines") is a 1990 definition by the US Institute of Medicine (27):

> *"Systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances."*

Though guidelines have existed for much longer, some arguing as far back as Hippocrates (28, 29), I will use this starting point to consider the current role of guidelines in British general practice. Though the obvious and purported role of guidelines is to make clinical recommendations based on summaries of the best available evidence, there are a variety of motivations for their production across the healthcare system. One major driver has been the desire of healthcare systems to reduce costs by discouraging ineffective, expensive interventions (30-32) employed by clinicians who vary significantly in their practice (33). In this case, guidelines could be viewed as an imposition on professionals by controlling administrators. Another interpretation of guidelines is that they are a response by the profession to maintain autonomy; if guidelines produced by doctors are evidence-based and cost-considered, then interfering policy makers will be happy. However, this shifts autonomy from the individual clinician to an elite group of guideline producers who prescribe what good practice looks like (34). In 2007, Weisz *et al* conceptualised the development of guideline-driven healthcare as a result of multiple actors responding to a need for and shift towards increased overall regulation in healthcare over the late 20th century. This was a move away from a system reliant on professional authority and autonomy with its attendant problems, the recognition of which has been one of the main drivers of EBM(35).

I will discuss criticisms of guidelines later in this chapter, but in this thesis I will take a pragmatic and straightforward view of guidelines: that they are a well-intentioned enterprise, designed to encourage good quality EBM, support practitioners and deliver good quality care to populations. The aim of the doctoral project is to produce a resource which will augment guidelines, building on their strengths, addressing some of their shortcomings, and recognising that they are now a fixed, central part of clinical practice.

The rise of the guideline over the last 30 years has been accompanied by research and standard-setting for guideline development. A series of papers in the BMJ in 1999 described the state-of-the-art at the time, outlining a process with many features still in place today. This included the identification and prioritisation of topics and research questions, formation of project committees, systematic identification and appraisal of evidence for quality and relevance, and a translation of the evidence into a guideline using a committee approach, acknowledging the role of judgement and opinion in this process (36).

The same authors published an update series in 2012 (37-39), reflecting on developments in the field. It reveals developing understanding of the challenges and complexity of guideline production, in contrast to the rather linear, idealised version described before. New elements included were:

- the identification of the audience for the guideline, acknowledging a wide range of potential users, both professional and lay
- the value of involving patients and public, and associated challenges
- consideration and management of conflicts of interest of committee members
- the development of analytical models to structure clinical questions and evidence searches
- the inescapable presence of value judgement: for example in regard to how to address situations of uncertain or low quality evidence, or finely balanced benefits and harms. They acknowledge a wide ranges of values and preferences among patients and propose the idea of shared-decision making as one solution to this.
- health economic considerations

- the wording of recommendations to increase clarity and specificity

International processes and organisations were established to foster high standards of guideline development:

- the Guidelines International Network (GIN), connecting organisations and individuals in the field (40)
- clearing houses to collate and support access to guidelines (41)
- the development of the AGREE tool, a guideline appraisal instrument(42)
- the establishment of large-scale guideline production organisations such as the National Institute for Health and Care Excellence (NICE) in the UK
- the development of the GRADE system for rating evidence certainty and strength of guideline recommendations (43)

Guideline production has become its own field of practice (sometimes called an industry (44)) with increasingly high standards and continued evolution. In the UK, this field is dominated by NICE and the Scottish Intercollegiate Guideline Network (SIGN). Most of the work relating to *GP Evidence* is based on NICE guidance, for reasons I will discuss in Chapter 5.

Since its establishment as part of the NHS plan of 1998, NICE has come to be regarded as world-leading in the field (45, 46). Its original mission was threefold: to publish evidence-based guidelines and appraisals of new technology, promote equal access to new treatments (an end to "postcode lottery"), and prevent expenditure on poor-value treatments by employing cost-effectiveness analysis (47). It has seen its role expanded since to include the development of quality standards and performance indicators, accessible clinical knowledge summaries, procurement of journals and databases for the NHS and maintenance of the

British National Formulary (48, 49) (50). NICE's technology appraisal system has become increasingly connected with strategy for the UK life sciences industry (51) and though its methods and judgements have been challenged in this area (52), NICE technology appraisals are now a common route for new treatments to enter practice.

NICE's outputs are undoubtedly now the strongest driver of clinical practice in the UK, from the level of the knowledge of an individual clinician to the large-scale planning of healthcare systems. At the heart of this is an acceptance that the process for evidence review, appraisal and development of recommendations is of high quality and trustworthiness. NICE's methods have evolved over the years, the most recent iteration being a 2022 update of "Developing NICE guidelines: the manual" (53), key features of which are summarised in appendix 1.1.

1.10.1 Criticism of EBM and Clinical Guidelines

Despite the transformative and positive effect of EBM and its promotion through clinical guidelines, many of the problems originally envisaged by early critics have at least partially come to pass, in addition to new ones which have arisen as the discipline has developed. The literature of the last 10-15 years highlights multiple issues.

Firstly, there are well-described problems with the scientific foundation on which the knowledge base is built. The risk of bias in some research has multiple origins: financial and professional conflicts of interest, publication bias, trial design which maximises the chance of showing benefit and is less good at detecting harms, manipulation of data to show benefit where little exists (e.g. outcome-switching, abuse of composite outcomes, post-hoc sub-group analyses) and outright research fraud. Though the EBM community has developed tools

such as formal critical appraisal and quality rating systems to detect and address these biases, they cannot exclude them entirely (54-57).

Criticisms of clinical guidelines are similarly multi-faceted. Poor quality and lack of rigour in guideline production persists around the world despite frameworks such as those set by the Guideline International Network and quality assessment tools such as AGREE (42, 58, 59). A key observation is that it is difficult for an average guideline user to tell good from bad quality guidelines (59). Guidelines do not necessarily undergo a peer-review process (60). Even unclear language in published guidelines has been shown to affect the likelihood of uptake of recommendations(39).

Conflicts of interest may bias guideline recommendations. Recommendation of new treatments or the lowering of treatment thresholds in major guidelines generates enormous income for pharmaceutical companies and private healthcare providers. Internationally, guideline panels have been shown to comprise heavily of members with financial conflicts of interest, as well as being targeted by sophisticated lobbying from industry (59-64). An over-reliance on super-specialists and their societies to form guideline committees, who may have financial or intellectual biases as well as motivations around power and prestige, has been criticised as promoting an “overspecialised worldview” (61).

The evidence base which informs the guidelines may not necessarily be applicable to all patients. For example, research may have been conducted in populations with severe disease in secondary or tertiary care settings and then extrapolated to generate recommendations for milder illness in primary care (58, 59). A 2014 study reviewed 22 primary care relevant

NICE guidelines; of 495 recommendations, only 38% were based on evidence derived from populations typical of primary care (58).

Even if the trial populations are a reasonable match for the patients receiving care, how can a guideline recommendation apply to an individual who may vary in any number of ways from the population average? This problem of guidelines being a "poor fit" for individuals is a recurring theme (55, 59, 65, 66) and is to a large extent unavoidable, even with the application of clinical judgement. Guideline design may exacerbate this through a loss of detail and nuance: a variety of treatment effect sizes and strength of evidence hide behind simplified, standardised language, which gives the impression that all recommendations are equally important or valuable (67-69). Linked to this is the (incorrect) perception that recommendations are "written in stone". A lack of transparency about the quality of, and uncertainties in evidence presents a similar issue (65, 68, 70, 71).

Clinicians have the task of using their clinical judgement to tailor guideline recommendations to individuals (59), but to do this requires critical information about the evidence base, which is lacking in most guidelines (59, 72-76). Such information (which is the central subject of this thesis) includes data about effect sizes, the study populations and strength of evidence.

Finally, there is a cultural shift whereby guidelines have come to be seen as rigid dogma, rules which must be followed, either because they are perceived to represent good-quality, high-value care, or because of a fear of negative consequences to a clinician who deviates from them (65, 77). This has been exacerbated by the parallel development of performance measures, audit and league tables (55, 65, 78, 79). This has threatened professional autonomy

and the doctor-patient relationship, introducing a policy-maker into the consultation room (55, 66). Mehta and Lehman summarise, writing in 2020 (67):

> *"Evidence-based medicine was originally intended to keep the expert out of the decision loop. The evidence, properly synthesized, would be allowed to tell its own story, while the rest of the discussion would focus on uncertainty, relative or absolute. Properly speaking, EBM guidelines should look like half-drawn maps, indicating where the good roads are but, equally importantly, showing where there are no good roads. With this map in front of them, clinicians and patients should be able to make much better decisions together. Expert guidelines are not like this: in their maps, uncertainty is filled in by opinion. They tell the patient and the primary health care professional that somebody knows better, even when they do not."*

1.10.2 Under-use of guideline recommended treatments: the implementation gap

The difficulty of getting evidence-based care to actually happen, even when supported by guideline recommendations, has been a long recognised problem. Writing in the Lancet in 2003, a decade or more into the EBM era, Grol and Grimshaw quote figures from the USA and Europe suggesting that 30-40% of patients do not receive care according to current scientific evidence(80), describe an extensive body of research into why this happens, and a range of implementation strategies which might help. These are summarised in figure 1.3. The effects of implementation strategies was variable, about half reported only "mixed" or "limited" effects.

Figure 1.3 Barriers to and strategies to support the implementation of evidence. Adapted from Grol 2003.

Examples of barriers to implementation of evidence	Examples of strategies for implementation of evidence
Practice environment (organisational context) • Financial disincentives • Organisational constraints, time limitation • Perception of liability • Patient's expectations **Prevailing opinion (social context)** • Normative standards of practice • Opinion leader (dis)agreement • Obsolete knowledge from medical training • Advocacy e.g. by pharmaceutical companies **Knowledge and attitudes (professional context)** • Clinical uncertainty • Sense of competence to implement recommendation • Compulsion to act • Information overload, inability to appraise evidence	• Educational materials and strategies • Conferences, courses • Small group meetings • Educational outreach visits • Use of opinion leaders • Performance feedback • Reminders, computerised decision support • Task substitution, multi-professional collaboration • Media campaigns • Financial interventions • Patient-mediated interventions

The following years show continued efforts to improve uptake of guideline recommendations by GPs across a range of clinical areas. Examples are: an educational package to support guidelines for antibiotic stewardship in Belgium (2004) (81), academic detailing to improve hypertension and cholesterol prescribing in the US (2005) (82), education and provision of tools to implement a guideline on low back pain in the Netherlands (2005) (83), a multi-faceted intervention with educational outreach and computer support to improve hypertension and lipid prescribing in Norway (2006) (84), a video education package to improve chlamydia screening in Belgium (2005) (85) and computerised prompts to increase oral anticoagulant prescribing in the UK (2017) (86). All of these studies show only modest at best improvements in guideline adherence.

The implementation problem has been conceptualised in a number of ways. An example is the “research-to-practice pipeline” described by Glaziou and Haynes, early EBM leaders. This takes a linear, clinician-centred perspective, identifying stages of awareness, agreement and action necessary for translation of evidence into care (87). Taking a more complex, system-wide view, Greenhalgh *et al* proposed a model for the diffusion of innovations in service organisations which includes a multiplicity of factors relating to the innovation itself (an example of which could be a single guideline recommendation), the context into which the innovation will land, actors involved, design, and implementation process (88).

The field of implementation science attempts to research and promote methods to support the translation of research findings into practice by developing guiding frameworks. A 2018 review article described 10 different theoretical approaches (89), one of which was itself a consolidation of 19 different models (90), the complexity of this field reflecting the complexity of the problem.

The largest guideline-implementation intervention in the UK was the introduction of the Quality and Outcomes Framework (QOF) into the GP contract in 2004 (91). It provided a financial incentive for GPs to monitor and treat specified long term conditions, and resulted in an improvement in measures of care quality for diabetes, asthma and heart disease (92). There is some evidence of benefit to long-term meaningful outcomes, with relative reductions in hospital admissions of approximately 10% for incentivised conditions, though no evidence of benefit on total mortality (93, 94). Whatever its clinical effect, the pressure to conform to QOF has powerfully shaped the culture of UK general practice, encouraging a systematic approach to clinical care with acknowledged advantages, but negative effects including a

reduction in continuity of care and a perception of "tick-box" and target driven practice among clinicians and patients (92, 94-96). It can be argued that one effect of QOF was to force target-driven care on individuals, encouraging healthcare professionals to *persuade* people to take up guideline recommended treatments, in contrast to the ideal of individualised, shared decision-making. This was observed in an ethnographic study of cardiovascular care in an English practice: whilst the staff were speaking the language of shared decision-making, their actions and attitudes were more about ensuring guideline concordance (97).

Even in the post-QOF era, there are ongoing areas of under-treatment of long-term conditions. NHS England's national audit on cardiovascular disease, "CVD Prevent" presented data from 2020 on the apparent under-use of: anticoagulation in atrial fibrillation (80-90% of eligible patients), poor blood pressure control in hypertension (60-70% of patients reaching target) and statins in patients with chronic kidney disease (50-80% of eligible patients) (98). Variation in care standards and achievement of treatment targets remains a stubborn problem internationally (99, 100). A 2021 systematic review found evidence of under-use of anticoagulation in atrial fibrillation, with rates of prescribing ranging from only 10-80% in different settings over the last decade (101). Recently, the NHS England 2023-24 Direct Enhanced Service for Primary Care Networks includes requirements to improve the treatment of atrial fibrillation, hypertension, hyperlipidaemia and heart failure (102).

It is important to note that these descriptions of "under-treatment" take a population-level perspective, assuming that all or most eligible people (as defined by current thresholds and targets) should ideally be on a particular treatment, and that it is the job of the healthcare system to achieve this. Whether this is correct view or not is debatable and will be considered

in the rest of this chapter and chapter 2. However, it is probable that many of these "under-treated" patients would indeed potentially benefit from treatment, and choose to accept it if offered.

1.10.3 The role of guidelines in over-treatment

The problem of over-medicalisation is well described. It is complex and multi-faceted, affecting all areas of healthcare, with origins in society and culture as well as individual-level medical activity and wider policy (1, 3, 103). However, I wish to focus on the specific issue of over-treatment for long-term conditions which might arise as a direct result of the application of clinical guidelines: where a clinician may be intending to provide good quality evidence-based care, but the treatment may not be ideal for a particular patient at a particular time. To place this in context, I will start by briefly describing current and recent work in the area of prescribing quality in primary care.

Prescribing incentive schemes were introduced in the early 2000s to improve prescribing quality and safety, reduce variation in practice and promote cost-effective prescribing. Practices would receive moderate financial rewards for work in this area, which tended to focus mostly on cost-containment (104-106). Simultaneously, medicines management and medicines optimisation work aims to help people get the most from their medicines, maximising benefits and minimising harms whilst aiming for clinical and cost-effectiveness. The focus is typically on practical aspects of managing medicines: how to conduct medication reviews, supporting patients to understand and adhere to medicines, healthcare providers' communication systems (with each other and with their patients), medicines reconciliation

(tidying up complex lists and packets of medication and streamlining medicine ordering processes), and issues of cross-organisational interaction (107, 108).

Prescribing safety – the reduction of prescribing errors and practice which may cause harm, has received much attention and been supported by computerisation which allows identification of high risk prescribing and action to avoid this. This type of work often forms part of prescribing incentive schemes and medicines optimisation activity. Prescribing patterns associated with risk of harm have been identified and developed into indicators for assessing safety in practice (109). These patterns may involve co-prescription of drugs with potential interactions, inappropriate dosing, inadequate monitoring or prescription in the context of co-morbidities or risk-conferring biomarkers (110). Trials testing the identification of high risk prescribing coupled with multi-professional interventions have shown improvements in the amount of hazardous prescribing and reductions in the rate of hospital admissions for medicine-related harms such as gastro-intestinal bleeding and heart failure (111, 112).

In the consulting room, primary care computer systems flag up safety warnings such as allergies, drug interactions, drug monitoring alerts, drug-morbidity interactions (e.g. beta-blockers for someone with a record of asthma) and general safety alerts. These provide much detailed information for clinicians, though may have limited effect due to “alert fatigue” and a variety of human factors (113, 114).

In 2021 the Department for Health and Social Care published the findings of the National Overprescribing review for England (115). It describes the harms associated with over-prescribing and a network of causal factors. Causation was separated into two categories:

systemic and cultural. Systemic factors included those which have been described above regarding clinical guidelines and medicines management/optimisation issues. Cultural aspects considered were: power imbalances between clinicians and patients, communication and knowledge sharing, and tacit, incorrect assumptions about the value of medicines. A wide-reaching consultation process showed much congruence between patients' and clinicians' views of the problem, with shared frustrations about the status quo (116, 117).

Solutions to these problems are proposed in the report, many of which are already established but require ongoing efforts and improvements, such as medicines optimisation work driven by local and national organisations, increasing the use of structured medication reviews (SMRs) and data analytics, as well as the roll-out of social prescribing (118). Specific projects are underway to reduce antimicrobial prescribing (119) and dependency forming medicines (120). Two important concepts which have evolved in the last decade and relate to this work are the SMR and deprescribing:

A structured medication review is defined by NICE as

> *'a critical examination of a person's medicines with the objective of reaching an agreement with the person about treatment, optimising the impact of medicines, minimising the number of medication related problems and reducing waste.'(121)*

Though the concept of reviewing a patient's medicines has long been part of primary care practice (122), this process has taken some time to formalise. Conducting an annual medication review for selected patients was included in the QOF in the mid-2000s, though what this should entail was not defined. I remember participating in a national medicines-management project at this time, with much debate about how deep or complex medication

review needed to be. A number of tools, guidelines and proposed structures to support medicines review have now been developed and are summarised in appendix 1.1. The expansion of the role of pharmacists in primary care is a key plank of support for this work, which is hoped to evolve into a more structured, high-quality part of clinical practice (123).

Deprescribing, the process of actively stopping medicines to avoid harm and reduce treatment burden, is a concept which has also evolved over this period, is suggested in many of models of medication review and supported by the same set of tools. There is no clear definition of what a deprescribing process is, but it stems from the idea that modern medical practice is highly geared towards starting medicines, but that stopping them is rarer, more challenging and somewhat counter-cultural (124, 125). Some medicines may be relatively uncontroversial to stop: the painkiller that is no longer needed, a blood pressure lowering drug in someone experiencing symptomatic hypotension, or a time-limited intervention such as dual anti-platelet therapy following a coronary stent insertion. Other scenarios present more difficult questions, such as how to balance the potential harms (giddiness, falls, or renal impairment) from multi-drug heart failure regimes with their potential benefits on survival. Systematic reviews of trials of deprescribing interventions have shown on average, only modest reductions in prescribing levels compared to controls (approximately 1 fewer drug prescribed for every 2-3 people receiving the intervention) (126). These reviews also show no difference in mortality or risk of hospital admission, a finding which can be viewed either as in support of deprescribing, or as a measure of lack of success. As with a SMR, deprescribing can be complex. The findings of a 2020 systematic review (124) on the barriers and facilitators to deprescribing in primary care are summarised in figure 1.4.

Figure 1.4 Barriers and facilitators to deprescribing in primary care. Adapted from Doherty et al BJGP 2020.

Barriers to deprescribing	Facilitators of deprescribing
Cultural "Pill for every ill" Preventive medicine Lack of financial incentives to deprescribe **Organisational** Disease-specific guidelines Poorly resourced care for the elderly or multimorbid. Time constraints for professionals **Interpersonal** Problematic inter-professional communication Communication challenges between patients and clinicians. Medico-legal fear among clinicians **Individual** Patient's reluctance to stop treatment Inequalities of healthcare: language, culture, mental health, socio-economic deprivation.	**Cultural** "Prudent prescribing" culture **Organisational** Guidance on multimorbidity, polypharmacy. Inclusion of deprescribing guidance in treatment guidelines. Tools to support deprescribing Availability of non-pharmacological treatments **Interpersonal** Trusted clinician-patient relationship Continuity of care Multidisciplinary working Patient involvement in decision making **Individual** Better information for clinicians on deprescribing and benefits and risks of treatment.

Returning to the topic of this section – the role of guidelines in driving overprescribing, it is interesting to observe how it is discussed in the literature. Though the critiques of EBM and guidelines imply that this is the case, and it is acknowledged in the literature on multimorbidity and polypharmacy as well as in the discourse on over-medicalisation, the issue is relatively absent in current guidance on medication reviews, deprescribing and indeed the national overprescribing review.

Perhaps this is because it is a difficult problem. How can a guideline-recommended treatment for a long-term condition be defined as an over-prescription or unnecessary medication, let alone be deprescribed? I have not seen this exact question asked in this way, but suggest the following answers in figure 1.5.

Figure 1.5 My suggested definitions of over-treatment with respect to guideline-recommended treatments for long-term conditions.

When is a guideline recommended treatment for a long-term condition an over-prescription?

- When a guideline recommendation **extrapolates evidence** from populations with severe illness and treatments are applied to those at much lower risk.

  Example

  ACE inhibitors for "nephroprotection" in patients with type 2 diabetes and albuminuria. Strong evidence exists for benefit in reducing the risk of end-stage renal disease (ESRD) in severe cases, but most treatment-eligible patients are not destined to develop ESRD and therefore will not benefit from treatment.
- When **for an individual, the potential benefits of treatments may be outweighed by harms**, despite good evidence of benefit in clinical trials. This usually (but not exclusively) relates to increased risk of harm.

  Examples

  - Multiple co-morbidities, very old age or frailty may increase risk of harms in a context of reduced chance of treatment benefits due to shorter life expectancy and competing risks.
  - Polypharmacy, where drug-drug interactions increase risk of harm.
- When **an individual would not choose to accept the offer of treatment** were they aware of the absolute likely benefits or risks to them – a question of individual preference despite perhaps high-quality clinical evidence.

  Example

  A patient at low-moderate baseline cardiovascular risk declines a statin because they do not think an absolute risk reduction in cardiovascular events over 10 years of (say) 3% is worthwhile.

The challenge here is that all these scenarios will vary on a case-by-case basis; evidence to inform decisions on questions of risk of harm, life expectancy, baseline risk or prognosis is

usually lacking. Decision-making in this area unavoidably requires clinical judgement and relates to individual preferences and characteristics.

If guidelines promote blanket one-size-fits-all treatment (assuming they are being implemented), then overtreatment is bound to occur. However, it is impossible to measure this at population level. Even at an individual level, any decision or definition of overtreatment is more or less subjective and arguable. This is in contrast to measurable under-treatment in populations, which therefore receives more attention, funding and political concern.

1.11 Multimorbidity and Polypharmacy

The challenges I have described regarding evidence-based practice (section 1.9) and clinical guidelines (section 1.10) are magnified and crystallised in the context of multimorbidity and polypharmacy.

1.11.1 Epidemiology

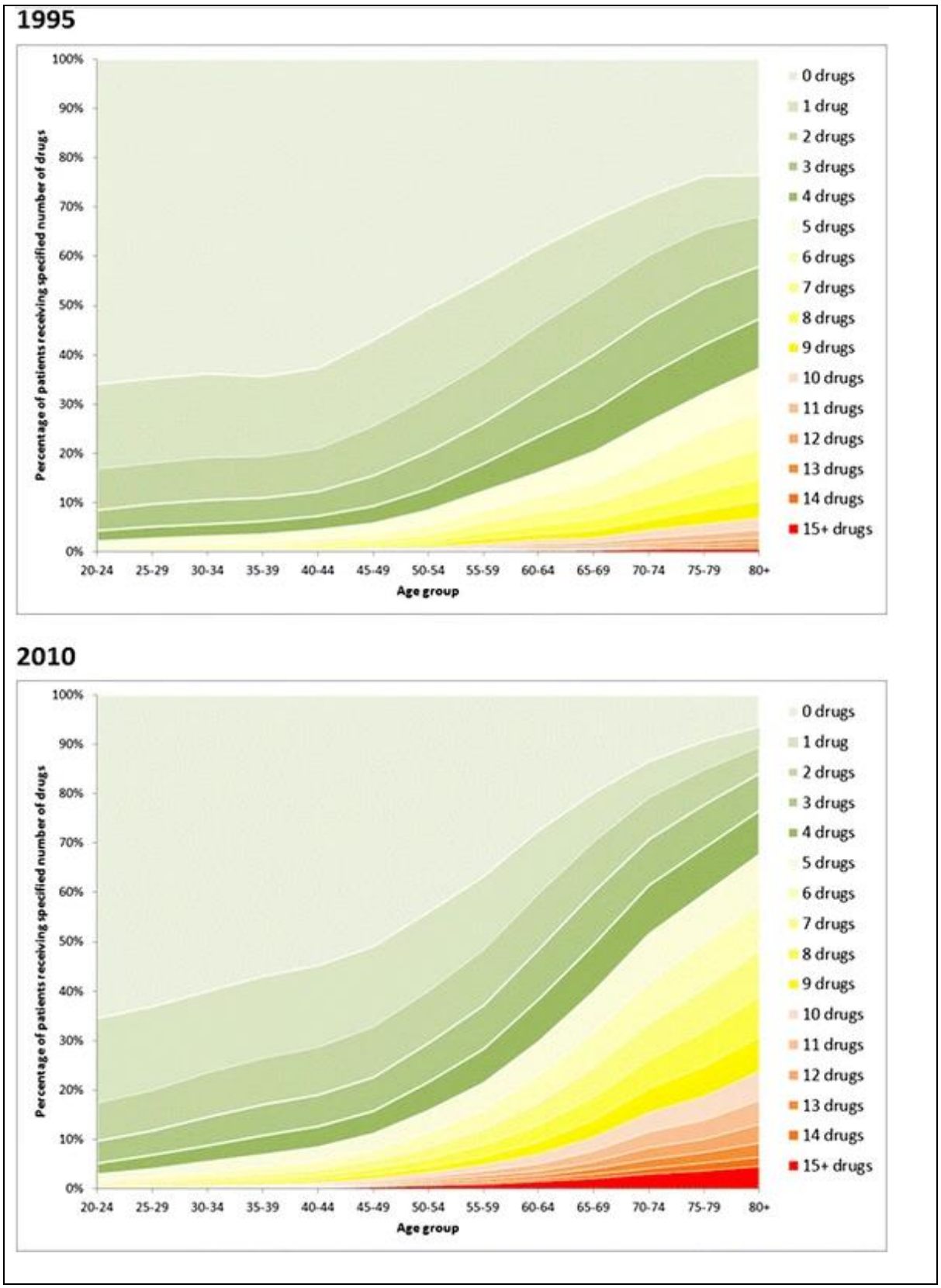
Multimorbidity means the presence of 2 or more long-term conditions (LTCs), although other definitions requiring more LTCs have been used (127, 128). There has been a steep increase in prevalence over recent decades attributed to the aging population, lifestyle factors resulting in higher rates of non-communicable diseases, advancing medical coverage increasing disease detection, as well as the re-definition of disease at lower thresholds (62, 127, 129-131).

A key UK study revealing the scale of multimorbidity was published in 2012 (132). It drew data from Scottish primary care, finding an overall prevalence of multimorbidity of 23%, increasing

with age and social deprivation. Clustering of LTCs was described, as well as the association between mental health and physical conditions. Since then, similar data has been gathered worldwide. A 2023 meta-analysis (127) reported a global prevalence of multimorbidity of 37%, with 51% in those over the age of 60. A 2018 study using English general practice databases found an overall prevalence of 27%, with variation by sex (30% in women, 24% in men) and socio-economic deprivation (30% in lowest quintile, 26% in highest) (133). This study also explored healthcare utilisation associated with multimorbidity. 53% of GP consultations, 78% of prescriptions and 56% of hospitalisations occurred in patients with multimorbidity.

The prevalence of polypharmacy shows a similar trajectory, due to the rise in multimorbidity and treatment options for LTCs. Definitions of polypharmacy have changed over time, reflecting what “normal” levels of prescribing look like. Thresholds are inherently arbitrary, ranging from 2 or more drugs to >5 or >10 (134, 135). Recent literature provides more granular information such as is shown in Fig 1.6, derived from prescribing data from Tayside between 1995 and 2010 (136). A 2022 meta-analysis of observational studies reported a pooled prevalence of polypharmacy of 37% (using a range of definitions) with older age and out-patient or hospital settings as risk-factors for polypharmacy (134).

Figure 1.6 Prevalence of polypharmacy by age and number of drugs prescribed. Y axis shows percentage of patients receiving specified number of drugs. From Guthrie et al 2015. Published by Springer Nature under Creative Commons Attribution 4.0 International License



1.11.2 Impact on patients

Living with multimorbidity has a significant impact, being associated with increasing disability, functional decline and mortality. It impacts negatively on general wellbeing, and quality of life (137). Qualitative research reveals that important issues for patients are the impact on living with chronic pain or mobility difficulties, shrinking social networks, difficulties sustaining working life and a sense of low mental well-being (138).

A major issue with both multimorbidity and polypharmacy is treatment burden. Large expenditures of time, energy and resources are needed to manage medicines, appointments and interactions with professionals (137, 139). One analysis of this problem called for "minimally disruptive medicine" to reduce the additional harms caused to people by their medical care (140). If patients feel overwhelmed, this may impair their ability to adhere to complex treatment regimens. These burdens are also shared by carers for people with multimorbidity (141, 142).

Polypharmacy is part of this burden of multimorbidity, and is also associated with risk of harm from drug side effects and interactions ranging from mild to severe or fatal. Increasing numbers of drugs are associated with increased chances of being prescribed high risk drugs, adverse drug reactions, hospitalisations and death (143, 144). However, it is hard to prove that these associations are all causal; higher levels of prescribing occur in those with more numerous or more severe health problems which themselves increase risk of adverse outcomes. A study linking primary care and hospital data in Scotland showed that higher rates of hospitalisation associated with increasing polypharmacy were at least partly accounted for by the risk associated with increasing multimorbidity (144).

An influential report on polypharmacy from The Kings Fund in 2013 (135) proposed a distinction between appropriate polypharmacy,

> *"…prescribing for an individual for complex conditions or for multiple conditions in circumstances where medicines use has been optimised and where the medicines are prescribed according to best evidence."*

and problematic polypharmacy,

> *"…the prescribing of multiple medications inappropriately, or where the intended benefit of the medication is not realised."*

This distinction has gained traction in the literature and serves to highlight the important point that much polypharmacy represents good medical practice and is likely to offer more benefit than harm.

1.11.3 Challenges for clinicians

Multimorbidity and polypharmacy present challenges for clinicians too. The qualitative literature exploring GPs' persepctives describes that they can feel poorly equipped (145) or even "helpless" (146) in this context. Similar themes to those in the literature on guidelines arise: difficulty applying single-condition guidelines to multimorbid patients, making and sharing decisions in the absence of applicable evidence, and a significant amount of medico-legal fear surrounding decisions to personalise care which may deviate from guideline-approved best practice (147). Coping strategies described include "satisficing" (combining hunches, best guesses and negotiating compromise) in order to personalise care (148).

1.11.4 Strategies to improve care in polypharmacy

Well considered strategies to guide the management of polypharmacy have been produced by national bodies in Scotland (149), Wales (150) and England (151) over the last 12 years. There is considerable congruence across these. The complexity of polypharmacy is acknowledged, as well as the distinction between appropriate and potentially problematic medicines. Consultation and process models are suggested to support practice with a variety of tools and information resources. The content of these are very similar to that of the SMR summarised in appendix 1.2. Indeed, SMRs are the major tool proposed to address polypharmacy, although long-term care with opportunistic review of some or all medicines is acknowledged as an important (and perhaps more realistic) mechanism.

Whereas there is consensus about the theoretical ideal of polypharmacy management, barriers to its implementation are well researched and described (152-154). Again, there is much overlap with the literature on deprescribing as part of SMRs (appendix 1.2).

1.11.4.1 An important omission in guidance on polypharmacy?

Returning to the issue of guideline-directed treatments in this context, I propose that there is a significant omission in current guidance on polypharmacy management.

In these models and tools to support SMRs, there are instructions to consider whether individual medicines are indicated, likely to offer benefit and are supported by evidence. This seems straightforward, but the problem is that any treatment for a LTC which is supported by a guideline recommendation will fit these criteria. There is little exploration of when such a treatment might not be suitable for a patient (as I propose in figure 1.5). Though SMR tools highlight potentially risky individual drugs and combinations of medicines (e.g. diuretics + ACE

inhibitor + NSAID[1]) for consideration of dose reduction or deprescribing, a clinician thinking about this may face an opposing imperative to continue a drug with an apparent strong (guideline-recommended) indication. What are they to do? How do they understand and balance potential benefits and harms?

1.11.5 Strategies to improve care in multimorbidity

A similar degree of consensus has been reached regarding the management of multimorbidity (128, 131, 137), though uncertainty remains and there are large gaps in the evidence base to guide treatment strategies in these complex scenarios (155). How best to support quality of life for individuals is still unknown; answers lie across the domains of family, community, social care, healthcare organisation as well as individual clinical interactions (137, 138).

In 2016, NICE published guidance on the management of multimorbidity (128). This was a major publication, not just because it was much needed, but because it specifically described the problem of applying single-conditions guidelines, derived from an evidence base of carefully selected trial participants, to the complex scenario of managing multiple long term conditions and polypharmacy. It highlighted the potential harms and treatment burden which may occur by following single-condition guidelines to the letter and made explicit that it was acceptable practice to deviate from these when appropriate. Speaking at the Royal College of General Practitioners' annual conference that year, Chair of NICE Professor David Haslam

[1] Non-steroidal anti-inflammatory drugs. All these drugs can cause kidney damage and the combination increases this risk. Other example include combinations of drugs which all lower blood pressure, or drugs which all have anticholinergic side effects.

said, “Good care does not look like following eight guidelines to the letter”[2]. A summary article written by members of the guideline committee (156) highlighted key messages:

- Guidelines on single health conditions may not be applicable
- Aggressive management of risk factors for future disease is often a major treatment burden and can be inappropriate
- Assess whether patients may benefit from an approach to care that takes account of their multimorbidity
- Consider all conditions and treatments simultaneously
- Easier access to data about the absolute benefit of commonly prescribed treatments is needed

The guideline made wide-ranging recommendations about an approach to care in multimorbidity, emphasising the need to establish patients’ priorities, makes recommendations on how to identify or select those who would benefit from this approach and the need to take a holistic view.

A key issue mentioned in both these summaries, and highlighted above in the discussion about the omission in polypharmacy guidance, is encapsulated in this recommendation from the guideline which frames this doctoral research and the *GP Evidence* project.

“Review medicines and other treatments taking into account evidence of likely benefits and harms for the individual patient and outcomes important to the person.”

[2] Personal observation JT

2 Chapter 2: Background

This chapter summarises literature relevant to clinicians' understanding of the benefits and harms of treatments. I summarise research on GPs' relationship with clinical guidelines, outline work over the last two decades in the field of shared-decision making, and what is known about patients' preferences in this area. I describe patient and public involvement in the research and conclude by summarising the case for the development of *GP Evidence* and the aims and scope of the project.

2.1 General practitioners and guidelines

The literature describing the views of general practitioners' attitudes to guidelines produces similar themes to the general critiques of guidelines discussed in the previous chapter. Two early reviews examined this issue from an implementation perspective – addressing concerns that guidelines were not being used or not having the anticipated impact on practice.

A 1999 review by Cabana *et al (157)* sought to summarise the barriers to guideline use among physicians, though the majority of participants were GPs. From 76 publications, most of which were surveys, they identified awareness of guideline existence as the commonest barrier reported, followed by low familiarity and agreement with a guideline, lack of confidence that desired outcomes would be achieved, a perceived inability to enact recommendations, and clinical inertia. Cumbersome and confusing guidelines were a problem as well as practical issues about time and resources. Interestingly, factors relating to patient preference appear quite low on the list at this time.

In 2002 Farquhar *et al* (158) reviewed surveys of clinicians' attitudes to guidelines and found 153 studies, again mainly on GPs. This revealed positive and negative attitudes. Over 70% of respondents agreed that guidelines were helpful sources of advice and were intended to improve quality. However, 30-50% considered guidelines too rigid to apply to patients, impractical, over-simplifying and a challenge to autonomy. Two-thirds felt they were intended to cut healthcare costs.

Some years later, in 2007 Carlsen et al published the first qualitative synthesis of GPs' attitudes to guidelines from 12 studies from GPs in the UK, Netherlands and Canada. The authors grouped the findings into six themes:

- **Questioning of guidelines:** the applicability of research to a wider patient population, reliability of research, changing evidence-base, concern about narrow specialist perspectives and the agenda of cost-containment.
- **Clinical experience:** tensions between simple guidelines and the complexity of individuals and difficulties with multiple diagnoses and patient preferences.
- **Doctor-patient relationship:** when guidelines restricted access to healthcare, for example investigations of imaging, guidelines were perceived as helpful to displace responsibility from the GP onto the system, thereby protecting the relationship.
- **Professional responsibility:** varying responses, where some GPs perceived a need to "override" guidelines, compared to those with a view that guidelines were there to be adapted.
- **Practical issues:** time and convenience issues, skills with new procedures.

- **Guideline format:** problems with usability and clarity, preference for clear and simple format.

The authors of this review conducted further studies with GPs in Norway and Sweden, revealing similar issues, but a mostly positive view of guidelines in general (159, 160). In a subsequent postal survey of over 1000 Norwegian doctors (161), they found that GPs were significantly more uncertain than specialists about the accessibility of evidence in guidelines and their legal status. They were more likely to have concerns about practical issues, applicability to individuals and clinical autonomy. Further studies from 2011 to 2021 in France (162), Germany (163), UAE (164), Sweden (165) and Norway (166) contain similar findings, with no new major themes described.

2.2 How could clinicians consider the benefits and harms of treatments?

The recommendation from the NICE multimorbidity guideline (section 1.11.5) to *"review medicines and other treatments taking into account evidence of likely benefits and harms for the individual patient and outcomes important to the person"*, presents a number of challenges. It asks that we place ourselves in the centre of the EBM Venn diagram (fig 1.1) whilst making multiple complex decisions with our patients. However, the evidence suggests that GPs do not feel well supported to do this by clinical guidelines. NICE do make further recommendations in support of this, reproduced in figure 2.1.

Figure 2.1 Selected guidelines from NICE Guideline (NG56) Clinical assessment and management of multimorbidity 2016.

- When reviewing medicines and other treatments, use the database of treatment effects* to find information on:
    - the effectiveness of treatments
    - the duration of treatment trials
    - the populations included in treatment trials

- Take into account the possibility of lower overall benefit of continuing treatments that aim to offer prognostic benefit, particularly in people with limited life expectancy or frailty.

- Discuss with people who have multimorbidity and limited life expectancy or frailty whether they wish to continue treatments recommended in guidance on single health conditions which may offer them limited overall benefit.

- Discuss any changes to treatments that aim to offer prognostic benefit with the person, taking into account:
    - their views on the likely benefits and harms from individual treatments
    - what is important to them in terms of personal goals, values and priorities

- Ask the person if treatments intended to relieve symptoms are providing benefits or causing harms. If the person is unsure of benefit or is experiencing harms from a treatment:
    - discuss reducing or stopping the treatment
    - plan a review to monitor effects of any changes made and decide whether any further changes to treatments are needed (including restarting a treatment)

*the "database of treatment effects" is explained in section 2.5

Whilst the final recommendation regarding treatments for symptomatic benefit is straightforward (it can be carried out with standard clinical knowledge, and subjective judgements are appropriate), the others are much more challenging. Firstly, they require a detailed knowledge of the evidence base supporting a treatment. While GPs understand *why* they are prescribing (for example, 'the blood pressure drugs reduce the chance of a stroke'), their understanding of exactly how likely a patient is to benefit from a treatment or experience harm may be lacking. If they are to consider questions such as: 'If I take more tablets for my diabetes, how much does that protect me from diabetes complications?', or 'If I take this anticoagulant drug, what is the chance of it causing dangerous bleeding?' they need to have an at least approximate numerical idea.

This quantification of the chance of benefit or harms (derived from clinical evidence) can be expressed in a number of ways: absolute risk reduction (ARR), number needed to treat (NNT),

relative risk reduction (RRR), or natural frequencies (plain language) (167). For example, for someone with a baseline 10-year cardiovascular risk of 20%, taking a statin reduces future cardiovascular events by: 7% ARR, NNT (10y) 14, RRR 37%. This means that for every 100 people who take a statin, 13 will have a cardiovascular event over 10 years compared with 20 people out of 100 who do not take a statin.

Knowledge of this quantitative information is the essential starting point from which clinicians might begin to resolve some of the problems and tensions inherent in single-condition-guideline-driven-EBM described in chapter 1. However, once summary numerical information is available, it also requires complementary information about the evidence from which it is derived: trial duration, study populations, quality of evidence and details about interventions may all be needed to support the process of clinical judgement required to make a treatment decision.

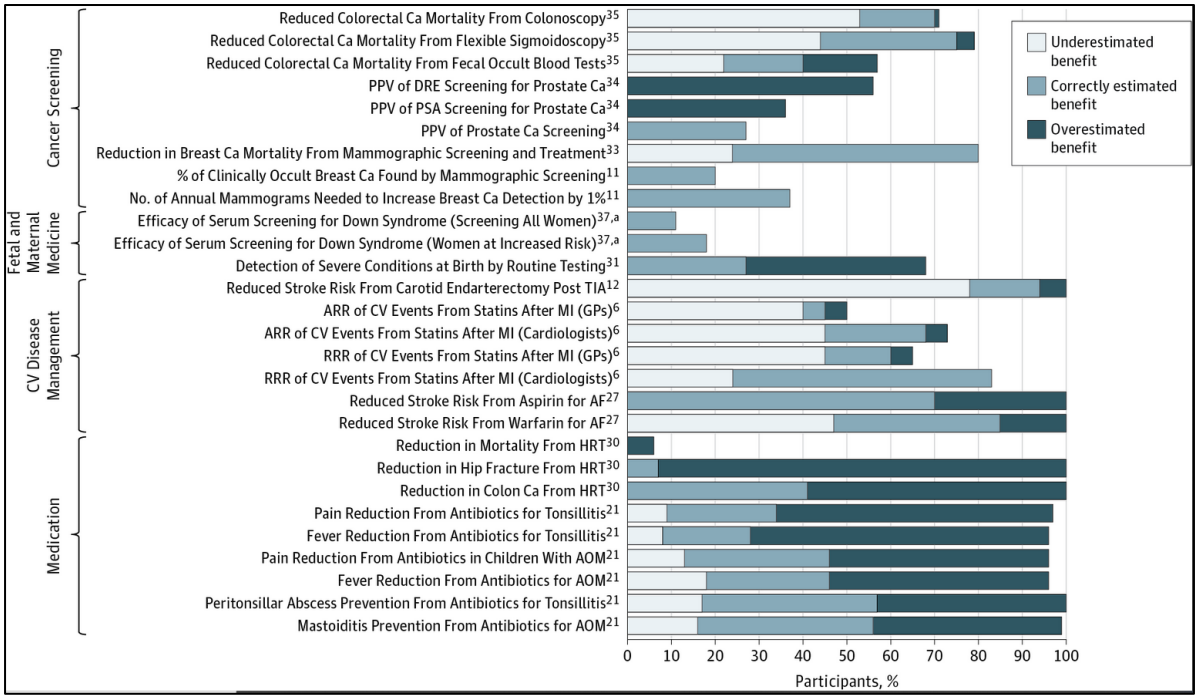
2.3 Clinicians' understanding of the benefits and harms of treatments

How good are doctors' quantitative estimates or knowledge of the benefits and harms of treatments? A 2017 systematic review collated studies which had tested this (168). It included 48 studies, 20 regarding treatments or interventions and 28 regarding tests or screening, conducted between 1981 and 2015 in 17 countries employing a variety of approaches and parameters. Of the 20 studies on treatments, 14 involved specialists (a wide range including cardiologists, paediatricians, gynaecologists, general physicians and liver transplant surgeons) and 6 with general practitioners.

Overall, the results show marked inaccuracies in clinicians' understanding across the range of medical specialities, interventions and tests with a tendency for clinicians to overestimate

benefits and underestimate harms. A summary figure for benefit outcomes is reproduced in figure 2.2.

Figure 2.2 Percentage of clinician participants providing a correct estimate, under or over-estimate of intervention benefits. From Hoffman and Del Mar 2017. Reproduced with permission of the American Medical Association.



2.4 British general practitioners' understanding of the benefits and harms of treatments for long-term conditions

Despite the major role that GPs in the UK play in the management of LTCs, there were no research studies in the systematic review involving this group as participants. In addition, I perceived a need for more detailed data about their levels of understanding. In 2018 I conducted an online survey of British GPs as a dissertation project for the MSc in Evidence-based Health Care which preceded my doctoral research. The findings were published in BJGP

Open in 2020 (169), reproduced here in appendix 1.1. I will briefly summarise the methods and major findings[3].

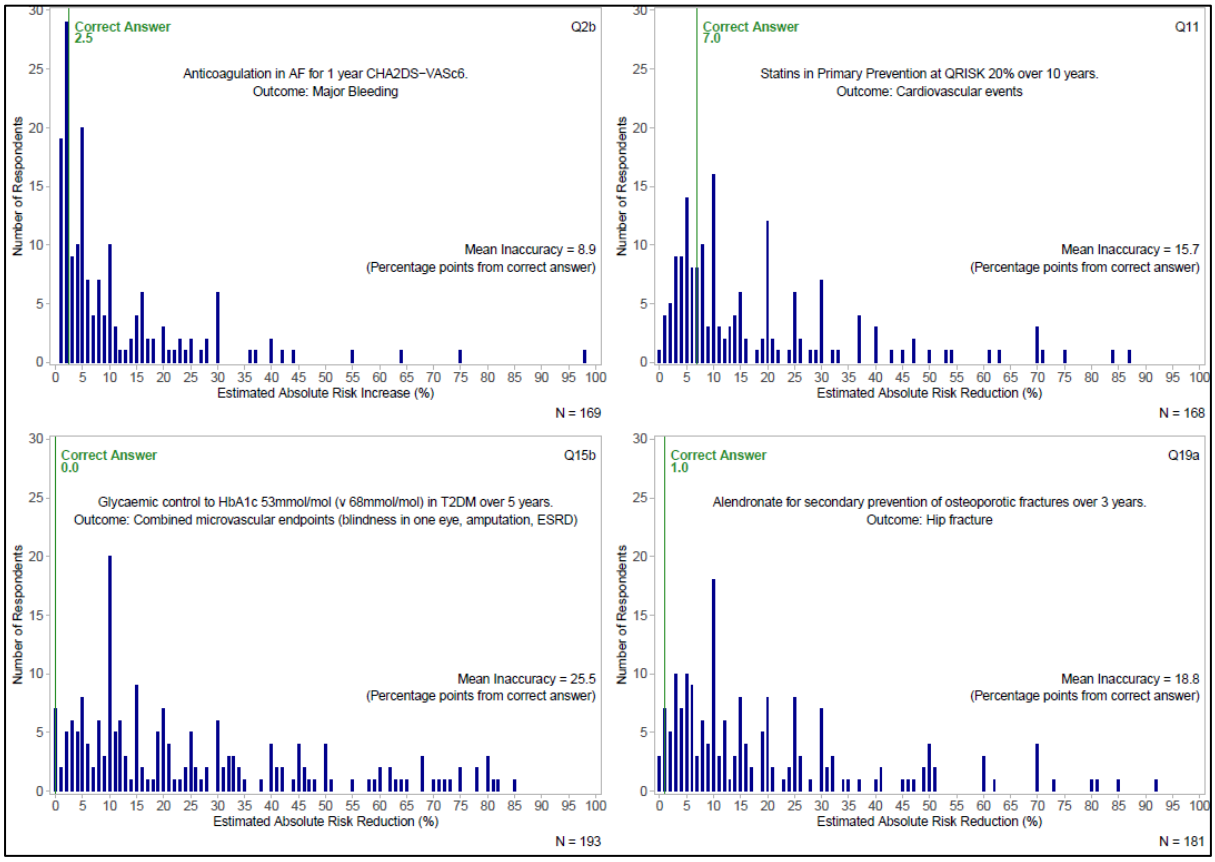
The survey was distributed through a variety of generalist networks and 433 GPs' responses were included in the analysis. Their demographic and professional characteristics were broadly representative of the national GP population. The survey consisted of questions presented as clinical vignettes and asked participants' estimates of absolute risk reduction (ARR) or side effect risks for treatments for long-term conditions. Key findings were:

- an average of 11% of responses were correct allowing for +/- 1% margin in absolute risk estimates
- 23% were correct when allowing a +/- 3% margin
- 88% of responses overestimated and
- 9% of responses underestimated treatment effects
- 65% of GPs self-reported low to very low confidence in their knowledge, only 5% self-reported "quite good" or "very good" levels of confidence.

Figure 2.3 shows graphical illustrations of answers to four of the 35 questions.

[3] Some small sections of text in this section are copied/adapted from the published paper. These particular sections were originally written by JT and had little or no alteration by co-authors.

Figure 2.3 Four selected questions from GP survey. Number of respondents submitting answers by percentage estimates of absolute risk reduction. Summary of clinical vignette presented in centre of each image. Green line represents the "correct" evidence-based answer.



The results paint a picture of significant and wide-ranging levels of misunderstanding of the size of treatment benefits and harms among GPs. Though there are limitations in this data which are discussed in the published paper, it reasonable to assume this reflects reality in the GP population closely enough. The mis-estimations are of a magnitude likely to meaningfully affect clinical decision-making and impede conversations with patients regarding treatment choices.

An argument is that this doesn't really matter, GPs have guidelines which they can follow and use their clinical judgement to make decisions with patients. However, imagine a patient at low cardiovascular risk wishing to discuss whether to take drug treatment for stage 1

hypertension. The conversation will be very different if the patient talks to a doctor who understands the benefit to be a 1% absolute risk reduction in cardiovascular events over 10 years rather than one who thinks it is 20%. Consider also how a doctor might encourage a patient with type 2 diabetes to take treatment for tight glycaemic control if they believe it will reduce the chance of hard microvascular outcomes by 10% over 5 years (as 20% of responders did in Q15b) rather than 0% (what the key trial showed). Underestimation by clinicians of the benefits of anticoagulant therapy for stroke prevention in atrial fibrillation coupled with an overestimation of bleeding risk is known to be a factor in the under-use of this highly effective treatment (170, 171).

The responses to the main questions and the confidence question are in one sense unsurprising in view of the findings of the systematic review described in the previous section, and my personal and anecdotal experience. However, I was taken aback by the scale of inaccuracy and extremely low levels of confidence. This is core GP work affecting tens of millions of people. Are we providing the person-centred, evidence-based care to which we aspire?

2.5 The case for improving GPs' understanding of the benefits and harms of treatments for long-term conditions

There is an obvious need to address these shortcomings in our collective knowledge-base and practice which has been echoed in policy and the literature over the last decade. The NICE multmorbidity guideline (128) was accompanied by a "database of treatment effects" which provided summary data from clinical trials presented as NNTs and ARR, acknowledging that this information was necessary to support its recommendations. However, this was presented

as a complex and unintuitive Excel spreadsheet. The guideline authors highlighted the need for better access to information on the absolute benefits of treatments in their BMJ summary paper (156).

A 2012 paper led by the Chair of the multimorbidity guideline committee entitled "Adapting clinical guidelines to take account of multimorbidity"(78) considered ways in which clinical guidelines might be improved:

> *"Providing meaningful comparative data on likely benefit and harm in an individual is not straightforward, but this is again an argument for systematically embedding what data exist in individualised guidelines to support clinicians and patients in decision making rather than providing virtually no data on comparative effectiveness as currently happens."*

In a letter to the BMJ in 2014 in the wake of the controversy surrounding the updated NICE guideline on lipid management that year, Goldacre wrote, "NICE must do better at summarising and communicating evidence on statins" (172, 173). Highlighting that information on the absolute benefits of statins was essential for conversations between doctors and patients, he suggested:

> *"Clear summaries of information on benefit and risk are the bedrock of informed patient choice. They should be our highest priority, not a poor second cousin. A simple table in NICE's own summaries—giving NNTs we know, and highlighting those we don't—might be a good place to start. I hope others will have more suggestions."*

The MAGIC (Making Grade the Irresistible Choice) programme was established in 2009 by a group of European and North American academics and clinicians. They aim to address a number of the problems with EBM and guidelines, including the absence of information on effect sizes and strength of evidence (72, 174). Major outputs have been new formats and tools for presenting evidence one of which is the BMJ "Rapid Recommendations" portfolio (175) which will be reviewed in chapter 6.

Pathirana et al identified reformation of clinical guidelines as one solution to the problems of overdiagnosis and overtreatment in 2017 (103). Likewise, when suggesting solutions for the challenges facing EBM, Greenhalgh et al wrote in 2014 (55) :

> *"…evidence must be individualised for the patient. This requires that research findings be expressed in ways that most people will understand (such as the number needed to treat, number needed to harm, and number needed to screen) and that practitioners, together with their patients, are free to make appropriate care decisions that may not match what "best (average) evidence" seems to suggest."*

2.6 Shared decision-making

The idea of shared decision-making" (SDM) has been evolving over the last 25 years or so, aiming to integrate this kind of information about clinical evidence into practice and consultations. It is a mature field, with international networks, conferences, libraries of decision aids, methodological standards, consultation models and implementation research (176-182). More recently, NICE produced a guideline on SDM (183) and NHS England has made this a focus of activity with a programme of work via the Personalised Care Institute.

Both aim to educate professionals in the ethos and skills required to support SDM (184), as well as the production of a suite of decision support (185, 186).

This work has typically involved the production of patient-facing information resources or tools, designed either for patients to read alone, or to share together in a consultation with a professional. The tools have a variety of formats: textual, diagrammatic or graphic with degrees of interaction with online resources. They are structured around quantitative and qualitative information about the benefits or harms of interventions for a single condition, and may be supported by a consultation model to enable their optimal use.

Randomised controlled trials of the use of decision aids have shown that use of decision aids improves patient knowledge and a sense of being well informed, risk perceptions and their level of engagement in decision making (187). However, uptake of this strategy globally has been poor. Barriers described include institutional and organisational issues, time limitation and complex interpersonal dynamics between clinicians and patients (188-190) .

My personal sense is that they represent an idealised concept of clinical practice, where professionals and patients have adequate time, interest and capabilities to engage in rational, deliberative decision making about single questions which is simply not reflected in the real world. However, the work and progress in this area has been extremely influential and informative for this project – this will be explored more in section 6.1.1.2 where I report on GPs' reactions to some decision aids as part of the design-research.

2.7 What do patients want?

Does the numerical value of a long-term treatment matter to patients - how much risk reduction or potential benefit would make it worthwhile for someone to take medicines? Do patients want to "know the numbers?"

A number of studies have explored the question of the minimal amount of treatment effect or risk reduction required to make individuals consider taking a long-term treatment. These usually involve a questionnaire or interview method and employ a variety of modes of communicating treatment effects (such as ARR, NNT, RRR). A 2011 study reported the responses of 114 patients referred for bone density measurement when asked what level ARR they would desire to warrant taking a tablet to prevent osteoporotic fractures. The mean response was a 50% ARR, with an inter-quartile range of 25-60% and total range of 0-100% (191). A 2012 review of patients' requirements for ARR for stroke associated with anticoagulant treatment found answers ranging from 1.65% to 6%(192). A 2017 systematic review explored the minimum ARR of a cardiovascular event that patients feel would justify daily treatment. It included 22 studies and found wide variation in responses, ranging from a 5-year ARR of less than 1% to 10% (193).

These studies are limited by the artificial nature of enquiry – might patients' preferences change when faced with a real decision about their own health, or at a different age? However, it is clear that there is a wide range of perceived value about long-term treatments – we cannot assume all people would want the treatments recommended by a guideline committee or their clinician. Even when best methods of evidence generation and appraisal

are applied, there is still an individual value judgement to be made about whether a treatment is “right for me”. This should happen in the consulting room, not in a committee room.

The question of whether patients want to know numerical data about treatments is much less explored. A systematic review of qualitative studies of patients’ attitudes to taking cardiovascular preventive medicine revealed a number of themes: questioning drugs, preferences, trust in healthcare professionals, influence of family, friends and the media, and perceived benefits and risks (194). However, even within the theme of benefits and risks, there was nothing on the *quantification* of these, views were expressed in terms of heuristics or qualitative aspects: “…prevention to be better than a cure” or “if the drugs extended lifespan…or lowered the risk of having a heart attack”.

Though the shared decision-making literature describes a clear desire for patients to be involved in their care and decision making in general terms (188, 195, 196), there seems to be little direct research about whether patients actually want the numbers specifically. A review of barriers and facilitators to SDM did identify an “informational paradox” whereby information provision alone does not seem to encourage SDM. The interpretation was that imbalances of power and expectation over-ride information availability when making joint decisions (188).

A large US survey conducted in 2002 assessed people’s preferences for participation in medical decision making. 96% of respondents preferred to be asked their opinions and offered choices, but 52% preferred to leave the final decisions to their doctor (197). A 2012 survey and interview study of patients in the UK, exploring decision making around prescribing revealed a similar paradox: 85% of patients agreed with the statement “I prefer

to take an active role in consultations with my doctor", however 65% agreed with the statement "I want my doctor to tell me what to do"(66).

2.8 Patient and Public Involvement

To further explore the views of patients and the public, I conducted an initial phase of patient and public involvement (PPI) work, which started during the design of the online GP survey. Via the "Thames Valley Patients Active in Research" group, academic and personal contacts I recruited 14 members of the public, engaging at two in-person meetings, one Skype call and one e-mail exchange. Following publication of the survey results, an in-person feedback meeting was held with roughly half of the participants.

Invitations to participate included a plain language summary on the background and aims of my research. Volunteers were adults over 60 who had taken long term medication.
I sought their views on the importance of the project in general, and in particular on their perceptions of GPs' knowledge deficits and priorities regarding the benefits and harms of treatments.

Key themes from discussions were:

- surprise (and alarm) regarding (a) the variable magnitude of benefits of common treatments and (b) evidence of variable knowledge among clinicians
- an expectation that their doctor makes decisions in their interests - cognisant of benefits and harms
- unsatisfactory attempts at choice discussions with GPs

- emphasis on the importance of side effects and harms
- no easy consensus on “how much benefit” makes a treatment worthwhile
- strong support for the proposal to improve GPs’ understanding

In addition, I have drawn on the “Public dialogue on medical evidence” report by Ipsos MORI for the Academy of Medical Sciences (2016)(198), exploring how the public view medical evidence and make decisions about medicine. In this exercise, four public/professional facilitated workshops with >100 participants reported findings including:

- a desire for medicines to work for them as individuals
- trust in GPs to make decisions about treatment, regarding them as gatekeepers to information about medical evidence
- concern about multiple medicines, risk of side effects and interactions

I took the PPI findings as expressing clear support for the project, both in agreeing that there is a problem and a need for improvement. Two particular messages which shaped the project were:

- the need for a strong focus on side effects and harms – a clear priority for patients
- the need and value to engage PPI members in the co-design phase, to ensure that information on treatment benefits aligns with their needs, not just those of GPs.

The other finding of interest was support for the idea alluded to at the end of the last section: that patients wish to be involved in decision making but still want their doctors to play an active role in making final decisions and wish to trust in their knowledge of medical evidence.

2.9 Summary

This background literature supports the idea of a clinician-facing information resource to improve the understanding of the benefits and harms of treatments for long-term conditions. Filling this gap in information availability has the potential to at least partly address many of the tensions and problems described in chapters 1 and 2:

- the lack of trust in and perceived rigidity of guidelines
- the difficulties of applying guideline recommendations to individuals
- the undermining of doctor and patient autonomy by over-mechanistic EBM
- the twin problems of over- and under-use of evidence-based treatments - might a better awareness of treatment values make an impact?

Such a resource would provide a missing element in managing polypharmacy, multimorbidity and medicines optimisation. It has the potential to re-locate value judgements about treatments closer to the patient and their clinician.

That it should be doctor-facing rather than patient-facing is important. Firstly, there is a clear knowledge gap among GPs. Though patient-facing resources do have the potential to improve clinicians' knowledge, indeed I have heard decision aids described as "Trojan horses" to educate clinicians, their use is not widespread enough to make a significant impact. Secondly, the relative failure of shared decision-making (built around patient-facing tools) to change the landscape is instructive, coupled with *some* information suggesting many patients do not want to take on quantitative, final decision-making, suggests a different approach is needed. There are a handful of such online resources already, though they have important limitations. These will be described further in chapters 4 and 6.

2.10 Aim and Scope of this project

This is an appropriate point to clarify the goal of the *GP Evidence* website: to improve GPs' understanding of the benefits and harms of treatments for long-term conditions.

What then happens with this new knowledge – what will subsequently happen in the consulting room, or whether or how much this knowledge will help address the problems discussed in these introductory chapters, can only be answered once GPs have access to this information. Though I undertake an exploratory evaluation study (chapter 8) to assess how well a group of GPs understood the website content and how they imagined this might affect their practice, these further, broad questions are outside the scope of this project.

The aim of the website is to provide information on the benefits and harms of treatments in a way that can be understood by GPs and potentially integrated into their practice, filling an important gap in the knowledge available to them. This could provide an enhanced opportunity to combine the best available evidence with their clinical experience and patients' preferences.

2.11 Assumptions, potential pitfalls and the need for this research

The idea that the simple provision of online information will improve GPs' understanding, let alone change their practice, involves a number of assumptions and raises important questions:

- Do GPs even want this information anyway?
- Do they have capacity to absorb it in terms of time and cognitive load?
- Do they have adequate confidence in interpreting the results of clinical studies?

- Can this information be communicated adequately, managing tensions between simplicity, clarity, accuracy and detail?
- Where might this information sit within their complex practice?

The answers to these questions were unknown. I sought to answer them over the course of the doctoral research, whilst at the same time producing a resource which could made available to practising GPs. Employing participatory co-design as the central research method allowed both these to happen in parallel – I will describe my choice of this method in detail in Chapter 4.

The next chapter will describe the first phase of research: interviews with GPs.

3 Chapter 3: General Practitioner Interviews

3.1 Aims

In chapter 2, I outlined the theoretical basis supporting the idea for a website to deliver quantitative information on the benefits and harms of treatments, as well as important assumptions and potential pitfalls.

This phase of research relates to the first overarching research question:

- What are GPs' needs and preferences for accessing evidence-based quantitative information on the benefits and harms of treatments?

I undertook a series of one-to-one interviews with GPs with two aims:

Aim 1) To understand how GPs currently think, reason and feel about quantitative information on the benefits and harms of treatments.

Where does this type of information currently sit in their clinical practice? Do they currently use it, if so, how and if not, why not? How do they feel about the idea of using more quantitative information and what benefits or challenges do they imagine this would bring? There is a gap in the evidence base on EBM, multimorbidity and polypharmacy regarding these specific questions. This aim also relates to a basic imperative of design research: to understand your user. I will discuss this further in Chapter 4.

Aim 2) To explore GPs' preferences when viewing online information, in order to inform the design of the website: how to deliver complex information in a way that is understandable, useful and usable in practice.

3.2 Research questions for GP interview study

- What are GPs' attitudes to and understanding of the quantitative benefits and harms of treatments for long term conditions?
- What are GPs' design preferences for access to evidence-based information on the quantitative benefits and harms of long term treatments?

3.3 Presentation of interview methods and findings within the structure of this thesis

In this chapter, I will describe the methods relating to answering the first research question and how those findings contribute to the medical literature. In Chapter 6 I will describe how these findings were used to inform the design process.

The second research question relates directly to the design process, so I mention the methods used to answer this only briefly in this chapter, describing them in detail in chapter 4 alongside other design methods. In Chapter 6 I describe the findings, which more directly inform the prototype design.

3.4 Ethical approval

Ethical approval for this study was granted by the University of Oxford Medical Sciences Interdivisional Research Ethics Committee. Reference: R55459/R0001 (appendix 3.1). I judged this to be a low risk study from an ethical perspective; details of my considerations regarding this are given in appendix 3.2.

3.5 Methods and approach

3.5.1 One-to-one interviews

I chose one-to-one interviews as a method because I wished to explore a complex issue in substantial depth. My survey described in section 2.4 had revealed low levels of knowledge

and confidence in this area and I anticipated that detailed and perhaps probing questioning would be needed to find out what was underlying this. In addition, participants would be talking about something which might reveal a deficit in their professional knowledge or skills. For these reasons, I abandoned an initial idea I had to conduct focus groups, thinking that vital insights may be missed if groups of outwardly confident and perhaps competitive doctors wished to hide their real thoughts and feelings from each other. Some of the questions I wished to ask addressed an individual GP's management plan for a patient, or reactions to existing websites, which are best addressed through one-to-one conversations.

Another option I considered was some form of ethnographic research, observing GPs in practice with a focus on their use of quantitative information in consultations. The maxim "don't believe what people tell you they do, believe what you see them do" drove this idea. However, knowing that GPs' knowledge and confidence is low in this area, it would seem probable that they are only using this information infrequently in practice, so it may have been difficult to observe. Other researchers have found EBM decision making to be rarely observable in GP practice. ref

3.5.2 Analysis: the framework method

I chose the framework method, a form of thematic analysis with origins in social policy research, for this study for a number of reasons. It is a flexible tool, not aligned with a particular theoretical approach and is suited to applied or policy research with practical aims and objectives (199). The process of analysis allows for both a deductive and inductive approach, suitable for this study to which my own experience and the background literature were brought to bear as well as the empirical findings (200).

The defining feature of the method is the construction of “framework matrices”, structures of rows and columns creating cells which contain data and interpretation. This allows the consideration of data within and across interviews whilst organising data into categories and themes. It provides an audit trail linking data, codes and themes to support transparency and trustworthiness of the analysis.

The method was developed by Ritchie and Spencer (201) in the 1980s and is described in a number of qualitative methods textbooks (202-204). These vary slightly in their terminology and description. Drawing on these I have summarised the key stages in table 3.1, using terminology described in a pragmatic overview by Gale *et al* (199) and updated methodological advice from Ritchie and Lewis (204).

Table 3.1 Summary of the framework method

Stage of Analysis	Description
Transcription	Audio-recordings of interviews are transcribed verbatim into electronic documents.
Familiarisation	The researcher undertakes an initial reading of the interview data.
Coding	The data needs to be organised by topic. This is initially done using “codes” - categorisations of types of data at a “fine grain” level. Initially, a set of deductive codes are created, derived from researchers’ prior knowledge, in anticipation of content that is likely to be present in the data. This results in the creation of a: Coding framework (sometimes called a code book). A list of codes which can be applied to the interview data. Each code has a written definition to support consistency and clarity of application. Inductive coding (sometimes called open coding). As the transcripts are read and coding applied, new categories of data will become apparent and new codes are created to describe and contain them. Indexing. The researcher reads through the transcripts and organises the data by applying codes to sections of text. More than one code may be applied to one piece of text.

	Iteration. This coding process is highly iterative, with the development of new codes, retirement of redundant ones and re-definition of codes as needed as the data is explored and understood.
Matrix Creation	During the process of coding, the researcher will observe commonalities and themes within the data, begin to develop insights and organise their understanding of the data. This allows the creation of a Framework Matrix, one for each major theme. Each code (sometimes referred to as a sub-theme by this stage) is allocated to one or more matrix. Each matrix is organised by "case" (i.e. one interview) which is ascribed a row on the matrix and by codes which are ascribed columns. This process creates individual "cells" which contain data from one interview regarding one code or sub-theme. The analytic process can generate more than one matrix, depending on how many major themes are created.
Charting	Using Computer Assisted Qualitative Data Analysis Software, the matrices are populated with research data organised by case and code (sub-theme). The researcher then reviews the content of each cell, considering the meaning of the data, recording any interpretation and linking any specific text to this.
Further analysis	Once the matrices have been populated and reviewed, there is the opportunity for further analysis. Exactly how this is done does not appear to be definitively prescribed. Gale *et al* (199) say: *"Gradually, characteristics of and differences between the data are identified, perhaps generating typologies, interrogating theoretical concepts (either prior concepts or ones emerging from the data) or mapping connections between categories to explore relationships and/or causality. If the data are rich enough, the findings generated through this process can go beyond description of particular cases to explanation[s]…"* Ritchie and Lewis (204) describe further processes: The creation of "Elements" drawn out of matrix content, their categorisation, then high-order classification and linkage of content across themes and cases to support interpretation.

3.6 Recruitment and Sampling

Interviewees were recruited from a pool of 213 GPs who had completed the online survey described in section 2.4, having been originally recruited via widely distributed email

invitations. The survey respondents had demographic characteristics broadly representative of the UK GP population (appendix 3.3).

They were sent an email invitation (appendix 3.4) on 13th March 2019 with a reminder sent 10 days later. This was accompanied by a full study information sheet and copy of a consent form (appendices 3.5 and 3.6). Volunteers followed a link to a Survey Monkey questionnaire to confirm their eligibility, enter demographic details and re-submit their contact email address. They were offered a £75 voucher as an incentive to participate and a certificate of participation for their Continuing Professional Development (CPD) records.

The inclusion criterion was to be a:

- GP currently practising in the NHS.

Exclusion criteria were:

- GPs who had not been in NHS clinical practice within the last year.
- GPs still completing specialist training (ST1-5)
- GPs with further training or extended roles which may confer expertise about the subject matter. For example, an academic GP or a CCG prescribing advisor. Participants were asked to provide a free-text description of any additional roles and a judgment was made by myself.

28 GPs volunteered, one was excluded due to additional expertise in prescribing practice. Purposive sampling of 15 interviewees was used to achieve maximum variation with regard to age, sex, geographical region, rural/urban setting, GP role (principle, salaried, locum) and deprivation index of GP practice.

I anticipated that this relatively small sample size would generate adequate data. In considering sample size, I referred to guidance from Morse (205), who describes characteristics of research studies which may influence required sample sizes. My study had characteristics suggesting a smaller number of participants would be adequate: a well-defined topic, participants likely to be articulate and deliver high quality data, a relatively homogenous group of participants and interviews of relatively long duration. I planned to use the principle of data saturation (206) to determine when enough interviews had been conducted, with contingency for further recruitment if this was not the case.

I chose to interview GPs in Scotland and Wales, anticipating geographical differences to England due to polypharmacy guidance which had been published in these countries a few years prior to the interviews (149, 150).

3.7 Interviews: Process and data collection

The two research questions to be addressed in this study are related, but required different types of enquiry.

The first - *What are GPs' attitudes to and understanding of the quantitative benefits and harms of treatments for long term conditions?* - required a traditional interview approach combining open and more closed, probing questions. I prepared two fictional vignettes (appendix 3.7) describing patients with multimorbidity and polypharmacy which could be presented on paper to interviewees to prompt discussion if needed.

The second - *What are GPs' design preferences for access to evidence-based information on the quantitative benefits and harms of long term treatments?* – required a supplementary

approach to review websites with interviewees: the “think aloud” method which I describe in section 4.4.2.

Though planning to address these two questions separately in the interviews, I anticipated there would be considerable overlap in conversation flow and content and wished to allow this to happen naturally. I therefore chose to undertake semi-structured interviews with an interview guide in two parts to be employed flexibly.

3.7.1 Development of interview guide – Part 1

I wished to explore how GPs currently practice, reason and feel about quantitative information on the benefits and harms of treatments. I anticipated that there might be areas of their practice where this was being applied which would be useful to explore and also areas where it was not. I was interested in how they were making prescribing decisions in the absence of this information, in order to understand current clinical practice for its own sake, but also to gain insight into their needs as future users of a website.

I designed this part of the interview guide (focusing on the first research question), from experience of my own practice, knowledge of the background literature and consideration of my research aims. Two academic supervisors gave feedback leading to minor amendments and additions.

3.7.2 Iteration of the interview guide

I decided not to set up formal pilot interviews to test the guide, judging that as the interviews were so long, this would waste valuable data and be an unethical use of participants’ time. However, after two interviews I revised the questions to improve the flow of conversation, creating greater differentiation between open and closed questions to improve focus. The

final version of the interview guide is reproduced in Figure 3.1. The early version is shown in appendix 3.8.

Figure 3.1 Interview Guide for GP interviews - part 1

Part 1. Perspectives

When you are prescribing treatments for patients with (or without) multiple chronic conditions, how do you consider the balance between benefit and harm?

When you consider the balance between benefit and harm, what is it you are thinking about?

Do you think about the size of treatment benefits or risks in your daily practice? I mean in numerical terms, like absolute risk reduction and NNTs. If so, how and why?

Does it (the quantitative aspect) come up very often, or not much at all or never? Either in your mind as something you'd like to know or actually include in discussions with patients?

If you're not using these sorts of numbers, how do you or your patient make a value judgment about the pros and cons of treatment?

Do you feel confident about making informed decisions about benefit-harm balance? If not- what would you need to feel confident?

Thinking about making these prescribing decisions in general: at what points in your day do you find yourself thinking about them (inside or outside the consultation)?

Can you describe an example of how this has come up in your practice?

What was your consideration/query?

What did you do?

Did you look for additional information to help?

What kind of information? How, where, what search terms, etcetera?

Was the information helpful?

What kind of information has been helpful to you?

If we just think purely about quantitative/numerical information about absolute benefits and harms…. How important do you think this kind of information is for you and your patients?

How you think about this? In numerical terms, like, say percentages or in other ways?

Information on the benefits and harms of treatments are sometimes presented using statistical terms like: Relative risk reduction, absolute risk reduction, numbers needed to treat. Are you confident about the meaning of these or not? Without making this into an exam - could you describe what these terms mean to you?

If GP replies they are not using it much:

Why do you think we don't use this kind of information much?

How do you feel about the fact we don't use this information much?

Would it be better if we were, or not really? How good do you think our access is to this kind of information? Has it formed part of your education (at any stage of your career)? Do you perceive any shortcomings in the current resources available to us? I mean guidelines, websites, resources, etcetera? Let's imagine there was a new resource to give us more information on benefits and harms of treatments. Would that be useful, or not really?

3.7.3 Interviews – Part 2

I wished to explore how GPs might prefer to view quantitative information on the benefits and harms of treatments online. I expected some information about this to emerge in the first part of the interview, but this second part addressed this specifically.

I selected five online resources which deliver this kind of information and asked the interviewees to either explore them freely or try and seek a particular piece of information from them. They were asked to "think aloud" as they did this, describing their reactions during the process. I will describe this method in detail in chapter 4, and the findings in chapter 6, where they align directly to the design research process.

3.7.4 Interview process

Participants were contacted by email and a time arranged to interview them in their surgeries. They were asked to be available for up to two hours. Conducting the interviews in their workplace allowed access to their usual electronic and paper resources as well as revealing some context about their practice. After introductions, they were given the opportunity to ask any questions before signing the consent form. Interviews were audio-recorded using a dedicated, non-web-enabled digital recorder (Olympus WS-853).

A pre-prepared introductory statement was used to start the interview (fig 3.2). The second paragraph was added after the first two interviews as I realised my questioning needed to be quite probing and I did not wish to make the participants uncomfortable or defensive.

Figure 3.2 Introductory statement for GP interviews

> Thanks again for agreeing to participate. I'm hoping to cover quite a bit of ground during the interview so thought it would be helpful to outline the plan briefly. You remember the questionnaire and the sort of clinical questions that were asked in that: the percentage chances of benefit or harm with various treatments? We're going to be talking about that subject matter – in two parts if possible. Firstly, what you think about that sort of information currently, and secondly about better ways we could find to communicate that sort of information to GPs. I'm sure our conversation will move around the two things quite a bit but I have a loose structure and some questions here to guide us.
>
> I want to add that it's important to realise that this interview isn't in any way a test of your knowledge – the reason we are doing this research is because we know that as GPs, we don't have good information at our fingertips about all this. So, if I'm asking probing questions it's just to try and get to the heart of the matter – not because you haven't given the "right" answer!

The interview script was followed in a flexible manner, with GPs often turning to their computers, or pulling books or paperwork from shelves and noticeboards. The separation between Part 1 and Part 2 of the interviews held up naturally for the most part, with occasional overlap which did not interrupt the process. Immediately after the interviews, I recorded some brief field notes on paper with observations about the participant, their attitudes and contents of the interview that had not been demonstrated clearly in verbal form.

3.7.5 Data handling and transcription

Data handling and storage was conducted according to University of Oxford policy. Detailed description of this process is reproduced in Appendix 3.9. Audio files were downloaded onto secure University computers. Transcription was performed by a University approved

transcriber with data transferred via encrypted files on an approved network (OxFile). Transcriptions were stored on Microsoft Word files.

Participants' personal data was stored separately from research data. Research data (including audio files) were pseudonymised by allocating a number to each interview. During the familiarisation phase of analysis, interview transcripts were edited to remove identifying data such as names of people, places or institutions. Pseudonymised transcripts were uploaded into NVivo(v12) Computer Assisted Qualitative Data Analysis Software to manage the process of analysis.

3.8 Analysis

3.8.1 Familiarisation

Once transcripts had been received, I read through them in full to re-familiarise myself with the interview content. I corrected spelling mistakes and transcription errors, filled in blanks and removed identifying data. During this process I wrote notes on thoughts and insights which arose, which informed iterations of the coding framework.

I developed insight during this process, seeing things in the transcripts that had not been apparent to me during the interviews. When undertaking coding and indexing in the next step of the analysis, I was aware that this familiarisation phase had been valuable to form a deeper understanding of what was being said. For example, the context of a particular piece of transcript with respect to both the language and meaning of what had been said, the individual GP and their environment.

3.8.2 Coding

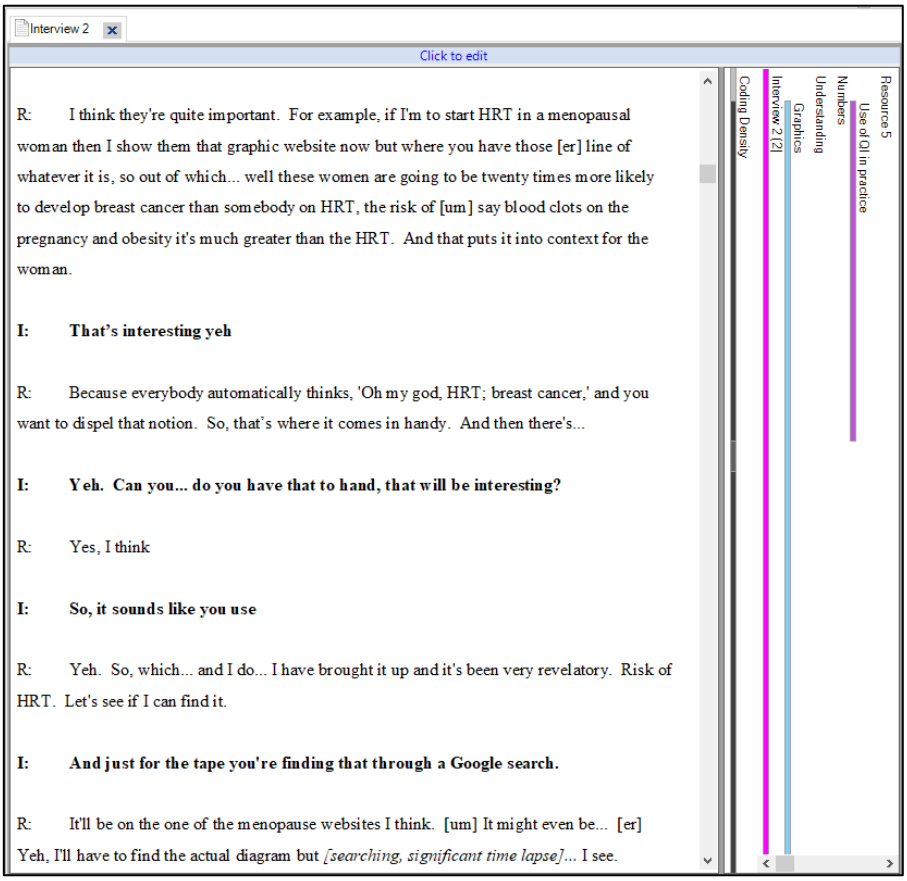
I developed an initial coding framework before conducting any interviews, using a deductive approach drawing on my experience and background literature, anticipating what participants might say and areas on which I knew I would focus my questions. Screenshots of this initial list (from a Microsoft Excel file) are shown in appendix 3.10. Initially I created a hierarchy of: Codes, Category and Meta-Category (this final classification to distinguish the two parts of my interviews). In the end, this hierarchy became irrelevant with code labels being applied to all text simply as codes, though some data suited more fine-grained labelling and some suited broader labelling of the type I initially called Category.

For the first two interviews, joint coding was undertaken. For Interview 1, with one of my academic supervisors (JC) and Interview 2 with a post-doctoral researcher from my department (Dr Alex Rushforth). For the first interview, I coded the transcript alone, then compared with JC who had also coded independently. For the second, AR and I sat in the same room and coded independently for short periods of time, then compared results. Variations and disagreements in coding were resolved by discussion.

This process of joint coding was valuable in a number of ways: support in the technique of coding, fresh perspectives on the interview content, challenging my interpretation and assumptions, stimulus to greater clarity regarding the use of codes and their definitions, and new insights into the data. Joint coding is recognised as a possible method of improving rigour in qualitative research (207). After two joint coding sessions I judged (after discussion with my supervisors) that my approach to and structure of coding was established enough to

complete the coding of the rest of the interviews independently. A screen shot of a section of coded interview is shown in figure 3.3.

Figure 3.3 Screenshot of a section of coded interview from NVivo v.12



I adjusted the coding framework in an iterative process until the ninth interview was coded. Some new codes were introduced (inductive coding arising from interpretation of the data), some codes were merged or retired and definitions of some were refined. The final version of the coding framework is available in appendix 3.11.

3.8.3 Preliminary themes, framework matrix development and charting

Once coding was completed on all interviews, I was able to identify themes and then create and populate my framework matrices. Data regarding participants' future preferences for an

information resources and their “think-aloud” feedback on websites was not included in the framework analysis (though I had coded this content). I will describe how this data was used in section 5.1.

By the time coding was complete, I had thought of four overarching themes within which the data would sit (these are defined in table 3.3). This insight had occurred naturally after spending so much time with the data, rather than through a formal process. For each theme, which would become an individual matrix, I allocated a set of codes pertaining to the theme which would be used to populate the matrix. Some codes were relevant to more than one theme and I anticipated overlap between themes. Figure 3.4 shows an excerpt from the draft schema of codes and themes with provisional thoughts. The framework method allows for iterative adjustment of the titles, structure and content of the matrices. In the end, this was not needed, though a final phase of further analysis dealt with commonalities and linkages between themes.

Figure 3.4 Draft schema of codes and themes. Screenshot from research diary (QI=quantitative information on the benefits and harms of treatments).

	Codes	Themes/ comments
2	Ben,Har,Thr,Com, KS,SU, ExUQI,UQp,	**Current use of QI**. What are they actually doing, how are they using it, what effect for patients, positives and negatives
3		
4		
5	UQp, Ben, Har, Ti, diffSU, ownSU, ExUQ	**Current non-use of QI.** Descriptions, feelings/opinions, Explanations. Will overlap with theme 3 but emphasis is on WHY not or HOW are they FEELING ABOUT IT?
6		
7		
8	PF Hu Inf KS Co	HOW are they **making decisions if/when NOT using QI**. Overlap with theme 2 but emphasis is on how they're functioning/doing the best they can for patients.
9		
10	{	**Feelings regarding the future use of QI**. NB Whilst half way through charting, I added in SU to this as confidence (or not) affects if and how they might use Qi in future. Also ExUQ for similar reasons
11	Qinv, SU, ExUQ	
12	{	
13		
14		
15		
16		Story? GPs are not using QI much - reasons and attitudes
17		Instead, they use work arounds, rule following, mindlines etc
18		They would like to use more QI (with some reservations and identified barriers) but this desire is not yet a fully formed concept. It is hidden away from our conciousness - too difficult?

Once the matrices were created in NVivo(v.12), the software auto-populated them with coded sections of text, creating content in each cell of the matrix which pertained to a particular interview and a particular theme. I was then able to begin the charting process. This involved examining each populated cell on a matrix and reviewing the text and code associated with its content. Then, writing a brief summary of the data, its meaning and any interpretive comments into the cell.

Ritchie and Lewis (204) advise that the summaries of data should be brief enough to be manageable, but rich enough to convey meaning. Interviewees' original language, key terms and expressions should be adhered to as much as possible. Researcher interpretation should be minimal, and be recognisably different from the data. Linkage of this new cell content to

original data is maintained by the software. An illustration of some completed matrix cells is shown in figure 3.5.

Figure 3.5 Populated cells on a Framework Matrix.

Matrix 1 Use of QI in practice | Matrix 4 Future use of QI | Matrix 2 Not using QI in practic

	A : Benefits	B : Communication	C : Thresholds
8 : Interview 2 (2)		Describes using QI directly in communication with patient. And its value - dispelling an incorrect perception of risk about HRT.	Using a specific figure of CV risk to make a treatment decision. Expresses discomfort.
9 : Interview 3	Refs 2-4 in this cell: discussion about diabetes meds. Aware that most drugs don't have hard end point or cardiovascular benefits but doesn't have quantitative level of detail. I reviewed the whole interview text again to see if this was specified- it wasn't but I think would have come out if she DID have this info	Describes using Qrisk calculator to communicate benefits. Though strictly speaking is incorrect	Shows a knowledge of the 10% CVD risk threshold for statin prescribing, but then expresses uncertainty about its value. So, is using QI but with some discomfort/low confidence.

Rows=Interview number, Columns=Code, Plain text =data summary, Underlined text=researcher comment or interpretation, Highlighted text=software link to specific piece of textual data.

Completion of all four matrices thus resulted in a collation of findings, organised by theme and sub-theme (code) complete with abbreviated summaries and researcher reflections. The process of performing this involved spending much time with the data, re-reading and reflecting, which generated new insights and understanding.

3.8.4 Further analysis

I was now ready to draw the findings together into a meaningful narrative. Though such a narrative was reasonably clear to me, I recognised that it would involve drawing on data which sat across themes and that some data initially thought to be most relevant to one theme might be better placed in another.

My first step was to read through the matrices again, making freehand notes which linked quotes and themes to the narrative. Using cell identifiers (e.g. 6I, 11A) I could closely relate the notes to the matrices and used different coloured inks to create cross links across themes.

My next step was to create a new table in Microsoft Word of themes, sub-themes and data (quotes) to be used in the narrative. The four overarching themes defined at the stage of matrix creation still worked well. New sub-themes were developed drawing on the previous step of analysis, with corresponding quotes and matrix cells recorded alongside to link the new sub-themes to evidence. These sub-themes and their supporting evidence are described in the results section.

3.8.4.1 Member Checking

Member checking is a technique to improve quality and rigour in qualitative research. Recognising the potential for researcher bias, this process aims "*to validate, verify, or assess the trustworthiness of qualitative results"* (208) by seeking and integrating the views of research participants into analysis. It may be conducted at varying levels of intensity, from joint discussions of interview transcripts and coding with participants, conducting focus groups during the analysis, or sharing fully analysed data for feedback (209). I chose to adopt the latter approach. I emailed the 14 included participants inviting them to read a near-final

version of the manuscript in preparation for journal submission. They were invited to make general comments and also to say if they felt if their particular quote had been misinterpreted in any way (they were individually told their interview number to identify their quotes).

9 replies were received. Replies were all positive and no changes were made to the manuscript. Two participants responded to an aspect in the limitations section to state that they felt the interviews has been conducted in a way which allowed honesty and was not leading (participants 6 and 11).

Appendix 3.12 shows the invitation email, participants' comments and my own notes.

3.9 Results

15 GPs were interviewed between May and August 2019. One interview was excluded as the participant had not declared significant subject expertise. A diverse sample was achieved. Their characteristics are listed in Table 3.2. Data saturation was judged to have been reached and no further interviewees were recruited.

Table 3.2 Participant characteristics for GP interview study

Characteristics (n = 14)	
Sex	
Female	8
Male	6
Age, years	
Under 30	2
30-39	3
40-49	5
50-59	4
GP Role	
GP Principal	5
Salaried GP	5
Locum GP	4

Place of original medical degree	
UK	12
Non-UK	2
Geographical Region	
North-East England	1
Yorkshire and Humber	1
East of England	1
Greater London	3
South-East England	1
South-West England	4
North Wales	1
East of Scotland	2
GP description of practice	
Urban	6
Rural	2
Mixed Urban-Rural	4
n/a (locum)	2
Decile of index of multiple deprivation (by practice postcode) 1=most deprived 10 = least deprived	
1	1
2	2
3	3
4	2
5	1
6	1
8	1
9	1
n/a	3

An abbreviation – QIRx

During the write up of this study for publication, it became apparent that the long and clumsy phrase "quantitative information on the benefits and harms of treatments for long-term conditions" would need to be used frequently. After discussion with co-authors of the published paper, we were unable to come up with a briefer alternative. I suggested QIRx (**Q**uantitative **I**nformation + the abbreviation for treatment: **Rx**) which was used to submit to the journal. Use of a new abbreviation was questioned by a peer reviewer. We acknowledged

that this was an unusual thing to do, but could not think of a better alternative and would be open to editorial suggestions. In the end, it was accepted and I will repeat its use here.

The findings were summarised within four themes and related sub-themes, presented in table 3.3 and described below. Further supporting quotes are provided in appendix 3.13.

Table 3.3 GPs' use and understanding of quantitative information on the benefits and harms of treatments (QIRx): Themes and subthemes

Theme	**Subthemes**
Descriptions of GPs' current use of QIRx	• Examples of successful use • "Partial" use of QIRx: risk scores as thresholds but without knowledge of subsequent risk reduction • Internalised non-numerical ideas of the value of treatments
Discussion of the lack of use of QIRx	• Awareness of a knowledge deficit • An absence of accessible information in a context of information overload • Low confidence in statistical terminology • Competing drivers to clinical practice
Making decisions in the absence of QIRx	• Working with non-numerical, internalised "knowledge fragments" • Using knowledge of physiology and extremes of age or risk • Using non-numerical heuristics • Taking into account patients' characteristics • Employing qualitative communication styles to convey non-numerical estimates of risk
GPs' attitudes and feelings about the use or non-use of QIRx	• Positive expressions of the value of QIRx in current practice • Relative contentment with not using QIRx for some • Negative impact on patient care due to a lack of understanding of QIRx • Negative emotions arising from challenges in this area
GPs' attitudes about possibly increasing the use of QIRx in the future	• Imagined benefits of increasing the use of QIRx • Anticipated barriers to increasing the use of QIRx

3.9.1 GPs' current use of quantitative information on the benefits and harms of treatments

Only a small number of examples were given of the use of QIRx, typically only one or two per GP. The prescribing of statins for the prevention of cardiovascular disease was the only area where any participants confidently described a numerical risk reduction (treatment benefit) for a patient. This information was acquired from decision support tools (often featuring infographics) either on paper, downloaded, or built into clinical systems.

> *"Well, I usually say things like.., 'if a hundred people like you took this tablet for ten years. Eighty of them wouldn't be going to have a heart attack anyway but twenty of them would but five people will not have a heart attack that they would have had and we don't know if you're one of those people…'" (Interviewee 1)*

Quantitative information about specific treatment harms was described by a few, for example: major risks associated with hormone replacement therapy, or bleeding risks associated with anticoagulants. This information was usually derived from, and delivered to the patient, using an external resource such as a patient information leaflet. GPs using these seemed to value the information and confidently apply it in practice.

> *"But now I tend to frame it differently and will say, you know, 'These adverse effects are really quite rare; here's a leaflet to explain it' ". (Interviewee 11, discussing how information on the risk of osteonecrosis of the jaw with bisphosphonates was in fact lower than their previous perception. This had led to a change in practice and apparently greater uptake of this treatment.)*

Numerical risk thresholds such as QRISK or FRAX (numerical scores calculated to estimate patients' baseline risk of future health problems) were discussed by many. These could be regarded as a partial aspect of QIRx, but were often discussed without a subsequent

understanding of how treatment might reduce this risk, i.e. the actual benefit of the treatment. Instead, they functioned as a simple treatment threshold.

> *"NICE… talk about if your risk is over ten percent you are likely to benefit…So, I tend not to take an ownership of the recommendation. I say 'this is what's recommended'" (Interviewee 4)*

One GP described using a risk threshold to support a decision they had already made on the (non-numerical) basis of various risk factors. Some described using internalised ideas about the value of treatments which were non-numerical or imprecise.

> *"…so I think, 'well, can I take him off his simvastatin?' … I would think number needed to treat, five hundred or something, whatever it is… On the other hand, there's atrial fibrillation, risk of stroke and that's extremely high…so, I would be a lot more trigger happy to start on a NOAC [Novel Oral Anticoagulant] or a warfarin, and that's got a very low NNT, which I can't remember." (Interviewee 2)*

3.9.2 The lack of use of quantitative information on the benefits and harms of treatments

GPs described their awareness of a lack of this specific kind of knowledge, one framing it as "missing" information:

> *"I know that I'm a little bit in the…or perhaps massively in the dark about this." (Interviewee 1)*

> *"But actually… I think it is missing…if you're making a decision about starting someone on a tablet for potentially the rest of their life… it's important you get that decision right." (Interviewee 6)*

Descriptions were given of a lack of QIRx in the context of particular conditions and treatments, for example hypertension, type 2 diabetes and regarding the use of bisphosphonates for fracture prevention.

A number of barriers to the use of QIRx were described. Importantly, a lack of easily available information. One GP described an online tool on the benefits of statins and speculated about the possibility of a similar tool for hypertension treatment.

"I haven't used these [decision aids] for hypertension though…I'm not sure they're there; they might be…but they might not be as clear cut." (Interviewee 1)

On the other hand, the challenge of retaining such information in a context of information overload was described.

Many GPs reported low confidence in statistical terminology (such as ARR and NNT), sometimes implying that they imagined this to be too difficult, or too specialist to be part of core GP skills.

"I did know it from my MRCGP[Membership of the Royal College of General Practitioners examination]…but I'm not convinced I could give a watertight definition now…Because I don't do it day to day…it feels like it's all part of that maths struggle." (Interviewee 13)

"…it's funny because stats were my thing before…it wasn't an area that frightened me, like I like numbers…so it's a bit shocking when you think about it…but I can't even remember what the terms mean." (Interviewee 3)

Wider drivers to clinical practice were described which shape decisions in the absence of QIRx, or might act as barriers to its use were QIRx available. These included clinical guidelines, performance measures such as the quality and outcomes framework, a desire to conform to normative practice and fear of adverse outcomes, including medico-legal fear.

I'm looking at their numbers before they come in and thinking, 'Oh, right we need to get that one better.' Part of the decision making is driven by QOF and the numbers that they are set in in my computer. (Interviewee 1)

"I…have been beautifully indoctrinated for twenty years that 'thou must have a lower blood pressure'" (Interviewee 14)

I suppose in the back of my mind someone somewhere has said "this is a good medication." I don't want to be the one to go against that; him have another heart attack and someone say, 'Well why did the GP stop that?' (Interviewee 12)

3.9.3 Making decisions in the absence of quantitative information on the benefits and harms of treatments

The GPs are still required to make decisions with patients about treatments even in the absence of knowledge about QIRx. A variety of strategies were reported:

Drawing on non-numerical internalised "knowledge fragments" relating to the value of treatments.

> *"So, after a STEMI [ST-elevation myocardial infarction], I think bisoprolol is the most evidence based…again, I can't remember the numbers but…exam question…bisoprolol was the one on the multiple choice that you tick is the most useful" (Interviewee 12)*

> *"So I tell my patients…their most important drug is metformin, out of your diabetes drugs…that's the one that's not only helping your sugars but it's also helping your heart and your vessels..." (Interviewee 3)*

Some drew on their knowledge of physiological mechanisms to support decisions. For example, managing the urate raising effect of thiazides in a patient with gout and hypertension. Another considered the relative short term risks of stopping anticoagulant or blood pressure lowering drugs.

> *"So, if I stop the [anticoagulant] drug and [the blood] becomes thicker…maybe it won't get through. That might have a more imminent immediate effect. But I feel like the blood pressure has more of a longer term effect." (Interviewee 1)*

Framing thinking around extremes of age or risk (and therefore higher and lower chance of benefit or harm) was a mechanism described by many:

> *"So, if you're looking at a ninety year old who's on a statin, you know, the chances are they're not going to live for very many years…the gains are always going to be marginal." (Interviewee 6)*

Some described using non-numerical risk comparators as communication tools. Examples were: comparing the small risk of breast cancer from hormone replacement therapy (HRT) to

the larger risk associated with moderate alcohol consumption, or the risk of osteonecrosis of the jaw from bisphosphonates with getting hit by a car. Some described internalised heuristics from previous teaching or experience:

> *"…it'll be more gut instinct which is really non-numerical and not very medical, but it's a synthesis of the information and experience you've had I guess over the years." (Interviewee 14)*

The GPs' understanding of their patient's characteristics and medical history could be integrated to guide treatment choices. In one example, the GP considered an individual's tendency to develop side effects. In another, the risks associated with a gliclazide for a frail patient with diabetes, and decides (without numerical estimates) that these outweigh the benefits, taking into account the possible non-applicability of trial evidence to this individual*:*

> *"You have to base it on the patient in front of you, not on population studies…she's approaching her eighties … causing hypos is a bit more risky than running a bit high on the sugar at her age maybe. You know, she's living on her own and she has a hypo, she falls, breaks her hip…So I'd be very keen to, at some point, tail off the gliclazide." (Interviewee 2)*

Sometimes, a lack of explicit quantitative information was hidden within qualitative communication styles, such as one example where the GP discusses the choice to treat mild hypertension:

> *I'd probably [say] 'Overall, over the next ten/twenty years if we keep this controlled we're likely to reduce the risk of…' but I'm not going to give them numbers because I haven't … got easy numbers to give them. And I would try to convey … 'There's definitely a benefit, but if you don't want to we can monitor things ....' And it's all … fluffy and communicative…very much non-scientific and non-numerical… because I haven't got the numbers on the tip of my tongue." (Interviewee 14)*

3.9.4 GPs' attitudes and feelings about the use or non-use of quantitative information on the benefits and harms of treatments

Some GPs discussed positive aspects of their current use of QIRx, such as supporting informed choice or medico-legal confidence. Regarding the status quo of relatively little use of QIRx, some expressed a reasonable degree of comfort, being happy to trust and follow guidelines or accepted practice without taking further ownership of decision making.

> *"…even if you told me what the number needed to treat or number needed to harm for any given drug is, I'd forget it. So I just need to rely on guidelines and formularies…to help guide me" (Interviewee 9)*

Others felt less comfortable, expressing concerns about over-treatment or a lack of personalised care:

> *"…my feeling is that we treat a lot of people according to thought-free algorithms because they've got a condition that feeds in the top end. But we do a lot of that without really having a sense of how important what we're doing is" (Interviewee 14)*

> *"…it's very apparent that for people…who are getting older and frailer and have co-morbidities, suddenly you're way over-treating or massively increasing the complexity of their life." (Interviewee 14)*

Negative emotions clearly arose for GPs when thinking about this aspect of their practice, including fear of adverse health outcomes and a degree of shame about their perceived knowledge levels.

> *"What's going through my head is: you take off beta blockers … 'Oh my God, what if I induce another heart attack?'" (Interviewee 2)*

> *"It makes GPs sound dumb, doesn't it? … Listening to myself talking about this… it doesn't sound like I've had a university education…" (Interviewee 3)*

> *"I think there's a big thing amongst GPs about how they're 'just GPs' and there's a kind of collective hidden shame in not knowing about this stuff…and just don't assume anything about the level of competence of GPs because we've forgotten*

everything. I'm terrified about how little I know having [been interviewed] today" (Interviewee 13)

3.9.4.1 GPs' views on possibly increasing the use of QIRx in the future

GPs were asked their views on whether a hypothetical new information resource delivering information on QIRx would be helpful, including whether they would like to increase their use of QIRx at all. Many expressed positive views, imagining benefits for their patients and themselves.

"…the information's helpful for some patients because they want it, but it would also be helpful for me, so I feel a bit more like I'm on firmer ground about what I'm actually suggesting." (Interviewee 15)

"…so the patient perspective might be, 'Well, that doesn't change my risk much,' whereas going from thirty percent likely to have a stroke to twenty percent…they might see that as more of a significant finding" (Interviewee 12)

However, concerns were raised about the suitability of sharing QIRx with patients, and whether it would actually affect their choices. Other concerns included introducing QIRx into the workflow, the potential for distraction or its place amongst conflicting priorities.

"I know some of my patients… '1 in 270' might not be an easy concept for them to digest…you could communicate the risk but it's the individual who decides…eventually…they might say, 'Even if it is that, it doesn't matter, I would rather have that' " (Interviewee 4)

"I don't want to search for something during the consultation because it just breaks the flow…it's just not practical to do it unless you really know your stuff" (Interviewee 3)

"The more I do this, I actually think we respond to expectation: number one of patients; number two of hospital consultants, number three of what we think we can manage, number four…government and CCG controls. And the rationality in medicine is probably number five." (Interviewee 5)

3.10 Discussion

3.10.1 Summary

The GPs interviewed only described using QIRx for a few treatments: considering the treatment benefits of statins and for some specific treatment risks. They were aware of their knowledge and confidence deficit, with mixed attitudes regarding this. Some perceived an important gap in their ability to provide optimal care. Others were content to follow guidelines. Often, an individual GP would have both these perspectives. Instead they used a variety of strategies to make treatment decisions, drawing on their clinical knowledge and understanding of individual patients.

Regarding the idea of increasing their use of QIRx, most were positive, imagining benefits for patients and themselves. However, barriers to such a change in practice were described. These included pressure to conform to clinical guidelines and performance measures, perceptions of normative practice and medico-legal fear. They need accessible, understandable information on QIRx that can be integrated within their complex, time-poor practice.

3.10.2 Strengths and Limitations

To my knowledge, this is the first qualitative study to explore GPs' understanding and use of QIRx.

The sample of participants had a broad range of characteristics reflective of the wider GP population. Some bias may have occurred due to participant self-selection, attracting participants with an above average level of interest in the topic. The sample did not appear

to be unusually confident in their use of QIRx, but all GPs might not share the same degree of positivity imagining an increase its use in practice.

That I as the interviewer was also a GP supported spontaneous expression and understanding. A limitation might be that given my interest in the subject, participants may have shaped their answers to what they imagined were "correct". I was mindful that their prior assumptions might affect the analysis. These potential sources of bias were mitigated by an interview guide employing positive and negative framing of questions, dual-coding of early transcripts with non-clinicians, member-checking, discussions with academic supervisors and ongoing reflection. The framework method involved repeated cycles of analysis and created an audit trail linking data to conclusions.

3.10.3 Comparison with existing literature

Existing explanations for why knowledge deficits on QIRx exist include biased or over-simplified information from researchers, industry or guideline producers (168, 169). Many also arose in this study, including: lack of access to information, the dominance of system drivers such as performance measures, and medico-legal concern.

The qualitative literature on GPs' management of multimorbidity and polypharmacy described in the previous chapter is also reflected here. Difficulties applying single-condition guidelines to individual patients and sharing decisions about treatment in the absence of applicable evidence are common themes, as is medico-legal fear. One strategy described to deal with this is "satisficing"(148) - combining hunches, best guesses and negotiating compromise in an attempt to offer optimal personalised care (210). These findings are echoed in this interview study.

Similarly, there is much in common with the literature on GPs' relationship with and reservations about guidelines, including: doubt about their applicability to individuals, tension between doctor experience, patient preferences and guideline recommendations as well as time and communication constraints (157, 211)(154, 208).

A notable omission in the literature on GPs' practice is a specific exploration of the role of the understanding of QIRx. This interview study addresses this gap.

3.10.4 Implications for *GP Evidence*

The feasibility of the idea that GPs might wish to acquire knowledge about QIRx and integrate it into their practice is broadly supported by these interview findings. I drew encouragement that GPs might be enthusiastic about this, from both their statements and examples of current practice.

Use of QIRx was described with regard to statin prescribing and for some specific treatment risks. What unites these examples is that GPs had access to usable information, and that the option for patients to take these treatments (statins, bisphosphonates and HRT) or not has been the subject of mainstream debate. Given good information and a sense of "permission" to offer choice, they found a way to integrate this information into their consultations.

Their practice of and enthusiasm for shaping decisions based on even imprecise ideas of benefit and harm, drawing on their tacit or explicit clinical knowledge and understanding of individual patients suggests a genuine interest in personalised care - there may be an appetite to enhance this with a sharper understanding of QIRx. The GPs' frustrations with the status quo regarding both sub-standard patient care, and negative feelings about their knowledge and confidence, may act as a driver to improve and adapt their practice with the right support.

However, they imagined a number of barriers to the use of QIRx: time constraints, interruption to consultation flow, and variable applicability of QIRx to patients, especially where there is clinical complexity. Communication challenges were anticipated, there may be a need to further develop risk-communication and consultation skills to underpin the process of shared decision-making(212, 213) (209, 210). The quantitative information itself is only one element of many informing the choices patients may make in partnership with their doctors. Establishing patients' values and preferences is critical to shared decision making (214) and individuals may have priorities which override rational risk-based decision making. Though wishing to exercise choice and have their opinions valued, many patients still wish their doctor to make final treatment decisions (66, 197). All these elements will need to be considered when using this new information.

Consideration of how GPs acquire and use knowledge is important when thinking about how this new knowledge might be made available and presented. The idea that GPs derive information and change practice directly and systematically from clinical guidelines was challenged by Gabbay and Le May, who conducted ethnographic research on British GPs in practice in the early 2000s (215). They found that rather than seeking out, learning and adhering to guidelines directly, the GPs based their decisions on what the researchers termed "mindlines":

> *'collectively reinforced, internalised tacit guidelines…informed by brief reading…their interactions with each other and with opinion leaders, patients, and pharmaceutical representatives and by other sources of largely tacit knowledge that built on their early training and their own and their colleagues' experience'*

The GPs in this interview study clearly demonstrated their “mindlines”, in descriptions of various pieces of knowledge and their origins. Information on QIRx is likely to be best delivered in a way which recognises this way of operating, allowing “knowledge fragments” to be sought and used at the point of need, then integrated into their mindlines. It is known that GPs use many online information resources in their practice (216). Foraging for information online is well suited to the process of updating and maintaining mindlines.

The GPs reported low levels of confidence with statistical terminology. This aligns with the literature showing poor statistical and risk literacy among doctors in general (217-219). Clearly this is a major obstacle to the understanding of QIRx – the design of the website will need to address this as a fundamental challenge.

Absorbing new, potentially challenging information is impeded by time constraints, a barrier identified in this study and repeatedly in the literature on clinical guidelines described in section 2.1. This could be solved by the presentation of simplified, summary information, but the inclusion of nuance and detail about clinical evidence is needed to support clinical judgment when applying evidence to individuals. This tension between brevity and detail will need to be resolved by the design of the website.

The challenges created by the context of current clinical practice, normative standards, regulation, performance measures and medico-legal anxiety came through strongly in these interviews. How might the GPs be able to use new information which might lead them to deviate from what they perceive as “correct” practice, such as following a guideline to the

letter? These themes arise frequently in the literature on multimorbidity, polypharmacy and EBM described in chapters 1 and 2. Presenting information on QIRx in a way which situates it *alongside* ideas of current good practice as a natural partner, rather than as something in opposition to it, may be an approach which helps with this challenge.

3.11 Conclusion

A lack of knowledge of and confidence in using QIRx in practice is recognised by GPs. Improving this is likely to appeal to many, with potential benefits for patients and GPs. However, a number of challenges and barriers exist at informational, practice and system levels which need to be considered when designing solutions for information delivery or implementation into practice.

In chapter 6 I will describe how these findings were integrated in to the co-design process.

4 Part 2. Chapter 4: Methods part 1. Co-design and Design Research – principles and methods

4.1 Choice of co-design as a method

The background literature and findings from the GP interview study highlight a number of challenges to be addressed when seeking to provide information on the benefits and harms of treatments for long-term conditions to GPs, in a way which might improve their understanding and support clinical practice.

Tensions exists between the amount and complexity of scientific information available and the time, skills and confidence a GP may have to take it in. Population-level research findings drive the evidence base, yet must be applied to individuals. Any sense of enthusiasm and new clinical freedom acquired by an improvement in understanding of the evidence will be tempered by a regulatory working culture and expectations of normative practice – cognitive dissonance is a real risk. Once knowledge or understanding has been acquired by the doctor, it may then need communicating to their patient: another challenge.

My understanding of these challenges had developed over the years leading up to this project, and the idea that good design would be essential to a new website became obvious in the broadest sense. I was aware from a general knowledge perspective that websites (and other artefacts) are designed with user involvement and that there was a discipline of user-testing in the technology industry, but had not explored it systematically. My original thinking about the website had revolved around critical appraisal of evidence and how this could be done, rather than as a design question, reflecting an assumption I held that my fellow GPs had the same level of interest and skills in EBM as I did.

During my MSc, I completed a module entitled “Knowledge into Action”, focussed mainly on implementation science. This in itself had been a revelation: as a clinician who had been on the receiving end of any number of badly implemented initiatives, it was fascinating to hear how these could have been done better and that there was a discipline which was learning how to do this. The key message was about complexity and the need for all people involved and mechanisms operating in a system to be carefully understood for a project to succeed. One session was devoted to co-design, where I began to understand how useful this approach might be for my project. A parallel realisation was about just how *badly designed* many of the information systems for clinicians are: who is expecting us to read chunks of prose during our working day, track down the right original paper or, conversely, operate in a patient-centred way whilst using an oversimplified flow-chart?

Co-design was established as a central method during the planning of the doctoral research. At the outset, I had only broad ideas that this would involve interviews, workshops, lots of coloured sticky notes and blue-sky thinking, but had absorbed the central tenet of placing your user at the centre of the design process and listening very carefully.

4.2 Establishing methods

I sought out literature on co-design and design in general to seek the most appropriate methods and formulate a plan for my research. This proved quite challenging, compared to finding the best methods to undertake a systematic review for example. Design methods and research have their origins outside academia and healthcare. The literature describing innovation and best practice in information technology design is formally located in papers of conference proceedings such as the international Conference of Human-Computer

Interaction (220) and was not easily understandable with my lay knowledge. However, the diffusion of this knowledge occurs through informal writings by industry professionals: blogs, books and short courses abound describing methods and techniques.

The medical literature contains many examples of the use of various design methods applied in healthcare, and provided many field- or project-specific examples from which to draw – yet still there did not appear to be an "approved" set of methods ready for me to use. The Equator network (221) does not list reporting guidelines for co-design research for example.

My exploration of this diverse literature was informal rather than systematic: reading papers, online material, following citations and links, attending courses and conferences and picking up advice from my supervisors and contacts. Out of this process, which continued throughout the project, I accumulated two types of knowledge. The first was a collection of examples, principles and understandings about design and co-design which shaped and influenced my approach in a general sense as well as providing specific, small lessons which could be directly applied to a research choice. The second was a set of more clearly defined methods which I chose, adapted and applied in a structured way.

In the rest of this chapter, I will describe these two sets of knowledge and methods. In Chapter 6 I will show how they were used and the subsequent results.

4.3 Part 1: Principles and ideas

In this section I will describe a number of principles and ideas which informed the project: sometimes in specific ways, sometimes operating by creating a shift in my understanding, attitude or perspective which is trickier to define. I will link some of them here to specific examples in the design of *GP Evidence*, and in later chapters will refer back to them when

appropriate. They are ordered in a way intended to make for easier reading, rather than the order in which they were used.

4.3.1 What is co-design?

Reading about co-design led to discussion of a range of related ideas and methods: participatory co-design, co-creation, co-production, action research, human-centred design and more. No single agreed definition of co-design seems to exist, reflecting its evolution across disciplines over five decades or so. A commonly cited origin of co-design is in Scandinavian industry in the 1970s, where trade unions had negotiated worker participation in the design of inclusion of computer systems in the workplace (222).

What these terms have in common is a recognition of the importance of inclusion of workers, end-users, stakeholders, system players and in healthcare, patients in the design of objects or systems. This is in response to a recognition that top-down or expert-led design has a powerful tendency to create inadequate or harmful results due to the limited perspectives of these designers. A 1971 paper on design participation (223) put this bluntly:

"For the layman, who is on the receiving end of the planning and design processes, much of what the various professionals hand down to him must seem a very mixed blessing. Every development seems to hold as many threats of harmful side-effects as it holds promises for the enhancement of society. Too frequently, the most that the threatened layman can do is to protest when it is already too late. Not only is he not consulted even about proposed developments in his own neighbourhood, but planning and decision- making at all levels are often deliberately kept secret.

Yet the professional designers in every field have failed in their assumed responsibility to

predict and to design-out the adverse side effects of their projects. These harmful side effects can no longer be tolerated…"

Proponents of one definition or another usually argue that their preferred method places more or less emphasis on user-involvement, such as at what stage in a process they are involved or how much power they have (224-232). In table 4.1, I present three examples of definitions and descriptions of co-design from the healthcare literature (with further examples in appendix 4.05).

Table 4.1 Selected definitions, quotes and comments on co-design and related methods.

Definition / quote	**Source, context and comments**
"[Co-design is] **active collaboration between stakeholders** in the design of solutions to a pre-specified problem."	Vargas et al. 2020. (225) Australian public health academics writing on definitions and perspectives on design approaches. They contrast co-design with related methods, providing these definitions: **Co-creation:** the collaborative approach of creative problem solving between diverse stakeholders at all stages of an initiative, from the problem identification and solution generation through to implementation and evaluation **Co-production** refers to implementing previously determined solutions to a previously agreed problem with emphasis on the most efficient use of existing resources and assets
"…application of co-design principles aims to ensure that technologies *and* the services in which they are embedded **co-evolve** in a way that is grounded in **the lived experience of users, who are fully engaged** in the design process."	Wherton et al. 2015. (227) British health services researchers reporting on outcomes of co-design workshops for assisted living technologies.
"…**co-design has different meanings**; for some, it means testing technologies in development and making changes based on user feedback; for others, it denotes in-depth, deliberative engagement with users' routines and priorities, alongside mutual adaptation of technologies within sociotechnical practices."	Papoutsi et al. 2021. (228) British health services researchers reporting case studies of co-design in digital health. They argue that: "Co-design becomes **more productive when viewed as an iterative process of development, value creation, and knowledge generation.**"

These various definitions and explanations all place a high value on the role of users, and suggest a greater value or increasing likelihood of good quality design with earlier, integrated, iterative and ongoing involvement of users in a design process.

They have much more in common than the variety of terms would suggest. There is a sense that what matters is the quality with which co-design processes are conducted, what methods and level of user involvement suits a particular project, and what attitudes project leaders bring in terms of genuine interest and belief in what users have to offer.

4.3.2 Design Principles

The Design of Everyday Things is a now classic book (233) written by Don Norman, an academic psychologist who became an early pioneer and global leader of user-centred design. In it, he describes the interaction between objects or systems (which must be designed) and the psychology of the people using them. Poor design ignores the psychological mechanisms which are employed as we interact with things, resulting in problematic design which is at best irritating (beautiful bathroom taps which don't tell you if they're hot or cold) and at worst, disastrous (unintuitive control panels at the Three Mile Island nuclear reactor). Though much of his work relates to machines and systems with which people much interact and control (rather than a simple information-giving website), some core ideas were relevant to my understanding and informed the design of *GP Evidence* – see figure 4.1.

Figure 4.1 Discoverability and design for human cognition and emotion. Adapted from The Design of Everyday Things (234).

- Discoverability: the degree to which it is obvious what features a design has, or how to work it. This has four elements.
    - Affordances: the relationship between the properties of an object and the capabilities of a user. This combination defines "what can be done" with a design.
    - Signifiers: aspects of a design which indicate what the purpose of a design feature is and/or how to use it.
    - Mapping: good visuospatial arrangement within a design informs users about the correct action to take.
    - Feedback: users need to rapidly know if an action they have chosen is correct, or will lead to where they intend.
- Design for human cognition and emotion operates at three levels:
    - Visceral: instant psycho-physiological reactions. This might be a sense of delight triggered by a pleasing design or successful interaction with an artifact, or a negative emotion such as irritation or frustration with a poorly functioning design.
    - Behavioral: the learned actions and responses users acquire through repeated interaction with a design.
    - Reflective: a longer-term reaction based on conscious consideration, reflection and anticipation about an interaction with a design artifact.

Understanding your user

This is a foundational principle of design. In the first year of the doctoral research, I undertook a short online course entitled "Insights for Innovation", produced by IDEO, a leading global design and consulting firm (234). This introduced the aims and methods of designers as they seek to understand a user-group in order to generate design ideas. They propose four important stages, presented in figure 4.2.

Figure 4.2 Four stages to develop user insights. Adapted from IDEO "Insights for Innovation" course.

- **Observation** of actions, environments, interactions, objects and users. Suggestions for fruitful aspects to observe are: behaviour prompts, adaptations and workarounds, body language and patterns of behaviour and to see what people *care* about.
- **Learning from extremes.** Much may be learned from situations or people at the extreme end of any variable. For example, how does a frail elderly person find their way around a newly designed supermarket? How does a beginner cope with a new technology compared to an expert?
- **Deepen understanding** by purposeful research such as interviewing, employing artefacts or tasks and challenges with users.
- **Develop empathy.** Recognising that the designer is likely not to have shared experience with their user, direct participation in activities by the designer can yield insights: be in a wheelchair for a day, "do it yourself" e.g. try to seek social housing by going through an application process, walk around a toy store on your knees to get a child's eye view.

It is easy to see parallels with qualitative healthcare research: ethnographic research, purposive sampling and interview studies represent similar approaches.

User as designer

Given my role in this project as both a designer *and* target user of *GP Evidence*, how much of this was necessary? Didn't I already understand my users? The received wisdom in the design world appears to be that users (i.e. ordinary people) as designers can indeed generate successful innovations. The concept of a "Lead user" was developed in the 1980s by an economist, Eric von Hippel who wrote from the perspective of business and marketing about how enthusiastic members of the public develop design solutions before companies do. An example of this is children in the 1970s adapting their bicycles to look like motorbikes, which inspired the manufacture of highly popular "chopper" bikes (235). This evolved into a method for business to identify and draw on such innovators (236).

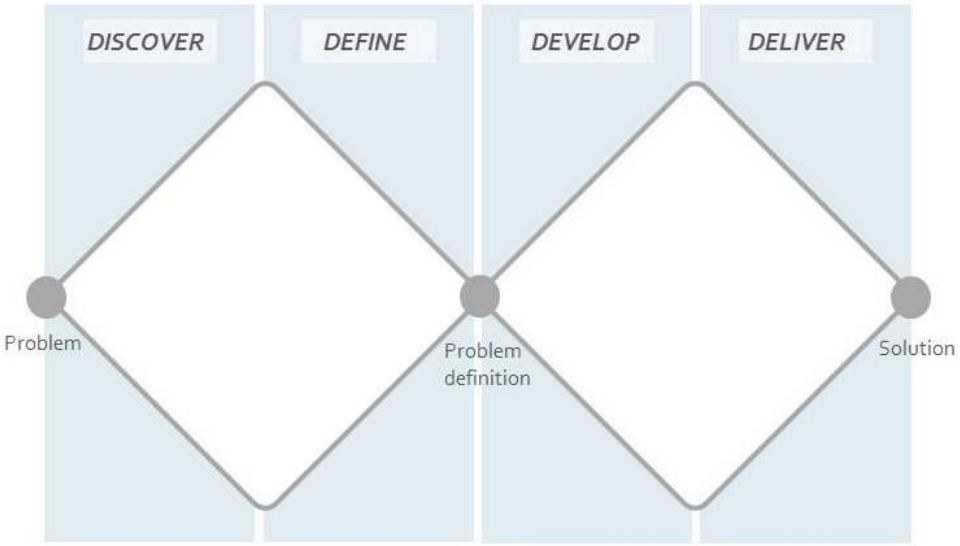
Writing about the designer-user gap in 2008 (237), Jakob Nielsen, co-founder of the Nielsen Norman design research group, proposed three categories of distance between the designer and user. The furthest is where the designer is operating in a completely foreign domain, an intermediate level where the designer has *some* knowledge of a product or field, and the closest being where the designer is the user designing something for themselves. This latter situation he describes as a best case scenario, but cautions that these user-designers will always know far more about their product than anyone an average target user.

It is in this latter category that I found myself – with lots of insight and expertise, but a very different perspective from my peers due to my special interest and further education.

4.3.2.1 Iteration, exploration and refinement

The design process embraces exploration, error and uncertainty as necessary and positive aspects of research and creation. This is reflected in models of the design process: two examples are described in figures 4.3 and 4.4.

Figure 4.3 The Design Council Double Diamond process model



This model was generated by a research process examining excellence in design by leading global companies. It describes four stages in a design process, two each of divergence and convergence (238).

The first diamond is about "designing the right thing", the second is about "designing the thing right".

Discover Where the nature of a problem and user needs are identified and understood.

Define Where objectives are created and agreed regarding the nature and goals of the finished product with respect to the problem it needs to solve.

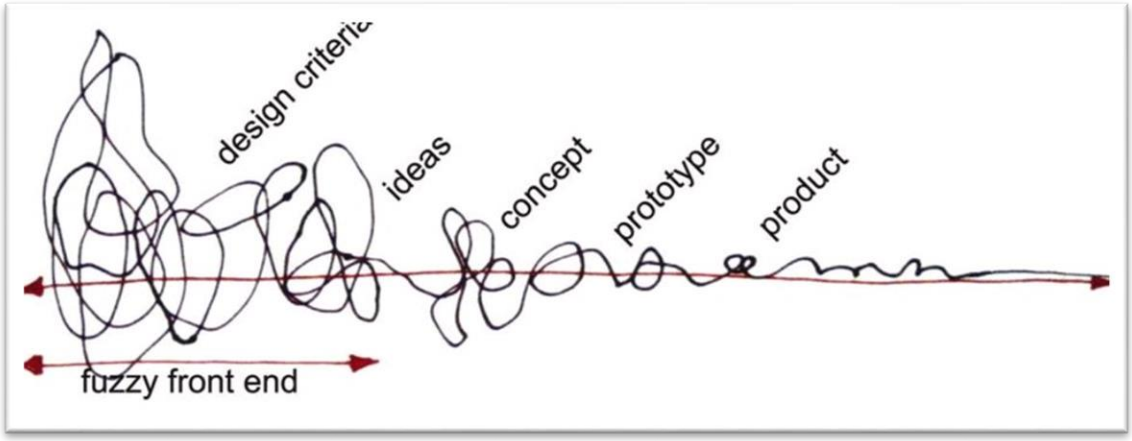
> For this project, this diamond relates to my own identification of a problem, stakeholder interaction, literature review, GP interviews and exploration of other resources in the field.

Develop Early phase of creative design or prototype development which is deliberately exploratory and broad, likely to generate multiple ideas which do not make it through to the final design.

Deliver Refinement of the resulting design idea, user-testing, iterative development, evaluation and launch.

> For this project, this diamond relates to the co-design process and evaluation study – the following chapters in this thesis.

Figure 4.4 The fuzzy front end



The fuzzy front end comes from the Front End Innovation model (239) which originated in business in the 1990s. It emphasises the value of and describes the exploratory process prior to product development – an approximate equivalent to the first diamond in the double diamond model.

The "fuzziness" relates to the importance and value of broad multi-method exploration when defining a problem or need for innovation. This process may feel messy and lead down blind alleys, but is necessary to achieve good ideas before refining a product.

This was a highly valuable model for me throughout this project, giving a sense of permission to try things out and be less concerned when feeling lost.

"Fail often, fail fast" is a phrase which crops up frequently in design writing, though I have been unable to identify its source. It relates to the exploratory nature of the design process, the likelihood that designers will get much wrong, and the value of user-involvement as early as possible. The users will tell you what is wrong and it is better to find this out sooner rather than later, in order not to waste time and resources refining something which has a fundamental flaw. This concept informs design methods which support the development of "lo-fidelity" prototypes requiring little resource, but with the ability to illustrate a concept well enough to be tested with users. For example, wire-frame illustrations of apps, or mock-ups of real physical spaces.

One synthesis of these principles is the concept of "Design Thinking", an idea popularised in a 2008 Harvard Business Review article (240), but originating in 1979, when Bruce Archer, Professor of Design Research at the Royal College of Art wrote,

> *"there exists a designerly way of thinking and communicating that is both different from scientific and scholarly ways of thinking and communication, and as powerful as scientific and scholarly methods of inquiry" (241)*

Design thinking has been described as a method, a way of thinking, a philosophy or set of personality traits. Detailed descriptions place it as a human-centred approach and reveal a structure virtually identical to the double diamond model; it includes processes such as empathy, ideation, definition, prototyping and testing (242, 243).

4.3.3 Design for computing and websites

Human Computer Interaction (HCI) is a field of research and practice involving a range of disciplines including computer science, psychology, ergonomics, software design, graphic arts, human factors, interaction design and user-experience. It evolved organically from the 1980s, merging autonomous disciplines united by an interest in blending the new power of computing with human experience (244).

A major output of HCI was the development of principles to support the usability of computer systems. Two recurrent references appeared in my reading: Schneiderman's "8 Golden Rules of Interface Design" (245) and Nielsen's "Ten Usability Heuristics" (246). Scheiderman was a professor of computer science who published his Golden Rules in 1986. Nielsen's career had been industry-based, going on to become a world leader in web-usability. Nielsen's list, published in 1989, builds on Schneiderman's and formed the basis for a method to evaluate website usability (247). Many of their recommendations relate to human-computer

interaction where the user has a task to complete, or needs to interact with the system. Examples might include a website where a consumer makes a purchase and has to successfully navigate processes of choice and payment, or a control system for a piece of complex machinery. This is in contrast to a website such as *GP Evidence*, which requires limited user-input. However, about half of their principles were highly relevant to the design of *GP Evidence*, in terms of both website design and content. These are summarised in table 4.2, with notes on their relevance to this project.

Table 4.2 Summary adaptation and notes on Schneider's 8 golden rules and Neilsen's 10 usability heuristics for interface design.

Golden Rule / Heuristic	**Relevance to *GP Evidence***
Match between the system and real world The design should "speak the users' language", using words phrases and concepts with which they feel comfortable.	Central, related to: 1) Information architecture. That the structure of subjects mapped to existing mental models and drivers of normative practice like NICE guidelines. Presentation of information with "take home message" first to map with GPs' thought processes. 2) Language and written content would be designed to match everyday language of GP users rather than that of evidence producers.
Aesthetic and minimalist design Avoid irrelevant information or content. This *competes with* the relevant information.	Central. Lots of open space in the design and simple colour schemes would minimise distraction and cognitive labour. Deliberate non-inclusion of unnecessary background information or detail.
Aim for consistency of appearance and process Menus, information hierarchies, actions across the interface should be consistent. Users then develop a sense of ease and confidence.	Important. Page and links to secondary layers of information would be consistent from page to page. Similar types of information content would appear in consistent places in a page.

Recognition over recall Use design and structure to reduce need for effort of memory, aim for intuitive understanding of content.	Important. Design of "Condition" pages would offer a menu which also serve to remind GPs of the options on offer. In selected places, summaries of background information which may be harder to remember would be included (such as levels of left-ventricular function relating to heart failure severity).
Help and instructions Ideally, design so no instructions are needed. However, this may be necessary.	Important. The design would aim to be "visit and use", with the intention that users could dip-in and find pieces of information rapidly with little explanation or exploration. However, some pages in the website were written giving some supporting information on "Key Concepts".
Visibility of system status A user should be able to tell easily where they are in a process by appropriate methods of **feedback**.	These last four elements are not so relevant to *GP Evidence*, mainly referring to interfaces for systems where there is active user-interaction. However, the website was designed so that users could easily keep track of where they were in the site and navigate easily. Examples of this would be the design of pop-up information boxes, and links opening in new tabs.
Give users a sense of control and freedom This includes an obvious ability to correct mistakes or backtrack in a process: **"error recovery"**.	
Prevent errors When users need to control an interface, make it harder for them to do the wrong thing.	
Flexibility and efficiency of user for experienced users For frequent users, build in options for customisable options and shortcuts which will increase efficiency (even if not immediately obvious to a novice user)	

4.3.3.1 Accessible design

At a conference on user-experience (UX) design which I attended in the first year of the project, I attended a talk by Sarah Winters[4], former head of content for the Government Digital Service. She had been responsible for creation and implementation of content strategy for the GOV.UK website, which would become a world-leading exemplar of accessible content design, their methods later adopted by NHS Digital.

The principles underpinning this work, a new method and discipline called "Content Design" became central to this project. I will describe these in detail in the next part of this chapter, but wish to reflect here on an important insight for me which powerfully shaped *GP Evidence*.

The talk was about "accessible design", the principles and methods for opening up web-content to those who have physical or mental disabilities:

> *"Imagine you have a visual impairment which means that your visual field is only large enough to read two words at a time. How would you want text to be written? Every word should count. Every unnecessary word would make life more difficult. What about someone with arthritis in their hands for whom every mouse-click causes pain? They would want you to have thought very hard about where and when information is placed, to be able to get efficiently through your content"*[5]

Her discussion moved onto the idea that if content is designed to be accessible for those with these kinds of limitations, then it produces *great* design for everyone. We all benefit from clarity, clutter removal and effort-avoidance.

While listening to this, it struck me that GPs (and all healthcare professionals) are effectively *disabled* in their working environments. By fatigue, digital and human distraction, multiple

[4] Sarah Winters changed her name over the course of this project, her surname had previously been Richards. I will use Winters throughout this text, though some references retain her previous surname.
[5] Paraphrase from personal notes. Sarah Richards, Accessible Design. Camp Digital conference. Manchester 12/09/19

agendas competing for attention, emotional transference and time pressure. All this serves to reduce the time and cognitive space available to take in and use new information in practice. Understanding this and designing accordingly would be essential to the success of the project.

Accessible design includes a wide range of technical principles and detail, ranging from consideration of font size and colour contrast, to tools such as screen readers and screen magnifiers. Worldwide technical recommendations are defined in the Web Content Accessibility Guidelines (WCAG 2.1)(248) . For *GP Evidence*, technical aspects of this were taken care of by my web-developers who had a special interest in accessible design. Aspects of this will be detailed in chapter 6.

4.3.3.2 Minimal clicks v foraging

The principles of accessible design suggest that efforts should be made to minimise the amount of work which users need to do to find content. Indeed, a repeated refrain from GPs in interviews and workshops early in this project was that information should be available "in a couple of clicks" or similar. This is valid, but not the only priority regarding how users seek information online.

"Information foraging theory" was described in 1999 by Pirolli and Card (249). As it sounds, it was inspired by established behaviour models of animals foraging for food. The key insight is that people have an information goal and will make conscious or unconscious calculations to judge whether a source will provide (a) the right information and (b) how much effort will need to be expended to extract it.

When choosing an information "patch" (the analogy being that a website corresponds to a patch of land in an animal's habitat), users will pick up a "scent of information" which will encourage them to keep exploring that patch. If the scent is attractive enough, they will continue foraging in that patch, and be willing to spend a certain amount of energy (clicks, scrolling, reading/scanning) in the process (250).

The implication for website design is that the "scent of information" should be strong from the beginning. This can be achieved by a well-designed home page which clearly communicates the type of information available, and that as users navigate through the site seeking particular details, the "scent" remains strong. This can be achieved with logical information structure, clear headings, images or other devices to draw them to an information destination.

This idea influenced various aspects of design in *GP Evidence*: the home page, the prominent use of infographics and heading titles and structures. These will be detailed in chapter 6.

4.4 Part 2 – Design Methods

In this section, I will describe specific design methods and tools which were used, including project-specific adaptations. The results and findings arising from these methods are described in Chapter 6. They are listed here in approximately the order in which they were used, detailed project flow is described in Chapter 6.

4.4.1 Partnership with professional web-design company

To build *GP Evidence*, I commissioned professional web-developers to work with me on the project. Nexer Digital[6] is a research, design and development company who employ a person-centred approach and have a special interest in accessible design. They have worked extensively with charities, governmental, health and social care organisations and produced appealing and intuitive websites for a diverse range of users.

Their role in this project was principally to deliver the technical design and production of the website, though the small team with whom I worked also reviewed and gave advice on my research plans and co-led a design workshop which I will describe in detail in this chapter. In this methods chapter and chapter 6 where I describe the results of the co-design process, I will make clear in detail what work was mine and what was done by Nexer. For the thesis I will use the phrase "my web-developers" to refer to this organisation.

4.4.2 Think-aloud method

This method is a staple of design research, particularly with respect to computer and web interfaces. It seeks to capture information about what users' experiences, reactions and behaviours are when interacting with a product. Findings may be about function (e.g. whether

[6] www.nexerdigital.com

users can find a link or button on a web-page), or emotional responses (confusion, frustration or delight). It may be used to test a new product at any stage of development, or to review pre-existing products as part of exploratory research (251-253).

During a testing session or interview, participants are given a task to perform, which might be quite general, such as “Have a look at and explore this website and tell me what you think”, or more usually, focussed to test a specific function of an interface. They are asked to express their thoughts and feelings as they complete the task. Researchers are generally advised to intervene as little as possible in the process, so as not to disturb or influence user behaviour, though may pause the process to gather particular insights, or save questions till the end of a session.

Advantages of think-aloud are that

- it allows simultaneous capture of functional and emotional information
- minor verbal and non-verbal information can be observed which may give insight into the nature of problems which would be missed by less direct methods
- underlying issues such as user misconceptions and confusion are revealed
- it is relatively simple, cheap and adaptable to many phases of the design process
- easy to learn and produces usable findings even in the hands of the relatively inexperienced

Disadvantages are that

- some users may find it difficult or intrusive to perform
- researchers may influence behaviour with ill-judged intervention

- the controlled nature of the process may increase user concentration and lead to different behaviour than would be seen in real life

4.4.2.1 Think-aloud method in GP Interviews

The think-aloud method was used in two places in this project. Firstly, as a component of the interview study described in chapter 3: I will describe this immediately below. Secondly, as a component of user-testing the *GP Evidence* website: I will discuss this later in this chapter.

The second aim and research question of the interview study is repeated in figure 4.5.

Figure 4.5 Second aim and research question of GP interview study.

Aim (2)

To understand GPs' preferences with regard to viewing online information, in order to inform the design of the web-site: how to deliver complex information in a way that is understandable, useful and usable in practice.

Research Question (2)

What are GPs' design preferences for access to evidence-based information on the quantitative benefits and harms of long term treatments?

To address these, I explored GPs' reactions to established resources and websites which aim to deliver quantitative information on the benefits and harms of treatments to clinicians or patients. I selected websites which adopted a range of designs and approaches. This would enable me to see the effect of a variety of styles of information presentation. These websites and their attributes are listed in figure 4.6. Images of relevant sections are shown in appendices 6.1-6.5.

Figure 4.6 Websites used for think-aloud testing in GP interview study

NHS Scotland Polypharmacy website > For Healthcare Professionals > Efficacy (NNT)

- Modern design using logos and graphics
- Clinician-facing
- Menu structure by condition or drug
- Information summaries by short sentences, notes and tables
- Data presentation by clinical study findings

NHS Scotland Polypharmacy website > For Patients and Carers > Decision-making tools

- Modern design using logos and graphics
- Patient-facing
- Menu structure by condition or drug
- Presents "How likely is [drug] to help me?" data using infographics and plain language to convey chance of benefits or harms
- Data presentation by clinical study findings

NICE Statin Decision Aid (2013)

- PDF available online accompanying NICE guideline
- Patient-facing
- Produced according to best practice in decision-aid development
- Employs graphics, plain language and variable framing of data
- Data presentation by baseline risk categories and application of relative risk reduction

BMJ Rapid Recommendations

- Very modern interactive design
- Clinician-facing
- Structured by single clinical question
- Employs graphics, text and icons
- Data presentation by baseline risk categories and application of relative risk reduction
- Includes information on strength of evidence and components of evidence quality

thennt.com

- Simple, slightly dated website design
- Clinician-facing
- Structure by clinical question, long pick-lists subdivided by clinical category
- Data presentation by NNT or ARR derived from single studies, supplementary text sections with detail

In Chapter 3 I described the methods for this interview study and presented an interview guide and results for the first aim and research question. The think-aloud part of the interview had its own interview guide which was commenced when a natural resolution to the first part occurred. This is shown in figure 4.7. GPs were invited to share any information resources they currently used and these were explored before looking at the resources in figure 4.6. After the think-aloud section, some more focussed, theoretical questions about information preferences were asked. This interview structure provided three different ways to explore the research question, aiming to maximise insight.

Figure 4.7 Interview guide part 2.

Part 2. Preferences

A) GPs' own preferred resources

I'd like to move on if that's OK and ask you some more detail about any information resources you currently use for this kind of information.

Can you tell me about any favourites and what you like about them?

Can you show me?

Invite participants to share on-line or paper based information resources and talk me through how they use them and any likes and dislikes.

B) Think-aloud method

Now I would like to show you some websites which try to help us in this area.

What I would like us to do is something called a "think-aloud" exercise.

This means that I will ask you to look on the website for a particular piece of information, and will watch while you do it to see how you get on.

While you are looking, I would like you to tell me what's going on in your head: let me know especially about difficulties or things that you particularly like.

Does something appeal to you? Does the information make sense? If not, why not?

Feel free to say anything you like and remember, this is not a test of your ability to use the website – it's a test of how well the website has been designed.

- GPs were then shown the five websites in turn and given specific tasks or pieces of information to gather, for example:
    - "Imagine you are interested in finding out the benefits and harms of alendronate, can you have a look here?"
    - "Here is an evidence summary of the benefits of corticosteroids for sore throat. Can you explore and tell me what you think?"

C) Theoretical questions on information preferences

Let's imagine there was a new resource to give us more information on benefits and harms of treatments.

Would you have a preference for the kind of information you'd like to see? What would your priorities be?

We've been talking a lot about numbers so far. Is there any other information you'd like to know about what's behind the numbers?

This question could be followed up with more specifics:

Do you want to know if advice is based on expert opinion or hard evidence?

How about the evidence? Was it good quality, how many people in studies, what kind of research?

How much do you care about: Confidence Intervals, population size in the clinical trials?

Is there any other additional information you would like to see included alongside what we've talked about so far?

We've covered a lot of ground, thank you. Is there anything else which has come up for you while we've been talking?

4.4.2.1.1 Data capture and analysis

This part of the interview was recorded, transcribed and coded as described in section 3.8. The final coding framework including codes for this second part can be seen in appendix 3.11. However, this content was not subject to framework analysis. What was required was a simple list of likes and dislikes, problematic or successful design features, and observations of GPs' reactions to bring forward to inform prototype design. The coding framework allowed me to search the transcripts for data on particular issues, design features or a particular website.

The think-aloud method proved so effective at highlighting what worked and what didn't, that my insights were mostly instantaneous and also highly memorable. I decided on a three-stage process to ensure the capture of relevant data which would be efficient yet robust:

Stage 1) A re-examination of the websites on my own, writing down insights which I recalled from the interviews.

Stage 2) Creation of a framework matrix on the NVivo software (where the transcripts were coded and the previous framework analysis had been done) entitled "Interviews x Resources". This summarised all data pertaining to a particular website for each participant. I then re-read this set of interview data closely and noted any additional insights that had not been remembered in step 1.

There was a very low additional yield from this step, reflecting the relative simplicity of the information gathered and my interest and experience as "user as designer" which had enabled me to have good, memorable insights during the interviews.

Stage 3) Creation of another framework matrix on NVivo entitled "Interviews x Specific Preferences" which generated a collection of responses from individual participants to sections A and C of the interview guide (figure 4.7), organised by code and category. This captured various design preferences which had been expressed by participants.

The findings from this part of the interview will be described in Chapter 6.

4.4.3 Prototyping methods

The first phase of design research aimed to produce a low-fidelity prototype of the website. The starting point of prototyping work is from a position of understanding of the needs of

users and the context in which they work. This understanding can then be used to develop the prototype, applying design principles and methods. I employed three classic design methods during this phase of the project. The aim was to develop early ideas and low-fidelity prototypes to inform the subsequent design.

Job stories

These are a simple, clarifying tool which anchor the design process to a context, a user-need and a desired outcome (254). They are written in the first person and have three components with standard stems (in bold below), for example:

When I am ordering a shirt from an online store…

I want to have easy access to a sizing guide…

So that I don't get the wrong size and have to return the garment.

This format is felt by many designers to be superior to earlier tools (255) such as:

- **Personas.** Where an imaginary (but research-based) profile of a user or customer type is created in order to describe or imagine their characteristics and needs in order to stimulate design ideas. This method has been criticised for the tendency to encourage researcher assumption and bias. Also, it takes little account of context or desired outcome.
- **User stories.** Almost identical to job stories, the first stem statement is changed to

  As a customer…

  This lacks the specificity of a task, and may result in inadequate focus on the problem and solution.

Job stories help designers understand their users and the desired outcome of their product. This was less important for me (being a user-as-designer) but helped clarify the objective and was valuable in communicating with my web-developers. The job story and its use will be described in chapter 6.

4.4.3.1 Story boards

Story boards are cartoon-strip style drawings which aim to capture social, environmental and practical aspects of the imagined use of the designed product (251). Like job stories, they help designers understand their users and how a design fits into the real world. Again, this was less useful for my personal understanding, but creating one was helpful for my web-developers. The story board and its use will be described in chapter 6.

4.4.3.2 Low-fidelity prototyping

One definition of prototyping is

> *"…the tangible creation of artefacts at various levels of resolution, for development and testing of ideas within design teams and with clients and users."(251)*

This is a key stage in the conversion of exploratory research into a design, creating a visible manifestation of the insights and ideas generated by research, which can be used to communicate research findings, act as a starting point for team-based discussion, as an object to be tested with users, or as a foundation for iterative development. In the design of a physical object, a prototype may well be an object or partly working machine, but for web-design, early prototypes tend to be hand sketches or software generated images.

I created four different image prototypes at different points in the project to share with the web-developers to inform later-stage prototype building and refinement. These will be shown and discussed in chapter 6.

4.4.4 Joint application design workshop

This was a pivotal event in the project, representing the second apex of the double-diamond (figure 4.3) 18 months into a four-year process, where the exploratory research and early prototyping were transformed into a substantial near-final version of the website.

A Joint Application Design workshop (a "JAD") has a variety of definitions, the term having evolved in the 1970s in industry to define an approach to software or systems development within an organisation. It brought together engineers with managers and executives as a means of achieving innovation and change more successfully, overcoming problems related to silo-based working and thinking (256). For this project, the event had been proposed by my web-developers, who used the following description:

> *"A JAD provides an opportunity for **stakeholders**, **customers**, and **users** — who have different perspectives and types of knowledge — to come together to **understand** business **requirements** and **brainstorm** potential **solutions** to a given challenge."*

A more understandable term would perhaps be a "Participatory Co-Design Workshop". In short, it was a day-long meeting with five GPs recruited as user-co-designers, a UX expert, a technical design expert and myself. We started with a wire-frame prototype of the website amending and refining it over the course of the day.

In the few months leading up to this, I had been synthesising my research findings and prototype ideas, sharing them with my web-developers. I will describe this in Chapter 6.

4.4.4.1 Ethical approval

A submission for ethical approval for the workshop was made to the Central University Ethics Committee (CUREC). After review, the committee decided that formal approval was not needed. Copies of the application form and their email response are provided in appendices 4.1 and 4.2.

4.4.4.2 Recruitment of GP participants

In section 2.4, 3.6 and appendix 3.3 I describe a pool of 213 GPs who had expressed an interest in participating in research for this project. I planned to use this group again to select participants for this workshop. However, for this exercise, it was important to have participants who would have a high enough level of interest to engage constructively (they were being asked to generate valuable ideas), whilst also not being so interested or expert that the solutions they proposed were not appropriate for an average GP. My experience with recruiting from this pool for the interview study had been positive, identifying GPs who were clinically active and interested in improving the quality of their decision-making but did not for the most part have a special interest in, or advanced training in, EBM. However, among the 15 participants in the interview study, I had noted a spectrum of interest and depth of thought. Those less interested in this topic (whilst making ideal interview subjects) may have been less likely to make a valuable contribution to the JAD.

Therefore I decided to adopt a selective approach to recruitment, first inviting those who I had interviewed and who on reflection seemed like candidates with the right level of insight and enthusiasm. All of the interviewees had offered to participate in further research and signed a statement to that effect on their consent forms at the time.

I also wished to achieve as diverse a group as possible within the small sample of five attending the JAD - this influenced my choices about whom to approach first. I selected and invited three participants who had been interviewed, and the final two were recruited via an invitation to the wider recruitment pool. Inclusion and exclusion criteria are shown in figure 4.8.

Invitations were sent via email, with replies via direct email to my university email address. Participants were offered £500 remuneration for the day, plus travel and accommodation expenses. A copy of the invitation email and the full participant information sheet are shown in appendix 4.3. Informed consent was obtained through the participant information sheet, opportunity to ask questions via email and confirmed on the day with a written informed consent form copied in appendix 4.4.

Figure 4.8 Inclusion and exclusion criteria for JAD workshop participants

Inclusion criterion:

- To be a GP currently practising in the NHS.

Exclusion criteria:

- GPs who have not been in NHS Clinical practice within the last year.
- GPs still completing specialist training (ST1-5)
- GPs with further training or extended roles which may confer expertise about the subject matter. For example, an academic GP or a CCG prescribing advisor.

Seven GPs volunteered from the invitation to the wider group. I first invited those whose demographics would contribute most to group diversity, with the final two being available to attend on the day. Participant characteristics are summarised in Table 4.3.

Table 4.3 Characteristics of GP participants for JAD workshop

Gender	Age	GP Role	Location	Undergraduate degree in UK?
F	43	Partner	West England	Yes
M	50	Locum	East of England	No
F	31	Salaried	South England	Yes
F	37	Partner	North Wales	Yes
M	29	Locum	South England	Yes

4.4.4.3 Practicalities

The workshop was held in London, at the headquarters of the Royal College of General Practitioners. The only requirement for the meeting room was that it has a large wall-mounted monitor/TV to display presentations and the website wire-frame. Refreshments and lunch were provided throughout the meeting by in-house caterers.

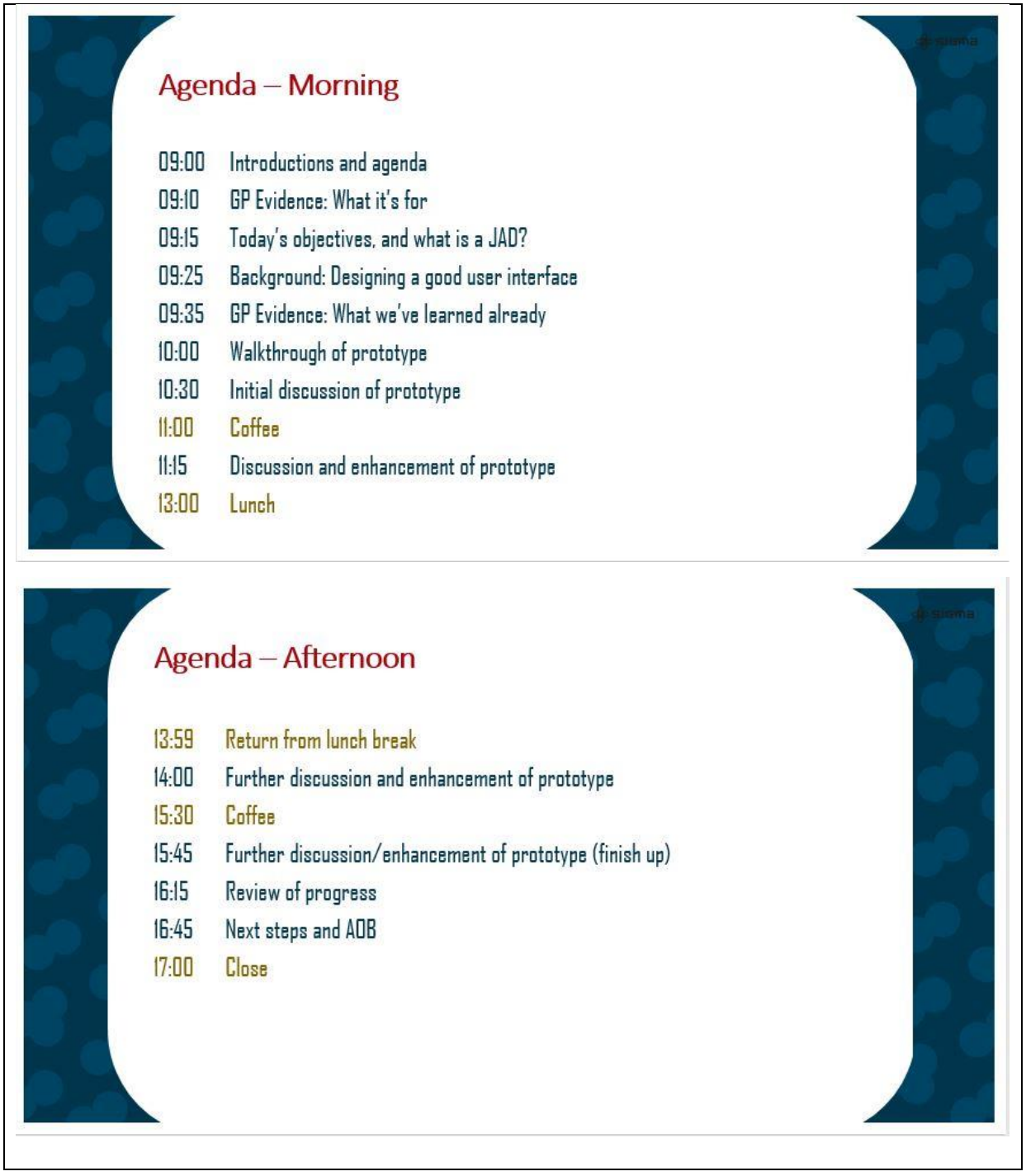
PowerPoint presentations were used to open and direct the meeting. The web-developers brought their own computer with software, on which they had developed an interactive wire-frame of the prototype website. This generated a large projected screen image which looked exactly like a website, including responses to clicks and scrolling but which could be rapidly edited in real-time to experiment and amend design and content. The technical web-designer controlled this throughout the day, responding to suggestions and decisions by rapidly producing demonstration images for immediate feedback. This same person would continue to be the technical lead for the rest of the project.

4.4.4.4 Workshop plan and methods

The workshop was mainly planned by the web-developers, after our series of meetings as described in section 4.4.3.. I had been expecting some quite complex methods, perhaps using post-it notes, wall-charts, card-sorting and diagramming but the web-developers advised that "most of it happens on screen". An agenda for the day is shown in figure 4.9.

The first half of the morning was dedicated to introductory material to orient participants and prepare them for the active work. My part of this was limited to 15 minutes on "GP Evidence: what's it for?", the web-developers being keen to lead the process, to avoid my point of view as a "lead GP user" polluting participants' opinions at the start.

Figure 4.9 Agenda for JAD workshop.



The rest of the day was devoted to a review of the prototype website with active discussion, requests for suggestions, feedback on changes and informal consensus. There was enough

structure to ensure the major aspects of the site design were covered, but not so much that free-flowing discussion was impeded.

The UX expert chaired the day. During the prototype discussion sections, I played a very active role, ensuring that important questions were answered, acting as a "bridge" between the GPs and designers especially when thinking about practice and clinical implications of design choices. Throughout, I was careful to avoid bringing too much of my own opinion to the table, to allow the voices of the participants to be heard and have power and impact over the design choices.

4.4.4.5 Data collection

I had made provision to capture data in a variety of ways: photography, field notes, and opportunistic voice recording using a portable device, paper records on post-it notes, wall charts or cards. This had been included on participant information sheets and consent forms. Ultimately however, all the findings were captured in the finalised website design itself, some field notes and a small number of photos and screen-grabs.

The results of the workshop are presented in chapter 6.

4.4.5 Content design

Content design is a method for producing online content *"…to give the audience what they need, at the time they need it and in a way they expect"* (257). It refers to the process of writing and producing text, but may include all forms of digital content such as images, videos or interactive forms.

It was developed as a method and introduced as a new discipline in the digital world by Sarah Winters in the early 2010s whilst leading content re-design for GOV.UK, the web-portal for

the public to interact with the UK government. This includes such essential services as benefits claims, immigration, passports and vehicle licensing. To create online services which would be usable and accessible to a wide range of people whilst simultaneously delivering enough complexity of information and the ability for users to interact and complete administrative tasks, a design oriented approach was adopted. This method was iterated over a number of years (258) into a seven-component process, summarised in figure 4.10.

Figure 4.10 Components of Content Design. Adapted from Content Design London.

- Background research
- Identify and understand user needs
- Channel and journey mapping
- Language and emotion
- Content creation
- Sharing
- Iteration

The usability and functional success of GOV.UK was a significant breakthrough in user-design and is now a global exemplar. The GOV.UK content design and style guide continue to be updated and are available online (259), and are the basis for the NHS Digital Services Manual content style guide (260).

I recounted in section 4.4.5 on accessible design that I had heard about content design at a conference where Sarah Winters was speaking. It seemed like such an intuitively good fit as a method for GP Evidence that I explored it further. Content Design London (CDL) is a consultancy and website developed by Winters and her team. They have published a book (261) and set of online resources to describe and promote the methodology (257). I looked at these in detail and saw that it could provide a powerful and feasible method to develop

content, in particular with respect to dealing with some of the tensions I had identified from my research: the need for simple, rapidly understandable information on complex topics, and how to structure information in a user-oriented, appealing and efficient way.

I attended a two-day training course with CDL early in the project, shortly before I started writing website content, and subsequently joined their online "Content club" community where I was able to meet with professional content designers, hear talks from industry professionals and later arrange a one-to-one mentoring session for some specific advice.

I adopted these principles and methods when writing and refining content throughout the project. This was in the context of writing textual content into the website architecture which had already been designed at the JAD.

With regards to content design methods, I will describe here how I used them in the GP Evidence project. Figure 4.12, lists the 7 components of the content design process. The first three: research, user-needs and journey-mapping, relate to the design-research process prior to content production. These are essentially the same as apply in design-research in general, and which I have already described with respect to this project so I will not expand on the CDL version. However, it is important to realise that this research-first, user needs-focussed approach of the content design method is critical to its power, and that the subsequent four components (when content is generated) are intimately linked to these preliminary findings.

With regard to content creation, the last four steps in CDL's framework fall naturally into two pairs:

- Language/emotion and content creation relate to a set of principles which guide *how to* write (or design other content) in a way which is likely to be appealing and successful.
- Sharing and Iteration relate to the process of refinement and improvement of draft content by sharing it with others in group settings or a one-to-one process called "pair writing".

These two aspects are not entirely separate, each may inform the other.

4.4.5.1 First draft writing

For the *GP Evidence* project, I was working mainly alone as opposed to in a team like many content developers. Therefore, my process was that I created draft content to as high a standard as I could before sharing it in a "pair writing" process to refine it. I wrote draft content following a collection of principles and recommendations synthesised from the CDL website, book, course and "content club" discussions. I had made notes and selected aspects which were particularly relevant to *GP Evidence* to create a tailored content design guide to use when writing. This summary is reproduced in figure 4.11, including some explanatory comments.

Figure 4.11 Principles for content design for GP Evidence. Selected and adapted from Content Design London.

Link the writing to your users' needs

- Serve your audience, not an imaginary specialist reviewer
- Simple clear writing is not "dumbing down", it is "opening up"
- Remember the job story. Make the content serve this.
- Choose a tone or voice in the writing which supports the needs and goals of your users and is likely to result in a positive emotional response.

Online reading behaviour and writing with this in mind

- The “80/20 rule”. 80% of readers will only read 20% of content. Prioritise the most important messages.
- On average, people only read 20-28% of a web-page.
- People read web-pages by scanning. Either in an “F-shape” across the page or from top-left to bottom right with a tendency for content in the bottom left of pages to be ignored. Page layout and menu structure can affect this. “Banner–blindness” is common.
- People will be “foraging” for information. Use structure and headlines to support this.

- Blank space has a positive emotional effect and makes reading easy.

- Use familiar, common words. Sentences as short as possible. About half of the word count of conventional writing.
- Write in plain English, minimise the use of jargon except where appropriate for your user.
- Write with an “inverted pyramid” structure: conclusion and most important message first, gradually drilling down to detail (which will not be read by most of your readers).
- Break up text and use sub-headings to support scanning, information foraging and comprehension. If sub-headings tell a story this is helpful.
- “Front-load” headings and text to highlight the main subject. Make headings as short as possible. Example:
    - “How to help your school to reduce bullying”, becomes
    - “Bullying: spot it and stop it”
- The first paragraph is important. It is likely to be scan-read during information foraging and will determine whether readers go further.
- Avoid “hyping” language which is exaggerated, boastful or opinionated. Neutral, objective language increases trust.

- Bullet-point lists are easy to read and effective, they
    - support scanning
    - break up information into digestible chunks
- **Bold type** is good for
    - supporting **scanning and foraging**
    - **highlighting** important pieces of content
- Variable **font size** can be a good way to separate essential content from optional detail.

- Capital letters make it harder to read and should be kept to a minimum.
    - this also applies to bullet points which look nicer without them
    - However, they may sometimes be appropriate if there is a particularly structured sentence being used.
    - using upper case to highlight important words INCREASES READING STRAIN, as does capitalising within sentences
- **use digits** rather than words for numbers in text
- **everyday punctuation rules** such as full-stops at the end of every sentence do not necessarily need to be followed

One question which arose when choosing to adopt these methods was whether they were evidence-based. Has it been proven for example, that using bullet points well or that use of blank space improves readability and understanding? Some of these elements do have some empirical research to support them. The Nielsen Norman Group have conducted quantitative research showing usability improvements associated with concise text, scannable layout and objective language in isolation or in combination (262). Readability recommendations from CDL are heavily referenced with sources from industry, government and academia and were developed using a Delphi-style consensus exercise involving international experts in writing for the web (263) . However, some of the content in figure 4.11 seems to have arisen as a set of learnings from the GOV.UK project, informed by large amounts of user-testing, though not published as formal research (261).

I considered whether I should be seeking and critically appraising more empirical evidence of the validity of these recommendations, but decided after exploration that this was firstly not feasible and secondly, that CDL as a “best-in-field” organisation was likely to be producing the best available advice.

4.4.5.2 Pair-writing

Like the editorial process in conventional writing, content design places a high value on the sharing and iteration of content, with an expectation and desire that content will improve through co-writing or review by another person. In an industry context, this process may happen with a variety of people, for example users, technical experts, peers or stakeholders depending on the objective. The content designer will produce draft content and then meet in-person or online with a partner who will then read, constructively criticise and suggest changes. The two people will then edit the content together to create an improved version. An alternative model is where the two people write content together from scratch (264).

The key difference between this and a standard editorial process is the live, in-person nature of the review and re-write as opposed to a mechanism of written feedback. This is felt to generate content more likely to address user needs (265, 266) in addition to supporting cohesion of purpose in an organisation (261). Underlying this is a design-thinking understanding that any writer or designer is unlikely to have the perspective of a user and that pair-writing functions in a similar way to a user-test. A second motivation for pair-writing is to check the factual accuracy of topic-specific content with an expert or stakeholder (264, 266).

Methodologically, pair-writing is very simple with no fixed rules or process. “How-to” advice varies, but common themes are:

- ensure both writers have a clear understanding of user-needs and project objectives
- create protected time, 1-2 hours maximum duration to avoid fatigue
- agree a working process, usually with one person as lead

- ensure clear visual access to text for both parties and a mechanism for rapid text updates (261, 264-267)

4.4.5.3 Pair-writing for GP Evidence

A key question for me was: who should act as a partner for pair-writing? My task was to produce written clinical content for the website, functioning as both a topic expert who had understood the background information, whilst also being the content designer making this accessible and usable. In addition I had a perspective as user-as-designer. What should be my priority be when considering what type of person with whom to write?

I took this question to the "content club" group and was able to get advice from professional content designers. Their answer was that there is no correct or incorrect pairing of writers, rather that this choice should be dictated by the needs of the project.

After consideration, I decided that there were two priorities. Firstly, to have as much GP user input as possible as there was a high risk of not getting the presentation of complex scientific information right, despite good intentions and using content design principles in the writing process. Secondly, I felt that some input from an experienced content designer would be likely to improve the quality of my writing. Therefore, I set up a pair-writing process with both these groups.

4.4.5.4 Pair-writing with GP users - pilot

I tested pair-writing with three GPs as a pilot exercise early in the project as an add-on to user-testing sessions to be described in the next section. The GP and I were speaking over an online video call, screen sharing to look at the website.

Having familiarised themselves with the website through the user-testing process, I then asked that they have a look at a short item of text. We then had an open discussion about the section of text, before switching to a shared Google Document to co-edit the writing. The script used to introduce the exercise is shown in figure 4.12.

Figure 4.12 Introductory script to GP pair-writing exercise

Now that you have had a look around the website, I would like to try getting another form of feedback from you – about the text on the website. We want the site to be easy to read, clear and understandable and have tried to write it that way, but it is almost certain that it could be done better. What I'd like to do is for you to read this piece of writing, take plenty of time and then we will talk about whether it makes sense or is confusing in any way. Could it be simpler or is there some important missing detail? Is the language over-technical? Is it easy to read or hard work? Then we will try and re-write it together to improve it. The idea behind this exercise is that we have probably not written this perfectly – don't feel shy about giving critical feedback, that's exactly what's needed. If it doesn't make sense to you, it's our fault not yours.

4.4.5.5 Pair-writing with GP users – main phase

In the final months before the launch of GP Evidence I had completed most of the clinical content for the website, employing content design principles throughout, and was ready for a substantial editorial process using pair-writing. I had anticipated the need for this work and had built in recruitment and consent for pair-writing sessions into the overall user-testing phase. Section 4.4.7 on user-testing describes this recruitment process in detail. In summary, GPs had agreed to participate in one hour, one-to-one online sessions and they knew in advance whether it would be a user-testing session or a pair-writing session. For pair-writing

sessions, they had received an email explaining the process beforehand (shown in appendix 4.7).

4.4.5.5.1 Considerations of sample size and content coverage

An important question was whether all the website content should be subject to pair-writing and how many GPs should review any particular section of text. From my reading on pair-writing methods and conversations with content designers I had found that there were no specific recommendations on this, with practice being dictated by context and available resources. At this stage in the project, I had time and budget to cover approximately 20 one-hour pair-writing sessions. I has also planned and budgeted for several days' work with a professional editor as co-content designer (described below), who I anticipated would bring expertise in readability and language but would not have clinical knowledge or a user-perspective.

I had decided that all content should be read by at least one GP. Having completed the draft content it was obvious that however well I had attempted to create accessible, clear content it would be likely to be improved by review with a user. Creating simple content to communicate complex information had been challenging. I was well aware that in the process of reviewing and synthesising the clinical evidence, my perspective had shifted towards that of evidence-expert rather than evidence-user. Input from real users would be essential.

Having decided that, the next question was how long would pair-writing take? There were 12 long term conditions with multiple treatments to review, and I was unsure how much pair-writing would be enough for a section of content, or if multiple readings would be needed. These were unknowns, and faced with limited time and resource I followed a pragmatic

strategy whereby each section of the website would be reviewed in one GP pair-writing session before moving onto the next topic. I kept notes of areas which were particularly problematic, to prioritise for a second round of pair-writing with a different GP which would take place after all content had been reviewed at least once.

This strategy worked well; all content was reviewed by an average of two GPs across a total of 18 pair-writing sessions with 14 GPs.

4.4.5.5.2 Method of GP pair-writing sessions

One-hour long, online sessions took place using Microsoft Teams. Participants were given access to the website and asked to share their screen so both people could follow what was being read.

GPs who had not seen the website before were given a brief introduction and an opportunity to browse for five minutes to familiarise themselves with the format. Then the introductory script which had worked well in the pilot sessions (figure 4.12) was used to focus attention on a particular section of text. An open discussion style was encouraged, with participants invited to either highlight problems, suggest changes or both. Changes to content were made live during the sessions by myself using the website editor, then the site was refreshed to review the amended content. Content was re-examined and re-edited until the participant was happy with the readability and clarity of content.

4.4.5.5.3 Data collection

Results of the pair-writing process were captured in two ways:

- hand-written field notes taken during the session and augmented immediately after the session with reflective content where necessary

- before-and-after screen grabs of content to record changes

4.4.5.6 Pair-writing with professional editor as co-content designer

I explored the possibility of hiring a professional content designer, with the idea that I would adopt the role of subject expert and explore what happened to my draft content in the hands of a professional. This would have been feasible but expensive and only possible to do on a small sample of content. This may have been a valuable exercise, in which I obtained feedback and improved skills to re-edit my own content. While I was considering this, I received an offer of editorial help from a member of the project steering committee (this committee is explained in the next chapter), which transpired to be an ideal solution.

Mandy Payne is a patient member of the project steering committee who also has a role as editor of a newsletter for a health advocacy charity. She had a background in scientific communications, journalism and expertise in bioethics. Now working part-time as a freelance writer-editor, she had become increasingly enthusiastic about the *GP Evidence* project during her involvement and offered her help (at low cost) to review website content.

It struck me that she might be an ideal person to pair-write with. Though not a content designer, she had extensive editorial experience and would bring insights from her professional perspectives as well as providing a very close patient and public perspective to the development of website text. She was familiar with the aims of the project, user-needs and issues which were arising around the presentation of scientific content.

We discussed the idea of working together as pair-writers and I described content design principles and pair-writing methods. She also read and viewed material on content design to familiarise herself with the process. We agreed that this would work well, and set up some

joint sessions of in-person pair-writing over two blocks of three days each. We worked 10am to 4pm, taking breaks as needed, working through the online content in a systematic manner and editing live. Changes made to content were captured by hand-written and electronic field notes kept by both of us, as well as before-and-after screen grabs. Some of these sessions occurred before GP pair-writing on particular content, some after.

4.4.5.6.1 Style guide

During this process we developed a style guide for the website, standardising language, punctuation rules and structure. This was an iterative process, drawing on content design principles and incorporating our subjective judgements. Examples from the guide are shown in table 4.4, the full style guide is shown in appendix 4.5.

Table 4.4 Examples from GP Evidence style guide

Content	Notes
abbreviations e.g., AF	If GPs use these consistently then no need to write full versions/define (JT judgement). Otherwise, spell out in full with abbreviation first time – no need to do this elsewhere in same condition section. If a condition commonly abbreviated is written in full, start each word with capital letter.
email	rather than e-mail, E-mail
treatment(s)	rather than drug/intervention etc.
drug names	generally not capital letters for drug names
referring to a term	use quote marks. E.g., NICE use the term "Acute Coronary Syndromes"
quality of evidence	always capitalise LOW, MEDIUM, HIGH when referring to evidence, even in text
regimen	rather than "regime"

Results from the pair-writing process are described in section 6.4.2.

4.4.6 Principles of risk communication

Methods of presentation of numerical information to communicate risk have been developed with particular reference to healthcare over the last 20 years. I drew on this body of work (217, 268-274) to produce a set of principles and guidance to apply to the development of the prototype design of *GP Evidence*, shown in figure 4.13.

Figure 4.13 Key messages from the literature on how to communicate risk.

Numerical information can be hard to understand: where possible, provide written content alongside numbers.

Relative risk reduction has low levels of meaning for an individual, use absolute risk instead.

- Numbers-needed-to-treat is a form of absolute risk communication with regard to treatments. Many doctors like this mode, but it is less well understood by patients.

Absolute risks can be presented as percentages, or in plain language such as "Out of one hundred people, ten will experience a heart attack".

- The evidence is unclear whether percentages or plain language are better understood.
- Use everyday, plain language.

Graphical illustrations support the understanding of risk and are the preferred design for many.

Varying denominators can be confusing, "1 in 10 people avoid a heart attack but 1 in 37 get a side effect"

- However, some research shows this is not necessarily the case for all.

Different people have varying preferences for how information is presented. Using multiple modes can be helpful.

4.4.7 User-testing

User-testing is a well-established method in website design. It is a process in which a user-participant is observed by a researcher whilst interacting with a web-interface, commonly after being given a task to perform, in order to test its functionality. It aims to identify problematic aspects of a prototype design, or which cause frustration or difficulty. These can then be corrected and an updated version tested with new users.

Direct observation of a user interacting with the interface reveals non-verbal information such as time taken to navigate or perform tasks, which options for interaction on a page are being selected, which content is being viewed, as well the emotional reactions of the user (be they irritation or delight). Observation is complemented by use of the think-aloud method as described earlier in section 4.4.2, to reveal further information about participants' thought processes and reactions.

4.4.7.1 Aims of user-testing for GP Evidence

The JAD had resulted in the production of a working prototype website. I wished to test the design as early as possible, following the design principles of early iteration and "fail early, fail fast" (section 4.3.2.1). The aims of the user-testing sessions were:

- To test the overall appeal and clarity of the website. Can users tell what the purpose of the website is and rapidly get a sense of what kind of information they will find?
- To test the mechanisms of navigation. Is the information architecture working to enable them to find what they are looking for?
- Are various features of the website easily discoverable?

- Is the main graphical presentation of quantitative data easily understandable? Which elements of this design work and which do not?
- What are users' emotional reactions to the website?
- Are the formats for presentation of textual content broadly appealing?

4.4.7.2 Ethical approval

Ethical approval for the user-testing and pair-writing phase of research was granted by Oxford Central University Research Ethics Committee(CUREC). Reference R67851RE002. A copy of the application form is in appendix 4.6.

4.4.7.3 Recruitment

Invitations to participate were sent to the pool of 213 GP volunteers described in section 3.6. Though the demographic and professional characteristics of this group were broadly representative of the GP population, I was interested to test the website with more newly qualified GPs and so also sent invitations via personal contacts in the Royal College of General Practitioners' "First 5" group – a network of GPs within 5 years of qualification. I was unable to determine how many GPs received an invitation via this route.

Inclusion criteria were the same as for the JAD workshop shown in figure 4.8. I decided to allow participants who had already taken part in previous stages of the work (the interviews or JAD) to participate in the user-testing and pair-writing sessions. Though it is true that subjects who are not naïve to the project will have a different perspective at user-testing, I felt that the exposure they had had would not be enough to invalidate their reactions and that they might have something extra to contribute. Most of the participants for these studies however, were naïve to the project ensuring adequate user-testing with fresh eyes.

Copies of the email invitation and combined participant information and consent sheets which cover both the user-testing and pair-writing sessions are in appendices 4.7 and 4.8. Informed consent was confirmed by verbal confirmation at the start of the user-testing or pair-writing sessions, participants having had the opportunity to read the information and consent form at the point of recruitment and ask any questions via email in the intervening period.

Data handling and storage was conducted according to University of Oxford policy. Detailed description of this process is reproduced in Appendix 4.1. Participants' personal data was stored separately from research data. Research data were identified by allocating a number to each interview.

4.4.7.4 Sample size and participant characteristics

Methodological guidance for user-testing says that four to five participants are enough to identify the majority of usability problems with a website (252, 275). This refers to the testing of a particular aspect of a site, for example a section of information or a specific task which needs to be completed. I anticipated that as the range of features on *GP Evidence* was relatively small, the function of the whole site could be tested within one or two sessions. This would therefore require 5-10 participants for a first round of testing, then a similar number to re-test after any iterations. How many iterations and cycles of user-testing would be needed was unknown, but with advice from my web-developers settled with a plan for 20 user-testing sessions. In the end, user-testing was completed with 16 GPs over the course of the project. Their characteristics are shown in table 4.5, included here rather than the design results chapter in the interests of clarity.

Table 4.5 Participant Characteristics for user-testing sessions

Characteristic n out of 16	
Sex	
Female	10
Male	6
Age, years	
Under 30	0
30-39	6
40-49	6
50-59	3
60 or over	
GP Role	
GP Principal	4
Salaried GP	6
Locum GP	3
Retainer GP	3
Place of original medical degree	
UK	14
Non-UK	2
Geographical Region	
Yorkshire and Humber	1
East of England	1
South-East England	7
South-West England	3
Wales	1
Midlands	2
GP description of practice	
Urban	5
Rural	2
Mixed Urban-Rural	9

4.4.7.5 Methods

Methodological advice was derived from three key textbooks (251, 276, 277) and the Nielsen Norman Group website (278) and tailored for GP Evidence. Sessions lasted for one hour and were conducted by video call over Microsoft Teams. Participants were given a link and password to the website and asked to share their screen so that their activity on the website was visible. Their faces were still visible in a separate window and microphones active to hear speech. The researcher's camera was switched off during the user-test itself to avoid distraction.

The sessions were conducted in a semi-structured manner using a consistent introductory script with a subsequent set of tasks and questions which could be adapted according to participants' behaviour. The range of questions were designed to address the aims of the user-testing, employing open invitations to browse and comment, as well as using focussed tasks.

To minimise response bias from any tendency of participants to respond positively out of politeness, in the introductory script I emphasised a desire for critical feedback and clarified that any problems with the website were our fault, not theirs. In addition, the questions were framed with both positive and negative options. The introductory script and questions are shown in figure 4.14.

Figure 4.14 Introductory script and questions for user-testing sessions

Introduction

Thank you for joining me today. I will start with a little bit of background to explain where the project is up to, and then talk about what we're going to do.

The website is in the early stages of development, and this has been done from the start with lots of involvement of GPs, so that we make sure it will be something that is genuinely helpful and will work well in practice.

This step is to test the prototype, see what your reaction is and *especially* to find any problems, mistakes or irritations.

This input from you as a potential user is highly valuable to us. We *expect* there to be criticisms, problems and improvements to be made so you must feel free to be honest and don't worry about hurting anyone's feelings!

It's important to say that this is not in any way a test of *you* and your ability as a computer-user or doctor, it's a test of the website and whether we're getting it right or not.

As well as looking at the website in general, I'll be asking about your understanding of some the clinical content on the website. Again, this is a test of the website *not* your clinical knowledge. We are working with the understanding that this kind of clinical information is in unfamiliar territory for most of us. If you don't understand it, we need to improve the website, it doesn't mean you haven't got enough clinical knowledge or scientific skills!

Is there anything you would like to ask before we start?

User-testing begins after screen share set-up

Imagine a colleague has told you about a new website called GP Evidence which they say, "has got stuff about treatments for long term conditions – risk reduction, NNTs and side effects, stuff like that".

You decide to have a look.

The first thing I want to ask is just about your first impressions, so open up the homepage and look around. Feel free to scroll but don't click on anything else just yet.

Can you just tell me your immediate reaction?

Thanks, now what I'm going to do is give you an imaginary clinical situation to work with. This should send you on a hunt round the website for some information.

While you're looking, I want you to talk as you're going along to let me know what you're thinking. Especially if you get stuck or don't like something. If there's something you really like, that's good to know too. It's called a "think-aloud" exercise. Just say whatever comes naturally, some periods of silence are fine while you're thinking or reading. I will stay fairly quiet but might chip in with a question occasionally.

So, the clinical scenario is:

A 60 year old man comes to see you a few months after having an NHS health check. He has worked on lifestyle factors as much as he thinks he can. His blood pressure after confirmation with home monitoring is 152/94 in clinic. His QRISK has come out at 20% and the nurse has told him he might need blood pressure drugs and a statin. He's come to see you to discuss if he really needs to take these. He's interested in the benefits and harms of these treatment options.

You decide to have a look at this website to get a bit more information to help you have the conversation.

Now feel free to explore the website, take as much time as you like and remember to "think-aloud".

Targeted questions to ask as needed

Thinking about the homepage:

Did you get a sense of what the website is *for,* or what *kind* of information you might find here?

Did you have any immediate reactions to it – like or dislike?

If participants had difficulty finding content

I noticed you were hunting around a bit there, did you find the menus and structure tricky? Can you tell me what you were trying to do?

Looking at the main graphics sections:

How easy or difficult did you find it to understand the numbers?

Which bits of that information package did you find yourself going to, or what worked best for you?

Are there any bits which just didn't make sense?

Can I ask you to tell me what you've understood from that section- perhaps how you might now explain it to a patient?

In the main graphic sections there are some numbers/percentages. Do you understand what they mean?

Buttons under the graphics section:

When you looked at the range of buttons underneath there… was it easy to grasp what they were for, or do they look a bit confusing or unclear?

What did you think of the type and amount of content in the pop-up boxes? Any comments about the layout of that text?

In the "Harms" section:

The presentation of information there is quite different from the graphics. How did you find that?

Were you able to find the information you needed and understand it? How easy or hard was that?

Do you have any comments about how that information is laid out?

If participants have not clicked on a tab, button or link

I noticed you haven't clicked on [xyz]. Did you see it or spot it was there?

Could you have a look at that bit now?

4.4.7.6 Data collection

Findings were captured by handwritten notes during the session, then immediately written up onto Microsoft Word documents. For the first three user-testing sessions I had used a prepared structured data sheet with sections devoted to particular aspects of the website. However, this proved cumbersome and a blank sheet of paper worked better. After each phase of user-testing, summary notes were collated from each individual session record to generate a list of findings. These are reported in chapter 6.

4.4.7.7 User-testing sessions with independent researcher

Response bias influenced by my role as project lead and user-tester could be avoided if the tests were conducted by an independent researcher. However, as this was a doctoral research project where I was performing my own research and without a budget to out-source the user testing, this was not appropriate or feasible. However, to explore any effect of such response bias, I enlisted help from two colleagues in the department to perform three of the user tests without my presence.

These colleagues were academically trained, one a fellow doctoral student and the other a research data manager. I gave them background teaching about the project, the website and user-testing methods. They observed one of the sessions I was conducting with the consent of the participant. They then conducted three user-testing sessions following the script and question guide, one of them acting as interviewer, the other as a scribe. Their written summary notes were discussed with me and I wrote these up as documents on Microsoft Word.

Though this limited exercise could not prove or entirely correct for response bias, it provided some opportunity for responses to emerge from participants which might otherwise have not, and an opportunity to see if there was any general difference in tone or level of enthusiasm from respondents in this context.

5 Chapter 5: Methods part 2. Evidence Collation – Approach and Process

The previous chapter described the approach to web-design and content creation for *GP Evidence*. This chapter describes the development of an approach and process for selecting the clinical information to include in the website. This had three stages:

- **Stage 1 – reflection, review and formulation.** A reflection on the aims and likely informational needs of *GP Evidence*, which led to an exploration of evidence-gathering standards for patient decision aids (PDAs), a closely comparable field. I drew on these to develop a proposal for an approach for *GP Evidence*.
- **Stage 2 – Hosting a one-day workshop with relevant stakeholders.** This workshop was conducted with a group comprising lay and expert members, to reflect on and refine the proposal developed in stage 1. Members of this group went on to form a patient-expert steering committee for the duration of the project.
- **Stage 3 - Application** The resulting evidence collation strategy was applied to the content creation of the website over two years, with reference to the steering committee to support my decision-making about the selection and use of evidence in situations of uncertainty. This is described in chapter 7.

5.1 Aims and information needs for *GP Evidence*

The aim of *GP Evidence* is to provide GPs with information on the benefits and harms of treatments for common long-term conditions in primary care practice. This is in the context of the role of NICE guidance in defining best practice: *GP Evidence* does not seek to provide

recommendations or directive opinion, rather to bring to the surface useful elements of the scientific evidence which underlie guideline recommendations.

The key element of this information is the quantitative value of treatment benefits and harms, expressed, for example, by absolute risk reduction or numbers-needed-to-treat. These figures derive from randomised controlled trials (RCTs) or systematic review and meta-analysis of these trials. The systematic review and meta-analysis is established as the basis for the generation of NICE guideline recommendations (53), representing the highest level of evidence quality to answer clinical questions about the benefits and harms of treatments (279). Extracting quantitative data from a NICE guideline systematic review would therefore be a reliable method for deriving content for *GP Evidence*, drawing directly on their rigorous review process and ensuring compatibility of *GP Evidence* with NICE guidance.

My explorations leading up to this project, of the publicly available documentation of NICE guideline development,[7] had shown that this does contain much information which could be extracted to present in *GP Evidence*. However, sometimes the answer to a particular clinical question I had been asking was not contained in their reviews, or the analysis was presented in such a way as to make extracting usable quantitative information impossible. Sometimes, the evidence base of a very well established treatment (for example aspirin for the secondary prevention of myocardial infarction) is so historic that the evidence reviews were no longer available on the NICE website.

[7] Each NICE guideline has a dedicated collection of web-pages which contain extensive documentation of all stages of the guideline development process. These contain details of evidence searches and reviews, meta-analysis including GRADE tables and forest plots, notes on committee discussions and "evidence to recommendation" summaries.

If information were not available from a NICE evidence review, I considered a next step could be a literature search for other evidence sources. Finding a single RCT for example would provide useable data which could be formally appraised for its quality (280, 281), but may not represent to the totality of evidence for that intervention. Another systematic review and meta-analysis which had been conducted to methodologically high standards would generally be a more reliable source of evidence. However, even well conducted systematic reviews may obtain differing results when attempting to answer the same question. For example, early in the project I found three systematic reviews exploring the value of the drug treatment of mild hypertension which reached differing conclusions. One reported no benefit (282), the other a 20% RRR in cardiovascular events (283), and the third, a 12% RRR (284).

Therefore, important questions at the outset of the project were: what approach should I employ, if unable to find the information I needed to derive content for *GP Evidence* within a NICE guideline review? How would I successfully identify an evidence source of high quality, and how to choose between differing sources if they provided meaningfully different quantitative answers? Avoiding introducing my own personal biases into this process would be critical to ensure a reliable information resource.

There was another important consideration of feasibility. I would be responsible for the evidence collation process with a limited amount of available time. Any chosen strategy would need to be realistically possible whilst also being rigorous and transparent enough to produce reliable and trustworthy information.

5.2 Standards of clinical evidence in the development of patient decision-aids

Patient decision aids (PDAs) are tools which have been developed to support shared decision-making between clinicians and patients. As discussed in section 2.6 there is an established network of expertise in their development and use, with published consensus methods for their production. More recently, NICE and NHS England have produced their own decision aids.

PDAs usually contain quantitative information on the potential benefits and harms of interventions, contextualised with patient-facing information to support understanding. The purpose of these is the same as that as *GP Evidence*: to translate scientific information for use in practice. Though their design, target audience and eventual method of use is different, they also require identification and extraction of relevant, reliable quantitative information from clinical research. I sought information on the evidence-gathering strategies used for these decision aids in order to inform my strategy for *GP Evidence*, I will summarise these strategies here.

5.2.1 International Patient Decision Aid Standards (IPDAS) Collaboration

The IPDAS Collaboration was established in 2003 as a group of researchers, practitioners and stakeholders to produce evidence-informed frameworks to support the development of high quality patient decision-aids (PDAs)(285). A 2012 update of their standards document (286) contained recommendations on required standards of evidence, key components of which are presented in figure 5.1.

Figure 5.1 Extract from 2012 update of IPDAS standards document

"…decision aid developers will use evidence derived from systematic reviews that a) avoid selection bias; b) carefully and reproducibly assess the quality of the incorporated reviewed studies (i.e., the studies' protection from error and bias); c) summarize the estimated pertinent effects (ideally quantitatively in a meta-analysis); d) indicate the extent to which these estimates are trustworthy; and e) assess the extent to which selective reporting and publication bias may corrupt the body of evidence."

"Decision aid developers could make use of existing syntheses they find to be of high methodological quality and reasonably recent. This suggests that developers should be able to judge the quality of these reviews. Some tools exist to assist with this work[1]. If high quality syntheses were not available, then developers need to conduct their own reviews or commission their conduct to credible parties. In these cases, decision aid users would need to critically review the synthesis, which implies that the review must be fully reported and placed in the public domain and subject to peer review. This process is similar to that followed by rigorous guideline developers"

1) Shea B, Hamel C, Wells G et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology* 2009;62(10):1013-20

These recommendations support the use of high-quality systematic reviews and refer to established criteria to judge this quality such as the AMSTAR tool. NICE evidence reviews and Cochrane reviews are conducted to these standards (287, 288) and reviews published in major journals should also be of high quality and could be appraised using the AMSTAR criteria (289).

However, the IPDAS recommendations do not specify how such systematic reviews should be selected from a range of potential options. In an accompanying paper (290), the authors ask a closely related question: *"How might developers decide what evidence used to inform a decision might be pertinent to the patients who are the intended audience of the tool?".* The

discussion of this refers to the challenge of indirectness – if and how to present evidence from populations who may not be similar to a person using the PDA. They comment that *"The extent to which evidence is pertinent is subject to much judgement"*, and do not propose any rule or guidance to support such judgement, though suggest that transparency in the PDA itself to communicate any uncertainty is desirable. This acknowledgement of the necessary role of judgement even in the context of working with high-quality systematic reviews seems relevant to the selection of reviews themselves. In 2018, authors from the IPDAS collaboration proposed national standards for the development of PDAs to be applied in the US (291), which included the recommendation that *"The patient decision aid content is based on a rigorous and documented evidence synthesis method"*, but did not provide any further detail.

5.2.2 NICE decision aids

NICE have produced a suite of PDAs since 2014, collaborating with a range of stakeholders during their development (292). Their most recent process guide was published in 2018 (293) and contains guidance for sourcing evidence, summarised in figure 5.2. This strategy reflects a desire to anchor PDA material in NICE's own evidence reviews, but allows for the possibility of drawing on other sources.

Figure 5.2 Extract from NICE decision aids: process guide 2018

> "The primary source of information is the evidence reviews or submission used in the development process for the guidance. Where necessary, individual studies included in evidence reviews or submissions may be examined for further data. Studies excluded by the guidance development process will not be used. Additional information may be taken from other NICE guidance and advice, and from standard reference sources including (but not limited to) NHS Choices, NICE Clinical Knowledge Summaries (CKS), summaries of product characteristics (SPCs), BNF and BNFC. NICE decision aids use only evidence that is in the public domain. A user guide and data sources document is published for each NICE decision aid that explains how the decision aid relates to the guidance, how it was produced and the evidence on which it is based."

I reviewed the content of a number of NICE decision aids and their supporting documents to explore their approach. This revealed a range of strategies. For example, the PDA for the use for statins in the primary prevention of cardiovascular disease used a relative risk reduction derived from its own meta-analysis and applied this to a range of baseline cardiovascular risk scores. They acknowledged an assumption that the same relative risk reduction would be applicable across all patient groups (294). A PDA for abortion care used a more complex approach, combining effectiveness data from a Cochrane review with harms data from government statistics (295). A PDA for treatment options for hypertension contained quantitative estimates of drug side-effects derived from the British National Formulary, but did not offer any quantitative estimate of benefits on stroke or heart attack reduction because *"…the NICE evidence review did not find a constant relative risk reduction for any given degree of blood pressure reduction from any given starting blood pressure"* (296).

I spoke to members of the NICE PDA development team in the early stages of my project to discuss the NICE approach to the application of PDAs. They described a priority to link PDA content to as high-quality evidence as possible, whilst acknowledging that pragmatic choices might need to be made in order to produce a useful tool. These choices were made in partnership with the guideline developers and a PDA committee on a case-by-case basis.

5.2.3 NHS England decision aids

NHS England's Personalised Care Institute[8] commissioned a series of PDAs (which they call Decision Support Tools, DSTs) in 2022 (186). The statement in their commissioning guidance

[8] The Personalised Care Institute was established by NHS England in 2020 to support development and training and set quality standards for the delivery of personalised care within the NHS.

regarding evidence sourcing for DSTs is brief and broad[9]: *"DSTs must be informed by the findings from a well conducted and transparent process of evidence synthesis."*

5.2.4 Implication for *GP Evidence*

This review of how producers of PDAs approached evidence sourcing was useful. On the one hand, there was a consensus that a high-quality systematic review and meta-analysis provide the ideal evidence source, but on the other, a recognition that this ideal has limitations. There may be a need to draw on other evidence sources for particular data and use judgement to consider how evidence, even if high quality, might be applied to an imagined user of the PDA. There was an absence of any defined rule or structure to guide the choice between more than one high-quality evidence source.

I considered that an appropriate strategy to adopt for *GP Evidence* might be:

- Seek information from NICE guideline evidence reviews as a first choice.
- Next, seek a Cochrane review as a second evidence source if information not available from within a NICE review. Cochrane is a suitable second choice due to its position as an established, neutral, trustworthy source employing high-quality and transparent methods.
- If neither NICE or Cochrane provide data suitable for presentation in *GP Evidence*, search for other systematic reviews and meta-analyses which should be quality appraised using AMSTAR criteria.

[9] Personal communication, Personalised Care Institute, NHS England 2022.

- Single RCTs may be used, with quality appraisal using the Cochrane risk of bias tool (280) and consideration of where they sit within the totality of evidence on a clinical question.

This seemed like a straightforward hierarchy, however it is vital that *GP Evidence* be considered trustworthy and not open to major criticism regarding its factual content. I was concerned about a number of ways in which my selection of evidence might be less than ideal:

- Vulnerability to my own biases: for example, a tendency to scepticism about preventive interventions and a preference for "minimally disruptive medicine"(140) which might lead me to favour evidence showing smaller effect sizes.
- Given that I was unable to conduct literature searching to the quality that would be standard in a systematic review, a possibility of missing important evidence such as the most up-to-date systematic review or a new practice-changing RCT.
- That when making inevitable choices about which granular level of data to present, I might make inappropriate judgements and assumptions.

I therefore set-up a workshop and steering committee to support and guide me in this process.

5.3 Expert-patient workshop

5.3.1 Recruitment of participants

To refine the evidence-collation strategy, I sought input from two perspectives. First, from those with professional expertise in evidence-based practice and shared decision-making and secondly from those with a patient perspective. The expert perspective could highlight methodological issues and offer experience in the analysis of clinical data. The patient view

would be particularly important to provide perspective and challenge about inevitable value judgements which might need to be made, for example: what degree of difference in clinical findings between clinical studies is large enough to make a meaningful difference in decision-making in the context of *GP Evidence*?

I wished to convene a group of people who would bring a diverse perspective to a complex discussion. In addition, I realised that patient and public representatives would need to have a greater than average level of confidence and insight about medical information – the discussions would be inevitably complex and they would need to be able to understand these and contribute meaningfully. In anticipation of the workshop, I had been looking out for potential participants over the first year of the project and after discussion with my supervisors invited a group of people with a range of perspectives who agreed to participate. Their names and profiles are summarised in figure 5.3.

Figure 5.3 Members of expert-patient group for evidence collation workshop.

Jon Brassey. Director of Trip database, expertise in evidence searching and application to clinical practice.

Andy Hutchinson. Senior Medicines Adviser at NICE, closely involved in NICE's work on shared decision making, including the NICE guideline on SDM, and leads on the development of NICE's patient decision aids.

Emily Lam. Person with experience of living with and making treatment decisions for multiple long-term conditions. A lay member of NICE technology appraisal committees.

Richard Lehman. GP and Professor for the Shared Understanding of Medicine at Birmingham University, expertise in EBM and SDM.

Kamal Mahtani. GP and Professor of Evidence-Based Healthcare at Oxford University. Expertise in evidence synthesis, one of the academic supervisors for this research.

Mandy Payne. Person with experience of caring for people with long-term conditions. Editor for the newsletter of UK charity Healthsense, which advocates for science and integrity in healthcare.

Jackie Walumbe. Person with experience of living with and making treatment decisions for multiple long-term conditions.

5.3.2 Workshop organisation

Participants were approached informally, and after expressing an interest were sent a formal invitation including details of the workshop. A copy of this is included in appendix 5.1. Though the intention had been to hold an in-person workshop, the Covid-19 pandemic meant that it was held via videoconference on Microsoft Teams on 20th October 2020. The workshop ran from 10am – 4pm with a loosely structured agenda reproduced in figure 5.4.

Figure 5.4 Expert-patient workshop agenda.

Agenda
Expert-Patient workshop for "GP Evidence"
20th October 2020

10 am

Welcome and introductions

Presentation by JT Aims, background, tour of the prototype website, Q+A

Break 10-15 minutes

Discussion 1: Is the type and scope of content right?

"What would I like my doctor to know?"

Lunch break 12.30-1.30

Discussion 2: How and where to source evidence and clinical content?

JT will start with a short presentation on thoughts and research so far.

Break if needed

To conclude

If possible outline principles and methods to guide website development.

Opportunity for ongoing participation.

During the introductory presentation I introduced the group to the background issues and problems the website was trying to address. I then described the process of website design so far, and gave a demonstration tour of the website (the design resulting from the JAD workshop), to visualise what type of content would be presented and how it would be structured. I had populated some areas of the site with draft clinical content. A copy of the presentation is in appendix 5.2.

During the first discussion section, I sought feedback on the general design and purpose of the website, information classification and proposals for clinical content. I asked participants to think about the question *"What would I want my doctor to know?"* as a device to frame this discussion. This was to clarify the purpose of the site and ensure that my plans for the *type* of clinical content to include met with approval before moving onto the more technical discussion in the afternoon about exactly how to seek out content. We closed the meeting by agreeing an evidence collation strategy for GP Evidence, which I later circulated to participants for comment. While chairing the meeting, I ensured that all participants had an opportunity to speak, particularly encouraging patient members to ensure that confident experts did not dominate the discussion.

5.3.3 Outcome of workshop

I kept written notes during the meeting and wrote them up onto a Microsoft Word document immediately afterwards. This is shown in appendix 5.3. Summary findings from the workshop and a flow diagram of the resulting evidence collation strategy are shown in figures 5.5 and 5.6.

Figure 5.5 Summary of outcomes from expert-patient workshop

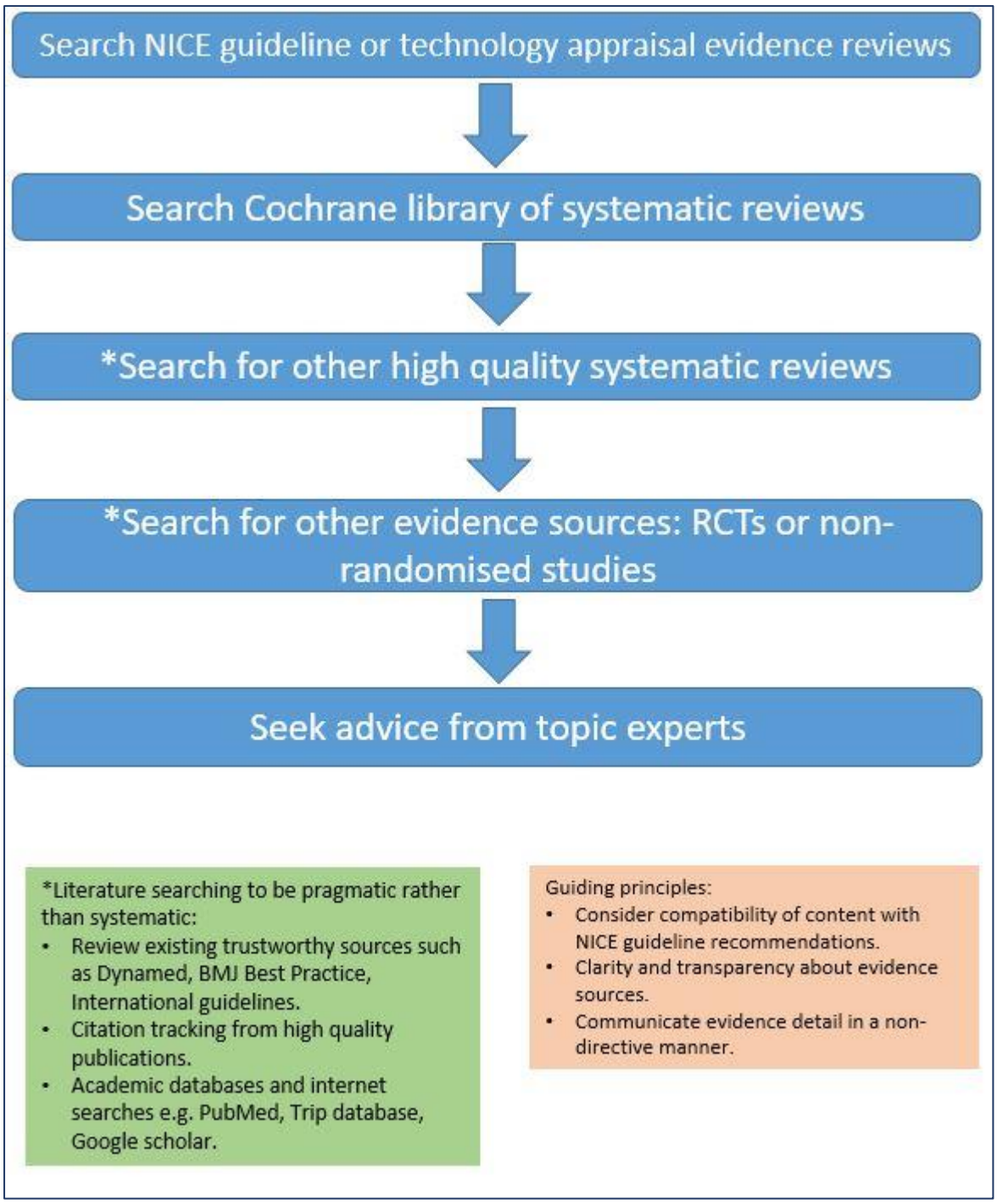
Is the proposed content and design broadly appropriate?

- Yes, subject to ongoing iterative testing with users.
- There is a range of possible extra content which would be valuable to patients (for example, details on practical aspects of med management or dose-related risk of side effects) but would present major challenges in terms of workload and maintenance. There is also a risk of stepping outside the key offering of the website thereby diminishing its impact.
- Plan to get the key content right: quantitative information on benefits and harms of treatment and the supporting evidence. Consider on case-by-case basis where extra information would be particularly valuable but not distract from the main function of the website, and be feasible to research and maintain.

Approach to evidence collation

- A pragmatic "good enough" strategy to find and use as high quality evidence as possible was unanimously felt to be appropriate rather than an extensive, formal systematic review style search strategy.
- This approach recognises that the function of GP Evidence is as a "translational tool" to support and sit alongside NICE guidance, rather than having the aim of producing independent recommendations.
- The proposed hierarchy of NICE reviews, Cochrane reviews, other systematic reviews and RCTs was agreed to be appropriate. Drawing on other trustworthy evidence sources as needed such Dynamed, BMJ Best Practice or international guidelines could be an additional strategy to identify evidence if needed.
- Seeking advice from external topic experts was agreed to be appropriate if needed, bearing in mind any intellectual or other conflicts of interests and not allowing opinion to over-ride evidence.
- Clarity and transparency about evidence sources as well as the highlighting of uncertainties and controversies on the website is essential.
- Bear in mind that website content needs to be robust to criticism, with NICE guidance providing a reference point.
- Avoid opinion and directive writing on the website to maintain a tone of neutrality.
- Bring uncertainties and dilemmas to the steering committee to seek consensus and to address any situations which have not been anticipated thus far.

Figure 5.6 Flow diagram of evidence sourcing strategy for GP Evidence



5.4 Reviewing of quantitative data extraction

When undertaking systematic reviews, it is considered best practice for two reviewers to extract data from studies to reduce error (297).The same risk of error would apply to

quantitative data I would extract from the variety of evidence sources and apply to GP Evidence. Therefore, I arranged for all the quantitative data on the website to be reviewed by a second reviewer; a group of DPhil student colleagues undertook this work (see acknowledgements).

When collating quantitative content, I kept a record on a Microsoft Excel spreadsheet of the evidence source, exact location of the specific data within a publication and a description of how that data had been handled where necessary. These spreadsheets were used in a series of in-person and online sessions with the reviewers, who compared content on the website with the original source, performed their own calculations when needed and informed me of errors, which I then double checked and corrected.

All quantitative information on the website went through this process. An example of a spreadsheet is in appendix 5.4. Notes on selection and sources of evidence including choices made and their justification were kept on Microsoft Word documents for all content.

5.5 Establishment of steering committee

Workshop participants were invited to continue their involvement with the project as members of a Steering Committee, to be convened at intervals during the iterative evidence collation and user-testing phase of the project. All agreed except RL who was unable to commit to the time but remained as an informal mentor. We agreed to adopt a flexible approach to the timing and frequency of committee meetings, allowing this to be led by the need for their input.

5.6 Implementation and results of the evidence collation strategy

The outworking of this strategy, including challenges which arose and the resulting content of the website will be demonstrated in the chapter 7 using examples of website content with descriptions of the process of evidence collation.

6 Chapter 6: Results part 1 – the design of *GP Evidence*

The design of *GP Evidence* drew on my clinical experience, the background literature, stakeholder input, GP interviews, design principles and research methods. These were brought together in an iterative process over three years of this part-time doctoral project. The process of collation, curation and content design of the clinical information happened in parallel, with interaction between user needs, design needs and the nature of clinical information available.

This was an exploratory, iterative process reminiscent of the "fuzzy front end" (figure 4.4); design and content changes happened incrementally. However, there were key events and periods of more concentrated activity on particular aspects which are illustrated in a timeline in figure 1. In this chapter, I use examples of content from the website itself to describe the application of principles and methods set-out in the previous chapters. Though it is unfeasible to report on the development of every piece of content on *GP Evidence*, a range of examples has been chosen to demonstrate the application of principles and methods of design.

This chapter has four sections:

- **Prototyping**
- **Joint application design** workshop, resulting in a working version of GP Evidence
- **User-testing** of this version with subsequent design changes
- **Content design**: fine development of textual content

Figure 6.1 Timeline for the GP Evidence project

6.1 Prototyping

As an essential prerequisite of prototyping, I achieved an understanding of GP user needs from the GP interviews and their findings described in chapter 3. These are summarised in the next section, alongside findings from the "think-aloud" part of these interviews.

As well as understanding user needs, an important element of design research is to understand the context in which a user is working and the interests and attitudes of stakeholders in that system(298). Methods and tools have been developed to understand and communicate with stakeholders, which are ideally implemented early in a design project(251). For this project, my understanding of the context, system and stakeholders had been developed from my personal experience and the range of work I had undertaken prior to the doctoral research (described in section 1.7). The learning from this which was most directly applied to the prototyping process was a collection of suggestions and recommendations from two conference workshops and one piece of committee work. These combined to form a valuable set of features and ideas to include in the prototype design of *GP Evidence*. These are summarised in Table 1.1(chapter 1).

6.1.1 Findings from the GP Interview study

6.1.1.1 First part of interview study

In sections 3.9 and 3.10 I described the findings from the first part of the GP interview study which explain how GPs currently think, reason and feel about quantitative information on the benefits and harms of treatments. To recap, the key findings from this were that the GPs were aware of a knowledge deficit in this area and perceived that this might impact their ability to offer personalised care. They were however, using what information they did have to try and

personalise care for their patients. They had an appetite for more information on the benefits and harms of treatments, but anticipated barriers to using this: the time available to find this information, their ability to understand research statistics, and how to communicate this to patients. They expressed concern about the pressures of normative practice: the expectation that they should follow guidelines, as well as medico-legal concerns regarding the implications of (even appropriately) deviating from guidelines.

This part of the interview study provided important information which shaped the design of *GP Evidence*:

- Simplicity and clarity in the communication of quantitative information is vital. Design should assume no or low confidence in statistical terminology.
- Access to information needs to be fast.
- Though the priority is for clear, quick information, GPs are also thinking in complex ways about their patients and how evidence applies to individuals. Contextualising information about the quantitative evidence presented will be needed to support its application to a patient.
- NICE guidelines are central to GPs' practice. The information that *GP Evidence* provides should be compatible with this and presented in a way which enables users to understand how it relates to guideline recommendations and normative practice.

6.1.1.2 Think-aloud part of interview study

In section 4.4.2 I described the aims and methods of the "think-aloud" part of the interviews, where five online resources which aim to deliver information on the benefits and harms of treatments to clinicians or patients were reviewed. This was to identify design features which

helped or hindered the understanding of this information by GP participants. Descriptions of the resources are in figure 4.6. Here, I will highlight relevant features and findings before describing the implications I drew for the design of GP Evidence.

6.1.1.2.1 Resource 1: NHS Scotland Polypharmacy website. Efficacy (NNT) page.

This page was designed for use by healthcare professionals. Screengrabs of the pages I describe are in appendix 6.1. Observations are summarised in figure 6.2.

Figure 6.2 Feedback on NHS Scotland polypharmacy website (NNT)

- The landing page for information on numbers-needed-to treat (NNT) provoked positive reactions. "NNT" and the two main menu options of "By Condition" or "By Drug" were in large print on a spacious page. Users responded positively and went to click on options rapidly.
- The "By Condition" page with only five menu options provoked similar positive reactions, with comments on its clarity and intuitive feel.
- However, the "By Medicine" page was less positively received, users reacting negatively to a very long menu with complicated headings including drug dosing and indication.
- One style of drug information page generated positive reactions from some, who honed in on a simply presented NNT figure of 88. However many then seemed to struggle to take in the rest of the contextual content about the study population and clinical scenario to which it referred. My impression was that orientating themselves required too much mental effort.
- Another style of drug information page was presented slightly differently. Again, users honed in on a number needed to harm (NNH) figure of 67, and understood that this related to the outcome of major bleeding. However, some were confused by the presentation of a time-frame of 1.8 years, or found the explanatory "Comments" text a strain to read.

My impression looking at this website and the GPs' reactions was that it had some positive features which helped with clarity, but that the information was presented in a way more compatible with the scientific papers which underpinned the data than the ways in which GPs thought about things. It didn't match their mental models, hence they were forced to

undertake laborious reading to get information about the scientific study to consider how they might make sense of the headline figure.

6.1.1.2.2 Resource 2: NHS Scotland Polypharmacy website. Decision-making tools page

Another section of this website was designed for patients, presenting data about the benefits and harms of treatments in a way designed to function as a decision aid. Screengrabs of the pages I describe are in appendix 6.2. Observations are summarised in figure 6.3

Figure 6.3 Feedback on NHS Scotland polypharmacy decision making tools

- It opened with a clear, spacious menu of individual drug names which appealed to the GP users. However, it was not clear what condition or indication these treatments were being applied to, this being revealed only on the next screen.
- The use of infographics and plain language statements in an example showing the benefits of metformin for people newly diagnosed with type 2 diabetes provoked clear enthusiasm. Many participants leant into the computer screen or made positive comments like "Ooh, that's nice". However, there was a mixed reaction regarding the degree of understanding they took away. Some managed to take in the fact that each infographic represented a particular clinical outcome, but others struggled. Some GPs said that the use of varying population sizes in the infographics was unintuitive and they would have preferred a consistent image size (i.e. a consistent denominator population).
- On another treatment page, infographics showing the benefits of alendronic acid for the prevention of osteoporotic fractures were supplemented by a series of plain language statements describing the variable effect of the drug in women of different ages. Some GPs welcomed the usefulness of this stratified data, saying it would be helpful to apply to individuals. Others found the variable denominators unintuitive, or the text too dense.

6.1.1.2.3 Resource 3: NICE Statin Decision Aid (2014)

This decision aid (299) had been produced by NICE to accompany its guideline on lipid modification (300). It was a lengthy PDF document with written content aimed at patients explaining concepts around cardiovascular risk and the role of statins. They provided a set of infographics to demonstrate the benefits of statins at varying degrees of baseline

cardiovascular risk. It was reaction to these graphics (shown in appendix 6.3) which I wished to test with the GPs.

Figure 6.4 Feedback on NICE statin decision aid

- GPs reacted extremely positively to these images, comments included:
  - How easy to understand they were.
  - How the presentation of outcomes with two graphics, one for *no* treatment and for *with* treatment was helpful to help explain benefits to patients.
  - The inclusion of written summaries alongside the graphics was helpful and reinforcing.

Most noticeable was the visible reaction to these images during the think-aloud exercise. The GPs seemed to have their attention grabbed, or visibly relax or smile. Some spontaneously described stories of using visual tools like this in consultation, or talked about how these helped them have conversations with patients. There had been a similar reaction to the infographics in the NHS Scotland polypharmacy site, but not as strong.

6.1.1.2.4 Resource 4: BMJ Rapid Recommendations

The British Medical Journal (BMJ) have published a series of systematic reviews and clinical guidelines accompanied by innovative graphic summaries (175). I asked GPs to explore one of these summaries relating to the use of corticosteroids for sore throat (301). Screengrabs of the pages I describe can be seen in appendix 6.4. Observations are summarised in figure 6.5.

Figure 6.5 Responses to BMJ rapid recommendations

- The opening graphic provoked mixed reactions. Many were attracted by the smart, coloured design and appeared interested and stimulated. Others found it too cluttered.
- There was generally a positive response to the structure, with graphics defining the type of patient and the side-by-side presentation of interventions being highlighted as a positive.
- Many GPs missed the white highlighting of the arrow in the final section indicating the conclusion that the authors were making a weak recommendation for the corticosteroids.

- A number of GPs commented on the phrase “We suggest short course steroids” being useful as an upfront key message which was visible on this start page, however, many did not appear to notice it.

An expanded section of information was available by clicking on a “details” box.

- Users seemed to appreciate the ability to expand the content and read more detail. Some found this format reasonably easy to interpret and extract a quantitative idea of the benefits and harms of treatment, whereas others found it too complex.

- Additional pieces of information in the grey boxes at the base of the graphic were commented on positively by some.

- It contained an evidence quality statement, to which some GPs responded positively finding the combination of small stars and single word easy to see and digest.
    - There was an additional facility to expand this section on evidence quality further providing a technical breakdown of the evidence quality. The GPs were less keen on this, finding it difficult to understand. It is written in the language of a systematic reviewer, not a clinician.
    - In this layer of information, there was an unexpected surprise: a plain language statement in a speech bubble delivering an easily digestible “take-home message”. A number of GPs commented positively on this and that they would have liked to have seen this in a more obvious position.

Overall, this online resource had succeeded in a number of ways, providing a stimulating interface and a satisfactory way to communicate information to many of the GP users. The GPs all displayed an appetite for information foraging, willingly clicking through to further information. However, the format was a little over-complex for many, both in terms of graphic design and content.

6.1.1.2.5 Resource 5: thennt.com website

This website provides evidence summaries for a wide range of interventions and is aimed at clinicians. Screengrabs are in appendix 6.5. Observations are summarised in figure 6.6.

Figure 6.6 Responses to thennt.com

- The long menu was structured by medical speciality. Though GPs navigated their way around this successfully, a number commented that it appeared cluttered.
- On selecting a treatment, a simple summary was displayed. GPs responded positively to the simple and obvious presentation of an NNT figure, in this case shown in the appendix, for the benefits and risks of statins in the prevention of heart disease.

- However, this single number is in fact scientifically unhelpful. The NNT for a statin will vary according to the individual and their baseline risk. It may be very significantly lower.
- Some GPs commented on this shortcoming, others did not – seemingly taking away this number as something which could be applied to all patients.

This was an important observation. I realised that information needed to be presented in a way which would encourage GP users to think about baseline risk or patient characteristics whilst at the same time not over-complicating the presentation of information.

- Below the simple summary, there are also multiple references and discussion text. All of the GP users said that they would be unlikely to read it, describing it as dense, off-putting or similar.

6.1.1.3 Main findings from the think-aloud study

A summary of findings from this study which informed the prototyping process are in figure 6.7.

Figure 6.7 Summary findings from think-aloud study

Infographics were very clearly a favoured feature

- They seemed to function in two ways. First, as an attractive attention grabber and communicator of what *kind* of information is on offer. The GPs instantly knew they were looking at information about the benefits and harms of treatments. Secondly, the graphics appeared effective in communicating quantitative information in a way that could be understood with little effort.
- The set of two 10 x 10 icon arrays format as used by NICE was preferred over a single variable denominator style as had been used on the NHS Scotland polypharmacy website.
- Small amounts of accompanying text reinforcing the graphic information seemed welcome and helpful, but too much text might overwhelm.

Provide a summary message or piece of information first (or obviously)

- The GPs tended to hone in on the key piece of information on a page. Sometimes this is all they would read, but even in those who read more widely they seemed to want to anchor themselves in a starting place of the "main message".

Menus should be clear, spacious and not too long

- Menus structured either by condition or treatment seemed to be satisfactory and GPs scanned and clicked rapidly on these.
- However, longer menus with cluttered, complex titles worked less well.

Textual content is valued but must be extremely clear and simple

- Short pieces of text as used by NICE, BMJ Rapid Recommendations and some examples from NHS Scotland polypharmacy website worked well.
- Dense or longer pieces of text were off-putting.

Smart design was appreciated, but not by all

- The BMJ Rapid Recommendations provided a good example of a design which caught the eye, appealed to some, but was too complex for others. Users reacted less dramatically, but more favourably to the simple style used by NHS Scotland polypharmacy website menus.

Providing contextual information alongside the key data is challenging

- Although a simple "take-home" single piece of information is key to engaging users, in the context of the clinical content of GP Evidence, this would always need contextualising to be meaningful and applicable. Careful thought would need to be given to how to do this.

Users *are* interested in more complex information

- The GPs actively explored these resources, demonstrating information foraging and an interest in learning more. They were happy to click, explore and expand content.
- Detailed clinical information could be welcome, for example the breakdown by age of benefits of alendronate shown in fig. 9, but could easily become too complex.

6.1.2 Use of prototyping tools

Following the GP Interviews and before the JAD workshop, I synthesised my findings and ideas using three design techniques: the job story, story boarding and lo-fidelity prototyping as described in section 4.4.3. I had the opportunity to do this whilst participating in two training courses on user-experience and content design[10] where I was able to develop ideas and receive feedback.

6.1.2.1 Job story

The job story I developed to frame the user-need of GP Evidence was:

When I am a GP thinking about a treatment for a long-term condition for a patient…

I want easy access to understandable information on the possible benefits and harms of that treatment…

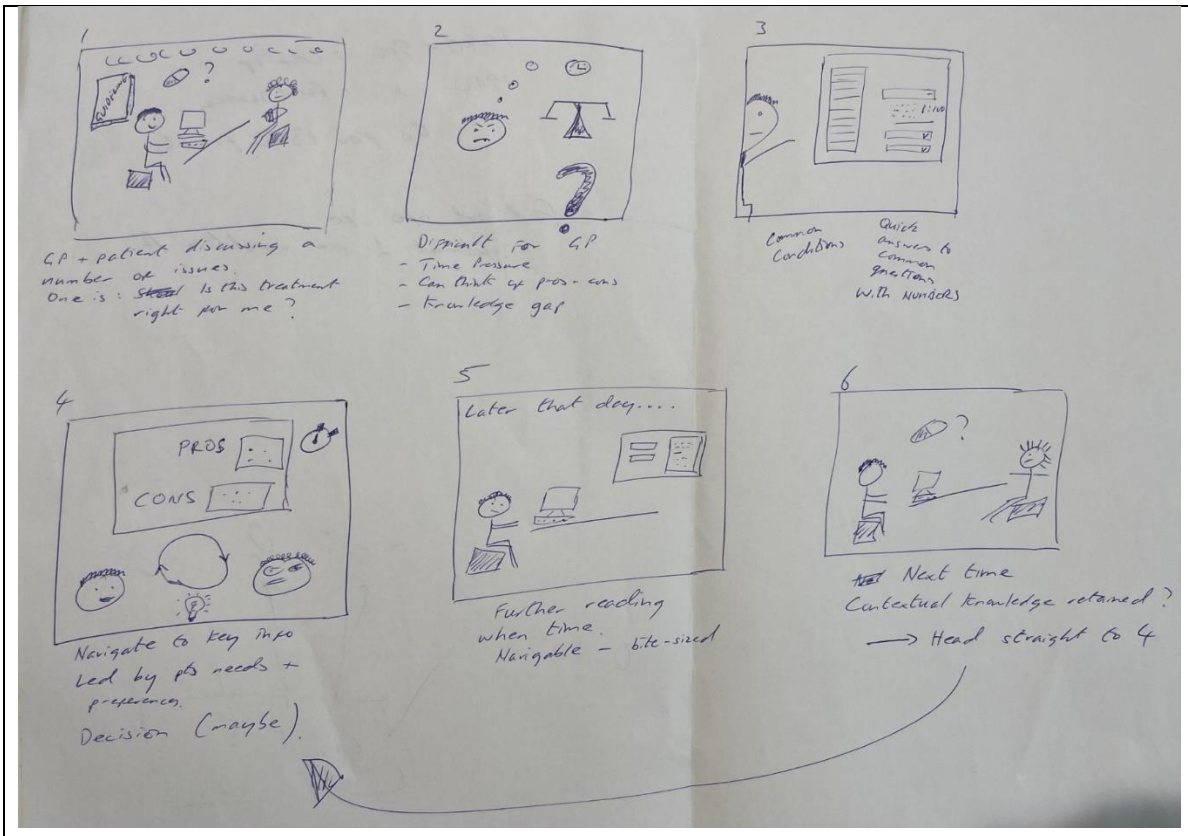
So that I can weigh up the pros and cons of the treatment to help make the decision with my patient.

[10] User Experience (UX) Design. University of the Arts, London. Chelsea College of Art. 28-30th August 2019
Foundation in Content Design. Content Design London. Waterloo Creative Studio. 25-26th February 2020

Storyboard

I created this sketch to imagine and clarify the use of GP Evidence in a clinical setting.

Figure 6.8 Storyboard for GP Evidence in practice

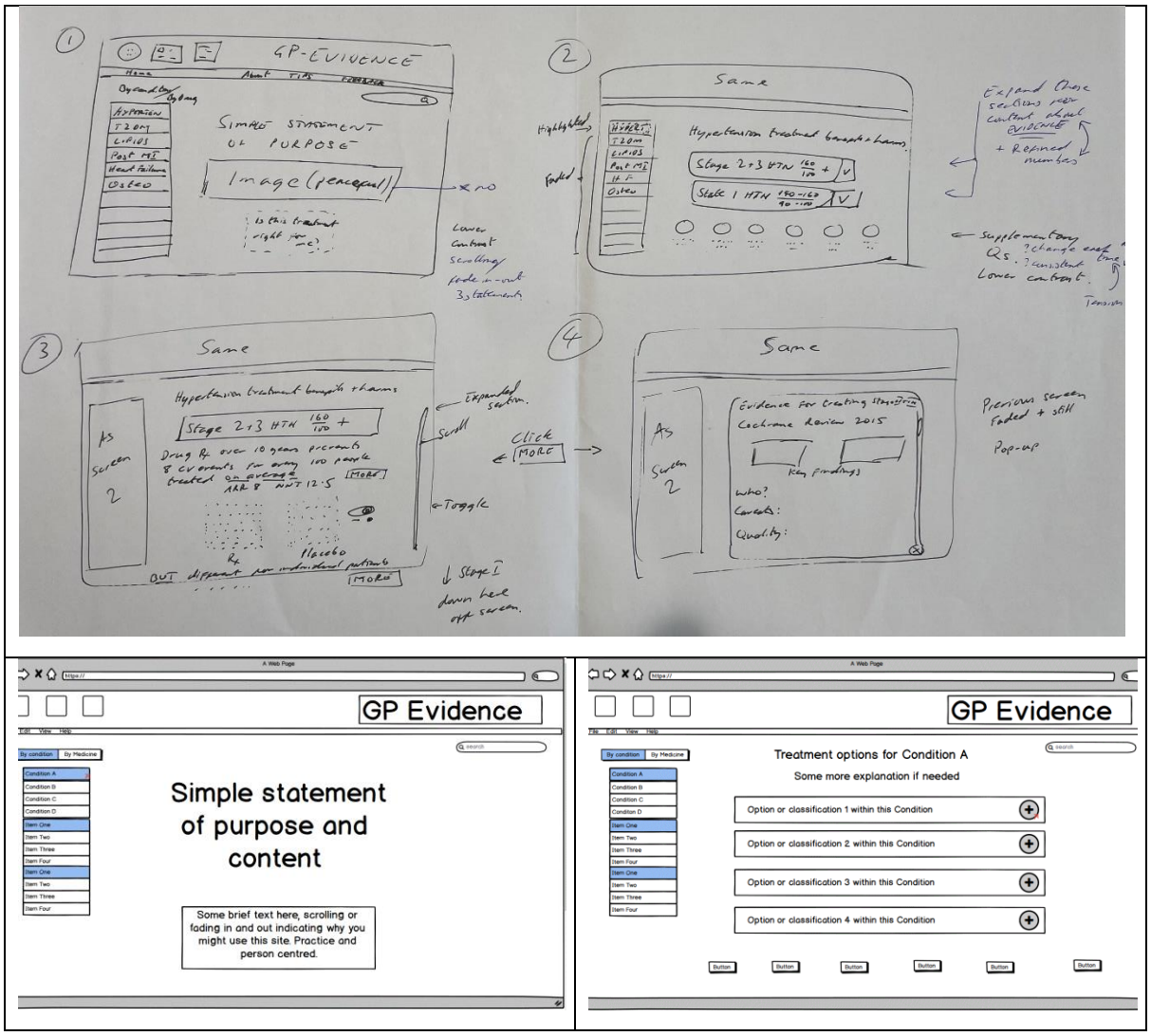


1. GP and patient discussing a number of issues. One of these is "Is this treatment right for me?"
2. GP has a knowledge gap, wants to know pros and cons. Is under time pressure.
3. GP knows there is a website which gives quick access to information about treatments for long-term conditions. Decides to have a look during the consultation.
4. GP quickly navigates to the desired information to inform the consultation and help reach a decision with the patient. They would spend 30 to 90 seconds getting high level information.
5. Later that day with no patient in the room, the GP has a chance to spend a few minutes reading some more content on the website, their interest having been stimulated earlier.
6. In the future, if the same question arises again, the GP may retain the knowledge and either not need to visit the website at all, or be able to access the information even faster a second time around.

6.1.2.2 Lo-fidelity prototype sketches

I created these at separate time points imagining how the website might look. They are not intended to be fully legible in this format, I will describe their key features and reasoning behind them below.

Figure 6.9 Lo-fidelity prototype sketch 1 with two digital mock-ups.



I produced the images in Figure 6.9 during the user-experience design course, using software provided at the time. Here, I was drawing on my research findings and design principles to imagine the overall structure of the site. Features were:

- Clear, spacious overall appearance with large font, contained menus and structure provided by expandable boxes across a page (termed “accordions”).
- Home page providing a simple statement of purpose, logos to show provenance and trustworthiness, menu of clinical conditions to make it obvious what the site is about.
- Clicking on a home-page menu item would lead to a “Condition” page, which would contain a list of treatment headings. These headings would be accordion bars, expanding to display treatment-specific information. This structure would situate information in a condition-treatment specific context without it having to be specified with extra text.
- On expanding a “Treatment” accordion, there would be a prominent infographic which would communicate the quantitative information on the benefits and harms of the treatment. This would be reached within two intuitive clicks from the home-page, providing rapid information as well as the interest-grabbing, content-affirming aspects of this design.
- The infographic would be surrounded with some (minimal) text, numbers or buttons where supplementary information may be found.

Figure 6.10 Lo-fidelity prototype sketch 2.

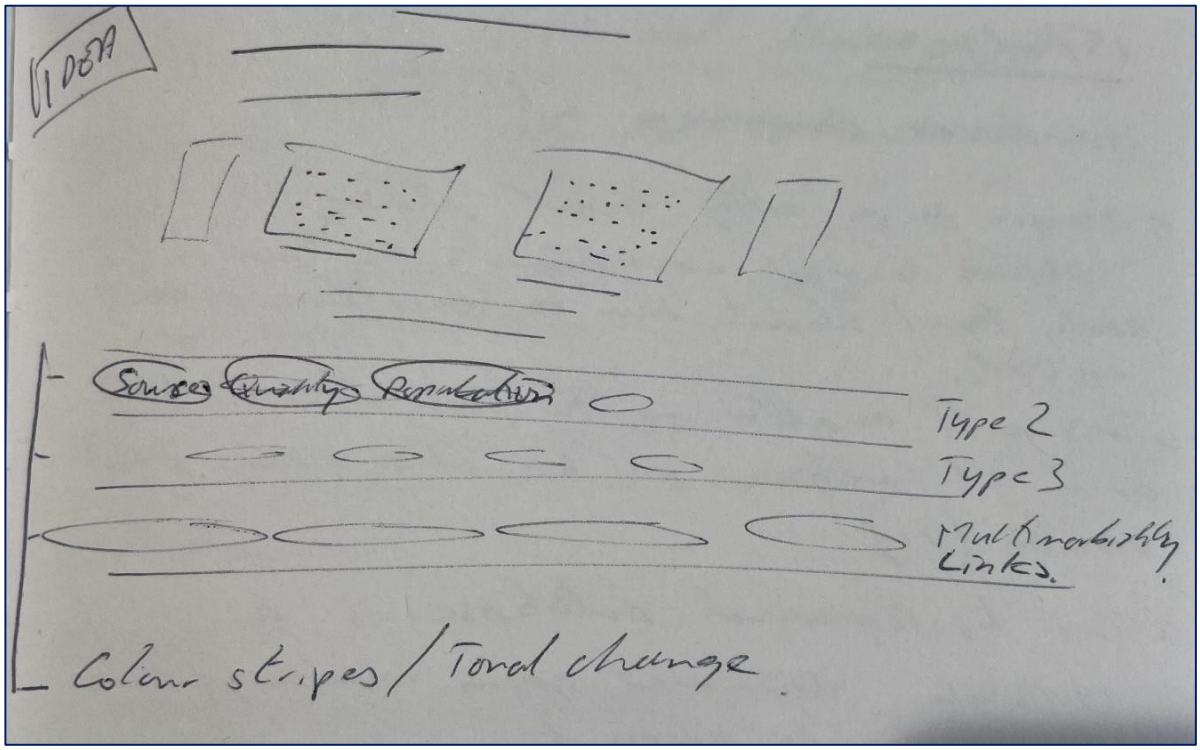


Figure 6.10 was produced a while later, at a content design course, where I gave more thought to the detail of how to present the quantitative data and its contextualising information. I judged that this information could be classified into four categories:

1) Numerical and textual versions of the quantitative data which support understanding of the infographic itself, bearing in mind the difficulty many GPs had described about understanding statistics. This might include presenting results as NNTs or ARR, or having a written explanation of the numbers.

2) Information about the clinical evidence itself which may be necessary to interpret it adequately, or apply it in practice. There would be three components of this, which would be the same for any piece of evidence:

- Evidence Source. Where had the data come from? What was the nature and size of the evidence base?
- Evidence Quality. To guide confidence in the quantitative data.
- Study Population. Information about the characteristics of the people who participated in the clinical trials, in order to consider the applicability of the results to an individual patient.

This choice of three components had arisen after consideration of the stakeholder work, GP interviews and an understanding of the principles of evidence-based practice.

3) Supplementary information which may be valuable to include on a context-specific basis. Examples I had in mind were:

- "Choice of drug" for hypertension, where the infographic data showed information for drug treatment in general, and users may be interested in drug class-specific effects.
- "Bisphosphonate holidays" to accompany information on this class of drug, where there is a specific issue about how long to stay on these drugs and some existing recommendations to have a period of time off them before restarting.

4) Links to information about co-morbidities, to support use of the website in situations of multimorbidity or polypharmacy. For example, a "Treatment" section for ACE-inhibitors on a "Condition" page for coronary heart disease could contain a link to the use of ACE-inhibitors on the "Condition" page for heart failure.

These four categories of information could be placed consistently throughout the site (following the usability heuristic of consistency), in expandable buttons or boxes with simple titles, allowing users to know that information was there to be accessed when needed, but without distracting form the core information. Categories could be separated or indicated by low-key colouring or shading, again minimising distraction from key content but being prominent enough to be noticeable.

6.1.3 Preparation for JAD workshop with professional web-design partners

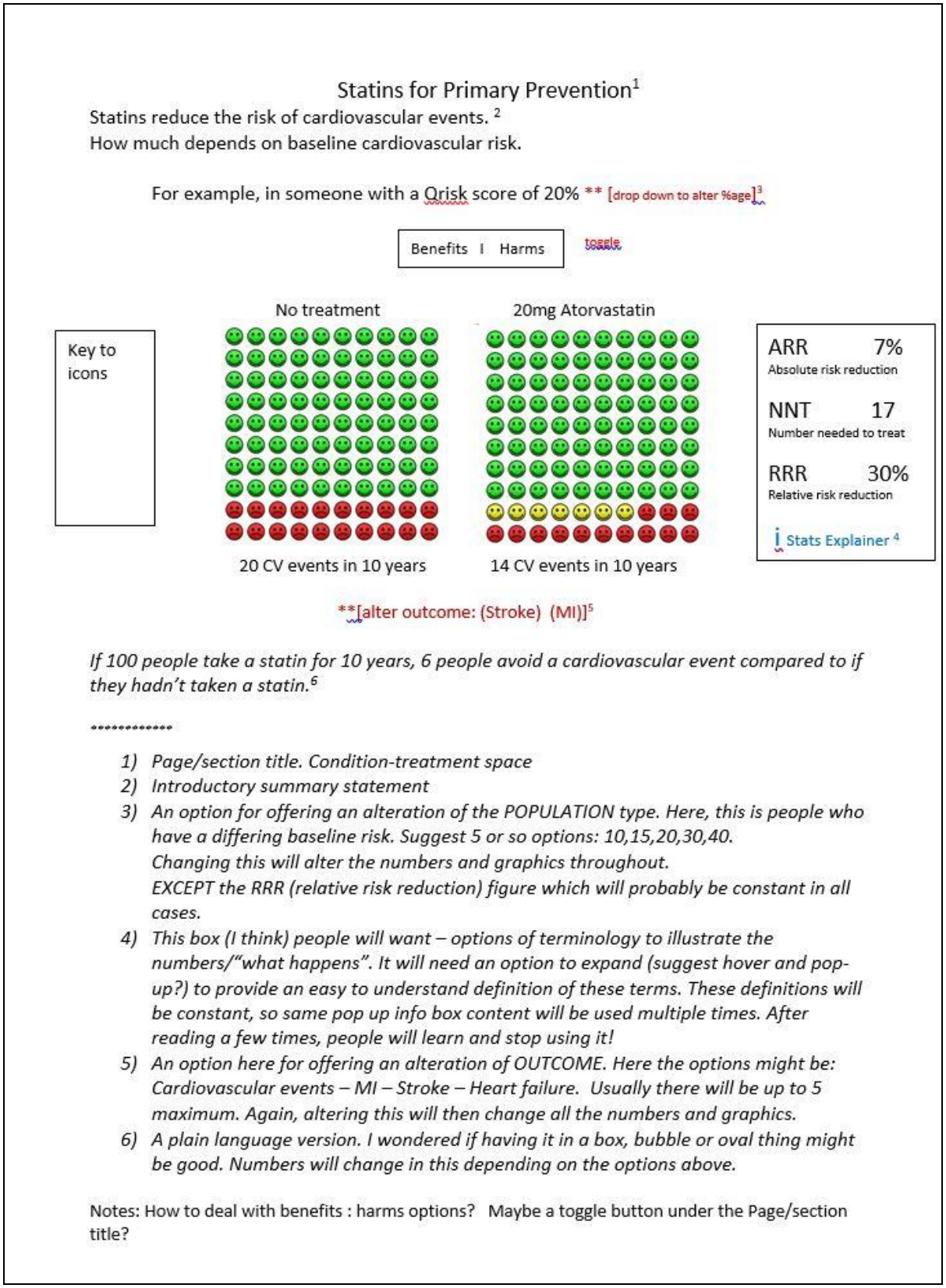
Over the course of the research process, I had been in discussion with my web-developers. We had had two meetings over the first year, where I had shared information on the GP Evidence project, providing updates from the interview study and considering how we should engage regarding design. They suggested a Joint Application Design (JAD) workshop, the methods for which I described in section 4.4.4. We planned a one-day meeting prior to this where I would provide an update on my research findings and propose design ideas for them to take away to develop an electronic prototype to use at the JAD.

For this meeting I synthesised my ideas into a Microsoft PowerPoint presentation which drove our discussion. This is copied in appendix 6.6. In addition I shared my prototype ideas as well as a further iteration of the infographic design (shown in figure 6.11) which drew on the literature on risk communication described in section 4.4.6. This final prototype included:

- five different modes of expression of the quantitative data, with the aim of maximising understanding whilst maintaining the central role of the infographic and not creating excessive visual clutter
- ideas for a drop down menu to alter baseline risk scores

- a toggle to switch between benefits and harms information
- a pop-up button to provide an explanation of the statistical abbreviations which were now included on the page

Figure 6.11 Additional prototype design with notes for Nexer Digital meeting

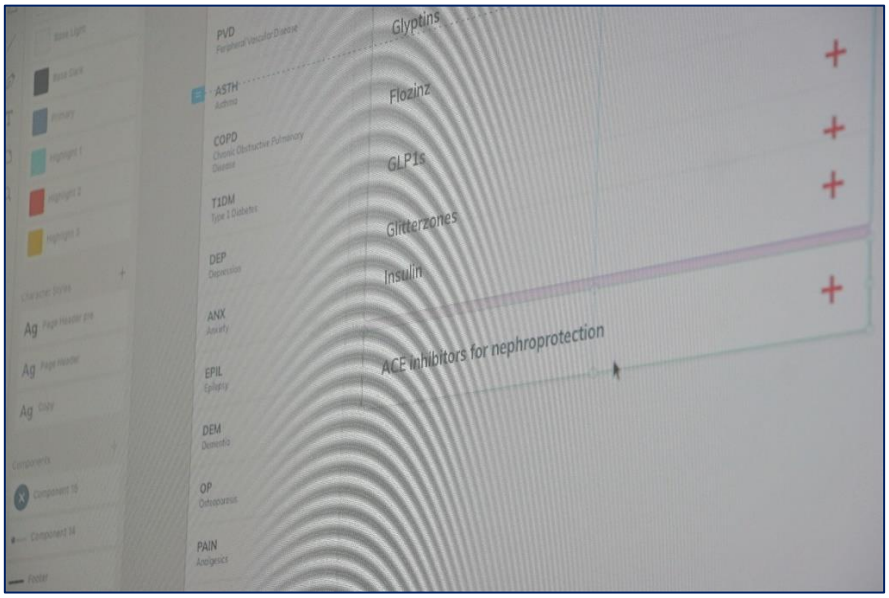


6.2 Joint Application Design workshop

The methods for the JAD were described in section 4.4.4. Here I will describe the results of the workshop using images of the resulting designs with notes summarising the discussion and decisions leading to them. The first two images show the process on the day:

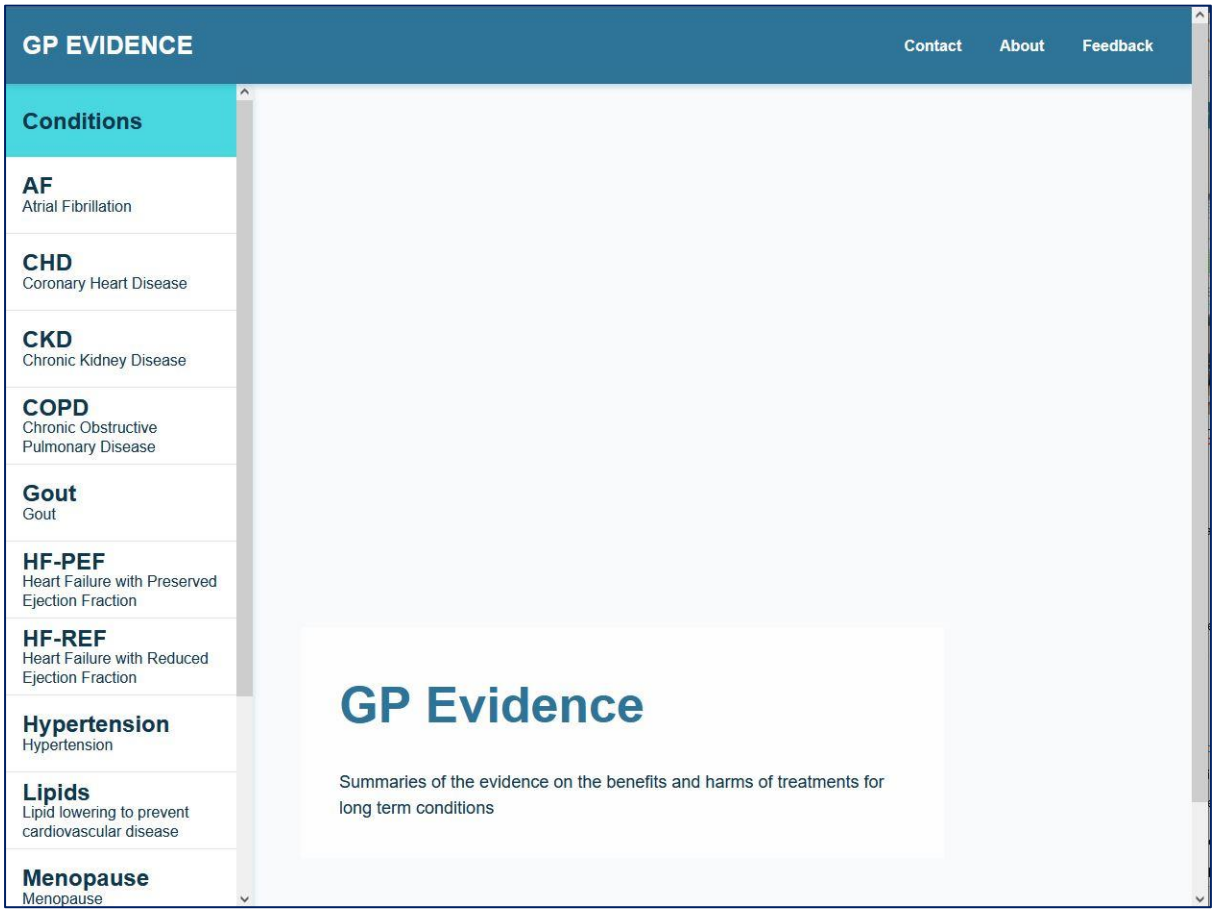
Figure 6.12 JAD participants sharing large screen for live rapid-prototyping.

Figure 6.13 Close up of large screen showing design software in use.



6.2.1 Home page

Figure 6.14 Initial design of GP Evidence home page



6.2.1.1 Design features and notes

6.2.1.1.1 Spacious, clear design, large font, minimal text, blue tones.

The starting point for the design looked much like this. Participants responded well to this simplicity and felt that the summary sentence combined with the list of long-term conditions would adequately communicate the purpose of the site.

The blue tones had been chosen by the web-developers as they provide a sense of calm, trustworthiness and in this context relate to the blues used by the NHS and NICE. Participants liked them and agreed with this overall effect.

6.2.1.2 Menu

6.2.1.2.1 Large-font abbreviations, small-font longer titles, spaciousness, left-hand placement with scroll bar

This menu layout had been chosen for its intrinsic simplicity and also that similar designs had appealed to the GP participants in the think-aloud study with other websites.

The role of abbreviations was discussed. GPs agreed that these were intuitive and helpful and allowed an at-a-glance understanding of website content. Importantly these were all abbreviations that all GPs should recognise, being in commonplace use for this group of people. Exceptions to this were HF-REF and HR-PEF, but this was overcome by the use of small-font full titles below the abbreviation. During the JAD, we were able to experiment with different font-sizes and the option of having full titles first or in larger font but settled for this as a final design.

There was discussion about the suitability of one long menu with a scroll bar. Would this be too effortful? Should the menu be categorised and collapsed? Given the number of long-term conditions planned for the website (only twelve at launch) it was decided that this design would work well, allowing space for large menu boxes and font-size whilst only requiring limited scrolling to see the entire menu. Had there been a longer menu, this decision would probably have been different as compromise between readability and findability would have been harder to achieve. We discussed the options of ordering the menu by clusters of long-term conditions (such as collating all cardiovascular content in a group) but decided after experimenting with re-ordering the menu that an alphabetical list was most intuitive. Again, this might have been different with a longer menu.

6.2.1.3 No search bar

The initial version of the home screen had contained a prominent search bar for users to enter free-text to find clinical content. GP participants felt this would be a good feature. However, when we began to discuss the possible terms users might enter into the search bar it transpired that this would be quite challenging. Users might try any drug name (brand or generic), any term or abbreviation for a long-term condition or any unpredictable words. Given that the website would contain information on a limited number of conditions, and that treatments would generally be dealt with by class (rather than individual drug) it would be highly likely that text searches would yield no results and frustrate users. The technical advice from the web-developers was that search bars can be programmed to deal with these kinds of issues, but that it is time-consuming and costly (beyond the budget of this project).

The idea of a search bar was therefore dropped, with some concern that it would disappoint some users. However, given the brief menu, it was felt users would be able to find content with relative ease.

6.2.2 Condition page

Figure 6.15 Condition page, Atrial Fibrillation.

6.2.2.1 Design features and notes

6.2.2.1.1 Spacious design, treatment options layered under condition page, brief introductory text

This straightforward design appealed to all participants and did not need much iteration. It was agreed that minimal text on this page was an advantage, so as not to distract from the treatments menu.

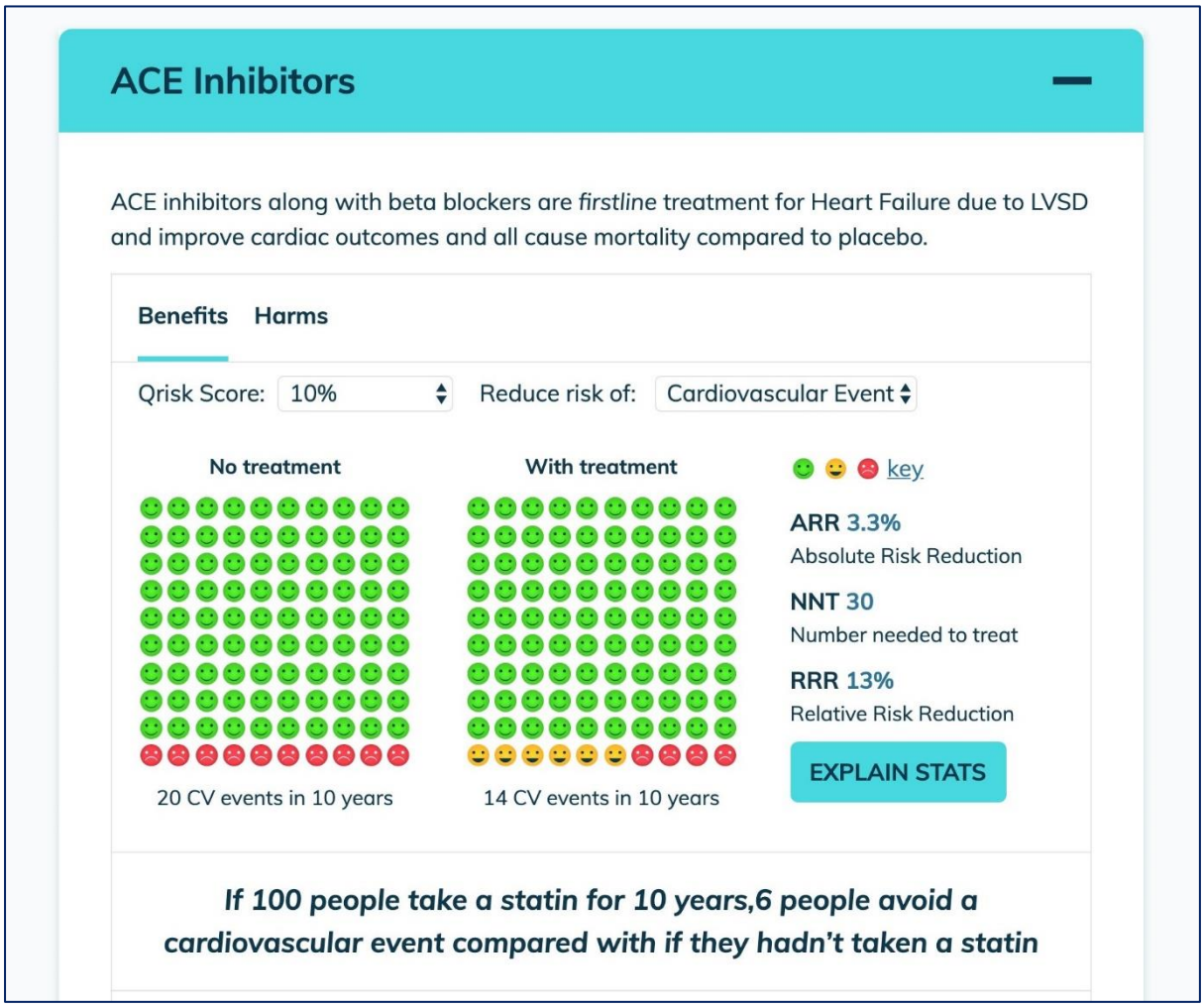
We discussed the ordering of treatment options. Should lifestyle and non-drug measures always be at the top? Should treatments be listed in order of effectiveness or by how commonly used they are? Advantages and disadvantages were seen with all these and we settled on a discretionary principle where the menus would aim to reflect GPs' mental models and practice, for example in the "Condition" section on Lipids, "Statins for primary prevention

of cardiovascular disease" would be at the top (being the most commonly used treatment), but statins would also feature elsewhere in "Condition" pages for long-term conditions where they had a specific role (such as coronary heart disease) but may not necessarily be at the top of the menu.

This discretionary principle was partly possible due to my role as user-as-designer, having good insight into how GPs think. These choices would also be subject to user-testing where menus could be rearranged if needed.

6.2.3 Infographic and quantitative information presentation

Figure 6.16 Infographic section design.



6.2.3.1 Design features and notes

6.2.3.1.1 Two 10 x 10 icon arrays, traffic light colours, smileys

We started with this formula which had been employed by NICE in their statin decision-aid and been so positively received by GPs in the think-aloud study in section 6.1 of this chapter. The web-developers applied their accessible-design expertise to adapt the facial expressions of the icons and ensure enough contrast between the colours to differentiate the icon types for those with colour vision deficiency.

6.2.3.1.2 Benefits and Harms tabs position

A number of arrangements of these tabs in relation to the infographic were tried. This design was chosen as it gave the two topics an equal degree of prominence. A default to show the benefits section first was chosen as it displayed the infographic and was felt to be more intuitive by the GPs.

6.2.3.1.3 Complementary data presentation

The positioning, font size and inclusion of ARR, NNT, RRR and plain language versions of data was experimented with to produce a balance which highlighted the most accessible formats (infographic and plain language) whilst making the numerical versions obvious but not too distracting. The format of large-font abbreviation and small-font full text was consistent with the menu structure and appealed to the GPs.

The large button for “Explain Stats” and smaller icons for “key” reflected a perceived need to highlight the availability of a pop-up with statistical explanatory information (shown in appendix 6.7).

6.2.3.1.4 Drop-down menus for baseline risk and clinical outcomes

The design of this feature revealed some tensions between usability and informational detail. Firstly, how to encourage or ensure that users selected an appropriate category for baseline risk or patient characteristic? Secondly, should users be forced to select a clinical outcome, or see a default such as a composite outcome? Too much forced selection might irritate, but some degree of selection may be essential for accurate clinical understanding and this would vary from topic to topic.

The group struggled with various versions of design until the user-experience (UX) expert had the idea to present a greyed-out infographic (see figure 6.20) wherever it was necessary to make a choice about baseline risk or patient characteristic. On selecting an option, the infographic would light-up with colour and show the data. This enforced the selection, but rewarded the user with the pleasant experience of a colourful image. The demonstration of the feature live at the JAD provoked an immediate, very positive response from all participants. This was a good example of the advantage of the rapid-prototyping method.

Regarding the drop-down menu for clinical outcomes, we decided to leave this as optional, thinking that users would by now have seen that there were drop-downs available and would intuitively use this option.

6.2.4 Supplementary buttons and links

Figure 6.17 Supplementary buttons below infographic.

6.2.4.1 Design features and notes

This arrangement of supplementary information, with categories described in section 6.1.2 was presented at the JAD but only subject to a short discussion due to limited time. The GPs

broadly understood and agreed with the layout which was taken forward to user-testing with an understanding that it might need further iteration.

The colour scheme was chosen after the JAD by the designer and myself. We aimed that the scheme should gently lead the users' eyes towards the content whilst not being too distracting from the primary focus of the infographic. Alternative, rejected schemes are shown in appendix 6.8.

6.2.5 Presentation of Harms information

Figure 6.18 Sample of website content presenting harms of statin treatment.

Myalgia and non-specific side effects

Muscle pains and general malaise are sometimes reported with statin use. Most of this (roughly 90%) is due to a nocebo* effect.

* an adverse effect experienced because the patient expects it, rather than as a result of the treatment itself

more

Myositis and rhabdomyolysis

2-22 excess cases per 10,000 person years for both combined[3].

more

Elevated liver enzymes

4 in every 1000 people will develop raised transaminases (ALT >3 x normal) due to taking a statin.

Statin-related raised transaminases up to 3 x the upper limit of normal are thought to be harmless and do not require statin therapy to be stopped if stable[1].

more

Serious liver disease

Acute hepatitis and liver failure associated with statins are so rare it is uncertain whether there is a causal effect.

more

New onset type 2 diabetes

1 excess case per 200 people over a 5-year period due to taking a statin.

However, the cardiovascular benefits of statins outweigh risks associated with glycaemic changes[5].

Statins do not need to be stopped in response to increases in blood glucose or HbA1c[1].

more

References

more

6.2.5.1 *Design features and notes*

The clinical data on treatment harms is generally more complex than for treatment benefits, with highly variable denominators and a variety of sources other than clinical trials and meta-analyses. Therefore, it does not lend itself to an infographic style of presentation as designed for treatment benefits.

For the JAD, the web-developers and I proposed the design shown in figure 6.18 with the following features:

- use of bold type for headings
- brief “take-home” messages using numbers or words as appropriate
- expandable content using a “more” option to provide detail

This aimed to provide a combination of easily accessible, understandable information in a flexible format.

Due to limited time, we were unable to iterate this substantially during the JAD, but the GPs found it acceptable and understandable, so it was taken forward to user-testing.

6.3 User-testing

Methods for user-testing with GPs were described in section 4.4.7. Here I will summarise the findings, collated from individual user test notes. Two examples of notes from testing sessions are presented in appendix 6.9.

6.3.1 Positive feedback on design

A majority of participants gave positive feedback or demonstrated satisfactory, low-effort use of the following features of the website:

- spacious, minimal design, large headings
- menu functions and navigation around Home, Condition and Treatment pages
- infographics were very well received
    - many positive comments and visible reactions
    - users demonstrated understanding of the quantitative information
    - suggestions that these could be shared with patients
- alternative data presentation formats (numerical and plain language) were noticed
    - plain language format received especially positive comments
    - "Explain Stats" button and content noticed by all and appreciated by some
- drop-down menus spotted by all without prompting and used mostly effectively
    - some exceptions to this, explained in next section
- presentation of "Harms" data was understood and acceptable to most
    - expandable "more" text feature spotted by all and felt to be valuable
    - some preferences expressed for alternatives, explained in next section
- supplementary buttons and links below the infographic were noticed and users read pop-up box content and successfully followed links

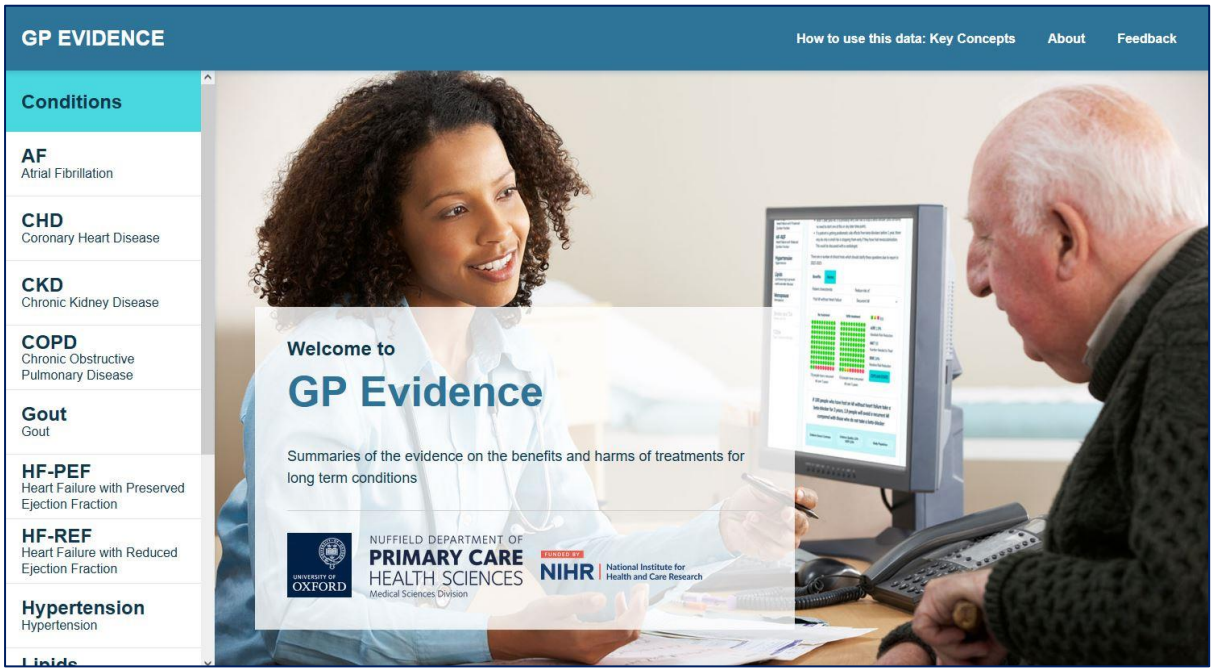
6.3.2 Design problems and improvements

6.3.2.1 Home page

Though users responded well to the spacious design shown in figure 6.14, a number of users commented that the site did look rather basic, like it was still in development and that they would not be sure whether to trust the content or not. Suggestions were given to add photos,

logos and more design content. This had been planned anyway, but these comments were taking into consideration when creating the final home page, shown in figure 6.19.

Figure 6.19 Final home page design

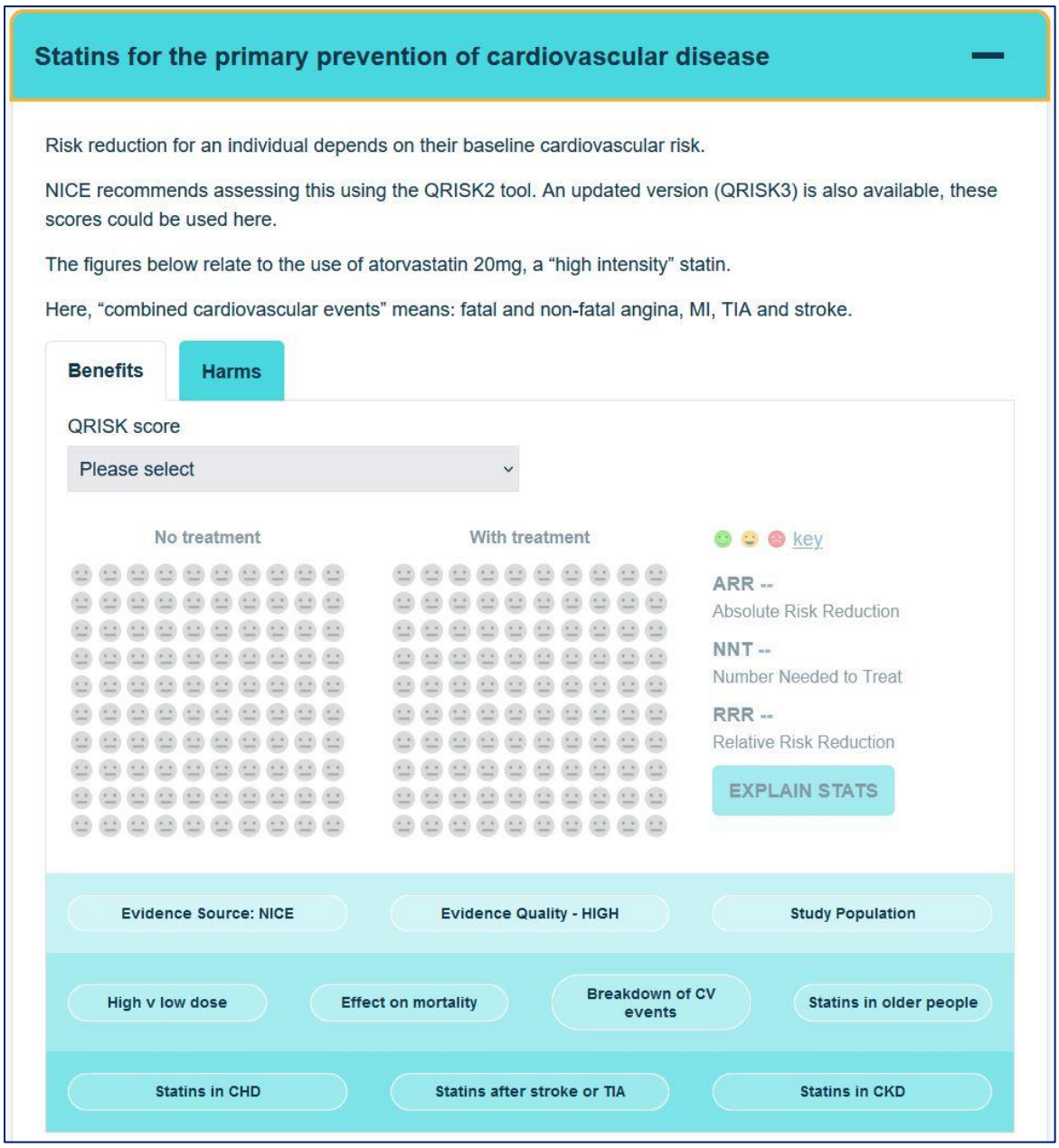


6.3.2.2 Harms tab going unnoticed

When viewing the design shown in figure 6.17, about half of the users failed to notice the "Harms" tab. Some would discover it alone, but after some active seeking, others would ask where the information about harms or side-effects was. This was an important problem, the aim was to present both benefits and harms with equal weight – a clear requirement to have emerged from stakeholder and PPI work.

The design was amended to produce a bolder "Harms" tab, shown in figure 6.20.

Figure 6.20 Amended design with bolder "Harms" tab. Image also demonstrates greyed out infographic.



6.3.2.3 Drop-down menus needing further development

Though users had managed to work the drop-down function to select baseline risk or clinical outcome, some took a while to get to grips with it, or commented that it wasn't all that obvious. Some discrete but impactful design changes were made, involving the addition of

shading and dynamic orange highlighted borders. This is best observed live on the website, but a summary is shown in appendix 6.10.

6.3.2.4 Some dissatisfaction with presentation of harms data

The format for presenting harms in figure 6.18 was understandable to users, but some commented that they would have liked to have had graphics for this area of information. One commented that it felt like to website was trying to "sell" benefits and underplay harms.

I explained the reasons for this design in the section below figure 6.18, but felt that this was important feedback, so put further consideration into how this might be addressed. The key challenge was how to deal with highly variable denominators: in figure 6.18 there are risks of harms of 1 in 200, 4 in 1000 and 2 in 10,000. What graphic could have this range yet show sufficient granularity to be intuitive, understandable and clear?

I sought advice from the Winton Centre for Risk and Evidence Communication[11] who were able to provide a "clinic" discussion with me and three members of their team. They found this communication problem challenging and were not aware of any research which showed an ideal way to graphically represent this range of numbers. However, we were able to draft an idea featuring a 2-part image: one part with a low-resolution dot-graphic showing large numbers featuring a magnifying glass logo in one corner which would blow up into a second part showing high resolution icons (matching those on the "Benefits" infographic) to provide granularity and consistency. Such an image could be tested with users.

[11] Based at Cambridge University, this Professorial unit/project ran from 2017-2022, conducting research and developing resources to support the communication of risk and evidence information to the public.

I discussed how this could be built into the website with my web-developers. It would have been possible, but involved a significant extra expense for which there was no budget, and in addition I was unsure I had time capacity to user-test and iterate it adequately. Therefore this development was not taken further at this stage but noted a future possibility.

6.3.2.5 Other suggestions

A number of users mentioned they would like a search bar function. However, they all successfully navigated using the menu function and did not report that this was too difficult. Some users asked about functionality on a mobile phone or an app. The use-scenario for the design was on a desktop computer in a consulting room and there was no budget for mobile or app development, but this suggestion was noted for possible future development. A few suggestions for patient-facing information or print-outs were made. These were not within the scope of this project but were also noted for possible future development.

6.3.2.6 Successful repeat user-testing of design changes

The three amended design features (home page, harms tab and drop-down menus) were tested by observation of use of the site during the pair-writing sessions in the latter phase of the project as well as the evaluation study described in chapter 7. The new home page received positive feedback, and 29 users in all were observed easily interacting with the harms tab and drop-down menus with no complaints expressed.

6.4 Content design

The principles and methods of content design and "pair writing" with GP users and a professional editor were described in section 4.4.5. Here, I show some results of that process with samples of content from *GP Evidence*.

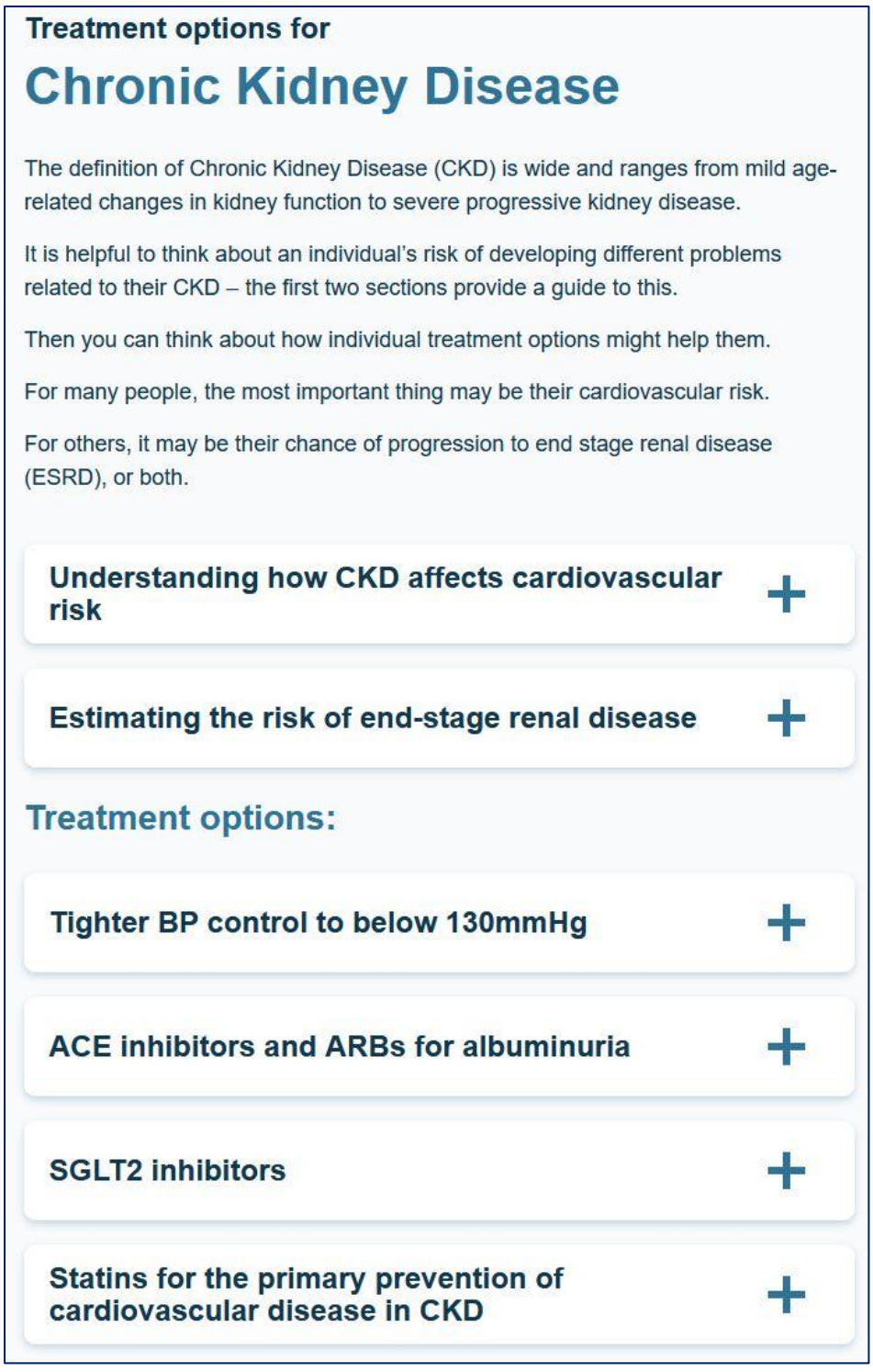
6.4.1 Part 1: examples of finished content

Examples of content after pair writing with notes on content design features:

Figure 6.21 Content sample from GP Evidence. Condition: Gout.

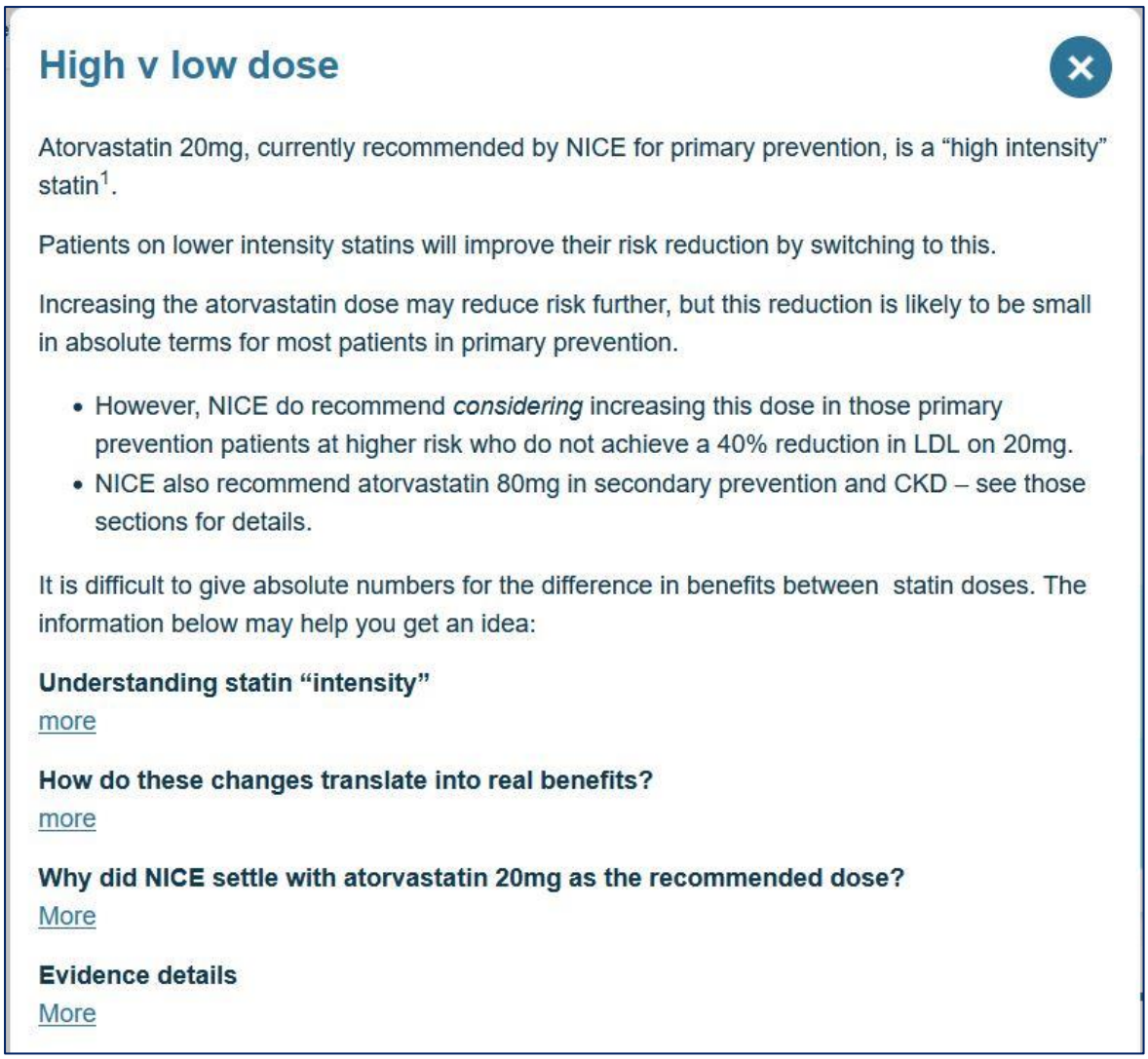
- brevity, space
- plain language introduction
- bold text to highlight key number
- smaller font for subsidiary message
- bullet points to highlight key messages

Figure 6.22 Content sample from GP Evidence. Condition: CKD homepage



- minimal but necessary introductory text to set up expectation of content
- clear division of topics into background and treatment sections
- ordering of content matches GP mental models

Figure 6.23 Content sample from GP Evidence: Statins for the primary prevention of cardiovascular disease



High v low dose

Atorvastatin 20mg, currently recommended by NICE for primary prevention, is a "high intensity" statin[1].

Patients on lower intensity statins will improve their risk reduction by switching to this.

Increasing the atorvastatin dose may reduce risk further, but this reduction is likely to be small in absolute terms for most patients in primary prevention.

- However, NICE do recommend *considering* increasing this dose in those primary prevention patients at higher risk who do not achieve a 40% reduction in LDL on 20mg.
- NICE also recommend atorvastatin 80mg in secondary prevention and CKD – see those sections for details.

It is difficult to give absolute numbers for the difference in benefits between statin doses. The information below may help you get an idea:

Understanding statin "intensity"
more

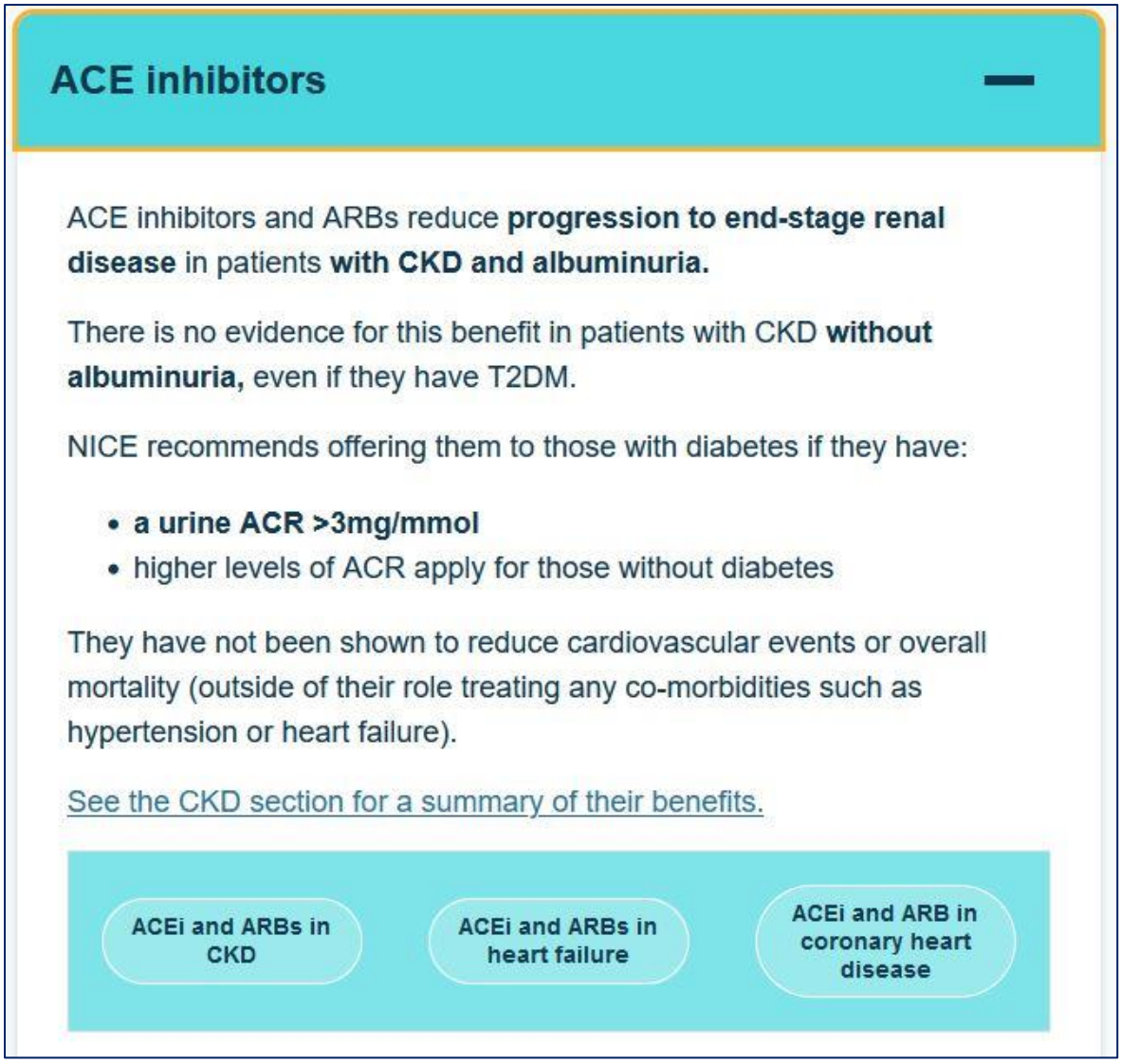
How do these changes translate into real benefits?
more

Why did NICE settle with atorvastatin 20mg as the recommended dose?
More

Evidence details
More

- plain language introduction with broad take-home message on second line
- plain language in 4th paragraph to introduce the idea of complex evidence
- bullet points for details
- use of headings to "tell a story"
- detailed information hidden in expandable content in order to not overwhelm the reader at first glance

Figure 6.24 Content sample from GP Evidence. Condition: Type 2 diabetes



- use of bold text to highlight key details about albuminuria
- links to further details (rather than copied content) in order to isolate these key messages relating to the role of these drugs in type 2 diabetes

6.4.2 Part 2: before and after pair-writing

Five examples over the next few pages show written content before and after pair-writing, highlighting the kind of changes this method produced. Before-and-after screengrabs of *GP Evidence* are shown, with notes on the changes made in-between:

Figure 6.25 Before-and-after pair writing: 1

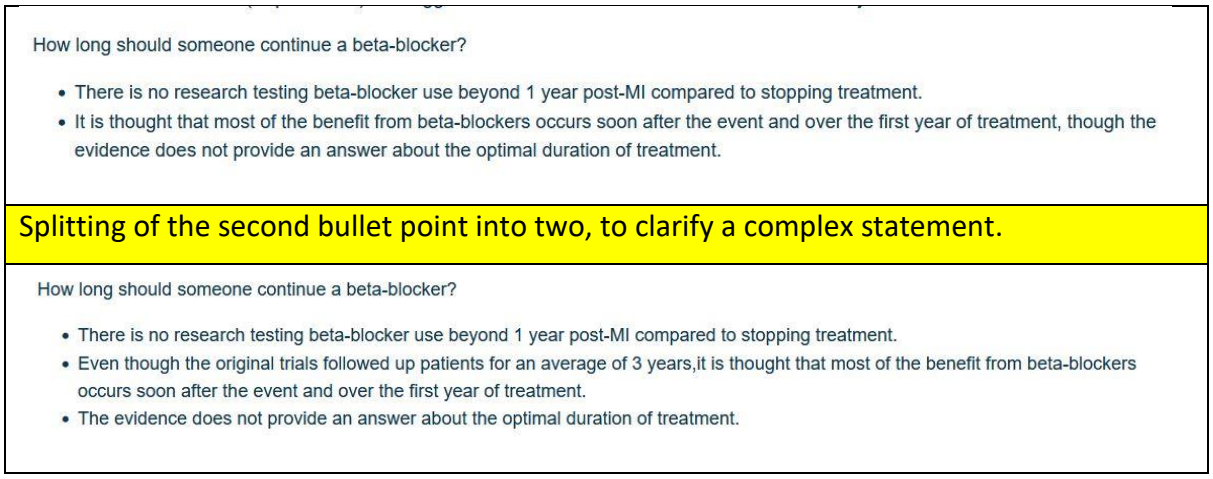


Figure 6.26 Before-and-after pair writing: 2

Rates of specific side-effects from BP lowering treatment over a 3 year period:

Outcome	Placebo or less intense treatment	Treatment or more intense treatment	Absolute Risk Increase	Number Needed to Harm	Relative Risk Increase*
All drug classes combined					
Hypotension	3.3%	6.1%	2.7%	36	1.97
Syncope	1.1%	1.3%	0.2%	500	1.28
Acute Kidney Injury	1.5%	2.1%	0.6%	171	1.18
Falls	No increased risk found				
ACE inhibitors and Angiotensin Receptor Blockers					
Hyperkalaemia	3.0%	4.8%	1.8%	54	2.03
Acute Kidney Injury	1.1%	1.5%	0.4%	250	1.26
Thiazide diuretics					
Hypokalaemia	0.8%	8.2%	7.4%	13	10.7[a]
Gout	0.2%	2.0%	1.7%	58	4.5[a]

Removal of all Relative Risk Increase columns from harms tables. These were not felt to add value and the tables were easier to read without it.

Rates of specific side-effects from BP lowering treatment over a 3 year period:

Outcome	Placebo or less intense treatment	Treatment or more intense treatment	Absolute Risk Increase	Number Needed to Harm
All drug classes combined				
Hypotension	3.3%	6.1%	2.7%	36
Syncope	1.1%	1.3%	0.2%	500
Acute Kidney Injury	1.5%	2.1%	0.6%	171
Falls	No increased risk found			
ACE inhibitors and Angiotensin Receptor Blockers				
Hyperkalaemia	3.0%	4.8%	1.8%	54
Acute Kidney Injury	1.1%	1.5%	0.4%	250
Thiazide diuretics				
Hypokalaemia	0.8%	8.2%	7.4%	13
Gout*	0.2%	2.0%	1.7%	58

Figure 6.27 Before-and-after pair writing: 3

NICE recommends **aspirin** as indefinite therapy following Acute Coronary Syndrome[1]

- data on aspirin v placebo is shown in the charts below

Dual antiplatelet therapy is recommended for 1 year using a P2Y12 inhibitor

- **ticagrelor or prasugrel** are now recommended as first choice over clopidgrel
- **clopidogrel** was the first P2Y12 inhibitor to show benefit in addition to aspirin – **this data is presented in charts below**
- **ticagrelor or prasugrel offer (very approximately) an additional 1-3% Absolute Risk Reduction** in re-infarction with no extra bleeding risk compared to clopidogrel[1]

Splitting text into two sections plus use of bold type to clarify the different statements in indefinite v temporary treatment.
Reordering the lower three bullet points to unite statements about ticagrelor and prasugrel. In addition, this reordering improved understanding of the graphics immediately below which referred to clopidogrel.

NICE recommends **aspirin** as **indefinite therapy** following Acute Coronary Syndrome[1]

- data on aspirin v placebo is shown in the charts below

Dual antiplatelet therapy

A P2Y12 inhibitor, in addition to aspirin, is recommended for 1 year post-MI.

- **Ticagrelor or prasugrel** are now recommended as first choice over clopidogrel.
- **Ticagrelor or prasugrel offer (very approximately) an additional 1-3% Absolute Risk Reduction** in re-infarction with no extra bleeding risk compared to clopidogrel[1].
- **Clopidogrel** was the first P2Y12 inhibitor to show benefit in addition to aspirin – **this data is presented in charts below** as the data for ticagrelor and prasugrel are too complex.

Figure 6.28 Before-and-after pair writing: 4

The chance of someone with AF having an ischaemic stroke,TIA or thrombus in 1 year can be estimated using the CHA_2DS_2-VASc score.

NICE recommends offering anticoagulation to those with a CHA_2DS_2-VASc score of 2 or more (and to consider it for men with a score of 1 or more).

Treatment with oral anticoagulants reduces stroke risk by approximately two-thirds.

Data details

The data below comes from trials of Warfarin v. Placebo.

Even though DOACs are now recommended as first line over warfarin (they may be marginally better), these figures are representative enough to apply to all oral anticoagulants for practical purposes.

Front-loading of the introductory sentence to highlight the subject: the CHA2DS2-VASc score.
Shorten final sentence for clarity and to front-load with the subject: oral anticoagulants
Extra line space before “Data detail” to reduce the effort needed to focus on the main content.

Use the CHA2DS2-VASc score to estimate the risk of someone with AF having an ischaemic stroke, TIA or thrombus in 1 year.

NICE recommends anticoagulation for those with a CHA2DS2-VASc score of 2 or more (and consider it for men with a score of 1 or more).

Oral anticoagulants reduce stroke risk by approximately two-thirds.

Data detail

The data below comes from trials of warfarin v placebo.

These figures are representative enough to apply to all oral anticoagulants for practical purposes.

Figure 6.29 Before-and-after pair writing: 5

Anticoagulants can cause haemorrhagic strokes and other intracranial bleeds, but the risk is mall and far outweighed by the benefit of preventing ischaemic strokes

Data details

- Intra-cranial haemorrage (ICH) rates in patients taking DOACs in clinical trials ranged from[1,2]
  - **0.2% to 0.6% per year**
- ICH rates for patients taking warfarin in the same trials were at least twice as high[3]

These trials lack a placebo control group, but for comparison,

- ICH rates for people with AF not taking anticoagulants in observational studies ranged from[2]
  - **0.1% to 0.8% per year**

It is still uncertain/unknown what the exact excess risk of ICH with DAOCs is, but these figures suggest it is likely to be a fraction of a percentage only per year.

Front-loading and shortening of first sentence. Designed to provide quick "take home message".
Second statement to function as an easy to understand summary of slightly tricky data below.
Use of abbreviation "ICH" to simplify appearance and create space.

The risk of intracranial haemorrhage (ICH) with anticoagulants is small and outweighed by the benefit of prevention of ischaemic strokes.

The exact excess risk is uncertain, but is likely to be a fraction of a percentage per year.

Data detail

- ICH rates in patients taking DOACs in clinical trials ranged from[1,2]
  - **0.2% to 0.6% per year**
- ICH rates in patients taking warfarin in the same trials were at least **twice as high**[3]

These trials lack a no-anticoagulation control group, but for comparison:

- ICH rates for people with AF not taking anticoagulants in observational studies[2] ranged from:
  - **0.1% to 0.8% per year**

7 Chapter 7: Results part 2 – selection, collation and presentation of clinical evidence

The approach and methods for the selection and collation of clinical evidence for GP Evidence were described in chapter 5. Here, I will illustrate the results of this approach using a range of examples from the website which illustrate a variety of issues which arose during this process. To aid understanding, in appendix 6.4 I have presented some copies of figures and tables from relevant publications to illustrate the original data sources.

It is not possible to describe the data collation process for every piece of content in *GP Evidence* in this thesis. However, every component of clinical content on the website is referenced on the site itself. In addition, I have records of the data collation process for all website content stored in Microsoft Word documents on a secure University server; further information about any particular piece of content can be provided.

7.1 Three examples of straightforward evidence selection

7.1.1 Statins for the primary prevention of cardiovascular disease

NICE had produced a patient decision-aid in 2014 to illustrate the effect of statins on the risk of cardiovascular events (299). Images from this are shown in appendix 6.3. It was accompanied by a supporting document[12] which explained how the relative risk reduction in total cardiovascular events used in the decision aid was calculated. NICE had taken risk ratios for separate cardiovascular outcomes from their systematic review of high dose statins compared to placebo (302) and applied a formula adjusting for the rates of individual

[12] This document is now no longer available on the NICE website, the decision aid having been updated in 2023. This description of their data is from my personal notes taken in 2021.

outcomes as a proportion of total cardiovascular events. This gave a relative risk reduction of 37% which I was able to apply to a range of baseline QRISK cardiovascular risk scores to generate the infographic data for *GP Evidence*.

NICE had rated this evidence as high quality and provided data on trial population sizes (302) which I was able to apply directly to GP Evidence. However, NICE provided no summary data on patient characteristics, so I used data from a 2013 Cochrane review on statins for the primary prevention of cardiovascular disease (303) to produce these figures, presented with a statement of explanation.

7.1.2 ACE-inhibitors and angiotensin receptor blockers (ARBs) for albuminuria in chronic kidney disease

NICE had conducted an evidence review (304) for a 2021 update of the Chronic Kidney Disease guideline (305). This contained meta-analyses of trials of ACE-inhibitors and ARBs which provided data on the absolute and relative risk reductions for end-stage renal disease. Accompanying GRADE tables provided evidence quality ratings. I was able to apply this information directly to the GP Evidence content. The NICE review did not summarise data on study population, therefore I went to the original studies to explore this data. In this case, this was feasible as there were only five trials and I was able to produce a descriptive summary of these heterogeneous populations for *GP Evidence*. Screengrabs of figures and tables from this NICE evidence review are in appendix 7.1 to illustrate the data source.

One issue which arose was that one of the trials (Ruggeneti 1999 (306)) had a 6-year duration, compared to the approximately 3-year duration of the other 4 trials. On *GP Evidence*, I had presented data with a 3-year time frame. Inclusion of a 6-year trial in the meta-analysis may

have rendered this incorrect as a longer trial would be more likely to show a greater absolute risk reduction. I considered separating out the data into individual trials, however, on closer inspection of the data, the 6-year trial had only contributed a few events to the total in the meta-analysis so i made a decision to leave the pooled meta-analysis data presented as 3-year outcomes. This was an example of making a pragmatic choice in the interests of presenting simple, understandable information. It was scientifically imperfect, but not of a magnitude which would be likely to mislead users of *GP Evidence* in a clinically important way.

7.1.3 Exercise-based cardiac rehabilitation for the secondary prevention of heart disease

NICE had merged and updated various guidelines on the subject into a combined guideline: Acute Coronary Syndromes in 2020 (307) and retained an old recommendation for exercise-based cardiac rehabilitation. However, no recent evidence review was available on the NICE website, therefore I used a 2021 Cochrane review on the same topic (308) which provided quantitative data, an evidence quality appraisal and enough information on the study populations to create a narrative description for GP Evidence. Screengrabs of the data tables from the Cochrane review are in appendix 7.2 to illustrate the data source.

The Cochrane review introduction provided a review of safety studies for this intervention including a large prospective cohort study (309) which provided data on the risk of adverse cardiac events during these rehabilitation programs. I performed a CASP quality appraisal (310) of this study, rated it as moderate quality and used their data for the harms section on *GP Evidence*. A copy of the completed CASP appraisal form is in appendix 7.3.

7.2 Two examples of using single trials to provide data for GP Evidence.

7.2.1 ACE inhibitors and ARBs for the secondary prevention of coronary heart disease

The most recent NICE evidence review for this intervention was in 2013 and included just two trials for patients without left-ventricular impairment (311). It provided quantitative data, a quality appraisal and a meta-analysis. However, meta-analysis only provided data on the outcome of cardiac mortality. In order to be able to present data on individual outcomes, I chose to use one of these trials (EUROPA) as it was the most applicable to a UK primary care population: the trial was conducted in Europe, had a high proportion of patients with a previous myocardial infarction and more patients taking preventive drugs which form part of current standard care (312).

7.2.2 SGLT2 inhibitors for type 2 diabetes

NICE conducted an updated evidence review for this intervention in 2022 which included four trials in a network meta-analysis (313). This provided a systematic search of clinical trials and an evidence quality rating, but no quantitative data which could be applied to the GP Evidence format. Two other recent high-quality systematic reviews and meta-analyses addressed the same question but neither of these provided usable data either (314, 315).

Therefore I selected a single trial from the NICE review (DECLARE-TIMI (316)) as it was only one of two trials which separated data by populations with and without existing cardiovascular disease – a distinction made in NICE guideline recommendations – and of those two it was the largest and of longer duration.

This trial provided usable data, although two issues arose:

1) The benefits on the outcomes of myocardial infarction and total mortality did not reach statistical significance in DECLARE-TIMI. However, they did in meta-analysis, so I used the trial data for the infographics and included an explanatory note in the "Evidence Source" section.

2) DECLARE-TIMI showed a lower benefit for total mortality than other trials. I was concerned this might under-represent the benefits of this class of drug, so added a button underneath the infographic "Mortality reduction with SGLT2i treatment" in which I added further information.

Screengrabs of the GP Evidence website showing these two explanatory pieces of content are shown in appendix 7.4.

In both of these examples, one potential criticism is that by selecting single trials, I have used lower-level evidence than their respective meta-analyses provide, diminishing the statistical power of the findings. However, my response is that the meta-analyses have been used to generate the guideline recommendations which themselves are undiminished. In the first example for ACE-inhibitors and ARBs, the absolute effect size of the single trial is broadly equivalent to the meta-analysis in any case. In the second, I overcame problems of statistical significance and effect size by adding transparent explanatory content.

7.3 Examples of combining data sources to provide useful information

7.3.1 Bleeding risk with anticoagulation

The 2021 NICE guideline for atrial fibrillation (AF) recommends the use of the ORBIT score to estimate the risk of major bleeding conferred by anticoagulant drugs prescribed for stroke prevention in this condition (317). ORBIT provides bleeding risk estimates per year of

treatment, stratified by patient characteristics (318). These estimates were straightforward to adapt to the *GP Evidence* infographic.

However, the ORBIT score is derived from an observational cohort study without a control group who were not taking an anticoagulant. Therefore there was no readily available data to present on bleeding risk with no treatment. An exploration of the literature on anticoagulants and AF (including other scoring systems) revealed two sets of data which provided useful, if imperfect information:

- bleeding rates in the control groups in clinical trials of warfarin vs placebo included in NICE's evidence review (319)
- a large cohort study from a New Zealand primary care database (320)

Both of these raised questions of applicability. The warfarin trials were old and may have contained participants with different characteristics to today's population with AF. The cohort study was conducted in another country and not in a selected population with AF. However, they both gave useful information which could help contextualise the bleeding risk generated by an ORBIT score, so I presented their data in a transparent way, highlighting their indirectness and emphasising that this information was to help understanding, rather than to be regarded as a direct comparator group. Screengrabs of the GP Evidence pages demonstrating the presentation of this data is in appendix 7.5.

7.3.2 Antiplatelets in coronary heart disease

NICE recommendations for antiplatelet therapy following an acute coronary syndrome are complex: indefinite aspirin plus a second antiplatelet for a year (ticagrelor or prasugrel) (307). Another antiplatelet, clopidogrel has also been in widespread recent use for this indication.

The most recent NICE evidence review on this topic was in 2020, a network meta-analysis comparing all these treatments (321). It found overall superiority for ticagrelor and prasugrel but did not provide suitable data on the benefits and harms of these compared to no treatment or aspirin alone. Previous NICE evidence reviews were not available online, so I used a 2017 Cochrane review for data on clopidogrel (322), extracting a single most applicable trial (323) and a 2002 Antithrombotic Treatment Trialists Collaboration meta-analysis for data on aspirin compared to placebo (324).

This combination allowed presentation of evidence showing incremental benefits of newer treatments over time. A screengrab showing how this was presented is in appendix 7.6. This allows readers to estimate the absolute benefits of the newer drugs. I contemplated producing some kind of summary estimate of, say, ticagrelor compared to placebo but judged that this would scientifically too inaccurate – a view which was supported by the steering committee when I presented this scenario to them.

7.4 Examples of managing uncertain evidence

7.4.1 Statins for the primary prevention of cardiovascular disease in older people

The value of statins for those in older age groups (variously described as ranging from the over-70s to over 80s) has long been an area of uncertainty and debate due to a lack of direct RCT evidence (325). Because this has been an important question for GPs (326), I included a section on *GP Evidence* within an "extra information" button. To present some illustrative evidence, I referred to the committee discussion section of the NICE guideline evidence review (327). A recommendation had been made to consider statin treatment for those over 85, with a specification that this might reduce the risk of non-fatal myocardial infarction, a

finding from a single trial in their evidence review which had included patients in this age group. I used data from this trial and also presented data from a high-profile meta-analysis from the Cholesterol Treatment Trialists Collaboration published after the NICE evidence review which suggested further benefits (328).

This information was presented in a way which reflected the ongoing uncertainty, offered as a suggestion to inform decisions. A screengrab is shown in appendix 7.7. This approach enabled the communication of valuable information from the evidence-base without presenting false certainty.

7.4.2 Drug treatment of stage 1 hypertension

This was the most challenging evidence-related decision of the project, so is described in some detail.

The evidence for the drug treatment of stage 1 (mild) hypertension in those who do not have established cardiovascular disease or type 2 diabetes has been a subject of debate over the last decade (329, 330) since a 2012 Cochrane review found no evidence of benefit and a paucity of clinical trials which could answer the question (282). A number of systematic reviews and meta-analyses over the following years had attempted to provide an answer by including trials including patients with cardiovascular disease and other co-morbidities such as chronic kidney disease and type 2 diabetes (331, 332). A large meta-analysis of 123 studies stratified treatment effect by a standard blood pressure drop of 5mmHg and used a meta-regression analysis (283). Another included 74 trials, excluding those at high risk of bias and analysing results by blood-pressure threshold whilst including a small proportion of patients with chronic kidney disease, established cardiovascular disease and many with type 2

diabetes (284). All reached differing estimates of relative risk reduction for cardiovascular events ranging from 12-22%.

This certainty of this evidence base sits in contrast to that of drug treatment of stage 2 and 3 hypertension which has much more direct evidence. A 2019 Cochrane review included 16 trials and provided moderate quality evidence of a 28% relative risk reduction in combined cardiovascular events in this population (333). This data was applied in GP Evidence in a straightforward manner.

I had hoped that the question of what evidence to use for stage 1 hypertension would be answered by a 2019 update to the NICE hypertension guideline (334). However, their evidence review (335) failed to reach a definitive answer about what amount of risk reduction is conferred by drug treatment in this group despite including two of the systematic reviews mentioned above (284, 331). The guideline recommendation to treat stage 1 hypertension in those who had a baseline 10-year cardiovascular risk of 10% or more was based on a committee decision which acknowledged that the evidence showed some benefit but was uncertain (335)^{p24}. However, for their cost-effectiveness analysis, NICE had used the 2017 meta-analysis by Brundstrom and Carlberg (284) which found a 12% relative risk reduction in cardiovascular events.

I therefore chose this analysis to present data for GP Evidence, aiming to align content as close as possible to the NICE guideline, but added an explanatory section entitled "Uncertainties and Controversies" outlining this evidence for transparency and to allow users to reflect on the evidence themselves.

Then in 2021, two large individual patient data (IPD) meta-analyses were published by the Blood Pressure Lowering Treatment Trialists' Collaboration which showed relative risk reductions of up to 24% for a sub-group aged under 55 with stage 1 hypertension without pre-existing cardiovascular disease (336, 337). Other subgroups showed smaller benefits, but these studies provide arguably stronger evidence than previous meta-analyses, though this is hard to judge given the “black-box” nature of analysis in IPD studies.

This publication presented a dilemma for *GP Evidence*. If the evidence from these IPD meta-analyses shows that the drug treatment for stage 1 hypertension is as effective as when used with higher levels of baseline blood pressure, then the presentation of data as described above would significantly under-estimate treatment benefits. On the other hand, these studies had not been reviewed by NICE, and would I risk communicating evidence which might later be found to be inadequate at a future NICE review?

I took this problem to the steering committee for advice. Expert members agreed that it was a challenging evidential problem and cautioned against independent inclusion of evidence which had not been reviewed by NICE for such a major topic. Patient members expressed a preference for transparency of uncertainty.

The agreed solution was to retain a separation of evidence for Stage 1 and Stage 2+3 hypertension but place a prominent alert highlighting the level of uncertainty to users and adding further explanation. The final version of the *GP Evidence* content is shown in appendix 7.8 and contains all the elements of the issues described here.

7.5 Heart failure with reduced ejection fraction: adding context to evidence

Heart failure with reduced ejection fraction (HF-REF) is a serious, common condition usually managed between primary and secondary care. Untreated, it has a poor prognosis but there is a large range of drug treatments available which improve this. In recent years, there has been a drive to maximise the use of these treatments, with GPs playing a key role (338, 339).

GP Evidence presents data for 7 drug classes available to treat this condition, which was straightforward to adapt entirely from NICE guideline and technology appraisal reviews. However, the application of this evidence to an individual patient is complex. Heart failure patients are commonly elderly with multiple co-morbidities. Complex drug regimens increase the risk of interactions and cumulative side effects such as hypotension and renal impairment. Some drug classes have been trialled and are indicated only in select groups of patients, for example with a particular degree of left-ventricular ejection fraction or New York Heart Association classification of disease severity. In addition, the value of risk modification and improvements in prognosis need to be considered in the context of prognosis of heart failure itself and overall life expectancy (340-342).

Therefore, I developed some introductory content to support understanding of the drug-specific evidence:

- Prognosis of heart failure, drawn from a series of UK primary care cohort studies and a systematic review which provided useful, applicable evidence.
- Classification of heart-failure and left-ventricular function, drawn from the cardiological literature.

- A short statement about the principles of drug treatment written as a first draft from my reading of the literature and NICE guidelines, then reviewed by three primary care academics with a special interest in heart-failure and subsequently by the steering committee. These extra reviews were to ensure I had achieved the correct balance in writing to promote active, prognosis enhancing prescribing, whist also encouraging appropriate caution and consideration of individual patient needs. These reviewers also checked text in the drug-specific sections where I had added particular comments about applicability or caution.

Samples of this writing with references are in appendix 7.9.

7.6 A wider range of evidence sources for treatment harms

Information about treatment harms is an important component of GP Evidence, and a clear priority from my patient involvement work. Sometimes systematic reviews and RCTs provided good quality, usable data which could be used in *GP Evidence*. However, often they did not. It is well recognised that RCTs provide less information about treatment harms due to study design, poor recording or reporting (343, 344). Therefore I had to draw on a variety of sources for further data. This varied according to the treatment: newer treatments or those with well-known or high profile safety concerns (for example statins or SGLTs inhibitors) would tend to prompt a more in depth literature search.

Over the range of GP Evidence content, information on treatment harms came from these sources:

- RCTs, systematic reviews and meta-analyses
- Observational studies

- High quality review articles
- Medicines and Healthcare products Regulatory Agency(MHRA)
- British National Formulary(BNF)

Studies which had not had quality appraisal by NICE or Cochrane were assessed using quality appraisal tools and reporting guidelines such as the Cochrane risk-of-bias tool for RCTs, PRISMA for systematic reviews or CASP for observational studies. Examples of these assessments are in appendices 7.10-7.13.

It is important to note that GP Evidence does not attempt to provide complete lists of potential drug harms. These are provided elsewhere, for example the BNF and electronic clinical record systems.

7.7 Explaining complex evidence: glycaemic control in type 2 diabetes

This was the largest single section of work for *GP Evidence*. It addressed the challenge of presenting estimates of the benefits and harms of “tight” or “very good” control of glucose compared with “standard” or “good” control in type 2 diabetes (T2DM). Other sections in the T2DM page on GP Evidence had provided drug-specific information on benefits and harms, though many of the individual drug classes so not have long-term data regarding significant end-points such as micro and macro-vascular complications.

Recently updated NICE guidelines for T2DM contained new recommendations to personalise glycaemic targets (measured by HbA1c) for people with type 2 diabetes according to their individual characteristics and preferences. These are summarised in figure 7.1.

Figure 7.1 Selected recommendations on glycaemic targets from NICE guideline NG28 2022

1.6.5 Discuss and agree an individual HbA1c target with adults with type 2 diabetes. Encourage them to reach their target and maintain it, unless any resulting adverse effects (including hypoglycaemia), or their efforts to achieve their target impair their quality of life.

1.6.7 For adults whose type 2 diabetes is managed either by lifestyle and diet, or lifestyle and diet combined with a single drug not associated with hypoglycaemia, support them to aim for an HbA1c level of 48 mmol/mol (6.5%). For adults on a drug associated with hypoglycaemia, support them to aim for an HbA1c level of 53 mmol/mol (7.0%).

1.6.8 In adults with type 2 diabetes, if HbA1c levels are not adequately controlled by a single drug and rise to 58 mmol/mol (7.5%) or higher:

- reinforce advice about diet, lifestyle and adherence to drug treatment **and**
- support the person to aim for an HbA1c level of 53 mmol/mol (7.0%) **and**
- intensify drug treatment.

1.6.9 Consider relaxing the target HbA1c level (see recommendations 1.6.7 and 1.6.8 and NICE's patient decision aid) on a case-by-case basis and in discussion with adults with type 2 diabetes, with particular consideration for people who are older or frailer, if:

- they are unlikely to achieve longer-term risk-reduction benefits, for example, people with a reduced life expectancy
- tight blood glucose control would put them at high risk if they developed hypoglycaemia, for example, if they are at risk of falling, they have impaired awareness of hypoglycaemia, or they drive or operate machinery as part of their job
- intensive management would not be appropriate, for example if they have significant comorbidities.

NICE had produced a decision aid to support these recommendations which contained qualitative information about possible risks and benefits of tight glycaemic control, but did not include quantitative estimations of these.

There was therefore an important piece of missing information. How could doctors and patients decide on the value of extra drugs to lower glucose without understanding how much this would actually reduce their risk of diabetic complications or add to their risk of side effects?

The context for this is an ongoing controversy about the clinical evidence. A view that the clinical evidence shows that "lower is better" with regard to glucose control has maintained a primacy in the medical consensus (345-347). However, this has been questioned for over 30 years (348, 349) and particularly since modern RCTs showed minimal or no benefit of very tight compared to moderately tight glycaemic control (2, 350). The argument against tight glycaemic control leans on small absolute effect sizes for selected outcomes, the risk of harms and that much of the evidence to support tight control relies on observational evidence. An observation I made whilst reviewing this literature was that researchers on each side of the argument were working and writing separately, with no major publication I could find attempting to present arguments from both points of view.

I therefore sought to extract from the evidence base the key findings of RCTs and observational studies upon which these views were based and present them in a balanced way to support interpretation of the NICE recommendations. This process involved:

- A literature review, starting with NICE and Cochrane systematic reviews and extending to other systematic reviews and individual trials and also observational studies (which had also been included in the 2022 NICE review).
- Consultation with four clinical academics with expertise in this area and perspectives from both sides.
- Writing of a summary piece for *GP Evidence* including tables and graphics to illustrate exemplar data, presented in a balanced way and in the context of NICE guidance.
- A review of this piece by the steering committee, with a request to consider whether the content was adequately balanced.

- A review by GP-users during pair-writing sessions, with a special request to comment on clarity and balance.

A copy of this is shown in appendix 7.14 with references. It was divided into sections to "unpack" the evidence in an easily understandable way for GP users:

- Introduction and summary take-home messages
- What can we learn from RCT data?
- What can we learn from observational data?
- What about harms of low glycaemic targets?
- How can we think about setting glycaemic targets? What does NICE recommend?

This process produced a (to my knowledge) unique summary of the evidence and debate on this subject in a way which is compatible with and supports NICE guidelines. An important observation was that the same evidence base is interpreted in different ways by different authors, so a neutral laying out of the evidence ended up feeling less controversial than I had anticipated at the outset.

7.8 Input from steering committee

Table 7.1 below summarises some more examples of input from the steering committee regarding challenging decisions about clinical evidence.

Table 7.1 Examples of input from steering commitee

Issue arising regarding clinical evidence	Committee advice
Breakdown of cardiovascular events prevented by statin therapy in primary prevention Providing data on this was difficult. I had used some epidemiological data used by NICE in their economic model to create some illustrative examples of rates	The committee agreed that this approach was reasonable, as the examples were clearly presented as illustrative and not intended to be taken as hard facts.

of stroke v myocardial infarction prevention as a proportions of all cardiovascular events.	As this was potentially useful information and better option was not available, this content was included.
Benefits of statins in older people I had created some adjusted NNT figures extrapolated over 5 years from NNT figures originally presented as over 1 year in the literature.	The committee felt that these figures introduced an inappropriate impression of certainty about the evidence which was not warranted by the overall strength of the evidence, so these were not used.
Decisions about searching for evidence of harms For two drug classes: (ACE-inhibitors/A2RAs and Statins) I had found only low quality evidence regarding potential harms. I asked the committee if I should be searching further.	We discussed where and how I had searched for evidence, what other sources might be available and whether there was likely to be better information available even with a further search. In both cases we decided the information I had found was likely to represent the only available evidence.
Including non-statistically significant results Two scenarios arose where potentially meaningful but non-statistically significant results from single RCTs were present within NICE evidence reviews. These were for (1) the harm of gout from thiazide diuretics for hypertension and (2) benefits of low blood pressure targets following a stroke. They seemed to represent useful information which might shape treatment decisions.	We discussed the pros and cons of these, statistically-insignificant results are generally not deemed to be reliable enough to inform decisions. However, as both of these on balance showed useful clinical information, they were included with appropriate caveats and transparency.

8 Part 3. Chapter 8: Exploratory evaluation study

8.1 Research question and aims

The third research question[13] of this project was:

- To what extent might such a resource be usable in practice?

To address this, I undertook an exploratory evaluation study with the following aim:

- To explore whether the clinical content contained in GP Evidence can be accessed and understood by GPs in a way which can be integrated into their clinical thinking and decision making.

8.2 Ethical approval

Ethical approval for this study was granted by the University of Oxford Medical Sciences Interdivisional Research Ethics Committee. Reference: R82282//RE001 (Appendix 8.1).

8.3 Methods

8.3.1 Online focus groups with embedded online survey

Four online focus groups were conducted with 15 GPs in total from November 2022 to January 2023. Each session lasted for 90 minutes. For the first half of the session, GPs were asked to work individually, in silence, on an online survey which employed clinical vignettes (figure 8.1) to explore their prescribing intentions in three scenarios. They were then given access to the *GP Evidence* website and asked to complete the same three questions again. Following this,

[13] The first and second research questions were: 1.What are GPs needs and preferences for accessing evidence-based quantitative information on the benefits and harms of medications when managing patients with multimorbidity and polypharmacy? 2. How can we draw on design principles and clinician and patient input to co-design and interactive information resource to support informed conversations and shared decision-making?

the online group was brought together for semi-structured focus group discussion. I designed this combined method to capture individual-level information before using the focus group discussion to draw further insights.

8.3.2 Pilot study

I trialled this method with a group of 11 GP trainers at an in-person meeting as part of a trainers' residential workshop[14]. It worked well, but some adjustments were made for the subsequent study: clarification of text for some of the online questions and development of a stricter, more structured approach to the sessions to ensure proper questionnaire completion and suppress premature discussion.

8.3.2.1 Recruitment

Participants were recruited from the pool of 213 GP volunteers initially recruited for the survey described in section 3.6 and appendix 3.3. I excluded GPs who had had previous involvement in the development of the website to ensure the evaluation was conducted with users who were new to *GP Evidence*. They were sent an email invitation (appendix 8.2) with a copy of the participant information sheet (appendix 8.3). Two reminders were sent after one and two weeks. They were then directed to an online survey tool (Jisc Online surveys) to collect contact and demographic data and complete an online consent form (appendix 8.4). Consent was confirmed verbally at the start of the focus group. They were offered a £75 gift voucher as an incentive to participate.

[14] GP trainers are experienced GPs who train junior doctors to become GPs. I had been invited to speak at a trainers' residential workshop in March 2022 and they kindly agreed to participate in this pilot study.

The inclusion criterion was to be a:

- GP currently practising in the NHS.

Exclusion criteria were:

- GPs who have not been in NHS Clinical practice within the last year.
- GPs still completing specialist training (ST1-5).
- GPs with further training which may confer expertise about the subject matter.
- GPs who had any previous involvement with the design of *GP Evidence.*

8.3.2.2 Sample size

I anticipated that a relatively small sample size would be appropriate for this study. Factors contributing to that decision were that it was an exploratory study with a relatively homogenous group on a specific topic. I aimed to involve 20 participants over four sessions.

8.3.3 Design of questionnaire

The questionnaire was designed to assess the comprehensibility of information from *GP Evidence* and its possible effect on prescribing behaviour in three different types of clinical situations:

1) An example where the quantitative benefits of common treatments may be variably known by GPs and where the effect of new knowledge on prescribing might be unpredictable. For this I chose the option of statins and blood pressure lowering treatment in the primary prevention of cardiovascular disease.

2) An example where the new information on *GP Evidence* would possibly be new to GPs and would be expected to change intended prescribing practice in many cases. For this I chose some prescribing examples for chronic kidney disease.

3) An example of multimorbidity and polypharmacy where GPs had a number of treatment options and a range of information to explore on *GP Evidence* which might or might not be expected to change intended prescribing practice.

For the first two questions, participants were asked to choose their likelihood of prescribing a treatment from a 3-point Likert scale and rate their confidence in these decisions from a 5-point Likert scale (this second scale had been piloted and used in my previous survey described in section 2.4). They were also invited to give free text responses. The third question invited a free text response only. The clinical vignettes and questions are shown in figure 8.1.

Figure 8.1 Clinical vignettes and questions from evaluation survey.

1) Cardiovascular prevention

A 68 year-old woman has had an NHS health check. She is previously well with no long term conditions, but has been found to have raised blood pressure and cholesterol:

BP 152/86, TC/HDL ratio 5.8.Her renal function, glucose, ECG and urinalysis are all normal. Her QRISK score is 15%

She has been working on lifestyle measures over the last 6 months with the help of the practice nurse and they feel there is no room for further improvement. The nurse has suggested she comes to see you and has told her she may need blood pressure lowering drugs and a statin. The patient is asking you if she really needs these.

a) Would you recommend blood pressure lowering medication? **Probably yes/not sure/probably no**

b) Would you recommend a statin? **Probably yes/not sure/probably no**

c) How confident do you feel about these decisions?
Very/I've got a pretty good feel about this/confident enough/quite uncertain/very unsure

Could you make some brief notes/comments about your answers? For example: Are you following a guideline? You might have a nugget of clinical knowledge supporting your decision. You might have a specific gap in clinical knowledge Do you feel some kind of conflict? What's on your mind?
Free text response

2) Prescribing in CKD

A 76 year-old man is in good health other than a history of:

Mild well-controlled hypertension, CKD 3a

BP is 136/88 mmHg, eGFR is 50 mL.min/1.73m2 (not declined significantly in 5 years), no proteinuria.

His regular medication is: amlodipine 10mg od, atorvastatin 20mg od

a) Would you prescribe to reach a lower systolic blood pressure target of < 130mmHg?

Probably yes/not sure/probably no

b) Would you prescribe an ACE-inhibitor for nephroprotection regardless of lowering his blood pressure (i.e. even if you didn't want to lower blood pressure any further for its own sake)?

Probably yes/not sure/probably no

c) How confident do you feel about these decisions?

Very/I've got a pretty good feel about this/confident enough/quite uncertain/very unsure

Could you make some brief notes/comments about your answers? For example: Are you following a guideline? You might have a nugget of clinical knowledge supporting your decision. You might have a specific gap in clinical knowledge Do you feel some kind of conflict? What's on your mind?

Free text response

3) Polypharmacy

A 77 year old woman has a number of long-term conditions and medications. She lives independently though you have observed she is beginning to look a bit frailer in the last year or two. She weighs 60 kilos, BMI 19 kg/m2.She is mildly short of breath walking up gentle slopes. You imagine she has a life expectancy of 5 years or more.

She is struggling to manage her medicines and wonders if the "cocktail" is making her feel tired. She is keen to keep taking meds that will help her "keep going", but also to avoid side effects.

Long-term conditions: Hypertension (most recent clinic BP 122/60 mmHg), type 2 diabetes (no microvascular complications, most recent HbA1c 63 mmol/L (7.9%)), CKD (eGFR 50 stable over last few years, urine ACR 10mg/mmol), atrial fibrillation, on a statin for primary prevention for several years, heart failure with reduced LVEF (Echo showed LVEF of 36%, no signs of fluid overload).

Repeat Prescriptions: Bisoprolol 7.5mg od, ramipril 10mg od, amlodipine 5mg od, furosemide 40mg od, metformin slow-release 2g od, atorvastatin 20mg od, edoxaban 30mg od

Would you start or stop (or would you be thinking about starting or stopping) any medications in this scenario?
Make some brief notes on what's going through your mind - there are plenty of options and likely to be some "maybes" here. Comments on any certainties/uncertainties would be especially interesting.

Free text response

After they had completed the first round of questions, they were invited to look at *GP Evidence* whilst attempting the same three questions in the second half of the survey slightly reworded (appendix 8.5) to explore any changes in prescribing intention or thinking. After all participants had completed their surveys, I conducted a semi-structured discussion using a guide shown in box 2. A researcher script for the entire focus group is in appendix 8.5

Figure 8.2 Focus group discussion guide

We are interested to know what you took away from using the website. Were there any particular new bits of information you learned, did the information make sense, might using the website make any difference to what you prescribe or the discussions you have with patients?

We'd like to know if you think the information was NOT very useful or if you think you couldn't use it in practice- positive and negative feedback are welcome.

We have done separate testing on the function of the website already – like navigation and structure, so don't need detailed feedback on that kind of thing today, but do feel free to mention anything you feel is important.

Would anyone like to start with a comment?

Specific questions and prompts to be used as needed
Can anyone give an example of a treatment decision which changed after looking at the website?
Can anyone give an example of where the information on the website was new but didn't change your decision but reinforced it, or increased your confidence in it?
Can anyone give an example of where new information led to you feeling confused or less certain about your treatment decision?
How do you think you might use this in practice, if at all?
What overall effect do you think this might have on your practice, if any?
Do you think this information might cause any harm or difficulties?

Plus, specific questions prompted by observations made by participants.

8.3.4 Data collection and handling

Data handling and storage was conducted according to University of Oxford policy (see appendix 8.6). Participants' personal data was stored separately from research data. Research data including questionnaire responses, discussion audio and transcripts were collected without identifiable data.

The focus groups were conducted over Microsoft Teams. Responses to the questionnaire were downloaded from Jisc Online surveys and stored on Microsoft Excel files on a secure university computer. Audio recordings of the discussion were captured using the recording and transcription function of Microsoft Teams. The computer generated transcripts were checked for errors, amended then saved on Microsoft Word documents on a secure university computer. The audio recordings were then deleted. Textual answers from the online survey and transcripts of the discussion were uploaded into NVivo(v12) computer assisted qualitative data analysis software to manage the process of analysis.

8.3.5 Analysis

As this was an exploratory study and time-limited within the overall research project, I undertook a descriptive, rather than analytic, thematic analysis and did not involve a second researcher or seek participant feedback. I employed thematic analysis methods described by Braun and Clarke (351). Data from transcripts and survey text were analysed together. Coding was inductive. A copy of the final set of codes with their definitions is in appendix 8.7. Categorisation of codes into themes was achieved using the “one-sheet-of-paper” technique described by Ziebland and McPherson (352) with further refinement of themes during the writing of the results section.

I applied reflexivity during the analytic process (353), paying attention to my own assumptions and perspective as well as the perspectives and voices of the participants. I had particular regard to my vested interest in seeing a positive response to the website, and the likelihood of participants responding to social instinct to give positive feedback to someone whose work they were being invited to critique.

8.4 Results

17 GPs responded to the invitation to participate and 15 were able to join a focus group. A response rate which was lower than the intended sample size precluded maximum variation sampling. Focus groups were created according to participant availability. The characteristics of participants are detailed in table 8.1.

Table 8.1 Participant characteristics for GP focus groups

Characteristic n out of 15	
Sex	
Female	10
Male	5
Age, years	
30-39	2
40-49	2
50-59	7
60 and over	4
GP Role	
GP Principal	3
Salaried GP	3
Locum GP	9
Place of original medical degree	
UK	14
Non-UK	1
Geographical Region	
North-West England	2
Greater London	3
South-East England	3
South-West England	6
Scotland	1
GP description of practice	
Urban	6
Rural	2
Mixed Urban-Rural	7

8.4.1 Qualitative findings

Findings were grouped into four themes described below. Themes, sub-themes and further supporting quotations are in table 8.2.

8.4.1.1 Theme 1. Positive reactions and effects

All participants expressed positivity about the value of GP Evidence as a new tool to support practice:

> *Yeah, this is completely game changing it is so good… and I'm not trying to be smarmy. For years I've loved the concepts of shared decision making and helping people with their goals and wanting to have these discussions. (Participant 4)*

Participants described acquiring new knowledge, which appeared to have been understood within the context of their existing knowledge and perspective. Mostly this appears to have been an easy process, with many participants commenting positively on the simplicity and clarity of design. There were no comments about needing too much time to gather single-treatment related information. Sometimes participants stated intentions to alter their imagined prescribing behaviour in the light of new knowledge:

> *I think I'd underestimated the power of the beta blocker over perhaps the ACE inhibitor…so looking at the numbers, I changed my opinion. (Participant 11)*

Others gave examples where new information on the website had informed or supported their preconceptions, without necessarily changing prescribing intention, but increasing confidence and understanding:

> *I think for me the main thing was that it made me more confident in my decisions...and perhaps made me more coherent in my reasoning. So sometimes I had gut feelings and then wasn't always totally sure how to back them up. (Participant 10)*

8.4.1.2 Theme 2. Patient interaction

A number of descriptions were offered about how information from *GP Evidence* might be used to support conversations with patients. A recurring idea was that the infographics could be shown directly to patients during a consultation. Other suggestions involved the use of pieces of information from the website as components of a discussion.

> *…now I'm clicking on the button and I think that would be so useful to share with patients, you know, the impact of aerobic exercise. And I'd quite like to compare that with the impact of medication, for example. (Participant 8)*

8.4.1.3 Theme 3. Complexity

Whereas finding and applying information on single conditions or treatments appeared to have been relatively straightforward, addressing the complexity of multimorbidity and polypharmacy was described as more challenging. Participants described thinking about multiple considerations and treatment options. Some seemed to regard this as almost too challenging; two quotes below illustrate differing reactions, one GP describes using gut feeling to support decisions, another speculates on how a potentially more advanced version of *GP Evidence* might help:

> *…it it's that really complex group where it's much harder to try and draw together, if we've got drugs which are treating two or three things. So if you've got cardiovascular disease and you're on bisoprolol and you've got some benefit for your heart, some benefit for your AF speed, some benefit for your heart failure, some benefit for blood pressure, but smaller…those are always the most complex bits of work…and then it's still almost impossible to get the data that you need to do anything other than a…gut feeling. (Participant 5)*

> *…for example, that patient was on a big dose of metformin, so potentially renal effects, if the dap [dapagliflozin] would give some benefit on the diabetes. So that maybe that [the metformin] could be reduced a little bit and they'd need less of their beta blocker in total. Can we end up with less tablets…and that's a kind of*

more convoluted pharmaceutical algorithm probably which…we're not there yet, but it's the kind of thing which you can imagine the development on a program like this would be able to say…(Participant 7)

Most free text answers in the online survey indicated specific learnings about individual treatment choices for the case example on multimorbidity. A few indicated specific changes in treatment intention. Here, in a discussion, a GP suggests that *GP Evidence* might help with this challenge of complexity:

…we're managing them without the specialists, without any specialist input. And we have to advise them on which you know which is the most important medication to start? You know, are we thinking about their blood pressure? We think about their diabetes. Are we thinking about their cholesterol? Where do we start? What's the side effects, how they're going to fit this around their work because they're mostly in work. So I see this as a great resource for managing that type of patient. (Participant 10)

A number of participants suggested that with increasing familiarity with the site they would be able to get more out of it, and that a gradual accumulation of knowledge may occur:

So I think that that is something that with longer to play with it, we'd get better at…utilizing that functionality. (Participant 5)

There were bits that were gaps in my knowledge that I could look up. Once I've looked them up I shouldn't need to look them up again. (Participant 12)

8.4.1.4 Theme 4. Barriers and negative perceptions

Confusion arose for some participants about particular aspects of content. One example was about the meaning or implication of a "low quality" evidence rating and how that should affect practice. Another described the challenge of balancing numbers-needed-to-treat and numbers-needed-to-harm; could one calculate a "net" effect? One GP in particular had found that the volume of information available in general was too great and it engendered a sense of panic.

Challenges were described regarding how to implement any changes in practice resulting from new knowledge. One participant raised the question of the diminishing role of the GP in making prescribing decisions in a multi-disciplinary context. A common issue was how information from *GP Evidence* might fit in with NICE guidelines or the Quality and Outcomes Framework(QOF), particularly where actual benefits of treatments were relatively low; I had a sense of a degree of cognitive dissonance arising.

> *The Blood pressure benefits/harms on website would clearly suggest not much benefit to treating but we are hampered by QOF/ NICE guidelines pushing into treatment, and cardiology! Was quite surprised. Creates an ethical dilemma re recommending. (Participant 11 – text response)*

8.4.2 Quantitative findings

The online questionnaire captured data on the number of participants changing (or not changing) their declared intention to prescribe a treatment after using *GP Evidence*, measured on a 3-point Likert scale completed before and after exposure to the website. They also recorded their confidence in their decision before and after exposure to the website on a 5-point Likert scale. Table 8.2 shows a summary of changes in participants' responses.

Table 8.2 Summary of participants' responses to intention to prescribe and confidence questions on online survey within the focus group.

Number of points change on Likert scale	**-2**	**-1**	**No change**	**+1**	**+2**
Q1a) Drug treatment for hypertension	4	3	7	0	0
Q1b) Statin	1	0	10	2	1
Q1c) Confidence	1		8	5	
Q2a) Lower blood pressure target	4	1	9	0	0
Q2b) ACE inhibitor	5	2	7	0	0
Q1c) Confidence	0	0	5	6	3

Full description of questions and Likert scales in box 2. Figure in each cell describes the number of participants with a particular degree of change in their answers before and after expose to *GP Evidence*. n = 14 (one participant was excluded due to an error in data entry).

38 answers out of a total of 84 revealed participants making some degree of change in their intention to prescribe, and a quarter of answers indicated an increase in confidence. Only one response reported a decrease in confidence.

Table 8.3 Themes, sub-themes and quotations from GP focus groups.

Positive reactions and effects
General I think it was a great as a website. It's the sort of thing that I love actually.*(Participant 7)* Yeah, this is completely game changing it is so good… and I'm not trying to be smarmy. For years I've loved the concepts of shared decision making and helping people with their goals and wanting to have these discussions. *(Participant 4)* It's fantastic to have a place where all the evidence is gathered together and I'm not having to be the person who reads all the latest guidance. *(Participant 7)* **New knowledge or change in prescribing intention** I think before the website I was more inclined to give her a blood pressure medication, less inclined to give her a statin. And having looked at the website, I was thinking probably either the other way around or give her both. *(Participant 6)* I think I'd underestimated the power of the beta blocker over perhaps the ACE inhibitor…so looking at the numbers, I changed my opinion*. (Participant 11)* There's a lot of positivity towards SGLT2, for heart failure as well as diabetes at the moment. And I was shocked to see what little impact the smiley faces showed me that they had. *(Participant 5)* Would aim for BP target 140/90 no albuminuria so I would not start an ACE[inhibitor]…I was able to come to these conclusion after using the website *(Participant 13)* **Improved confidence** The website sort of confirmed I think what I was feeling and gave me some evidence you know some more evidence behind it. So I thought that was really useful. *(Participant 1)* I think for me the main thing was that it made me more confident in my decisions...and perhaps made me more coherent in my reasoning. So sometimes I had gut feelings and then wasn't always totally sure how to back them up. *(Participant 10)*
Patient Interaction
The statin chart helped me realise the benefit and would help in consult to explain to patients. *(Participant 13 text response)* …now I'm clicking on the button and I think that would be so useful to share with patients, you know, the impact of aerobic exercise. And I'd quite like to compare that with the impact of medication, for example…*(Participant 8)*

About the infographics…there was options to change what you're reducing the risk of. So if you've got a patient that's got parents that have died of stroke and that's [what they're] most concerned about, you know that information is gonna be most relevant. Or…and hit home the most for them. *(Participant 9)*

Complexity

Multimorbidity and polypharmacy

…it it's that really complex group where it's much harder to try and draw together, if we've got drugs which are treating two or three things. So if you've got cardiovascular disease and you're on bisoprolol and you've got some benefit for your heart, some benefit for your AF speed, some benefit for your heart failure, some benefit for blood pressure, but smaller…those are always the most complex bits of work…and then it's still almost impossible to get the data that you need to do anything other than a…gut feeling. *(Participant 5)*

…or example, that patient was on a big dose of metformin, so potentially renal effects, if the DAP [dapagliflozin] would give some benefit on the diabetes. So that maybe that [the metformin] could be reduced a little bit and they'd need less of their beta blocker in total. Can we end up with less tablets…and that's a kind of more convoluted pharmaceutical algorithm probably which…we're not there yet, but it's the kind of thing which you can imagine the development on a program like this would be able to say…*(Participant 7)*

…we're managing them without the specialists, without any specialist input. And we have to advise them on which you know which is the most important medication to start? You know, are we thinking about their blood pressure? We think about their diabetes. Are we thinking about their cholesterol? Where do we start? What's the side effects, how they're going to fit this around their work because they're mostly in work. So I see this as a great resource for managing that type of patient. *(Participant 10)*

Further exploration of website

…playing with the site would give better familiarity with the links and the ways to sort of jump between the different sections. So I think that that is something that with longer to play with it, we'd get better at…utilizing that functionality. *(Participant 5)*

It's got a lot more detail on it. But I do need a lot more time to play around with it. *(Participant 2)*

Accumulating knowledge

There were bits that were gaps in my knowledge that I could look up. Once I've looked them up I shouldn't need to look them up again. *(Participant 12)*

In reality once the website is up and running, you would spend some time just looking at each section… without sort of doing it on the spot or with the patient and then you would be a bit more familiar. *(Participant 6)*

Barriers and negative perceptions

Confusion

I think the thing that will probably confuse or worry me would be the bit where it says say for this type of evidence "quality low to very low"… What does that actually mean in in practice? I'm sort of struggling…is that just an extra level of complication in this maybe? *(Participant 14)*

So I was trying to think if the numbers needed to harm are bigger than the numbers needed to treat. Does that mean that there's net harm? And then I thought…I'm not sure. *(Participant 10)*

…it looks like a good tool, but one would have to be using it a lot before it became a good tool, I'd say. And just using it straight off …it just engendered…like sense of panic and of uh, that makes my knowledge even worse *(Participant 15)*

Challenge to perceptions of normative practice

…knowing this does not make you happy…um, knowing the diabetes stuff…that all the pressure there is to treat and get sugars down is largely pointless…you then have a conflict of knowing that what you're actually doing is wrong. *(Participant 12)*

The Blood pressure benefits/harms on website would clearly suggest not much benefit to treating but we are hampered by QoF/ NICE guidelines pushing into treatment, and cardiology! Was quite surprised. Creates an ethical dilemma re recommending. *(Participant 11 – text response)*

But I think that what I found was difficult with the website was it gave you the information about what the NICE guidance says and then you actually look at the data for that and you think, "well actually that doesn't really match up that much in terms of the benefit for the patient". *(Participant 11)*

Barriers

If I've got heart failure to deal with or diabetes, I use the team. I don't do it myself. So my ability to influence all of this in both of these roles is markedly reduced compared to what it was five years ago. *(Participant 12)*

Positive responses to probes to potential negatives

Probe question: Does anyone feel that that it's maybe too complex or the information is a bit overwhelming …or it could be problematic…or any negative aspects of having this kind of stuff at your fingertips?

I think it's really positive. I mean, so I'm looking on the heart failure one now. And I think that is really helpful. You know, the fact that it goes, you know, it lists all the all the medicines…and you know where they're placed in the treatment of heart failure. I think that's really, really helpful. I think for me, it simplifies everything…which is otherwise, you know, very complex. *(Participant 8)*

I would say the same that I don't think it adds confusion I think….the scenario is confusing and it helps… *(Participant 7)*

8.5 Discussion

8.5.1 Summary

All participants in this study with perhaps one exception, responded positively to *GP Evidence*, perceiving it as a usable tool able to provide them with useful, understandable information which could support their practice and affect their interactions with patients. In hypothetical examples, the study showed that new knowledge acquired from the website could change

GPs' reported intentions for single prescribing decisions. An increase in confidence about prescribing decisions was reported in about half of the survey questions and reflected in the qualitative data.

Acquiring and using new information about treatments in the context of multimorbidity and polypharmacy appeared to be more challenging for GPs, who expressed different perspectives and levels of confidence about this.

Occasional confusion or uncertainty arose about content on *GP Evidence* in terms of both understanding and application to practice. Some barriers to application of new information in practice were described in terms of potential conflict with clinical guidelines or normative practice.

8.5.2 Strengths and Limitations

Strengths of the study are that the participant group contained GPs with a wide range of demographic and professional characteristics, though there were proportionally more female GPs and GP locums than the general GP population (354, 355). Though the recruitment target of 20 was not met, the sample of 15 generated substantial data. At this level of analysis, data saturation seemed to have been reached, with no new codes added after coding of the third focus group transcript.

Study design enabled the capture of individual GPs' responses unaffected by others, as well as wider ranging responses facilitated by focus-group discussion. Both qualitative and quantitative data were gathered. The questionnaire was designed to assess responses in both simple and complex clinical scenarios. Data analysis was performed in a structured manner following established methods of thematic analysis.

Limitations of the study are that there may be response bias driven by self-selection of participants likely to be interested or favourably disposed to *GP Evidence*. Participants may have given more positive responses to an interviewer who they knew had developed the website than to a neutral researcher; I aimed to reduce this potential bias by deliberately inviting criticism and negative feedback and clarifying that this was what was being sought.

The vignette-based questionnaire was designed to reflect everyday clinical work, but the artificial nature of answering theoretical questions within the protected time and context of the focus group limits the assumptions that can be made about what might happen in practice.

The data size and analysis were limited. I did not attempt a deeper analysis of issues which arose such as decision-making in complexity, the interaction of new knowledge with systemic drivers and barriers, or questions of power and professional freedom. Data are only able to describe the outcome for this study and are not generalisable to a wider GP population.

8.5.3 Meaning and implications

In this sample of participants, in the context of a formal study, *GP Evidence* successfully communicated information on the benefits and harms of treatments for long-term conditions to GPs.

Considering the informational needs and anticipated barriers described by GPs in the interview study in Chapter 3 and in the background literature, the design of *GP Evidence* appears to have been successful in addressing many of these for these participants. The barrier of statistical literacy was overcome by the design of information and content – these GPs were able to understand the quantitative information. Information appeared to be

regarded as accessible and understandable with occasional exceptions. The critical aspect of being able to access information in a short period of time has been addressed; at least for higher-level information on single conditions. This represents a success of the design strategy. GPs in the earlier interview study speculated that a better understanding of quantitative information and evidence about treatments would improve their confidence – this present study provides support for that idea.

The challenge of communicating with patients about quantitative information about the benefits and harms of treatments had been identified as a possible barrier in the previous interview study, but did not arise at all in this focus group study. By contrast, these participants described how they might use the website itself as a communication aid, showing infographics or particular numbers to patients, or alternatively describing how new information might shape their own understanding and behaviour without needing to directly communicate numerical information. This may be because the study was not specifically designed to explore this issue, or it may be that presented with well designed, understandable information, the GPs were easily able to imagine how they could integrate this into their thinking and consulting.

Using new information from *GP Evidence* in the complex context of multimorbidity and polypharmacy appeared more challenging for participants. There was evidence in the data (in response to the survey question on polypharmacy) of acquisition of new knowledge, consideration of how that might be used in practice, as well as some positive comments in focus group discussions about how *GP Evidence* might help in this scenario. However,

participants spoke at length about the challenges of making these complex treatment decisions.

One interpretation of this finding is that adding more information into the decision-making mix is simply too much for many clinicians to cope with. Another might be that given the degree of clinical complexity in multimorbidity and the context of patients' varied lives, values and preferences, the idea that quantitative judgements about the relative merits of individual drugs might play an important role in decision-making is naïve. Does an expectation that quantitative knowledge will make a difference place too much emphasis on ideas of Bayesian rationality (356), where clinicians and patients will consciously calculate or intuit individual chances of benefits and harms?

An alternative interpretation is that because decision-making in multimorbidity and polypharmacy is inherently so complex, a one-off introduction to some new data did not allow enough time for participants to integrate it into their thought processes. Acquiring new information entails significant cognitive work, engaging deliberative, logical, "slow" thought – described as "system 2" thinking by Kahneman (357). It is effortful and irritating, in contrast to "system 1" thinking - automatic, effortless, fast - the cognitive mode in which GPs (like all humans) operate most of the time. When addressing a very complex scenario, there is likely to be limited opportunity to engage system 2 thinking due to time pressure or cognitive overload. However, some participants in this study did describe acquiring individual pieces of new information when considering the multimorbidity question. Once new knowledge has been acquired (perhaps in the context of looking for single-treatment information in a non-complex scenario) it may be remembered and later applied in a more complex setting. This

integration of new knowledge into system 1 from system 2 has been described in a clinical context by Croskerry in a model of diagnostic reasoning (358). Familiarisation and repeated use of *GP Evidence* might lead to an accumulation of knowledge which can be later applied be that rationally or intuitively. Participants in this study talked about exploring the website further, gaining confidence and retaining information.

This idea, though optimistic, is supported by Gabbay and Le May's theory of mindlines, described in chapter 3.10.4. They showed that GPs construct internalised collections of knowledge built from multiple sources, which are retained as explicit or tacit knowledge, adapted over time, influenced by their communities of practice and applied flexibly with patients (215, 359).

This limited evaluation study cannot address these more complex questions, which are a potential subject for future research which I will discuss in the next chapter.

8.5.4 Conclusion

The findings of this study have limited generalisability due to the small sample size, potentially atypical participants and hypothetical nature of the exercise. However, they provide encouraging evidence of the success of the design of *GP Evidence* and its content to successfully communicate information on the benefits and harms of treatments for long-term conditions to GPs in a way which was seen as feasible, understandable and usable in practice. There were some exceptions to this for particular elements of information for particular clinicians. Using new information in the context of multimorbidity and polypharmacy presented greater challenges, though there was some suggestion that further use and familiarity with the website might enable this.

9 Chapter 9: Summary and Discussion

9.1 The launch of *GP Evidence*

GP Evidence was launched on the 1st February 2023 as a freely accessible website. The launch was supported by a news piece (Appendix 9.1) and Twitter campaign from the Nuffield Department of Primary Care Health Sciences in addition to my personal activity on Twitter and a range of dissemination methods summarised in figure 9.1.

Figure 9.1 Dissemination activity post-launch of GP Evidence

Royal College of General Practitioners: included a recommendation in the Chair's weekly newsletter

National Association of Sessional GPs: interview with subsequent piece included in their monthly newsletter

Local Medical Committees: emails to all LMCs to inform about GP evidence and request cascading to their members.

Conferences (all in 2023 including those scheduled in future):

- **Poster presentations** at: British Journal of General Practice annual conference, Society of Academic Primary Care (south-west) annual conference, Royal College of General Practitioners Annual Conference.
- **Oral presentations** at: Society of Academic Primary Care annual scientific meeting, Preventing Overdiagnosis conference.

Webinars: two for the Prescribing Quality Improvement Partnership.

Podcasts: *Aural Apothecary* podcast and *Richard Lehman on Evidence-based Medicine* podcast.

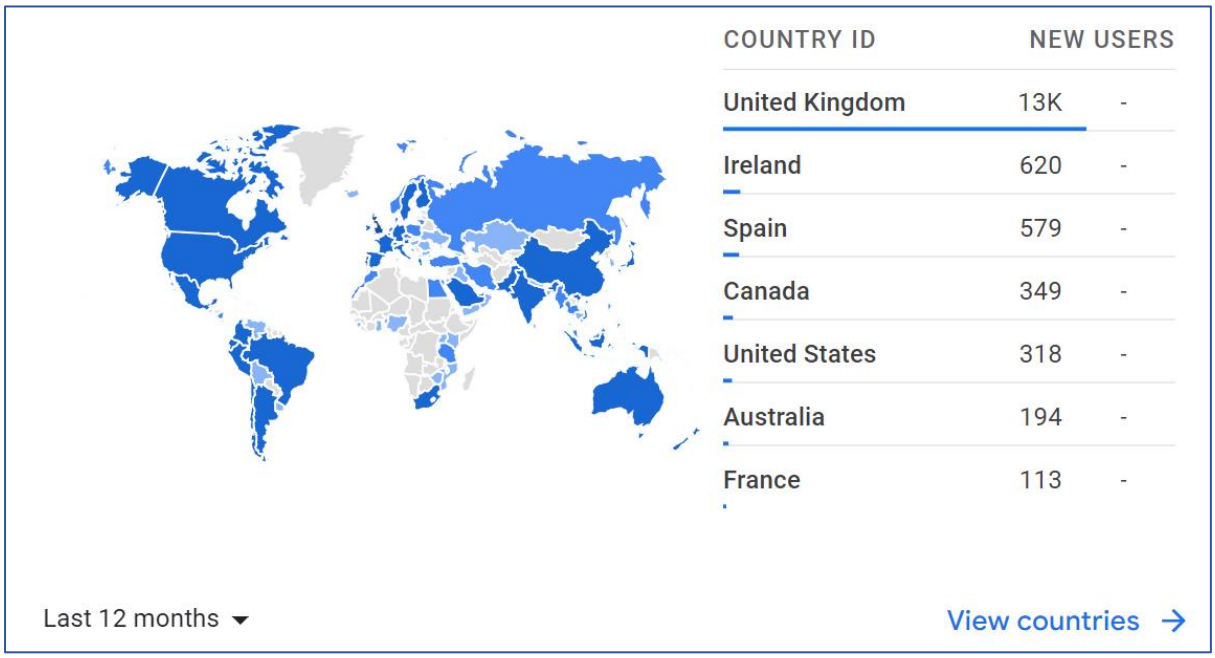
Independent GP educational organisations: Red Whale and NB Medical – commitments from these to disseminate awareness. Inclusion of links to *GP Evidence* in Red Whale online materials.

The response to the design and content of *GP Evidence* on social media has been extremely positive (some examples are shown in appendix 9.2), mainly from GPs, pharmacists,

academics interested in evidence and shared decision-making but also some members of the public. This has been very encouraging, but must be interpreted with the understanding that it reflects the reactions of a like-minded community. A number of users have responded directly through the "Feedback" facility on the website with positive comments and suggestions for future ideas about design and clinical content. To date, I have not received any criticism or negative feedback about the clinical content of *GP Evidence*. The only negative feedback about design has been regarding its current lack of compatibility with mobile phone sized screens.

On June 28th 2023, Google analytics showed that GP Evidence has had over 14,000 visitors with approximately 100-200 people visiting each weekday. Though the audience has been mainly in the UK, it has been viewed in 45 other countries (figure 9.2).

Figure 9.2 Google analytics data for GP Evidence 28th June 2023



9.2 Summary of research findings

9.2.1 Part 1: GP interviews

In part 1 of the thesis, I addressed the first research question: *What are GPs' needs and preferences for accessing evidence-based quantitative information on the benefits and harms of medications when managing patients with multi-morbidity and polypharmacy?*

Using a series of one-to-one interviews with 14 GPs with a framework analysis, my findings revealed new insights which were missing from the literature on GPs and their relationship to evidence-based medicine, clinical guidelines and the management of multimorbidity and polypharmacy (section 3.10). GPs were aware of a knowledge deficit about the quantitative benefits and harms of treatments (QIRx), they had found ways to try and individualise patient care without this knowledge, and were able to imagine how an improved understanding of QIRx might support their practice. They were keen to have access to information on QIRx but anticipated a number of barriers to its use: time limitations, lack of confidence in statistics and scientific terminology, competing drivers to practice and medico-legal concerns. These findings provided important insights for the design of *GP Evidence* (detailed in section 3.10.4 and 6.1.1.1) which were carried through into the next phase of this project.

In addition, these interviews included a "think-aloud" exercise where GPs looked at existing online resources which deliver information on QIRx. This identified important strengths and weaknesses of these resources, and a list of design features which worked well and which did not (section 6.1.1.3) which I was able to directly apply to the design of *GP Evidence*.

9.2.2 Part 2: Design and evidence

In part 2 of the thesis I addressed the research question: *How can we draw on design principles and clinician and patient input to co-design an interactive information resource to support informed conversations and shared decision-making?*

9.2.2.1 Design of GP Evidence

Using a range of design principles and methods (chapter 4) I conducted an iterative process of research through design over three years involving 20 GPs in total. The GPs' participation in the co-design process began with a joint application design workshop (sections 4.4.4 and 6.2) where they shaped the website design from a basic prototype. This was followed by a series of user-testing sessions to refine the design (sections 4.4.7 and 6.3), and pair-writing sessions where they reviewed and edited written content (sections 4.4.5 and 6.4).

This produced a functional website with clinical content on 12 common long-term conditions and their treatments. By this stage, feedback from the GPs was highly positive, but further evaluation was still required with a new group of GPs who were new to *GP Evidence*.

Zimmerman et al (9) propose four criteria for evaluating the results of research through design (RtD); these are summarised in table 9.1 with notes on this research project. I will expand further in the rest of this chapter.

Table 9.1 Criteria for evaluation of research through design. Adapted from Zimmerman et al 2007.

Evaluation criterion and summary	Commentary on *GP Evidence*
Process There is no expectation in RtD that reproducing methods will result in the same findings or artefact. Therefore, an assessment of *the appropriateness and rigor of application of the methods* used supports the validity of the findings. Documentation of rationale for and application of selected methods may be used to judge the quality of the work.	The selection and application of methods are described in chapters 4 and 5. The methods were selected and adapted for the particular needs of the *GP Evidence* project.
Invention The RtD contribution must constitute a significant invention. This means a novel integration of various subject matters to address a specific situation. Assessment of this can be judged from a detailed literature review which situates the work in context.	*GP Evidence* is a unique design which fulfils a specific need to support the clinical practice of GPs. The design drew on multiple sources: the literature discussed in chapters 1 and 2, and pre-existing work in similar fields (patient decision-aids, other online resources). It combines a novel design with an individual pragmatic approach to the sourcing and collation of clinical content.
Relevance RtD researchers must articulate the "preferred state in the world" their design attempts to achieve and provide support for why the community should consider this state to be preferred.	The aim of GP Evidence is summarised in section 2.10 in the context of the literature discussed in the first two chapters. In these terms, the "preferred state" is one in which the clinical practice of GPs is supported by a better understanding of the benefits and harms of treatments for long-term conditions. This will support informed, shared-decision making and better management of multimorbidity and polypharmacy.
Extensibility This is defined as the ability to build on the resulting outcomes of the research. This may be by employing the methods or process in a future design problem, or understanding and leveraging the knowledge created by the resulting artefact. The research must be documented in a way which allows the design community to leverage the knowledge derived from the work.	I believe there is potential to build further on this work, I will discuss this in sections 9.4 and 9.5.

9.2.2.2 The clinical content of GP Evidence

To ensure the quality and trustworthiness of the clinical content of *GP Evidence*, I developed a pragmatic approach to the collation of clinical evidence, tailored to the needs and resources

of this project. This was supported by a group of patient/public representatives and clinical experts who co-developed an evidence collation strategy (chapter 5) and supported my decision-making as a steering committee (chapter 7). Quantitative content underwent a process of double-checking by academic colleagues (section 5.4) and written content was reviewed by both a professional editor and GPs (sections 4.4.4.5 and 4.4.4.6).

I am confident that the clinical content is of high quality and compatible with NICE guidance and normative practice, whilst providing new insights for users regarding the evidence base which underpins guideline recommendations. I will discuss this further in section 9.3. However, it is important to acknowledge that this claim regarding quality of clinical content could only be confirmed by extensive, detailed review by subject experts, which was not possible within the resource limitations of this research. A feedback page on the website invites enquiry, corrections and criticism of the content on *GP Evidence* which aims to provide a mechanism to correct any errors and invite constructive challenge.

9.2.3 Part 3: Evaluation study

In part 3 of the thesis, I addressed the research question: *"To what extent might such a resource be usable in practice?"*

As an exploratory evaluation study, I conducted four online focus groups with a total of 16 GPs new to *GP Evidence,* with a descriptive thematic analysis of the discussions and results of an embedded online questionnaire study. This group of participants responded positively to *GP Evidence*, perceiving it as a usable tool able to provide them with useful, understandable information which could support their practice and affect their interactions with patients. They reported increases in their confidence in prescribing decisions, and in hypothetical

clinical examples, described a number of changes to their prescribing intentions. Some confusion or uncertainty about the content of *GP Evidence* was expressed, and the application of new clinical information to theoretical examples of multimorbidity and polypharmacy appeared to be more challenging.

Overall, this study provided positive confirmation of the success of the design and content of *GP Evidence*, though was limited by a small sample size and self-selection of participants who may have been predisposed to respond favourably. Conclusions, strengths and limitations are discussed in section 8.5.

9.3 Strengths and limitations of the overall programme of research

Strengths and limitations of the GP interview and focus group studies are discussed in detail in sections 3.10.2 and 8.5.2 respectively. For part 2 of the thesis, the research through design (RtD) phase, a brief evaluation is outlined in table 9.1 above.

The strengths of the RtD research were that it was grounded in a deep understanding of user needs and context which I had developed from: my own experience, exploratory work prior to the doctoral research (section 1.5), the academic literature (chapters 1 and 2), and GP interview study (chapter 3). GPs participating in the co-design process were recruited from (and were broadly representative of) the general GP population. They were involved meaningfully, early and continuously in the project and given genuine power to shape the outcome – factors known to improve the quality of co-design research (section 4.3.1). My own role as user-as designer (section 4.3.2) was valuable, bringing ready insights to bear on design choices.

The design of the website itself drew on recognised principles for human-computer interaction (section 4.3.3) and technical aspects were undertaken by experienced professional partners with expertise in person-centred, accessible design. A variety of methods were employed, selected for their appropriateness to this project. They are described in chapters 4 and 6 and included: prototyping methods, joint application design workshop, user testing, content design and pair writing. Throughout, the design was iterated and refined according to user feedback.

A central limitation of the project is that the GP users who participated were self-selecting and unavoidably represent a "type" of GP interested in this topic area and willing to devote time and mental energy to thinking about the detail of clinical evidence underpinning their practice. It is probable that this level of interest will not be shared by all GPs, for whom *GP Evidence* might not hold much appeal, or might prove to be too complex. The exploratory evaluation study is limited by sample size and self-selection of participants and was not capable of producing results generalisable to the whole GP population.

My role as user-as-designer whilst providing many strengths, may also have introduced some limitations. Despite meaningful involvement with GP users, patient and public participants and clinical experts, and active methods to manage my own biases and preconceptions the project was still driven heavily by myself. Much of the design and clinical content was developed from a starting point of my own thoughts and perspective. This may render the design and content less suitable for general users to some degree.

Limitations of time and resources will have affected the final results and product. Examples of areas which have been identified for improvement by user feedback are the lack of

graphical presentation of harms, text searching and mobile phone compatibility. Some areas of clinical content could doubtless be further refined and improved through repeated review and pair-writing.

9.4 Achievement of the aims of *GP Evidence*

As stated in sections 1.2 and 2.10, the aim of *GP Evidence* is to improve GPs understanding of the benefits and harms of treatments for long-term conditions. To achieve this, *GP Evidence* needs to provide information on the benefits and harms of treatments in a way that can be understood by GPs and potentially integrated into their practice, this new information providing an enhanced opportunity to combine the evidence with their clinical experience and patients' preferences to improve care.

Mindful of the limitations described in the previous section regarding the population of GPs who participated in the research, the aims of the *GP Evidence* project appear to have been achieved to the extent that can be assessed from the research so far. The co-design approach produced a novel design which communicated complex information on the benefits and harms of treatments in a way which could be easily understood by the participating GPs. In the evaluation study, they were able to describe how they might integrate this new information into their practice.

In terms of Zimmerman's criteria for the evaluation of research through design, *GP Evidence* demonstrates: (1) Invention - a novel design which addresses a specific situation and (2) Relevance - it offers a solution to (potentially) produce a "preferred state in the world" where patient care is improved through GPs' improved understanding of the benefits and harms of treatments.

There are important caveats to these conclusions of success. Firstly, user-testing and the evaluation study were conducted in controlled research contexts, with protected time. I cannot be sure if and how *GP Evidence* will be used in real-world practice with its distractions and time-pressures. Secondly, the application of new knowledge about the benefits and harms of treatments to address multimorbidity and polypharmacy proved more challenging in the evaluation study; I am only able to theorise that GPs may over time integrate new knowledge about individual treatments and conditions and be able to apply it in this context. Thirdly, it is not known what proportion of the GP population will find *GP Evidence* appealing and therefore how widely its potential benefits may be felt.

9.5 Reflections on methods and process

Overall, my approach and chosen methods worked well. The GP interview study provided new, rich, relevant information and in retrospect it provided all the information it could have to inform the next stage of the project.

The co-design process achieved its aims with considerable success, though being inherently an unpredictable and creative process it is impossible to say whether an even better design outcome could have been achieved (I have to assume this is the case however pleased I am with my efforts). The background research and stakeholder involvement supported the generation of a useful early prototype. The Joint Application Design (JAD) workshop was an incredibly effective and pivotal stage of the process, despite entering into it with relatively little structure and slightly unclear expectations. A measure of success of the JAD is how few design changes needed to be made after user-testing.

The refinement of website content via the editorial and pair-writing process was transformative. In fact, it was only at this stage that I truly got a sense of "I've got this right!", as the clarity of written content increased markedly, allowing the spaciousness of the design to really shine. Prior to this moment, I had still had doubts about whether it would all actually "work", perhaps I had only produced a pretty design which appealed to me and a few others.

The evaluation study was limited by time constraints, as was always anticipated at the outset of the project. Given this limitation, my chosen methods generated useful information and ideas for further research – I would use the same design again.

If I were to repeat the project, I would endeavour to obtain even more user input to enhance the quality of design. A second JAD may have been very valuable and provided the opportunity to explore better ways of communicating information on treatment harms, or acquire a better understanding of how users might like to access and structure supplementary information. More time pair-writing with GPs would almost certainly improve the quality of written content. As described above, this was a transformative process.

Regarding the approach to the selection and presentation of clinical evidence, this too worked well. At the outset of the project and during the development of the process for evidence selection (described in chapter 5), I had been anxious to have a quite detailed, rigorous framework to follow. This proved difficult to achieve due to a combination of the necessity to have room for flexibility and judgement in the process and the limitations in terms of resources to do highly structured evidence searches. The fairly loose approach which was arrived at (figure 5.6) in fact served the process well, in combination with support from the

steering committee. As the project progressed, I was aware that when there was a challenging decisions to be made about clinical evidence I was asking myself the following questions:

1) Is the evidence of high enough quality (or the best available quality) that I would feel comfortable making a clinical decision based on it?

2) Does the evidence support or broadly align with the relevant NICE guidance and insights available from the guideline development documentation?

3) Where there was a choice about an evidence source which might impact the treatment effect size presented on GP Evidence, was this of a magnitude which might make a meaningful difference to a consultation or patient's choice?

4) Would I be happy to be challenged about the evidence source I had used by a topic expert?

5) Can I present any complexity and uncertainty about the evidence adequately on *GP Evidence?*

For all of the dilemmas which arose, I found I was able to answer all of these questions in such a way[15] that I felt confident in the clinical content published on *GP Evidence.* These five questions were arrived at retrospectively, but could be applied prospectively as part of an updated process guide for writing new content, providing a framework for decision making when writing or reviewing content.

[15] This generally meant a "yes" for all these questions except Q3 which generally was a "no" answer. Where there was a meaningful difference in effect sizes to choose from, a "yes" answer to the other questions was adequate to support a decision.

9.5.1 Applicable insights from *GP Evidence*

Though the *GP Evidence* project had a specific aim and audience, and was not designed to produce widely generalisable findings, there were lessons learned and insights generated which may be of value to others working in related fields.

The first such audience is those who are interested in providing clinicians with new information in order to influence or improve their practice. Examples of this might be a clinical risk score for a condition upon which to base a management plan, or a simple awareness-raising piece of information such as warning signs of an important condition. In this situation, researchers or implementation specialists will wish to maximise the clinical impact of new information by delivering it to clinicians in a way which will be noticed, understood and will prompt behaviour change.

The second audience is those whose work in the field of knowledge translation. They may have a narrower scope, focussed more simply on provision of research findings to end-users. These might include clinical guideline developers, medical journals or educational content providers. Their agenda is simply to communicate information in an understandable and valuable way.

I have summarised key insights for these two groups from my experience during the doctoral research in figures 9.3 and 9.4. They have considerable overlap with general principles of design, implementation frameworks and content design but provide a context-specific emphasis. As these are derived from a single project, they cannot be considered to be fully generalisable, rather to be understood in their context and applied with consideration in another setting. I have framed them as questions, suggestions and statements in a way that might be used in an oral presentation or conversation, reflecting the intention that they are not intended to be perceived as dogmatic guidance.

Figure 9.3 Insights and recommendations for those working in implementation practice

- Is there really a need for this new information in practice?
    - Does this need genuinely come from the proposed users of the information, i.e. is it supported by research with and/or understanding of this group?
    - If the perceived need comes from a theoretical perspective about how new information could improve practice, that may well be valuable, but be aware that this may well not align with the perspectives of the people who will be using it in the real world.
    - Think seriously about how to match the needs of people in the system or context with the wider aims of your project.

- Do you understand the context into which your new information will arrive?
    - Do you understand details such as workflow, patient journeys, points of decision-making, the role of different staff, or how local information systems work?
    - Consider at what place in a complex system your new information might arrive, who will deal with it, what will be done with it and who will be involved in this.

- Do you understand the ability of your target user to understand and react appropriately to the new information?
    - Don't assume users will have the interest, time, knowledge or headspace to absorb and act on new information – they are probably already overloaded. They may well have less understanding of scientific details than you imagine.
    - Research early on to find this out.

- Participatory co-design provides a way to address these important considerations.
    - There are formal methods to design processes, products and systems with and around those involved which identify the issues above. Even informal consideration of these is much better than nothing.
    - New ideas and initiatives are much more likely to succeed if they respond to a genuine need among those who are expected to be involved in implementing them.
    - If the new information or idea comes from a theoretical perspective of its value, it way well still be worth implementing (sometimes people don't know what their needs are), but think how to make this align with their needs and interests.
    - Involving users early, with genuine power to shape things is highly valuable.
    - Iterative testing of concepts and designs with users over time reveals new information at each step which can be used to achieve better results.

- Simplicity in design should reign supreme.

Figure 9.4 Insights and recommendations for communication of scientific evidence to clinicians

- Be aware that as someone with expert knowledge about clinical evidence who is seeking to share this with clinicians, you will have a radically different level of understanding to your audience.
    - Do not assume clinicians have sophisticated understanding of the language and principles of research. Even though they may have been taught this, they are now likely to be pragmatic practitioners, using mental short-cuts and experience-based knowledge to inform daily decisions rather than thinking in an analytic fashion.
    - There is little time and cognitive space for clinicians to absorb information, they will be looking for usable, key pieces of information.
    - Complex numerical or textual information is likely to be highly off-putting and will probably result in the readers gaining zero information.
- Clinicians do want simplicity, but…
    - they are also interested in deeper understanding, to apply knowledge to individual patients and scenarios.
    - The simplicity of a clinical guideline recommendation has much value. These will be used and appreciated, but could be supported by further layers of information to be used when necessary.
    - There is unavoidable tension between this need for simplicity on the one hand, and more complex detail on another.
- Tension between simplicity and complexity can be resolved using design principles.
    - Understanding the information needs of users as a first principle, then co-developing content with users are critical aspects of resolving this tension.
    - These processes will reveal what information is important, in what order, and how best to present it. This information structure and content will almost certainly be different to what you as an evidence expert imagines is ideal.
    - It is better that users get **some** extra information that is useful to them rather than getting **no** extra information because it has been so badly presented it has not even been read or understood.
- Tensions exists between presenting information with full scientific detail and in a way which will be read and understood.
    - Be happy with simpler, understandable information even if this jars with your scientific instincts. Better to improve understanding to **some** degree with an approximation, than to **not improve understanding at all** with something which is not read.
    - Scientific details, qualifiers and caveats can all be included in supplementary levels of information which can be accessed by those who are interested.
    - In any case, small margins of quantitative differences in data are unlikely to influence real-life practice. Clinicians and patients are unlikely to change their intentions over a small difference. Also, even a very accurately presented result from (say) a meta-analysis, becomes immediately approximate when applied to an individual patient who will almost certainly vary from the population average in the research.

9.6 Further questions and research

Three main questions stand out as priorities for future research. Firstly, there may be value in more structured testing with larger numbers of GPs to assess the quality of understanding achieved by reading content from *GP Evidence*. Such research would aim to understand if users "get it right", i.e. after reading content do they draw a correct scientific conclusion about a treatment's benefits or harms in a theoretical sense. This could be conducted in a formal setting such as an online survey, focus group or in-person workshop.

One challenge would be how to define what a "correct understanding" of *GP Evidence* content is. With regard to the basic quantitative information provided in and around the infographics, a plain language version of a correct answer is already provided below the icon arrays. In a formal test setting, users should easily find and be able to simply repeat this. It would be possible to design an electronic survey which explores which aspects of information users have noticed and internalised, for example: what was the time-frame over which a treatment benefit could be expected? Such a question could be asked after participants had been given a website page to read and then had access removed before seeing the question. Other aspects of information uptake which could be tested in this way would be details such as "Patient Characteristic" or individual outcomes of interest (presented as "Reduce Risk of"). Figure 9.5 shows a screengrab illustrating these elements of information. This kind of informational understanding could be analysed quantitatively, as responses could be easily categorised as correct or incorrect.

Figure 9.5 Screengrab of GP Evidence website showing aspects of informational content

The understanding of the more complex information contained in *GP Evidence* (e.g. free text explanations or content of the pop up buttons relating to evidence source and quality) would be more challenging to assess. This information is intended to contextualise the quantitative information, in order to support clinical judgement. This necessarily implies the use of an individual's thought process and judgement to apply the information to a specific scenario, such as thinking about "does this evidence apply to my patient?". Therefore, it is reasonably subjective and difficult to define a universally "correct" answer which could be analysed

quantitatively. One approach could be to use clinical vignettes to prompt exploration of specific website content, then provide a question designed to probe understanding of this. Answers could be given in free text or orally. These could then be analysed by researchers and graded according to accuracy of understanding. This would require careful question design, thought about how to define grading, and work to develop inter-rater reliability when assessing answers.

Secondly, the question of how *GP Evidence* is actually used in practice and what effect it has on clinical consultations and decision-making needs to be explored. This is a much more challenging area to research. How one might capture use in practice of a resource which might only be used infrequently is a difficult question. Ideally, ethnographic "fly on the wall" methods would be applied, but realistically may be impossible without setting up special scenarios where the use of *GP Evidence* is pre-determined, thus limiting the validity of observations.

A more realistic approach might be to prospectively set up a method to capture opportunistic use of *GP Evidence* in practice. GP volunteers could consent to the use of IP address tracking software on their computers which would alert researchers when the *GP Evidence* website had been used by the GP. Shortly afterwards, researchers could contact the GP and conduct a brief interview to explore what information was used, how it was understood and how it affected their decision-making and discussions with patients. A time stamp on the IP address tracking software could be used by the GPs to refer to their appointment and clinical records to support recall. This method has been successfully used in a study exploring clinical information retrieval technology in Canadian family practice (360).

Subject to ethical considerations and approval, this study design could also include identifying and seeking feedback from patients in whose consultations *GP Evidence* had been used. Qualitative analysis of both clinician and patient data would provide a real-world understanding of the effect of *GP Evidence* in practice including the kind of information used, its interpretation, if and how this was shared with patients and whether it influenced prescribing choices. This method would allow the capture of spontaneous moments of clinical practice including the real-world context and barriers to use of the website.

Thirdly, researching the use of *GP Evidence* by other healthcare professionals. Specialist nurses and clinical pharmacists play an increasing role in the management of long-term conditions and often make treatment decisions for or with patients. Researching how they understand and interact with content on *GP Evidence* would be valuable.

Two approaches could be taken here. One would be to start from the ground up with a particular professional group, using similar methods to the *GP Evidence* project with a user-needs focus and participatory co-design of a new product. This would establish an understanding of their current practice, understanding and informational needs. It is probable that a resulting information resource would look quite different to *GP Evidence*. Ideally, members of the relevant profession would be deeply involved as researchers as well as user-participants. This would of course be a long-term project and require significant funding.

Another approach would be to take *GP Evidence* and present it to particular clinicians, for example a primary care pharmacist involved in conducting structured medication reviews or a specialist nurse running a diabetes clinic. Various methods could be used, such as short interviews following a brief introduction to the site, interactive educational workshops with

structured collection of qualitative data from participant feedback, or vignette-based studies similar to the evaluation study in chapter 8 of this thesis. These would be realistic and relatively low intensity studies in terms of resource, and could provide good insight into the information needs and understanding of different groups. It may transpire that the content and design of *GP Evidence* translates well for other professionals, or that some minor updates to the website would help in this regard (though design changes benefiting one group may reduce the design quality for another). On the other hand, such research might reveal a need for a more extensive design process to fulfil a particular need for a particular group by producing an entirely new product.

9.6.1 Further design development and research

Ideas for further design research are:

- Improving the presentation of treatment harms.
- Building ways in which to summarise or collate collections of information from around the website to produce patient-specific information.
- Exploring whether patient-facing content is needed or appropriate and then co-designing that with patients.
- In the very long-term exploring how *GP Evidence* might integrate into a clinical system.

Such developments represent Zimmerman's criteria of extensibility with regard to research through design.

9.7 Personal reflection

One of the consequences of working so closely for so long on this project has been the difficulty of maintaining an objective view of the work and what has been achieved. Especially

at the end of the write up of a thesis. That said, I am proud of what *GP Evidence* became and hope that its effect in the world will be to improve the experience of healthcare for patients and the experience of providing care for GPs. However, I am aware that this cannot be taken for granted at all – there is much more to do in raising awareness of its existence, learning how it is being used and understood, and improving what it offers. I look forward to seeing how it evolves.

References

1. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. BMJ. 2012;344:e3502.
2. Sleath JD. In pursuit of normoglycaemia: the overtreatment of type 2 diabetes in general practice. Br J Gen Pract. 2015;65(636):334-5.
3. Heath I. Role of fear in overdiagnosis and overtreatment—an essay by Iona Heath. BMJ : British Medical Journal. 2014;349.
4. Long KM, McDermott F, Meadows GN. Being pragmatic about healthcare complexity: our experiences applying complexity theory and pragmatism to health services research. BMC Medicine. 2018;16(1):94.
5. Greenhalgh T, Engebretsen E. The science-policy relationship in times of crisis: An urgent call for a pragmatist turn. Social Science & Medicine. 2022;306:115140.
6. McCaslin M. Pragmatism. In: Given L, editor. The SAGE Encyclopaedia of Qualitative Research Methods: SAGE Publications Inc.; 2012.
7. Douven I. Abduction. In: Edward N. Zalta, editor. The Stanford Encyclopaedia of Philosophy: Metaphysics Research Lab, Standofrd University, ; 2021.
8. White K, Willis E. Positivism resurgent: the epistemological foundations of evidence-based medicine. Health Sociology Review. 2002;11(1-2):5-15.
9. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; San Jose, California, USA. 1240704: ACM; 2007. p. 493-502.
10. Gaver W. What should we expect from research through design? Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Austin, Texas, USA. 2208538: ACM; 2012. p. 937-46.
11. Zimmerman J, Stolterman E, Forlizzi J. An analysis and critique of Research through Design: towards a formalization of a research approach. Proceedings of the 8th ACM Conference on Designing Interactive Systems; Aarhus, Denmark. 1858228: ACM; 2010. p. 310-9.
12. Stompff G, van Bruinessen T, Smulders F. The generative dance of design inquiry: Exploring Dewey's pragmatism for design research. Design Studies. 2022;83:101136.
13. Moynihan R, Heneghan C, Godlee F. Too Much Medicine: from evidence to action. BMJ : British Medical Journal. 2013;347:f7141.
14. National Instituite for Health and Care Excellence. Shared Decision Making: NICE; [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/shared-decision-making#shared-decision-making-collaborative.
15. Academy of Medical Sciences. Enhancing the use of scientific evidence to judge the potential benefits and harms of medicines. London: Academy of Medical Sciences; 2017.
16. Malhotra A, Maughan D, Ansell J, Lehman R, Henderson A, Gray M, et al. Choosing Wisely in the UK: the Academy of Medical Royal Colleges' initiative to reduce the harms of too much medicine. BMJ : British Medical Journal. 2015;350:h2308.
17. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71.
18. Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, et al. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. JAMA. 1992;268(17):2420-5.
19. Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayian K, et al. Sicily statement on evidence-based practice. BMC Medical Education. 2005;5(1):1.
20. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. BMJ. 1998;316(7128):361-5.

21. Galbraith K, Ward A, Heneghan C. A real-world approach to Evidence-Based Medicine in general practice: a competency framework derived from a systematic review and Delphi process. BMC Medical Education. 2017;17(1):78.
22. Young JM, Ward JE. Evidence-based medicine in general practice: beliefs and barriers among Australian GPs. Journal of Evaluation in Clinical Practice. 2001;7(2):201-10.
23. McKenna HP, Ashton S, Keeney S. Barriers to evidence-based practice in primary care. Journal of Advanced Nursing. 2004;45(2):178-89.
24. Upton D, Upton P. Knowledge and use of evidence-based practice of GPs and hospital doctors. Journal of Evaluation in Clinical Practice. 2006;12(3):376-84.
25. Irwig L, Isaacs A, Trevena LJ, Barratt A. GPs want tailored, user friendly evidence summaries: a cross sectional study in New South Wales. Australian Family Physician. 2007;36(12).
26. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9):e1000326.
27. Field MJ, Lohr KN. Clinical practice guidelines. Washington DC; 1990.
28. Kleisiaris CF, Sfakianakis C, Papathanasiou IV. Health care practices in ancient Greece: The Hippocratic ideal. J Med Ethics Hist Med. 2014;7:6.
29. Davies E. Can you trust your clinical guidelines? BMJ. 2013;346:f4003.
30. Durand-Zaleski I, Colin C, Blum-Boisgard C. An attempt to save money by using mandatory practice guidelines in france. BMJ. 1997;315(7113):943-6.
31. Kosimbei G, Hanson K, English M. Do clinical guidelines reduce clinician dependent costs? Health Res Policy Syst. 2011;9:24.
32. Schwartz JAT, Pearson SD. Cost Consideration in the Clinical Guidance Documents of Physician Specialty Societies in the United States. JAMA Internal Medicine. 2013;173(12):1091-7.
33. Wennberg JE. Time to tackle unwarranted variations in practice. BMJ. 2011;342:d1513.
34. Armstrong D. Clinical autonomy, individual and collective: the problem of changing doctors' behaviour. Social Science & Medicine. 2002;55(10):1771-7.
35. Weisz G, Cambrosio A, Keating P, Knaapen L, Schlich T, Tournay VJ. The Emergence of Clinical Practice Guidelines. The Milbank Quarterly. 2007;85(4):691-727.
36. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing guidelines. BMJ. 1999;318(7183):593-6.
37. Eccles MP, Grimshaw JM, Shekelle P, Schünemann HJ, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. Implementation Science. 2012;7(1):60.
38. Shekelle P, Woolf S, Grimshaw JM, Schünemann HJ, Eccles MP. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. Implementation Science. 2012;7(1):62.
39. Woolf S, Schünemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. Implementation Science. 2012;7(1):61.
40. Ollenschläger G, Marshall C, Qureshi S, Rosenbrand K, Burgers J, Mäkelä M, et al. Improving the quality of health care: using international collaboration to inform guideline programmes by founding the Guidelines International Network (G-I-N). Quality and Safety in Health Care. 2004;13(6):455.
41. Agenecy for Healthcare Research and Quality. Guidleines and Measures ahrg.gov.uk [Available from: https://www.ahrq.gov/gam/index.html.
42. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. Cmaj. 2010;182(18):E839-42.

43. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924.
44. De Leo A, Bloxsome D, Bayes S. Approaches to clinical guideline development in healthcare: a scoping review and document analysis. BMC Health Services Research. 2023;23(1):37.
45. Littlejohns P, Chalkidou K, Culyer AJ, Weale A, Rid A, Kieslich K, et al. National Institute for Health and Care Excellence, social values and healthcare priority setting. Journal of the Royal Society of Medicine. 2019;112(5):173-9.
46. Catchpole P, Barrett V. Keeping pace with pharmaceutical innovation: the importance of the NICE methods review. PharmacoEconomics. 2020;38:901-3.
47. Sheard S AP. Origins and Establishment of NICE. In: Department of Public Health and Policy, editor. Witness Seminar Series. online2020.
48. Smith R. The triumph of NICE. BMJ : British Medical Journal. 2004;329(7459):0-.
49. Timmins R. A Terrible Beauty: A Short History of NICE. Muang, Thailand.: Amarin Ltd.; 2016 January 2016.
50. National Institute of Health and Care Excellence. NICE Strategy 2021 to 2026. online: NICE; 2021 April 2021.
51. Government. H. Life Sciences Vision. London; 2021.
52. Charlton V, Lomas J, Mitchell P. NICE's new methods: putting innovation first, but at what cost? BMJ. 2022;379:e071974.
53. National Institue for Health and Care Excellence. Developing NICE guidelines: the manual2022. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction.
54. Goldacre B. Bad Pharma. London: Fourth Estate; 2013.
55. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? BMJ : British Medical Journal. 2014;348.
56. Heneghan C, Mahtani KR, Goldacre B, Godlee F, Macdonald H, Jarvies D. Evidence based medicine manifesto for better healthcare. BMJ. 2017;357.
57. Smith R. bmj.com: BMJ Publishing. 2021. Available from: https://blogs.bmj.com/bmj/2021/07/05/time-to-assume-that-health-research-is-fraudulent-until-proved-otherwise/.
58. Steel N, Abdelhamid A, Stokes T, Edwards H, Fleetcroft R, Howe A, et al. A review of clinical practice guidelines found that they were often based on evidence of uncertain relevance to primary care patients. Journal of Clinical Epidemiology. 2014;67(11):1251-7.
59. Elwyn G, Quinlan C, Mulley A, Agoritsas T, Vandvik PO, Guyatt G. Trustworthy guidelines - excellent; customized care tools - even better. BMC Med. 2015;13:199.
60. Lenzer J. Why we can't trust clinical guidelines. BMJ : British Medical Journal. 2013;346:f3830.
61. Ioannidis John PA. Professional Societies Should Abstain From Authorship of Guidelines and Disease Definition Statements. Circulation: Cardiovascular Quality and Outcomes. 2018;11(10):e004889.
62. Moynihan RN, Cooke GPE, Doust JA, Bero L, Hill S, Glasziou PP. Expanding Disease Definitions in Guidelines and Expert Panel Ties to Industry: A Cross-sectional Study of Common Conditions in the United States. PLOS Medicine. 2013;10(8):e1001500.
63. The L. Managing conflicts of interests in clinical guidelines. The Lancet. 2019;394(10200):710.
64. Parker L, Bero L. Managing risk from conflicts of interest in guideline development committees. BMJ. 2022;379:e072252.
65. McCartney M. Have we given guidelines too much power? BMJ. 2014;349:g:6027.
66. Solomon J, Raynor DK, Knapp P, Atkin K. The compatibility of prescribing guidelines and the doctor-patient partnership: a primary care mixed-methods study. British Journal of General Practice. 2012;62(597):e275.

67. Mehta R, Lehman R. A Solution for Guideline Overkill—More Guidelines or Shared Understanding? JAMA Internal Medicine. 2020.
68. Spellberg B, Wright WF, Shaneyfelt T, Centor RM. The Future of Medical Guidelines: Standardizing Clinical Care With the Humility of Uncertainty. Annals of Internal Medicine. 2021.
69. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Clinical guidelines and the pursuit of reducing epistemic uncertainty. An ethnographic study of guideline development panels in three countries. Social Science & Medicine. 2021;272:113702.
70. Lee DH, Vielemeyer O. Analysis of overall level of evidence behind Infectious Diseases Society of America practice guidelines. Arch Intern Med. 2011;171(1):18-22.
71. Fanaroff AC, Califf RM, Windecker S, Smith SC, Jr., Lopes RD. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. Jama. 2019;321(11):1069-80.
72. Vandvik PO, Brandt L, Alonso-Coello P, Treweek S, Akl EA, Kristiansen A, et al. Creating Clinical Practice Guidelines We Can Trust, Use, and Share: A New Era Is Imminent. Chest. 2013;144(2):381-9.
73. McCormack JP, Martin SA, Newman DH. Comment on Inzucchi et al. Management of Hyperglycemia in Type 2 Diabetes, 2015: A Patient-Centered Approach. Update to a Position Statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetes Care 2015;38:140–149. Diabetes Care. 2015;38(9):e141.
74. Morgott M, Heinmüller S, Hueber S, Schedlbauer A, Kühlein T. Do guidelines help us to deviate from their recommendations when appropriate for the individual patient? A systematic survey of clinical practice guidelines. J Eval Clin Pract. 2020;26(3):709-17.
75. Treadwell J. Coping with complexity: working beyond the guidelines for patients with multimorbidities. Journal of Comorbidity. 2015;5(1):11-4.
76. McCartney M, Treadwell J, Maskrey N, Lehman R. Making evidence based medicine work for individual patients. BMJ. 2016;353:i2452.
77. Pulse magazine. Pulse panel: guidelines or tramlines? pulsetoday.co.uk: Pulse; 2017 [Available from: https://www.pulsetoday.co.uk/analysis/clinical-areas/prescribing/pulse-panel-guidelines-or-tramlines/.
78. Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ : British Medical Journal. 2012;345:e6341.
79. Berwick DM. Era 3 for Medicine and Health Care. JAMA. 2016;315(13):1329-30.
80. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet. 2003;362(9391):1225-30.
81. Coenen S, Van Royen P, Michiels B, Denekens J. Optimizing antibiotic prescribing for acute cough in general practice: a cluster-randomized controlled trial. J Antimicrob Chemother. 2004;54(3):661-72.
82. Simon SR, Majumdar SR, Prosser LA, Salem-Schatz S, Warner C, Kleinman K, et al. Group versus individual academic detailing to improve the use of antihypertensive medications in primary care: a cluster-randomized controlled trial. Am J Med. 2005;118(5):521-8.
83. Engers AJ, Wensing M, van Tulder MW, Timmermans A, Oostendorp RA, Koes BW, et al. Implementation of the Dutch low back pain guideline for general practitioners: a cluster randomized controlled trial. Spine (Phila Pa 1976). 2005;30(6):559-600.
84. Fretheim A, Oxman AD, Havelsrud K, Treweek S, Kristoffersen DT, Bjorndal A. Rational prescribing in primary care (RaPP): a cluster randomized trial of a tailored intervention. PLoS Med. 2006;3(6):e134.
85. Verhoeven V, Avonts D, Vermeire E, Debaene L, Van Royen P. A short educational intervention on communication skills improves the quality of screening for Chlamydia in GPs in Belgium: a cluster randomised controlled trial. Patient Educ Couns. 2005;57(1):101-5.

86. Holt TA, Dalton A, Marshall T, Fay M, Qureshi N, Kirkpatrick S, et al. Automated Software System to Promote Anticoagulation and Reduce Stroke Risk. Stroke. 2017;48(3):787.
87. Glasziou P, Haynes B. The paths from research to improved health outcomes. Evidence Based Nursing. 2005;8(2):36.
88. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. The Milbank Quarterly. 2004;82(4):581-629.
89. Lynch EA, Mudge A, Knowles S, Kitson AL, Hunter SC, Harvey G. "There is nothing so practical as a good theory": a pragmatic guide for selecting theoretical approaches for implementation projects. BMC Health Services Research. 2018;18(1):857.
90. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Science. 2009;4(1):50.
91. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? BMJ. 2016;354:i4060.
92. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. N Engl J Med. 2009;361(4):368-78.
93. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. BMJ : British Medical Journal. 2014;349:g6423.
94. Marshall M, Roland M. The future of the Quality and Outcomes Framework in England. BMJ. 2017;359.
95. Heath I, Hippisley-Cox J, Smeeth L. Measuring performance and missing the point? Bmj. 2007;335(7629):1075-6.
96. Phillips R, Cro S, Wheeler G, Bond S, Morris TP, Creanor S, et al. Visualising harms in publications of randomised controlled trials: consensus and recommendations. 2022;377:e068983.
97. Cupit C, Rankin J, Armstrong N, Martin GP. Overruling uncertainty about preventative medications: the social organisation of healthcare professionals' knowledge and practices. Sociology of Health & Illness. 2019;n/a(n/a).
98. Office for Health Improvement and Disparities. CVD Prevent: First Annual Report. online: Healthcare Quality Improvement Partnership,; 2021.
99. Owen J. Audit paints "bleak picture" of diabetes care in England and Wales. BMJ. 2018;363:k4780.
100. McCoy RG, Lipska KJ, Van Houten HK, Shah ND. Paradox of glycemic management: multimorbidity, glycemic control, and high-risk medication use among adults with diabetes. BMJ Open Diabetes Res Care. 2020;8(1).
101. Sussman M, Barnes GD, Guo JD, Tao CY, Gillespie JA, Ferri M, et al. The burden of undertreatment and non-treatment among patients with non-valvular atrial fibrillation and elevated stroke risk: a systematic review. Curr Med Res Opin. 2022;38(1):7-18.
102. Department of Health and Social Care. Primary Medical Services (Directed Enhanced Services) Directions. online: gov.uk; 2023.
103. Pathirana T, Clark J, Moynihan R. Mapping the drivers of overdiagnosis to potential solutions. BMJ. 2017;358:j3879.
104. Ashworth M, Lea R, Gray H, Rowlands G, Gravelle H, Majeed A. How are primary care organizations using financial incentives to influence prescribing? J Public Health (Oxf). 2004;26(1):48-51.
105. Bateman DN, Campbell M, Donaldson LJ, Roberts SJ, Smith JM. A prescribing incentive scheme for non-fundholding general practices: an observational study. Bmj. 1996;313(7056):535-8.

106. Ashworth M, Lea R, Gray H, Gravelle H, Majeed A. The development of prescribing incentive schemes in primary care: a longitudinal survey. Br J Gen Pract. 2003;53(491):468-70.
107. National Instituite for Health and Care Excellence. Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes

NICE guideline [NG5]. online: NICE,; 2015.
108. Picton C WH. Medicines Optimisation: Helping patients to make the most of medicines. Good practice guidance for healthcare professionals in England. online: Royal Pharmaceutical Society; 2013.
109. Spencer R, Bell B, Avery AJ, Gookey G, Campbell SM. Identification of an updated set of prescribing--safety indicators for GPs. Br J Gen Pract. 2014;64(621):e181-90.
110. Bennie M, Kurdi A, MacBride-Stewart S, Avery T. Medication safety in primary care-from measurement to action. Drug Ther Bull. 2021;59(2):24-8.
111. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M, et al. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. The Lancet. 2012;379(9823):1310-9.
112. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer Prescribing--A Trial of Education, Informatics, and Financial Incentives. N Engl J Med. 2016;374(11):1053-64.
113. Schedlbauer A, Prasad V, Mulvaney C, Phansalkar S, Stanton W, Bates DW, et al. What Evidence Supports the Use of Computerized Alerts and Prompts to Improve Clinicians' Prescribing Behavior? Journal of the American Medical Informatics Association. 2009;16(4):531-8.
114. Olakotan OO, Mohd Yusof M. The appropriateness of clinical decision support systems alerts in supporting clinical workflows: A systematic review. Health Informatics Journal. 2021;27(2):14604582211007536.
115. Department of Health and Social Care. Good for you, good for us, good for everybody. A plan to reduce overprescribing to make patient care better and safer, support the NHS and reduce carbon emmissions. Report. GOV.UK; 2021.
116. Ogden J. NHS overprescribing: the national review recommendations. Prescriber. 2022;33(2):35-8.
117. Treadwell J. National overprescribing review for England: another step forward? Drug and Therapeutics Bulletin. 2022;60(2):18.
118. Tierney S, Wong G, Roberts N, Boylan A-M, Park S, Abrams R, et al. Supporting social prescribing in primary care by linking people to local assets: a realist review. BMC Medicine. 2020;18(1):49.
119. Moore M. Antibiotic stewardship: where next? British Journal of General Practice. 2023;73(728):100-1.
120. Chaplin S. Safe prescribing and withdrawal of dependence-forming medicines. Prescriber. 2022;33(7):29-30.
121. Medicines Optimisations Quality Standard [QS120] [Internet]. 2016. Available from: https://www.nice.org.uk/guidance/qs120.
122. Stott NC, Davis RH. The exceptional potential in each primary care consultation. J R Coll Gen Pract. 1979;29(201):201-5.
123. Polly D, Matthew JR, Deborah M, Bruce G, Christie C. Barriers and enablers to collaborative working between GPs and pharmacists: a qualitative interview study. British Journal of General Practice. 2020;70(692):e155.
124. Doherty AJ, Boland P, Reed J, Clegg AJ, Stephani A-M, Williams NH, et al. Barriers and facilitators to deprescribing in primary care: a systematic review. BJGP Open. 2020;4(3):bjgpopen20X101096.

125. Green AR, Aschmann H, Boyd CM, Schoenborn N. Assessment of Patient-Preferred Language to Achieve Goal-Aligned Deprescribing in Older Adults. JAMA Network Open. 2021;4(4):e212633-e.
126. Avery AJ, Bell BG. Rationalising medications through deprescribing. BMJ. 2019;364:l570.
127. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. eClinicalMedicine. 2023;57.
128. National Institiute for Health and Care Excellence. Multimorbidity: clinical assesment and management (NG56). 2016.
129. Kuan V, Denaxas S, Patalay P, Nitsch D, Mathur R, Gonzalez-Izquierdo A, et al. Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study. The Lancet Digital Health. 2023;5(1):e16-e27.
130. Doust JA, Bell KJL, Glasziou PP. Potential Consequences of Changing Disease Classifications. JAMA. 2020.
131. Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, et al. Multimorbidity. Nature Reviews Disease Primers. 2022;8(1):48.
132. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380(9836):37-43.
133. Cassell A, Edwards D, Harshfield A, Rhodes K, Brimicombe J, Payne R, et al. The epidemiology of multimorbidity in primary care: a retrospective cohort study. British Journal of General Practice. 2018;68(669):e245-e51.
134. Delara M, Murray L, Jafari B, Bahji A, Goodarzi Z, Kirkham J, et al. Prevalence and factors associated with polypharmacy: a systematic review and meta-analysis. BMC Geriatrics. 2022;22(1):601.
135. Deuerden M AT, Payne R. Polypharmacy and Medicines Optimisation. London: Kings Fund,; 2013.
136. Guthrie B, Makubate B, Hernandez-Santiago V, Dreischulte T. The rising tide of polypharmacy and drug-drug interactions: population database analysis 1995-2010. BMC Med. 2015;13:74.
137. Academy of Medical Sciences. Multimorbidity: a priority for global health research. online; 2018.
138. Charities RGo. "Just one thing after another". Living with multiple conditions. online; 2018.
139. Rosbach M, Andersen JS. Patient-experienced burden of treatment in patients with multimorbidity – A systematic review of qualitative data. PLOS ONE. 2017;12(6):e0179916.
140. May C, Montori VM, Mair FS. We need minimally disruptive medicine. BMJ. 2009;339:b2803.
141. Bruce M, Veronica N, Eleni E, Scott AM, Anne D, Cathy S, et al. 'My body's falling apart.' Understanding the experiences of patients with advanced multimorbidity to improve care: serial interviews with patients and carers. BMJ Supportive & Palliative Care. 2016;6(1):60.
142. Gill A, Kuluski K, Jaakkimainen L, Naganathan G, Upshur R, Wodchis WP. "Where do we go from here?" Health system frustrations expressed by patients with multimorbidity, their caregivers and family physicians. Healthc Policy. 2014;9(4):73-89.
143. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.
144. Payne RA, Abel GA, Avery AJ, Mercer SW, Roland MO. Is polypharmacy always hazardous? A retrospective cohort analysis using linked electronic health records from primary and secondary care. British Journal of Clinical Pharmacology. 2014;77(6):1073-82.
145. Bokhof B, Junius-Walker U. Reducing Polypharmacy from the Perspectives of General Practitioners and Older Patients: A Synthesis of Qualitative Studies. Drugs & Aging. 2016;33(4):249-66.

146. Anthierens S, Tansens A, Petrovic M, Christiaens T. Qualitative insights into general practitioners views on polypharmacy. BMC Family Practice. 2010;11(1):65.
147. Sinnott C, Mc Hugh S, Browne J, Bradley C. GPs' perspectives on the management of patients with multimorbidity: systematic review and synthesis of qualitative research. BMJ Open. 2013;3(9):e003610.
148. Djulbegovic B, Elqayam S, Dale W. Rational decision making in medicine: Implications for overuse and underuse. Journal of Evaluation in Clinical Practice. 2018;24(3):655-65.
149. Mair A, Wilson M, Dreischulte T. The polypharmacy programme in Scotland: realistic prescribing. Prescriber. 2019;30(8):10-6.
150. All Wales Medicines Strategy Group. Polypharmacy: Guidance for Prescribing. 2014.
151. Royal Pharmaceutical Society. Polypharmacy: Getting our medicines right. London: RPS on-line; 2019.
152. Laursen J, Kornholt J, Betzer C, Petersen TS, Christensen MB. General Practitioners' Barriers Toward Medication Reviews in Polymedicated Multimorbid Patients: How can a Focus on the Pharmacotherapy in an Outpatient Clinic Support GPs? Health Serv Res Manag Epidemiol. 2018;5:2333392818792169.
153. Hélène C, Anna Z, Aurélie B, Patrick V, Hélène V, Martin F, et al. GPs' management of polypharmacy and therapeutic dilemma in patients with multimorbidity: a cross-sectional survey of GPs in France. British Journal of General Practice. 2019;69(681):e270.
154. Kristen A, Danielle S, Christopher F, Ian S. Prescriber barriers and enablers to minimising potentially inappropriate medications in adults: a systematic review and thematic synthesis. BMJ Open. 2014;4(12):e006544.
155. Salisbury C, Man M-S, Bower P, Guthrie B, Chaplin K, Gaunt DM, et al. Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. The Lancet. 2018;392(10141):41-50.
156. Farmer C, Fenu E, O'Flynn N, Guthrie B. Clinical assessment and management of multimorbidity: summary of NICE guidance. BMJ. 2016;354.
157. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999;282(15):1458-65.
158. Farquhar CM, Kofa EW, Slutsky JR. Clinicians' attitudes to clinical practice guidelines: a systematic review. Med J Aust. 2002;177(9):502-6.
159. Carlsen B, Norheim OF. "What lies beneath it all?" – an interview study of GPs' attitudes to the use of guidelines. BMC Health Services Research. 2008;8(1):218.
160. Carlsen B, Kjellberg PK. Guidelines; from foe to friend? Comparative interviews with GPs in Norway and Denmark. BMC Health Services Research. 2010;10(1):17.
161. Carlsen B, Bringedal B. Attitudes to clinical guidelines—do GPs differ from other medical doctors? BMJ Quality & Safety. 2011;20(2):158-62.
162. Clerc I, Ventelou B, Guerville MA, Paraponaris A, Verger P. General practitioners and clinical practice guidelines: a reexamination. Med Care Res Rev. 2011;68(4):504-18.
163. Wangler J, Jansky M. What is the significance of guidelines in the primary care setting? Wiener Medizinische Wochenschrift. 2021;171(13):321-9.
164. Baynouna Al Ketbi LM, Zein Al Deen S. The attitudes and beliefs of general practitioners towards clinical practice guidelines: a qualitative study in Al Ain, United Arab Emirates. Asia Pac Fam Med. 2018;17:5.
165. Milos V, Westerlund T, Midlöv P, Strandberg EL. Swedish general practitioners' attitudes towards treatment guidelines - a qualitative study. BMC Fam Pract. 2014;15:199.
166. Austad B, Hetlevik I, Mjølstad BP, Helvik A-S. Applying clinical guidelines in general practice: a qualitative study of potential complications. BMC Family Practice. 2016;17(1):92.
167. Ahmed H, Naik G, Willoughby H, Edwards AG. Communicating risk. BMJ. 2012;344:e3996.

168. Hoffmann TC, Del Mar C. Clinicians' Expectations of the Benefits and Harms of Treatments, Screening, and Tests: A Systematic Review. JAMA Intern Med. 2017;177(3):407-19.
169. Treadwell JS, Wong G, Milburn-Curtis C, Feakins B, Greenhalgh T. GPs' understanding of the benefits and harms of treatments for long-term conditions: an online survey. BJGP Open. 2020:bjgpopen20X101016.
170. Jones NR, Hobbs FDR, Taylor CJ. Underuse of anticoagulation therapy for atrial fibrillation. Canadian Family Physician. 2017;63(12):943.
171. Mas Dalmau G, Sant Arderiu E, Enfedaque Montes MB, Solà I, Pequeño Saco S, Alonso Coello P. Patients' and physicians' perceptions and attitudes about oral anticoagulation and atrial fibrillation: a qualitative systematic review. BMC Family Practice. 2017;18(1):3.
172. Goldacre B, Smeeth L. Mass treatment with statins. BMJ : British Medical Journal. 2014;349.
173. Goldacre BM. NICE must do better at summarising and communicating evidence on statins. BMJ : British Medical Journal. 2014;349.
174. Agoritsas T, Heen AF, Brandt L, Alonso-Coello P, Kristiansen A, Akl EA, et al. Decision aids that really promote shared decision making: the pace quickens. The BMJ. 2015;350:g7624.
175. Siemieniuk RA, Agoritsas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. BMJ. 2016;354:i5191.
176. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared Decision Making: A Model for Clinical Practice. Journal of General Internal Medicine. 2012;27(10):1361-7.
177. Oshima Lee E, Emanuel EJ. Shared decision making to improve care and reduce costs. N Engl J Med. 2013;368(1):6-8.
178. Hoffmann TC, Montori VM, Del Mar C. The connection between evidence-based medicine and shared decision making. JAMA. 2014;312(13):1295-6.
179. Maskrey N, Gordon A. Shared understanding with patients. JAMA Internal Medicine. 2017.
180. Bonner C, Trevena LJ, Gaissmaier W, Han PKJ, Okan Y, Ozanne E, et al. Current Best Practice for Presenting Probabilities in Patient Decision Aids: Fundamental Principles. Medical Decision Making. 2021:0272989X21996328.
181. Grad R, Sandhu A, Ferrante M, D'Souza V, Puterman-Salzman L, Rahimi S, et al. Using incorpoRATE to examine clinician willingness to engage in shared decision making: a study of Family Medicine residents. Patient Education and Counseling. 2022.
182. Stacey D, Légaré F, Boland L, Lewis KB, Loiselle M-C, Hoefel L, et al. 20th Anniversary Ottawa Decision Support Framework: Part 3 Overview of Systematic Reviews and Updated Framework. Medical Decision Making. 2020;40(3):379-98.
183. Shared decision making NICE guideline [NG197] [Internet]. 2021 [cited 19/04/2023]. Available from: https://www.nice.org.uk/guidance/ng197.
184. Shared decision making learning package [Internet]. NICE. 2021. Available from: https://www.nice.org.uk/guidance/ng197/resources/shared-decision-making-learning-package-9142488109.
185. National Instituite for Health and Care Excellence. Patient Decision Aids (online collection) 2019 [17/06/2019]. Available from: https://www.evidence.nhs.uk/search?om=%5b%7b%22ety%22:%5b%22Patient%20Decision%20Aids%22%5d%7d,%7b%22srn%22:%5b%22National%20Institute%20for%20Health%20and%20Care%20Excellence%20-%20NICE%22%5d%7d%5d.
186. Institute PC. Decision Support Tools - online collection. online2022.
187. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database of Systematic Reviews. 2017(4).
188. Joseph-Williams N, Elwyn G, Edwards A. Knowledge is not power for patients: a systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. Patient Educ Couns. 2014;94(3):291-309.

189. Pel-Littel RE, Snaterse M, Teppich NM, Buurman BM, van Etten-Jamaludin FS, van Weert JCM, et al. Barriers and facilitators for shared decision making in older patients with multiple chronic conditions: a systematic review. BMC Geriatrics. 2021;21(1):112.
190. Coulter A, Collins A, Edwards A, Entwistle V, Finnikin S, Joseph-Williams N, et al. Implementing shared decision-making in UK: Progress 2017–2022. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2022;171:139-43.
191. Douglas F, Petrie KJ, Cundy T, Horne A, Gamble G, Grey A. Differing perceptions of intervention thresholds for fracture risk: a survey of patients and doctors. Osteoporosis International. 2012;23(8):2135-40.
192. MacLean S, Mulla S, Akl EA, Jankowski M, Vandvik PO, Ebrahim S, et al. Patient Values and Preferences in Decision Making for Antithrombotic Therapy: A Systematic Review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e1S-e23S.
193. Albarqouni L, Doust J, Glasziou P. Patient preferences for cardiovascular preventive medication: a systematic review. Heart. 2017.
194. Qadi O, Lufumpa N, Adderley N, Bem D, Marshall T, Kokab F. Patients' and health professionals' attitudes and perceptions towards the initiation of preventive drugs for primary prevention of cardiovascular disease: a systematic review of qualitative studies. BJGP Open. 2020:bjgpopen20X101087.
195. Deaprtment of Health. Liberating the NHS: No decision about me, without me - Further consultation on proposals to secure shared decsion making. 2012.
196. A. C. Making shared decision making a reality. online; 2011.
197. Levinson W, Kao A, Kuby A, Thisted RA. Not all patients want to participate in decision making. A national study of public preferences. J Gen Intern Med. 2005;20(6):531-5.
198. Ipsos MORI. Public dialogue on medical evidence. Report prepared for the Academy of Medical Scinces. 2016 June 2017.
199. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. 2013.
200. Pope CM, Nicholas. Qualitative Research in Health Care. 3rd ed. Malden,Mass.: Blackwell Pub.; 2006.
201. Ritchie J SL. Qualitative Data Analysis for Applied Policy Research. AnalQualData,. London: Routledge; 1994. p. 173-94.
202. Green J, Thorogood N. Qualitative methods for health research. Third ed. London: SAGE Publications Ltd; 2013.
203. Braun VC, V. Successful Qualitative Research. 1st ed. London: SAGE; 2013.
204. Ritchie J, Lewis J, McNaughton Nicholls C, Ormston R. Qualitative research practice : a guide for social science students and researchers. Second ed. London: SAGE Publications Ltd; 2013.
205. Morse JM. Determining Sample Size. Qualitative Health Research. 2000;10(1):3-5.
206. Given L. Data Saturation. 1-10. The SAGE Encyclopedia of Qualitative Research Methods: SAGE Publications Inc; 2008.
207. Mays N, Pope C. Qualitative Research: Rigour and qualitative research. BMJ. 1995;311(6997):109-12.
208. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? Qualitative Health Research. 2016;26(13):1802-11.
209. Given L. Member Check. The SAGE Encyclopedia of Qualitative Research Methods,. Thousand Oaks, California2008.
210. Sinnott C, Hugh SM, Boyce MB, Bradley CP. What to give the patient who has everything? A qualitative study of prescribing for multimorbidity in primary care. Br J Gen Pract. 2015;65(632):e184-91.

211. Carlsen B, Glenton C, Pope C. Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines. Br J Gen Pract. 2007;57(545):971-8.
212. Elwyn G, Vermunt NPCA. Goal-Based Shared Decision-Making: Developing an Integrated Model. Journal of Patient Experience. 2019:2374373519878604.
213. Maskrey N. Shared decision making: why the slow progress? An essay by Neal Maskrey. BMJ. 2019;367:l6762.
214. Mulley AG, Trimble C, Elwyn G. Stop the silent misdiagnosis: patients' preferences matter. BMJ : British Medical Journal. 2012;345:e6572.
215. Gabbay J, le May A. Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care. BMJ. 2004;329(7473):1013.
216. MacWalter G, McKay J, Bowie P. Utilisation of internet resources for continuing professional development: a cross-sectional survey of general practitioners in Scotland. BMC Medical Education. 2016;16(1):24.
217. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. Psychol Sci Public Interest. 2007;8(2):53-96.
218. Wegwarth O GG. Statistical Illiteracy in Doctors. In: Better Doctors,Better Patients,Better Decisions: Strungmann Forum Reports; 2011.
219. Wegwarth O, Gigerenzer G. The Barrier to Informed Choice in Cancer Screening: Statistical Illiteracy in Physicians and Patients. In: Goerling U, Mehnert A, editors. Psycho-Oncology. Cham: Springer International Publishing; 2018. p. 207-21.
220. ACM CHI Conference on Human Factors in Computing Systems.
221. Simera I. EQUATOR Network collates resources for good research. BMJ. 2008;337:a2471.
222. Bodker S. Creating Conditions for Participation: Conflicts and Resources in Systems Development. Human–Computer Interaction. 1996;11(3):215-36.
223. N. C, editor Design Participation. Proceedings of the Design Research Society International Conference; 1972; London: Design Research Society.
224. Mattias N. Co-design in a historical context [Internet]. Centre for Civic Media: Massahusetts Institute of Technology. 2011. [04/05/2023]. Available from: https://civic.mit.edu/index.html%3Fp=407.html.
225. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health – a perspective on definitions and distinctions. Public Health Research & Practice.
226. Beacon Strategies. 2023. [04/05/2023]. Available from: https://beaconstrategies.net/beacon-strategies-blog/key-elements-and-principles-of-co-design.
227. Wherton J, Sugarhood P, Procter R, Hinder S, Greenhalgh T. Co-production in practice: how people with assisted living needs can help design and evolve technologies and services. Implement Sci. 2015;10:75.
228. Papoutsi C, Wherton J, Shaw S, Morrison C, Greenhalgh T. Putting the social back into sociotechnical: Case studies of co-design in digital health. Journal of the American Medical Informatics Association. 2021;28(2):284-93.
229. Hensley J KB. mural.co. 2023. [04/04/2023]. Available from: https://www.mural.co/blog/co-design-method.
230. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory Design in Health Sciences: Using Cooperative Experimental Methods in Developing Health Services and Computer Technology. Qualitative Health Research. 2007;17(1):122-30.
231. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth Interventions With Children and Young People. Frontiers in Psychiatry. 2018;9.
232. Robert G. Participatory action research: using experience-based co-design to iprove the quality of healthcare services. In: Zeibland S CA, Calabrese J, Locock L., editor. Understanding and Using Health Experiences. Oxford: Oxford University Press; 2013. p. 141.

233. Norman DA. The design of everyday things. Revised and expanded ed. Cambridge, Massachusetts: MIT Press; 2013.
234. IDEO. Insights for Innovation, Learn to use insights to turn customer needs into human-centred solutions. Online course.2019. Available from: https://www.ideou.com/products/insights-for-innovation.
235. von Hippel E. Lead Users: A Source of Novel Product Concepts. Management Science. 1986;32(7):791-805.
236. Urban GL, von Hippel E. Lead User Analyses for the Development of New Industrial Products. Management Science. 1988;34(5):569-82.
237. J. N. Bridging the Designer-User Gap [Internet]: nngroup.com. 2008. [04/05/2023]. Available from: https://www.nngroup.com/articles/bridging-the-designer-user-gap/.
238. Ball J. The Double Diamond: A universally accepted depiction of the design process.2019. Available from: https://www.designcouncil.org.uk/our-work/news-opinion/double-diamond-universally-accepted-depiction-design-process/.
239. Pereira A FJ, Lopes A. Front End of Innovation: An Integrative Literature Review. JIM. 2017;5(1):22-39.
240. Brown T. Design thinking. Harv Bus Rev. 2008;86(6):84-141.
241. Archer B. Design as a discipline. Design Studies. 1979;1(1):17-20.
242. McLaughlin JE, Wolcott MD, Hubbard D, Umstead K, Rider TR. A qualitative review of the design thinking framework in health professions education. BMC Medical Education. 2019;19(1):98.
243. IDEO. Design Thinking Defined online [updated 05/05/2023. Available from: https://designthinking.ideo.com/.
244. J. C. Human Computer Interaction- a brief introduction. The Encyclopaedia of Human-Computer Interaction. 2nd ed: Interaction Design Foundation.
245. Shneiderman B, Plaisant C. Designing the user interface : strategies for effective human-computer interaction. 5th ed. Upper Saddle River, N.J.

London: Pearson Education; 2010.
246. Nielsen J. Ten usability heuristics. http://www. nngroup. com/articles/ten-usability-heuristics/(acc-essed …; 2005.
247. Nielsen J, Molich R, editors. Heuristic evaluation of user interfaces. Proceedings of the SIGCHI conference on Human factors in computing systems; 1990.
248. team. UGSS. Understanding WCAG 2.1 United Kingdom: UK Governement; [Available from: https://www.gov.uk/service-manual/helping-people-to-use-your-service/understanding-wcag.
249. Pirolli PC, S. Infomration Foraging. Psychological Review. 1999;106(4):643-75.
250. R. B. Nielsen Norman Group. 2019. Available from: https://www.nngroup.com/articles/information-foraging/.
251. Hanington B, Martin B. Universal Methods of Design : 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions. Osceola, UNITED STATES: Rockport Publishers; 2012.
252. Rubin J CD. Handbook of Usability Testing. Indianapolis, Indiana: Wiley Publishing Inc.; 2008.
253. Nielsen J. Thinking aloud: The #1 Usability Tool online: Nielsen Norman Group; 2012 [Available from: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/.
254. Lucassen G, van de Keuken M, Dalpiaz F, Brinkkemper S, Sloof GW, Schlingmann J, editors. Jobs-to-be-Done Oriented Requirements Engineering: A Method for Defining Job Stories. Requirements Engineering: Foundation for Software Quality; 2018 2018//; Cham: Springer International Publishing.
255. Vitali S. UX Collective. 2019. [12/05/2023]. Available from: https://uxdesign.cc/when-personas-and-user-stories-arent-what-you-need-c21842302dbb.

256. Andrews DC. JAD: A Crucial Dimension for Rapid Applications Development. Journal of Systems Management. 1991;42(3):23.
257. Winters S. Content Design London2019. [17/05/2023]. Available from: https://contentdesign.london/blog/what-is-content-design.
258. Winters S. Content Design London2016. [17/05/2023]. Available from: https://contentdesign.london/blog/history-of-content-design-in-the-uk-government.
259. Content design: planning writing and managing content. [Internet]. 2016 [cited 17/05/2023]. Available from: https://www.gov.uk/guidance/content-design/writing-for-gov-uk.
260. Digital N. NHS Digital Services Manual: Content style guide [Available from: https://service-manual.nhs.uk/content.
261. Winters S. Content Design. United Kingdom: Content Design London; 2017.
262. Nielsen J. How users read on the web: Nielsen Norman Group; 1997 [Available from: https://www.nngroup.com/articles/how-users-read-on-the-web/.
263. London. CD. Readability Guideliens Handbook. United Kingdom: Content Design London; 2019.
264. Bruce L. United Kingdom: gathercontent.com. 2020. [17/05/2023]. Available from: https://gathercontent.com/blog/pair-writing-with-subject-matter-experts-multi-disciplinary-team-members-or-content-designers?utm_sq=gjq74n7ppj&utm_campaign=Newsletter_Q3_2020&utm_medium=email&_hsmi=93374657&_hsenc=p2ANqtz-_Zt6d2-EV3vdSuf8qNrI4UWAZzAqW_U63yJNmnAPRE8UrdXGoUaW5-tkktEQTs0yMUtZY1TT6htZodeJwlbh0XBP3j6Q&utm_content=93361276&utm_source=hs_email.
265. Davis S. Government Digital Service. 2016. [17/05/2023]. Available from: https://gds.blog.gov.uk/2016/09/21/it-takes-2-how-we-use-pair-writing/.
266. Jackson H MK. University of Bristol. 2020. [17/05/2023]. Available from: https://digital-comms.blogs.bristol.ac.uk/2020/11/09/pair-writing-why-we-love-it/.
267. Bishop A. ONS Digital. 2021. [17/05/2023]. Available from: https://digitalblog.ons.gov.uk/2021/01/26/how-writing-in-pairs-improves-content/.
268. Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. BMJ. 2010;341:c4830.
269. Garcia-Retamero R, Hoffrage U. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. Social Science & Medicine. 2013;83(Supplement C):27-33.
270. Spiegelhauter D GJ. What Can Education Learn from Real-World Communication of Risk and Uncertainty? The Mathemeatics Enthusiast. 2015;12(1).
271. Academy of Medical Sciences. Perspectives on “Communicating Evidence about Medicines”. 2016.
272. Freeman ALJ. How to communicate evidence to patients. Drug and Therapeutics Bulletin. 2019;57(8):119.
273. Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. Ann Intern Med. 2011;155(2):87-96.
274. Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa RA, Mbuagbaw L, et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. J Clin Epidemiol. 2016;74:7-18.
275. Krug S. Rocket surgery made easy : the do-it-yourself guide to finding and fixing usability problems. Berkeley, Calif.: New Riders; 2010.
276. Rubin J, Chisnell D. Handbook of usability testing : how to plan, design, and conduct effective tests. 2nd ed. Indianapolis, IN: Wiley Pub.; 2008.
277. Krug S. Rocket Surgery Made Easy. Hoboken: Hoboken: Pearson Education, Limited; 2009.
278. Group. NN. User Testing [Available from: https://www.nngroup.com/topic/user-testing/.

279. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evidence Based Medicine. 2016;21(4):125-7.
280. RoB2 development group. Revised Cochrane risk-of-bias tool for randomised trials: Cochrane Collaboration.; 2019 [Available from: https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2.
281. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther. 2008;88(2):156-75.
282. Diao D, Wright JM, Cundiff DK, Gueyffier F. Pharmacotherapy for mild hypertension. Cochrane Database Syst Rev. 2012(8):CD006742.
283. Ettehad D, Emdin CA, Kiran A, Anderson SG, Callender T, Emberson J, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet. 2016;387(10022):957-67.
284. Brunström M, Carlberg B. Association of blood pressure lowering with mortality and cardiovascular disease across blood pressure levels: A systematic review and meta-analysis. JAMA Internal Medicine. 2017.
285. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. Bmj. 2006;333(7565):417.
286. Montori V LA, Buchholz A, Stilwell D, Tsapas A. . Basing information on comprehensive, critically appraised,

and up-to-date syntheses of the scientific evidence. H. VRL-T, editor2012.
287. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. Second ed. Hoboken, NJ.

Chichester, West Sussex: Wiley-Blackwell; 2019.
288. National Institiute for Health and Care Excellence. Developing NICE guidelines: the manual. 2020.
289. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.
290. Montori VM, LeBlanc A, Buchholz A, Stilwell DL, Tsapas A. Basing information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence: a quality dimension of the International Patient Decision Aid Standards. BMC Med Inform Decis Mak. 2013;13 Suppl 2:S5.
291. Elwyn G, Burstin H, Barry MJ, Corry MP, Durand MA, Lessler D, et al. A proposal for the development of national certification standards for patient decision aids in the US. Health Policy. 2018;122(7):703-6.
292. National Institiute for Health and Care Excellence. Shared Decision Making online [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/shared-decision-making.
293. National institiute for Health and Care Excellence. NICE decision aids: process guide. online2018.
294. National Institiute for Health and Care Excellence. Patient decision aid: user guide for healthcare professionals.

Implementing the NICE guideline on lipid modification (CG181) 2014 [Available from: https://www.nice.org.uk/guidance/cg181/resources/patient-decision-aid-user-guide-pdf-243780158.
295. National Institiute for Health and Care Excellence. Abortion: choosing between medical or surgical abortion. Decision aid: user guide and data sources online2023 [Available from:

https://www.nice.org.uk/guidance/ng140/resources/decision-aid-user-guide-and-data-sources-pdf-6906582253.
296. National Institiute for Health and Care Excellence. How do I control my blood pressure? Lifestyle options and choice of medicines. Patient decision aid: user guide and data sources 2019 [Available from: https://www.nice.org.uk/guidance/ng136/resources/how-do-i-control-my-blood-pressure-lifestyle-options-and-choice-of-medicines-patient-decision-aid-user-guide-and-data-sources-pdf-6899918222.
297. Higgins JPT. Cochrane handbook for systematic reviews of interventions. Second ed. Hoboken, New Jersey: Cochrane; 2019.
298. Gibbons S. UX Stakeholder Engagement 101: Nielsen Norman Group; 2023 [Available from: https://www.nngroup.com/articles/stakeholder-engagement/.
299. National Institiute for Health and Care Excellence. Taking a statin to reduce the risk of coronary heart disease and stroke. Patient decision aid. 2014.
300. Excellence. NIfHaC. Cardiovascular disease: Risk Assessment and reduction, including lipid modification 2014. CG 181. London2014.
301. Aertgeerts B, Agoritsas T, Siemieniuk RAC, Burgers J, Bekkering GE, Merglen A, et al. Corticosteroids for sore throat: a clinical practice guideline. BMJ. 2017;358:j4090.
302. Excellence NIfHaC. Lipid modification. Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. Methods, evidence and recommendations. online: NICE; 2014.
303. Taylor F, Huffman MD, Macedo AF, Moore THM, Burke M, Davey Smith G, et al. Statins for the primary prevention of cardiovascular disease. Cochrane Database of Systematic Reviews. 2013(1).
304. National Institiute for Health and Care Excellence. Chronic Kidney Disease. Appendix H. Evidence reviews for interventions to lower proteinuria. . online: NICE; 2021.
305. National Institiute for Health and Care Excellence. Chronic kidney disease: assessment and management/ NICE Guideline [NG203]. online: NICE; 2021.
306. Ruggenenti P, Perna A, Gherardi G, Garini G, Zoccali C, Salvadori M, et al. Renoprotective properties of ACE-inhibition in non-diabetic nephropathies with non-nephrotic proteinuria. The Lancet. 1999;354(9176):359-64.
307. National institiute for Health and Care Excellence. Acute coronary syndromes. NICE guideline [NG185] online: NICE; 2020 [Available from: https://www.nice.org.uk/guidance/ng185.
308. Dibben G, Faulkner J, Oldridge N, Rees K, Thompson DR, Zwisler AD, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database of Systematic Reviews. 2021(11).
309. Pavy B, Iliou MC, Meurin P, Tabet J-Y, Corone S, Evaluation F, et al. Safety of Exercise Training for Cardiac Patients: Results of the French Registry of Complications During Cardiac Rehabilitation. Archives of Internal Medicine. 2006;166(21):2329-34.
310. Critical Appraisal Skills Programme. CASP online2022 [Available from: https://casp-uk.net/images/checklist/documents/CASP-Cohort-Study-Checklist/CASP-Cohort-Study-Checklist-2018_fillable_form.pdf.
311. National Institiute for Health and Care Excellence. Secondary prevention in primary and secondary care for patients following a myocardial infarction. Partial update of NICE CG48. Methods, evidence and recommendations online: NICE; 2013 [Available from: https://www.nice.org.uk/guidance/ng185/evidence/j-rehabilitation-and-secondary-prevention-2013-evidence-reviews-pdf-8949422993.
312. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). The Lancet. 2003;362(9386):782-8.
313. National Institiute for Health and Care Excellence. Type 2 diabetes [B] Pharmacological therapies with cardiovascular and other benefits in people with type 2 diabetes

NICE guideline NG28 Evidence review underpinning recommendations 1.7.4-1.7.6 and 1.7.9-1.7.16 in the NICE guideline online: NICE; 2022 [Available from: https://www.nice.org.uk/guidance/ng28/evidence/b-pharmacological-therapies-with-cardiovascular-and-other-benefits-in-people-with-type-2-diabetes-pdf-10956473392.
314. McGuire DK, Shih WJ, Cosentino F, Charbonnel B, Cherney DZI, Dagogo-Jack S, et al. Association of SGLT2 Inhibitors With Cardiovascular and Kidney Outcomes in Patients With Type 2 Diabetes: A Meta-analysis. JAMA Cardiology. 2021;6(2):148-58.
315. Palmer SC, Tendal B, Mustafa RA, Vandvik PO, Li S, Hao Q, et al. Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. BMJ. 2021;372:m4573.
316. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. New England Journal of Medicine. 2018;380(4):347-57.
317. National Institiute for Health and Care Excellence. Atrial fibrillation: the management of atrial fibrillation 2021. NICE guideline [NG196] online: NICE; 2021 [Available from: https://www.nice.org.uk/guidance/ng196.
318. O'Brien EC, Simon DN, Thomas LE, Hylek EM, Gersh BJ, Ansell JE, et al. The ORBIT bleeding score: a simple bedside score to assess bleeding risk in atrial fibrillation. European Heart Journal. 2015;36(46):3258-64.
319. National Institiute for Health and Care Excellence. Management of atrial fibrillation. Update guideline, Appendices A - R online: NICE; 2014 [Available from: https://www.nice.org.uk/guidance/ng196/evidence/appendices-a-to-r-pdf-9081927327.
320. Selak V, Kerr A, Poppe K, Wu B, Harwood M, Grey C, et al. Annual Risk of Major Bleeding Among Persons Without Cardiovascular Disease Not Receiving Antiplatelet Therapy. JAMA. 2018;319(24):2507-20.
321. National Institiute for Health and Care Excellence. Network meta-analysis: Anti-platelets for managing unstable angina or NSTEMI or STEMI. NICE guideline NG185. Network meta-analysis report online: NICE; 2020 [Available from: https://www.nice.org.uk/guidance/ng185/evidence/network-metaanalysis-for-dual-antiplatelet-therapy-pdf-8903837485.
322. Squizzato A, Bellesini M, Takeda A, Middeldorp S, Donadini MP. Clopidogrel plus aspirin versus aspirin alone for preventing cardiovascular events. Cochrane Database of Systematic Reviews. 2017(12).
323. CURE Investigators. Effects of Clopidogrel in Addition to Aspirin in Patients with Acute Coronary Syndromes without ST-Segment Elevation. New England Journal of Medicine. 2001;345(7):494-502.
324. Antithrombotic Treatment Trialists Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ. 2002;324(7329):71-86.
325. Wang FM, Yebyo HG, Ballew SH, Cainzos-Achirica M, Boyd C, Puhan MA, et al. Older adult preferences regarding benefits and harms of statin and aspirin therapy for cardiovascular primary prevention. American Journal of Preventive Cardiology. 2023;13:100468.
326. Wade T, Jette Videbæk L, Peter H, Jesper Bo N, Line Bjørnskov P, Dorte Ejg J. Exploring how GPs discuss statin deprescribing with older people: a qualitative study. BJGP Open. 2020;4(1):bjgpopen20X101022.
327. National Institiute for Health and Care Excellence. Lipid modification Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease Clinical guideline Methods, evidence and recommendations online: NICE; 2014 [Available from: https://www.nice.org.uk/guidance/cg181/evidence/lipid-modification-update-full-guideline-pdf-243786637.

328. Armitage J, Baigent C, Barnes E, Betteridge DJ, Blackwell L, Blazing M, et al. Efficacy and safety of statin therapy in older people: a meta-analysis of individual participant data from 28 randomised controlled trials. The Lancet. 2019;393(10170):407-15.
329. Martin SA, Boucher M, Wright JM, Saini V. Mild hypertension in people at low risk. BMJ : British Medical Journal. 2014;349:g5432.
330. Viera AJ, Hawes EM. Management of mild hypertension in adults. BMJ. 2016;355.
331. Sundstrom J, Arima H, Jackson R, Turnbull F, Rahimi K, Chalmers J, et al. Effects of blood pressure reduction in mild hypertension: a systematic review and meta-analysis. Ann Intern Med. 2015;162(3):184-91.
332. Zanchetti A, Thomopoulos C, Parati G. Randomized controlled trials of blood pressure lowering in hypertension: a critical reappraisal. Circ Res. 2015;116(6):1058-73.
333. Musini VM, Tejani AM, Bassett K, Puil L, Wright JM. Pharmacotherapy for hypertension in adults 60 years or older. Cochrane Database of Systematic Reviews. 2019(6).
334. National Institiute for Health and Care Excellence. Hypertension in adults: diagnosis and management. NICE guideline [NG136] online: NICE; 2019 [Available from: https://www.nice.org.uk/guidance/ng136.
335. National Institiute for Health and Care Excellence. Hypertension in adults: diagnosis and management [C] Evidence review for initiating treatment NICE guideline NG136

Intervention evidence review underpinning recommendations

1.4.9 to 1.4.14 in the guideline online: NICE; 2019 [Available from: https://www.nice.org.uk/guidance/ng136/evidence/c-initiating-treatment-pdf-6896748208.
336. Rahimi K, Bidel Z, Nazarzadeh M, Copland E, Canoy D, Ramakrishnan R, et al. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. The Lancet. 2021;397(10285):1625-36.
337. Rahimi K, Bidel Z, Nazarzadeh M, Copland E, Canoy D, Wamil M, et al. Age-stratified and blood-pressure-stratified effects of blood-pressure-lowering pharmacotherapy for the prevention of cardiovascular disease and death: an individual participant-level data meta-analysis. The Lancet. 2021;398(10305):1053-64.
338. Shanmuganathan M, Goswami R, Dar O, A'Court C. Management of heart failure with reduced ejection fraction in 2021: an update for GPs. British Journal of General Practice. 2021;71(708):330.
339. Paul MH, Andrew SF. Management of heart failure with reduced ejection fraction. Heart. 2022;108(19):1571.
340. Ezekowitz JA, O'Meara E, McDonald MA, Abrams H, Chan M, Ducharme A, et al. 2017 Comprehensive Update of the Canadian Cardiovascular Society Guidelines for the Management of Heart Failure. Canadian Journal of Cardiology. 2017;33(11):1342-433.
341. Rosano GMC, Moura B, Metra M, Böhm M, Bauersachs J, Ben Gal T, et al. Patient profiling in heart failure for tailoring medical therapy. A consensus document of the Heart Failure Association of the European Society of Cardiology. European Journal of Heart Failure. 2021;23(6):872-81.
342. Kwak MJ, Cheng M, Goyal P, Kim DH, Hummel SL, Dhoble A, et al. Medication Complexity Among Older Adults with HF: How Can We Assess Better? Drugs & Aging. 2022;39(11):851-61.
343. Qureshi R, Mayo-Wilson E, Li T. Harms in Systematic Reviews Paper 1: An introduction to research on harms. Journal of Clinical Epidemiology. 2022;143:186-96.
344. Abdel Shaheed C, Maher CG, Furmage AM, Hoffmann T, McLachlan AJ. Strengthening the reporting of harms of all interventions in clinical trials. Med J Aust. 2022;217(10):502-4.
345. Gerich JE. The importance of tight glycemic control. The American Journal of Medicine. 2005;118(9, Supplement):7-11.

346. Kuo I DG, Barlow J,. Early and tight glycaemic control. Australian Journal for General Practitioners. 2012;41:681-4.
347. Bin Rakhis SA, Sr., AlDuwayhis NM, Aleid N, AlBarrak AN, Aloraini AA. Glycemic Control for Type 2 Diabetes Mellitus Patients: A Systematic Review. Cureus. 2022;14(6):e26180.
348. McCormack J, Greenhalgh T. Seeing what you want to see in randomised controlled trials: versions and perversions of UKPDS data. BMJ. 2000;320(7251):1720.
349. Lehman R, Krumholz HM. Tight control of blood glucose in long standing type 2 diabetes. BMJ. 2009;338:b800.
350. Rodriguez-Gutierrez R, Montori VM. Glycemic Control for Patients With Type 2 Diabetes Mellitus: Our Evolving Faith in the Face of Evidence. Circ Cardiovasc Qual Outcomes. 2016;9(5):504-12.
351. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
352. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). Medical Education. 2006;40(5):405-14.
353. Watt D. On Becoming a Qualitative Researcher: The Value of Reflexivity. The Qualitative Report. 2015.
354. General Medical Council. The state of medical education and practice in the UK.The workforce report. online: GMC; 2022.
355. National Association of Sessional GPs. How many GP locums work in the UK? online: NASGP; 2014 [Available from: https://www.nasgp.org.uk/faq/locums-how-many-gp-locums-in-the-uk/.
356. Oaksford M, Chater N. Précis of Bayesian Rationality: The Probabilistic Approach to Human Reasoning. Behavioral and Brain Sciences. 2009;32(1):69-84.
357. Kahneman D. Thinking, fast and slow. London: Penguin; 2012.
358. Croskerry P. A universal model of diagnostic reasoning. Acad Med. 2009;84(8):1022-8.
359. Gabbay J, Le May A. Practice-based evidence for healthcare [electronic resource] : clinical mindlines. Abingdon, Oxon: Routledge; 2011.
360. Grad RM, Pluye P, Meng Y, Segal B, Tamblyn R. Assessing the impact of clinical information-retrieval technology in a family practice residency. Journal of Evaluation in Clinical Practice. 2005;11(6):576-86.