

IMPROVING THE USER EXPERIENCE
OF EVIDENCE

A design approach to evidence-informed health care

Sarah Rosenbaum
PhD thesis

AHO
The Oslo School of Architecture and Design
2010

© Sarah Rosenbaum, 2010

ISBN-13: 978-82-547-0234-5
ISBN-10: 82-547-0234-9

CON-TEXT
Avhandling nr: 47

Akademisk doktorgrads-
avhandling avgitt ved
Arkitektur- og
designhøgskolen I Oslo

UTGIVER:
Arkitektur- og designhøgskolen I Oslo

ILLUSTRASJON OMSLAG
David Sacks/Getty Images

TRYKK:
Unipub Forlag AS

TEMPLATE DESIGN:
BMR

Table of contents

Table of contents	3
Preface	4
Summary	7
List of articles	10
Background	11
Aims and objectives	29
Underlying theories, frameworks and concepts	34
Methods	55
Results	74
Discussion	82
The end of the journey	120
Summary and conclusion	148
References	156
Appendix	170

Preface

This body of research represents a small part of an ongoing global effort to bring scientific evidence from health care research into clinical practice and policymaking. My advisor Arild Bjørndal, researchers Andy Oxman and Claire Glenton, plus a handful other colleagues at the Norwegian Knowledge Centre for the Health Services warmed to the idea that the challenges of communicating evidence into practice/policymaking might benefit from being framed as design problems, using methods and approaches from design practice. Without their support, I would not have had the opportunity to participate in or influence this work.

This text is based on three separate projects, all established with the aim of making the main results of Cochrane systematic reviews (syntheses of the best available research from individual trials on the effects of a health care treatment or intervention) more easily accessible and useful for health professionals and policy makers. In the first project we explored health professionals' experience of Cochrane Library website where these reviews are published. In the next two projects we developed and evaluated two different types of evidence summaries derived from Cochrane Reviews.

All three projects had both a pragmatic design dimension and a research dimension. Although the primary incentive behind these projects were to ultimately *improve specific products*, either already existing ones or proposals for new ones, we also had an aim to look for *transferable findings* in the data we collected that could be of use for others concerned with designing or publishing similar kinds of evidence-related artifacts.

My enrolment in the PhD program at the Oslo School of Architecture provided me the opportunity and incentive to carry out this work as full-fledged research rather than just as a set of design initiatives. One important reason for doing this, besides contributing to the quality of the work and to my schooling as a researcher, was to be able to communicate our findings in a credible way to research-oriented stakeholders as well as to a broader

audience involved with health care research dissemination. Therefore, instead of just getting on with the design work, we pursued the problems as research questions: choosing methods appropriate to the questions, working to eliminate bias or confounding factors, making sure the studies were carried out in a rigorous, ethical and transparent fashion, describing both methods and results in detail, and placing the findings in the context of earlier research and relevant theory.

This thesis also represents my participation in two parallel discourses – one within the knowledge translation community in health care research and one in the design research community. Hopefully this work contributes in a meaningful way to each domain. The questions relevant to the design discourse were not obvious to me five years ago when I began these studies. In a typically “designerly” fashion, the outline of these problems only became apparent through the process of muddling through it all, emerging clearly first at the end when I had time to step back and reflect on the body of work as a whole. Therefore, the parts of this thesis addressed to design theory (first and last sections) were not formulated at the beginning of the project, but written up at the end, and used to frame the other content like a set of parentheses. This wise move was suggested by Tone Bratteteig, my secondary advisor, who navigated me through the difficulties of discovering what this work meant seen from a design theoretical point of view.

Taking a multi-disciplinary approach in a PhD thesis is a risky business, as there always is the chance that the work may be found lacking seen from the perspective of each one of the disciplines in isolation. However it is my conviction that this kind of cross-disciplinary reflection is necessary in order to move the understandings of both fields forward.

Projects, roles, funding and audience

The text is divided in two parts – an overview and the four articles. I conceived and initiated the Cochrane Library web site study; the Summary of Findings Table project and the SUPPORT Summary project were both parts of larger work packages carried by international teams of researchers collaborating with a core team from my institute. In addition to participating as designer, I was responsible for the protocols regarding user testing, and had the central role in the collection and analysis of data in all studies with the exception of the randomized control trials. I am first author on all four articles included in this thesis. A full list of the roles of all authors for each article is included as an appendix. The projects have been funded through the Norwegian Knowledge Centre for the Health Services, The Cochrane Collaboration, Wiley-Blackwell and the EU.

The articles are written primarily for researchers, publishers or others who are concerned with dissemination of evidence from health care research (systematic reviews in particular) to clinicians or policy makers. The overview is also relevant for this audience, who I assume have little background knowledge in the field of design research and practice. But I have additionally written the overview with designers and design researchers in mind, who, on the other hand have little or no background knowledge related to evidence in health care.

Acknowledgements

I would like to thank four people without whom this work literally would not have happened. Arild Bjørndal had the vision to see that with my design background I might be able to contribute effectively to the domain of health care research dissemination. He also inspired me to embark on a doctorate degree, and kept me on track for many years with lots of hot coffee and sound advice. Claire Glenton effectively taught me the both the science of qualitative research and the art of scientific writing; without her my transition from designer to researcher would have been less rigorous, more painful, considerably lonelier and definitely less entertaining. Andy Oxman provided me the unique and unusual opportunity of participating in the Summary of Findings and Support Summary work which he led. His highly knowledgeable scientific guidance and inspiring personal encouragement has been invaluable. Tone Bratteteig steered me skilfully and patiently through the challenges of writing a doctoral thesis from a design research perspective while I was immersed (and sometimes lost) in the traditions and concerns of health care dissemination research.

I would also like to thank Hilde Kari Nylund, Jane Cracknell and all of the SUPPORT co-authors, as well as Irene Wiik Langengen for helping with user testing. Thanks to Simon Lewin for a thorough reading of the draft of this thesis before submission, and to Anne Mellbye for lively discussions and comments on the design-related writing. Additional thanks to Halina Dunin-Woyseth, Magnus Gustafsson and David Kaufmann for helpful guidance in different phases of my doctorate program.

Finally I especially thank Mats, Nora and Miriam who supported me (sometimes in mysterious ways) during the marathon of finishing this work.

Summary

Evidence about the benefits and harms of health care interventions has the potential to improve quality of care. Syntheses of the best available evidence - systematic reviews - can provide valuable information for health care clinicians and policy makers, but these publications are often too lengthy to be useful for decision makers in time-pressed contexts. Condensed summaries of reviews may help, but there is little research about how to design such summaries (what content to include and how to present it), how they will be experienced by users, and their effect. Earlier research to create summaries of Cochrane Reviews has also shown that condensing complex data can be challenging and lead to misrepresentation.

This thesis builds on a series of studies examining these issues from several perspectives, with the designed artifact as the main focal point. An underlying assumption for all of this work is that the design of artifacts in which evidence is embedded may shape or influence users' experiences of this evidence. Given that these artifacts are, in effect, mediating between the domains of evidence production and evidence use (by policymakers, clinicians or the public), they warrant far more attention than they appear to have received to date.

Multiple methods are employed, including a framework from design practice domain. The first study is a set of user tests, examining health professionals' user experience of the Cochrane Library (where Cochrane Reviews are published). The second study involves exploring user and stakeholder feedback to inform iterations of a Summary of Findings Table for Cochrane Reviews. The third study is an evaluation of the effect of including a Summary of Findings Table in a Cochrane Review (compared to a Review with no table) on user satisfaction, understanding and time spent to find key messages. The fourth study explores user and stakeholder feedback to inform the development of a template for short summaries of systematic reviews that are tailored for health policy makers in low and middle-income countries.

Findings from the user feedback uncovered many comprehension problems, for instance:

- Health professionals and policy makers may not know what a systematic review is or may confuse it with another type of report (e.g. single trial). This can lead to unrealistic expectations or critical misunderstanding of content.
- Unfamiliar language (both foreign language and jargon, abbreviations and unfamiliar terms) may lead to frustration and alienation, or cause critical barriers while reading and searching. Non-native English speakers are at an extra disadvantage.
- Outcome effects may be misunderstood when presented as absolute differences in a table; the use of absolute risks can solve this problem.
- Outcome effects presented as continuous outcomes are often difficult to understand when the scales that are used are unfamiliar.

However, in two small trials we demonstrated that correct comprehension of the main results in a Cochrane Review were improved by including a Summary of Findings Table (in comparison to a review with no table). These trials also showed that key findings were quicker to find in a review with a table.

User studies also helped us understand how summaries of evidence from reviews could be made more useful:

- The layering of the information, through use of a graded entry format, is well-received by policy makers
- Making summaries useful for policy makers also includes broadening the scope of information, such as adding author's comments about applicability or a wider range of references, as well as information about the information.
- Making summaries useful for clinicians includes better descriptions of interventions and risk groups.

One of the main challenges we encountered in development work was resolving the tension between stakeholders' concern for precision in data presentation and users' needs for simplicity. This can be reframed in light of fuzzy traces theory, where people are seen to process information along a gist-verbatim continuum, with a gist preference. Artifacts presenting evidence should be designed with the aim of enabling easy gist extraction.

Drawing on findings across the studies, I evaluated the suitability of the user experience framework we employed and present a revised version. I suggest that

user experience frameworks may need to be tailored for specific use contexts and product types.

At the end of the thesis I reflect on my experiences as a designer participating in this work. I discuss, among other things, the challenges to designer role and identity – such as when design-related tasks are increasingly shared with non-designers or when the product disappears out of sight. The results of design processes of complex digital information may not be strikingly visible, but rather subtle, hard to freeze in time and claim ownership to. I propose a move towards redefining design as more than a “reflective conversation with the materials”, and discuss how legitimate output from design should not be limited to the “things” that we make, but also include knowledge gained underway.

List of articles

ARTICLE 1

Rosenbaum SE, Glenton C, Cracknell J. **User experiences of evidence-based online resources for health professionals: user testing of The Cochrane Library.** BMC Med Inform Decis Mak 2008;8:34.

ARTICLE 2

Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. **User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings Tables for Cochrane reviews.** Journal of Clinical Epidemiology 2010; 63:607-19.

ARTICLE 3

Rosenbaum SE, Glenton C, Oxman A. **Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information.** Journal of Clinical Epidemiology 2010; 63:620-6.

ARTICLE 4

Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Migniini L, Young T, et al. **Evidence summaries tailored for health policymakers in low and middle-income countries.** WHO Bulletin, December 2010.

Background

POINT OF DEPARTURE: MOVING FROM DESIGN TO RESEARCH

...Newly arrived and quite ignorant of the languages of the Levant, Marco Polo could express himself only by drawing objects from his baggage – drums, salt fish, necklaces of wart hogs' teeth – and pointing to them with gestures, claps, cries of wonder or of horror, imitating the bay of the jackal, the hoot of the owl.

The connections between one element of the story and another were not always obvious to the emperor; the objects could have various meanings: a quiver filled with arrows could indicate the approach of war, or an abundance of game, or else an armorer's shop; an hourglass could mean time passing, or time past, or sand, or a place where hourglasses are made....

... As time went by, words began to replace objects and gestures in Marco's tales: first exclamations, isolated nouns, dry verbs, then phrases, ramified and leafy discourses, metaphors and tropes. The foreigner had learned to speak the emperor's language or the emperor to understand the language of the foreigner...

From "Invisible Cities", Italo Calvino

Someone once suggested that I might see myself as a designer who took a knapsack full of her skills and tools and traveled to a foreign country to spend time working with the people there and the issues that concerned them. This thesis could be seen as a narrative from that journey. To prepare the reader for the descriptions of where I traveled and what I found, it is helpful to first understand my place of origin, why I chose to wander into foreign territory, and what I brought with me in my baggage.

My professional training is in graphic design, which I have practiced for over 30 years. During much of that time I have worked with corporate identity, though my skills and interests have spanned a very broad range, from the visually oriented side of the field, such as illustration and photography, to the less visible utilitarian areas of information design. Over the last decade my work has increasingly revolved around designing for the Internet, and my areas of interest have expanded to including interaction with information accessed through digital media. I have worked closely with exceptional designers trained in interaction design, and learned much from their approaches to involving users in design processes.

I have always been interested in the user perspective, though in my field of graphic design, at least to my knowledge, there has not been a strong tradition in systematically involving users in the development of designed artifacts beyond the realms of advertising campaigns and large-scale branding initiatives. That situation has changed due to the advance of the Internet. The technological complexity of digital media creates new barriers for users and new challenges – as well as possibilities - for designers. For that reason, the approaches I describe in this thesis are predominantly derived from fields other than my own: Human Computer Interaction and Interaction Design. In these domains, methods to include user's perspectives are more developed and have a longer history and a more central position than in my own original field of graphic design.

My interest in medical information intensified when I personally experienced a serious disease and subsequently encountered the health system at close range over an extended period of time. Motivated by these experiences, I was drawn to the concept of working towards helping patients gain access to better information. In 2004 I began as senior designer in the Norwegian Knowledge Center for the Health Services (Kunnskapssenteret) and at the same time enrolled in the PhD program at Oslo School of Design and Architecture (AHO). I originally planned to combine design and research to develop innovative approaches to doctor-patient communication, particularly regarding patient-relevant evidence from medical research concerning

treatments. However, shortly after my arrival, Kunnskapssenteret made strategic decisions to narrow down their dissemination focus of research evidence to health professionals and policy makers; consequently projects involving lay communication eventually disappeared altogether from their agenda. So I rerouted my focus of study to the professionals in the health system and their interactions with evidence from health care research, which I found equally engaging. Doctors, nurses and health professionals are often a patient's primary trusted source of information about their condition and possible treatments. Also, these groups, along with administrative and political decision makers, make or influence treatment choices that affect both individuals and whole groups of patients.

Consequently, after 25 years of practicing as a professional designer, I found myself entering simultaneously the academic world of design research and the pragmatic and academic worlds of health care research and dissemination. What did I bring in my knapsack? To begin with, I brought my original professional identity. I was, and still am, a practicing designer. My skills have evolved over the years, my methods have changed somewhat, the media I design for has changed considerably, but my main objectives have remained rather constant: to design useful and meaningful things for other people. My interest in research is not to find a new career, but to try to better understand my own field of practice.

What is designing actually?

The term design comes originally from the Latin term “designare”: to designate, which means to specify, as in pointing out what to do. Many design theorists have contributed to the definition of design and what characterizes its practice. Herbert Simon, in “The Science of the Artificial” produced the much-cited description: “Everyone designs who devises courses of action aimed at *changing existing situations into preferred ones*” (my italics).[1] He also claimed that “Engineering, medicine, business, architecture and painting are concerned not with the necessary but with the contingent - *not with how things are but with how they might be* - in short, with design” (my italics).[1] These broad definitions challenge any one profession's claim to exclusive ownership of the concept ‘design’. Sless offers an equally broad definition, placing design as an activity between humans and our environment: “Designing is our most developed form of practical adaptation to our environment. It is the means by which we, as biological entities, change to meet the demands of our environment, and make changes to our environment to adapt it to our needs”.[2]

Others see design as an “in between” activity (or entity): between form and context;[3] between indeterminacy and determinacy.[4] Rosan Chow has written an interesting account of how many of these divergent descriptions of design have some basic things in common. She points out that on a macro level, design occurs as an interface between humans (including their intentions) and the environment; given a more micro perspective, design can be described as an interface between an (as yet unknown) artifact and a particular context.[5]

Design as an activity has often been described as a form of problem solving. Simon emphasized that design involves finding a good-enough solution (“satisficing”), in a process characterized by limited information, finite amount of time for decision-making, and cognitive limitations of our minds (“bounded rationality”).[1] Other researchers have demonstrated how design problem spaces are unique and separate from other types of cognitive problem spaces.[6] Schön, however, argued that the actual practice of design is not exclusively tied to problem solving, but that it also very much involves the process of “problem setting”. [7] He studied the practice of several professions, including doctors and architects, and describes designing as a reflective practice, engaging in a “conversation with the situation”. Others have also stressed this practical and cyclical nature of design activity, emphasizing the dialog with the possibilities and restraints in the situation and the environment.[8, 9] Gedenryd, influenced by among others Dewey and his philosophy of inquiry, argues that design (and cognition in general) is not exclusively an intermental process, nor does it resemble rational linear planning, but can be characterized as a constant interaction between the mind, actions and the world.[10] Chow, also informed by Dewey, makes a convincing theoretical argument for defining design as a form of inquiry. She says: “In brief, designing can be viewed as a process of periodically shifting and negotiating between goal, problem and solution and bringing the undetermined to the state of determinacy”.

Dewey’s philosophy of inquiry is a form of pragmatism,[11] which emphasizes the practical consequences of an ideology or proposition. Dewey sees inquiry as beginning with a problem. He maintains that when we face a problematic situation, our first task is to understand it by describing its elements and attempting to identify their relations. Inquiry continues through entertaining these hypotheses about the problem’s elements and relations, and testing them out.[12] Out of such inquiry emerges knowledge, which he views as temporary and contingent upon “its adequacy in providing a coherent understanding of the world as the basis for human action”. [13] His view of knowledge was that it did not serve acting (such as doing or making),

but rather that acting (i.e. testing the practical consequences of a hypothesis) is always central to knowing. In this way he challenged common assumption that theoretical beliefs and practical deliberations are separate processes. According to Dewey, all inquiry is practical, concerned with transforming and evaluating the features of (problematic) situations in which we find ourselves.

How does design differ from research?

Design, then, can be defined as a form of inquiry. But research is also a form of inquiry. What differentiates these two? Dewey maintained that his description of inquiry covered the acquisition of all types of knowledge, from common sense to findings from science. Chow, however, presents a modified view, maintaining that while design and research inquiry theoretically might share both subject matter and methods, the two differ fundamentally in object (or aim). This view builds on a proposal from Cross: that the object of science is a *description*, whereas the object of design is a *prescription* (or a proposal). Along the same line of thinking, science aims to produce knowledge, while design aims to produce products or artifacts. Chow offers a more fine-grained differentiation: that the object of science is to produce *generalizations through description*, while design aims to produce *specifications in context*. By this she means specific artifacts, such as a particular table (i.e. made of plastic) for particular contexts (i.e. for use by children). As I understand this, “specifications” refer not only to the finished product, but also the sketches, blueprints, prototypes, models, descriptions, etc, that are needed to produce it. These specifications in context may also be understood as “ultimate particulars”.^[14] User studies, Chow maintains, are techniques for better understanding context. By context she means the particular details regarding the particular design problem, the details of the “undetermined situation of concern”.

Krippendorff emphasizes one further important distinction between research and design. He stresses that whereas the practice of research attends to observing, measuring, and documenting already existing phenomena, the practice of design is concerned with creating “imaginable futures”, through new artifacts that do not yet exist.^[15]

If research and design are two different kinds of inquiry, what is design research?

Design research, in my understanding, is a systematic and transparent inquiry regarding some aspect of design. For instance, one may want to gain knowledge about the nature of creating certain artifacts or the characteristics of the material they are made of. One could explore design processes or

methods, the designers themselves, particular users and use contexts, or relations between artifacts and users. In order to gain some of these kinds of knowledge, it is often necessary to go through the phases of a designing and producing a product. Knowledge thus gained might be very specific and only relevant for development of that particular artifact. Or it might be general enough that it could be transferable to the development of similar artifacts, or similar contexts and users, or to similar types of design processes/design problems.

Attempts have been made to categorize design research in different ways, such as research *into*, *for* and *through* design.[16, 17] Sevaldson describes these as:

- “Research *into* design: Inquiry that looks at design from a distant perspective... Also called Research *about* design.
- Research *for* design: Research that serves design and is subservient to design.
- Research *through* design: Any research where the design practice is central in generating knowledge. Can also address tasks given by others.” [18]

He argues that these categories are too coarse and superficial, and proposes that more focus be given to a fourth mode - Research *by* Design (a subset of the category ‘research through design’): “a special research mode where the explorative, generative and innovative aspects of design are engaged and aligned in a systematic research inquiry...[this definition] implies that the design practices by themselves have an inherent element of investigation, innovation and knowledge generation...But it also indicates that these practices need to be complemented with a special dimension of reflection to qualify as research”.

The body of work in this thesis may be best understood as a form of research *through* design, where we have produced specific artifacts for particular groups of users and reflected on these processes in order to gain more general insights that could be used in future design work (either our own or others’). Through these efforts, we not only developed particular artifacts that were acceptable to both users and stakeholders; we also gained knowledge about many aspects of these types of artifacts and the task of designing them (i.e. better understanding of the nature of the problem, characteristics of the users experiences, knowledge about gaps between users and producers of the artifacts, characteristics of successful or unsuccessful solutions, strengths and weaknesses of the methods we used leading to suggestions about method

improvement.) Additionally, these projects provided the foundation for more general reflections about the nature of design work, written up at the end of the thesis. (For a schematic overview of the research questions and inquiries, and their placement within pragmatic design projects, see Figure 2.)

Design values

Another piece of baggage I brought with me into this work was my interest and focus on the end user, something I regard as both a professional platform as well as a personal ethical stance. *User-centered design* is a philosophical and methodological approach to design, emphasizing creating solutions catering to the needs of the user (rather than requiring the user to adapt to the solution of the developer) and involving users in development.[19] Krippendorff expands on this concept, from user-centered to *human-centered*, which he considers fundamental to design.[15] He maintains: “humans do not see and act on physical qualities but on what these qualities *mean* to them”. Therefore, a human-centered design approach entails understanding what the qualities of artifacts means. This perspective can permeate all phases of design work; for instance, identifying needs and limitations of users early in the work, collecting feedback users during development, and checking that final solutions actually work as intended.

Users may be defined in different ways, for instance primary users (persons who actually use the artifact), secondary users (persons who occasionally use it through an intermediary) and tertiary users (persons affected by the artifact’s use or make decisions about its acquisition).[20] But as Krippendorff points out, understanding other stakeholders’ perspectives is also a critical part of creating viable design solutions. Without broad stakeholder support, such as from people involved in production and marketing, good solutions and ideas might be ignored, carelessly altered or actively boycotted. The division between users and stakeholders is somewhat artificial - users might also be regarded as a sub-group of stakeholders. However, in the rest of this thesis I will often use the phrase “users and stakeholders”, in order to separate the primary end-users from the other groups of people who had an interest in the artifacts’ design.

Balancing the conflicting needs of both users and stakeholders is a fundamental challenge in most design work, and the projects in this thesis were no exception. However, in my experience from years of design practice, resolving user-stakeholder conflicts does not necessarily mean finding a compromise, but possibly can entail reframing the problem so that conflicts are reduced (or eliminated), or in convincing the stakeholders to modify their position to some degree for the benefit of the user. In the projects described

here, the stakeholders explicitly agreed that user-orientation was important. However, as scientists and scientific publishers, this group of stakeholders had their own strong need to maintain particular attributes of the artifacts that would secure credibility in a scientific context (an issue I take up again in the discussion).

Since the stakeholders were for the most part scientists, we had an extra incentive to organize our work as formal research, in order to generate research-based evidence of user feedback for scientists who might otherwise reject less rigorous documentation of users' needs. All in all, a user-centered/human-centered approach in this work led to explicitly illuminating the needs and perspectives of users and stakeholders alike, and then bringing this knowledge into stakeholder dialogs and design processes.

Summing up

In moving from my world of design practice to the more foreign territory of knowledge translation and dissemination of health care research, I brought first and foremost my own professional area of focus: attention to the details of the designed artifact and how these details might affect the artifacts' meaning, both for users and stakeholders. I also brought skills and methods for exploring users' and stakeholders' experiences with artifacts, and for investigating artifacts' possible futures (designing). I brought experience in resolving - through both design, reframing and dialog - conflicting interests and viewpoints in order to arrive at artifacts that might be more meaningful, particularly for those who are intended to use them. These are perspectives, methods and skills from both design practice and design research.

DESCRIPTION OF THE PROBLEM

Examples of the problem

Emma is a 45 year old, well-educated woman with a rather stressful job who has sought her GP to ask for advice on how to stay healthy. She is particularly interested in information about the value of mammography screening as a somewhat older friend recently was diagnosed with breast cancer. During the consultation (and repeat measurements in the weeks to come) it became clear that she also has a moderately elevated blood pressure and the doctor suggests medication.

Rachel is a newly appointed minister of health. She is known as a competent politician, but with no medical background. Only a few weeks after she has

taken up office, the national Cancer Fund launches a vivid, graphic and compelling campaign arguing that mammography screening should be recommended for all women over 40 and included as a part of the free national health plan. The pressure builds up quickly, and there seems to be a majority in Parliament in favor of saying yes. The ministry has no document ready that sums up the pros and cons.

Evidence in decision-making

In each of the examples above, people are faced with choices concerning the health care of others, either concerning individual patients or whole populations. What are the chances of successful outcomes, what are the risks of side effects or harm? Decisions of this kind make up a significant part of the practice of medicine or health policy development. Good decisions may save lives; poor decisions may cause unintentional harm.[21]

Part of making a good decision about a health care intervention or treatment is having knowledge about *the estimated effect*. Clinicians need knowledge of the effects (both beneficial and adverse) of available treatments in order to make appropriate choices for individual patients. Policy makers need robust knowledge about the expected effect of interventions on both health-related and other outcomes (such as equity of care) in order to best utilize their limited resources for the benefit of whole populations.[22, 23] They also need to know the effect of different arrangements for financing, delivering and governing services.[24] Both clinicians and policy makers also need to know when such knowledge is lacking. Knowing ‘what works’ (as well as ‘what may cause harm’ and ‘what we don’t know’) is a critical part of providing effective, efficient and equitable health services.[25-27]

What is evidence?

Knowledge about ‘what works’ may come from many different sources, such as personal experience, expert opinion or evidence. The Oxford American Dictionary captures a common definition of evidence: “anything that establishes a fact or gives reason for believing something.”[28] Another definition of evidence differentiates it from expert opinion: “evidence ‘concerns facts (actual or asserted) intended for use in support of a conclusion’; 2) facts are something known by experience or observation; 3) expert opinion is more than just evidence - it is the combination of facts, the interpretation of facts, and conclusions”.[26, 29]

Scientific evidence

Whereas the colloquial view of evidence described above is quite broad, the scientific view is more restrictive. Scientific evidence can be seen as

experiences or observations collected and analyzed through scientific methods that are systematic and transparent. Therefore, findings arising from scientific inquiry generally provide stronger support for conclusions than more haphazard and undocumented processes. Transparency of method allows others to contest the result. (From now on, when I write “evidence”, I am referring to scientific evidence.)

Quality of scientific evidence

Transparency of evidence is important, as not all research provides equally convincing results. Uncertainty may arise for a number of reasons, such as less than optimal choice of study design for question type, poor execution of the method, inconsistent or widely varying results, or the likelihood that bias has influenced the results. Evidence may also be susceptible to change in light of new observations or more rigorous inquiry. ‘Quality of evidence’ is a concept representing the *strength* of a claim resulting from scientific inquiry.[30] The higher the quality of the evidence, the more faith we can have in the end results.¹ Evidence by itself is not sufficient; decision makers also need sufficient information about the quality of that evidence to decide how much weight it should carry in their decision-making.

The role of evidence in decision-making frameworks

Health care practice cannot be determined by results from research studies alone. In a model of evidence-informed decision making, both the patient perspective as well as the clinician’s experience inform decisions together with the best available evidence from research.[31] Decisions hinge both upon knowledge (about the possible outcomes given a course of action) and on values (what do we want to achieve, how do we rank our priorities?) It is perfectly rational for Emma, given a moderately elevated blood pressure, to start treatment (“my father died early of heart disease, I will do all I can to reduce my own risk”) or not (“I do not want to take pills every day”). For her to be able to reach “a good decision”, however, she must be well informed – by the doctor or by other sources of information. Only when she is offered relevant and reliable information that she understands, about likely gains and

¹ People outside specific scientific communities may not share the same understanding of what ‘evidence’ or ‘quality of evidence’ means as the scientists within those communities. For instance, they may define ‘evidence’ in a broader sense rather than in a narrowly scientific sense; or they may employ other assessment measures to make a judgment about the quality of evidence than those used by the scientists within that community. They may not be aware of their own stance or how these may differ from a particular scientific approach. This is a factor that can disrupt communication, for instance between scientists and non-scientists, as well as between scientists from different disciplines.

risks, will she be able to integrate values and knowledge so as to form an evidence-informed decisions that is rooted in her own circumstances and values.

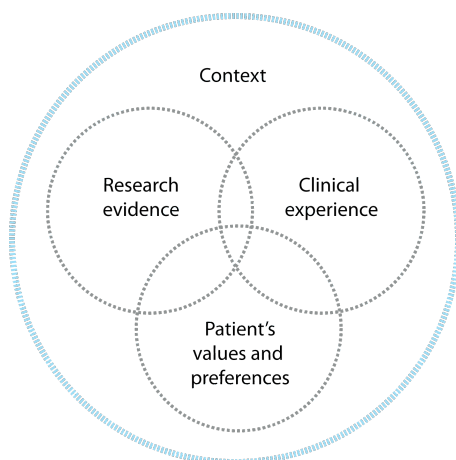


Figure 1. Framework for evidence-informed practice in health care[32]

Policy decisions are possibly even more complex, grounded on a spectrum of information from many diverse sources (e.g. research, media, think tanks) and influenced by values, beliefs and interests of a broad set of stakeholders. Contextual limitations such as availability and cost are also important to consider in both policy and clinical contexts. Lomas warns against viewing policy decisions as isolated events, but rather complex messy processes that stretch out in time.[33] But regardless of whether decision-making is conceptualized as a single simple event or a long complex process, up-to-date and reliable evidence from research is a critical ingredient in achieving well-informed policies or practices. The health minister described above would clearly benefit from a trustworthy and understandable presentation of the likely gains, side effects and costs of screening that could inform both the ongoing debate and decision process.

The research – practice/policy gap

Although the need for evidence in health care and policymaking is apparent, there is often a significant gap between existing evidence from research on the one hand, readily available information about that evidence for any decision-maker on the other hand, and finally what is actually carried out in health care practice.[34] Studies suggest that in the USA and the Netherlands about 30-40% of patients do not receive care according to present scientific evidence.[35] Balas and Boren found that the lag between the discovery of

more effective forms of treatment and their introduction into routine clinical care is between 15 and 20 years.[36] One study examining the development processes of WHO guidelines found that these rarely involve explicit use of evidence, but rely heavily on expert opinion.[37] Despite the enormous amount of resources poured into health care treatment and intervention studies, many decisions at the levels of both clinical practice and policy making remain uninformed by up-to-date research.

This kind of gap is not unique to the health care domain. A simple Google Scholar search² uncovers "research-to-practice gap" articles in a wide variety of fields: education, finances, human resources, librarianship, natural resources planning, and criminal rehabilitation. Even within the design domain this issue has gained recent attention.[38, 39] However, the implications of this gap in health care are more precarious than in most fields - at best leading to massive misuse of public resources and at worst leading to loss of lives.

EARLIER RESEARCH

Barriers and facilitators in clinical practice

If the need for research in clinical health care practice is evident, and evidence exists that could improve care in many contexts, why doesn't this knowledge make its way more readily into practice?

Many believed that increased Internet access would facilitate this process. However, though use of online information in health care is increasing, a review from 2006 showed that colleagues and textbooks are still the preferred source of information for clinicians with clinical questions.[40] In a study by Ely et al studying actual usage, online sources ranked third behind textbooks and colleagues.[41] In another study he documented 59 obstacles related to successful use pursuit of evidence in practice settings.[42] Literature reviews have indicated which of these many barriers seem to be the most significant. These are described briefly below.

Lack of time

According to two literature reviews from 2006 and 2007, lack of time to search for answers to clinical questions was the most salient finding in the studies reviewed.[40, 43] In one included study, 60% of the participants reported that time was a major issue.[44] Different studies report varying amounts of time needed to search for information, ranging from 53 minutes

² Google Scholar: search for "evidence practice gap"

to medians of 14.4 and 15 minutes.[44-46] A study from the US in 1999 reported observations that primary care physicians will spend on average less than two minutes on a searching activity[47] in another study physicians self-reported using on average less than twelve minutes, although these often did not use electronic sources.[48] In the UK, GP's patient appointments are often between 7 and 10 minutes, so that searching for information while the patient is present is not realistic in these contexts.[43]

Lack of skill

Lack of skill is related to lack of time, as an inexperienced searcher will need more time to find good answers to their questions. In a qualitative study of problems that doctors encountered answering questions about patient care with research evidence, six obstacles were particularly salient, all of which were related to searching and retrieving information[42]:

- Excessive time required to find information
- Difficulty modifying the original question to one suitable for searching the evidence base
- Difficulty selecting an optimal search strategy
- Failure of a seemingly appropriate resource to cover topic
- Uncertainty about when to stop the search (having uncovered all appropriate evidence)
- Inadequate synthesis of multiple bits of evidence into a clinically useful statement

Lack of awareness

A review of 76 studies looking at obstacles to the implementation of clinical guidelines found that lack of awareness was a particularly important barrier to use.[49] Systematic reviews in particular may suffer this fate: studies have shown that knowledge of databases of systematic reviews, such as the Cochrane Reviews database and DARE (Database of Abstracts of Reviews of Effectiveness), was low among physicians in the UK and Australia, though these numbers are rising.[43]

Lack of access

Access to research literature is dependant on several factors – for instance, having a computer, having an Internet connection and having access to the research literature. Some of these factors vary dramatically geographically. For instance, South Africa (one of the African countries with *highest* number of Internet users) had 8.2% Internet users in 2007, compared with 73% in the USA. Although Internet use is increasing world wide, most African countries were still under the 10% level in 2007.[50] Even though it might be argued that medical institutions and facilities might have a higher proportion of

Internet users than the population at large, the uneven distribution of connectivity is likely of similar proportions, also because it is in part due to country infrastructure. For instance, an study across African medical undergraduate schools from 2010 found an average 'computer : student' ratio of about 1 to 10, and that poor download speed was a significant barrier to use.[51] Even in countries with high connectivity, access to research literature at point of care may be prohibitive due to local infrastructure. For instance, although all clinicians in Norway have national access to a significant number of high quality medical journals and databases through the Norwegian Electronic Health Library, firewalls in hospital IT systems often prohibit the use of these resources at point of care.³

Facilitators - indicators of use

According to the review from 2006, the best indicators of primary physician's use of evidence were urgency of patient problem and the expectation that a clear answer existed in the literature.[40] An Australian study exploring differences of online resource usage between hospitals also found that team social and cultural factors such as the presence of champions and organizational cultures supporting use of evidence in clinical practice were the only clear discriminators between hospitals with high and low use, outweighing technical factors such as shortage of staff or ease of access and time.[52]⁴

Barriers and facilitators in health care policymaking

Policy makers experience many of the same problems as health care personnel, such as difficulty in finding relevant information. However, they are a different group of users in a different decision making context.

Timely and relevant research is hard to come by

One of the main facilitators for use of evidence in policy decisions is the timely retrieval of relevant research.[53] Policy makers and managers report struggling to retrieve research in general.[54] It tends to be published in academic sources, a factor that decreases its accessibility for this target group.[54] Systematic reviews can help make high quality evidence easier to find, but up-to-date reviews that are relevant for the topic and the setting may

³ This is not a documented example but comes from my own experience working design-related consequences of limited access to the Norwegian Electronic Health Library through my current employment position.

⁴ Interestingly enough, this study found that technical factors - such as long time to access information - were cited as major barriers by low use hospitals, but that no objective measurable differences in these factors were found to exist between high and low usage sites.

be lacking. Commissioning a new review can take from 6-24 months to produce, too long for many policy decisions.[55]

Additional challenges in low and middle-income countries

Policy makers in low and middle-income countries (LMIC) could face additional challenges finding existing studies or reviews, as their access to subscription-based journals or databases might be limited.[56, 57] Research on topics relevant to their primary disease burdens may be limited or may have been carried out in high-income countries, decreasing the applicability to their settings.[58-62] Also, as illustrated above, access to computers and speedy Internet connections tend to be poorer.

Insufficient content

Another problem facing policy makers is the scope of the information presented in a systematic review. Reviews typically answer a narrow question about “what is the effect of intervention x?”[63] (e.g.: “What is the effect of male circumcision for prevention of heterosexual acquisition of HIV in men?”). But policy makers tend to ask broader questions, such as “What do we know about problem x” or “what are the issues around doing action y?”[64] (e.g.: What do we know about HIV in our heterosexual male population? or “How socially acceptable would male circumcision as HIV prevention be in our setting?”) Policy makers may not only want estimates about the benefits of an intervention, but also additional information about harms, costs, uncertainties associated with these estimates, the applicability of the evidence to their setting, and impacts on equity.[22, 23, 54, 65-67]

Language, length and format barriers

Yet another barrier is use of language. Systematic reviews are full of medical and research terms. Use of unfamiliar language and jargon has been found to decrease the prospects for research use in policy decision.[54] The length of the document can also be a barrier due to time constraints.[68] In addition, important information, such as the quality of the evidence, may be difficult to extract from a systematic review for non-researchers. A study of public health decision makers in Canada found that two factors related to presentation - ease of use and help to overcome the barrier of limited appraisal skills - were among the four significant predictors of use of systematic reviews.[69]

EFFORTS TO “BRIDGE THE GAP”

New skills: evidence-based medicine

Much effort has been put into finding ways to bridge the evidence and practice/policy gaps. One of the most significant developments has been evidence-based medicine (EBM). I describe EBM – both its development and some of its limitations - in more detail in the next chapter. Here it will suffice to say that the basic EBM approach is to focus on changing the skill set and behavior of health professionals so that they "check the research" when faced with questions in their clinical practice. The five EBM steps are:

1. Convert information needs into answerable questions;
2. Search for the best evidence with which to answer them;
3. Critically appraise that evidence for its validity, importance and usefulness;
4. Integrate the appraisal with clinical expertise and apply the results in clinical practice;
5. Evaluate own performance.

Although this approach to closing the evidence-practice gap seems rather straightforward, it has some serious practical obstacles. Health professionals lack both the time and skill to search for evidence, as pointed out above. But even if they had time and skills, the sheer volume of published research literature itself is prohibitive.[42, 52, 53, 70] Between 2008 and 2009 over 700 000 new citations were added to the PubMed database; a search limited to research articles about cancer produced over 26 000 individual results from the same period. A skilled information-searcher would be strained to keep up with article publication output from health care research care, even within one specialty. To complicate the problem further, the quality of the individual studies varies, and sorting high from low quality takes both time and skill.

Secondary research: syntheses and summaries of individual studies

In response to these problems of volume and variation of quality, secondary research publications have emerged to render high quality evidence more easily accessible. In secondary publications, individual studies are collected, appraised and synthesized. The resulting evidence is then republished in a more condensed form. The *systematic review* is one type of secondary research publication. This is a summary of studies addressing a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise the relevant research, and to collect and analyze data from the included studies. Because reviews are syntheses of results from many individual trials, they save a reader much time and effort by collecting large amounts of quality appraised information in one document.[71]

Systematic reviews may also provide more robust data due to strict inclusion criteria and larger data sets.

Cochrane Reviews

Cochrane Reviews are high quality systematic reviews summarizing the effect (“what works”) of a health care intervention or treatment.[72] These reviews are produced by The Cochrane Collaboration, an international organization of volunteers dedicated to improving the quality and access to health care research.[73, 74] There are currently over 4300 reviews in the Cochrane Database of Systematic Reviews⁵, available online through the Cochrane Library.[75] Due in part to the rigorous methodological standards upheld by the Collaboration editorial groups, the quality of reporting in Cochrane Reviews tends to be much better than in non-Cochrane reviews.[76]

But despite the advantage of being synthesized and quality-appraised, a Cochrane Review – running often over 100 pages - is still too lengthy and complex a document to be truly practical in a many pragmatic clinical or policy contexts. In order to be as transparent as possible, they include much methodological detail such as search strategy, inclusion/exclusion criteria, as well as results from the wealth of outcomes from the individual trials. They are also filled with scientific language and concepts. This makes them poorly suited for use in time-pressured settings. Studies of information use among clinicians as well as policy makers return continuously to the same conclusions: if evidence is going to be used, it must be represented in forms that are easy to access and interpret quickly.[41, 69, 77-80] The document format of a Cochrane review, although itself a condensation of volumes of research, may still be too long and too complex for use in clinical and policy-making settings.

Other forms of synthesis and summary

Length is a not just a problem for Cochrane Reviews and other systematic reviews; single articles can also contain too much detail for quick access and use practical settings. To address problems of length in research publications, several condensed summary formats have emerged. Structured abstracts, originally developed to assist readers in retrieving, selecting and critically appraising relevant literature have now become common in biomedical journals.[81-83] These are often the only part of a study or review that readers view or use.[84, 85] More recently other forms of journal summaries have surfaced, such as the ELPS (Electronic long, paper short), Short Cut and

⁵ Accessed July 25, 2010

Pico formats developed by BMJ, motivated by the need for a better utilization of the respective advantages of paper versus online formats.[86, 87]

Secondary journals, such as ACP Journal Club, McMaster Plus and evidence-based journals, produce brief summaries of individual studies and reviews, selected in a systematic way for their clinical relevance and newsworthiness.[88-91] These quality-assessed resources aim to limit the number of journals one needs to access in order to keep abreast with new research and the amount of effort needed to spend on critical appraisal.

The 6-s pyramid describes a typology of increasingly condensed and clinically useful formats of summarized evidence: from studies (and their abstracts) to syntheses (systematic reviews), synopses (e.g. ACP Journal Club), summaries (e.g. Clinical Evidence) and systems (e.g. EPJ reminders) [81, 88, 92, 93]

The higher up the pyramid, the more tailored evidence becomes to clinical contexts, as the information is rewritten, reframed or re-contextualized through editorial treatment. The pyramid authors encourage practitioners to start at the top and search for the most condensed, summarized and clinically relevant forms of evidence rather than beginning at the bottom of the pyramid searching for individual research articles (where the volume is greatest and quality variation largest).

Despite all ongoing efforts to render evidence more easily accessible in condensed and summarized formats, little is known about how the summary formats (such as those closer to the top of 6-s pyramid) are *experienced* by clinicians or policy makers. Furthermore, the *effect* of reading a summary as opposed to a less condensed format, such as a whole systematic review, is unknown. In addition, summaries of evidence may be imprecise, as earlier work carried out within the Cochrane Collaboration has revealed. Summarization of complex statistical information is challenging work and can easily lead to misrepresentation of the original data.[94]

This is the starting point for the studies in this thesis.

Aims and objectives

The thesis is written on the basis of work carried out in a cross-disciplinary set of projects aiming to facilitate evidence-informed health care, involving development and evaluation of three separate information artifacts related to the use of Cochrane Reviews:

- ***The Cochrane Library***. This is the web site where Cochrane Reviews are published. It also houses other databases of systematic reviews, clinical trials, health technology assessments and economic evaluations, and is a central resource for evidence on the effect of interventions in health care.
- ***Summary of Findings Table template***. Summary of Findings tables summarize the main results of a Cochrane Review, presenting what we know and what we don't know about the effect of a treatment (including harms) and the quality of that evidence, in a 1-page format. The template forms a basis for producing tables through the Cochrane Review authoring and publishing systems.
- ***SUPPORT Summary template***. SUPPORT summaries are 5-8 page derivatives of systematic reviews specifically targeted at policy makers in low and middle-income countries. The template is a Word document for summary authors.

My contribution has been in the area of design, and the thesis is framed from a design perspective. Following is a more detailed description both of the pragmatic goals and the research questions underlying each of the four studies, as well as a description of the design research inquiry driving this project as a whole.

DESIGN RESEARCH INQUIRY

The overarching aim of this thesis is to explore how a design approach can help facilitate evidence-informed health care. This is done in two different ways: 1) through the design development of specific artifacts that lead to improved user experience of systematic reviews and 2) through design-

relevant knowledge from that emerged while developing these artifacts. What methods might we use; what results can we achieve; how might a design perspective contribute to a better understanding of the problems involved and – conversely – what might we learn about the nature of design while carrying out these inquiries?

STUDY 1: COCHRANE LIBRARY EVALUATION

In this project the pragmatic objective is to recommend improvements of the Cochrane Library to the Cochrane Collaboration Steering Group and Wiley's web site publishing team. The research objective is to answer the following questions:

Research question 1

How can we improve health professionals' user experience of the Cochrane Library web site?

- a) *How do health professionals experience the Cochrane Library?*
- b) *What design and content related improvements of the site do these findings suggest?*

STUDY 2: SUMMARY OF FINDINGS TABLE DEVELOPMENT

In this project the pragmatic objective is to develop a template for Summary of Findings Tables for use by health professionals that would summarize the main findings from a Cochrane Review in a condensed form. More specifically, the table should:

- Present what is known about benefits and harms, what is not known about benefits and harms, and the quality of the evidence
- Be compatible with the needs of a clinical audience
- Not oversimplify or incorrectly present the data
- Be acceptable to Cochrane stakeholders
- Be feasible to implement within the technical constraints of the Cochrane Review publishing system
- Contain enough information to be easily adapted as a stand-alone product
- Not lose vital functionality or information when reproduced in PDF format or printed out on paper.

The research objective is to answer the follow questions:

Research question 2

How can we improve health professionals' user experience of Summary of Findings Tables for Cochrane Reviews, while at the same time securing stakeholder acceptance and publishing feasibility?

- a) *How do health professionals experience the Summary of Findings Tables?*
- b) *What characterizes the stakeholder feedback?*
- c) *What design and content related improvements of the table do these findings suggest?*

STUDY 3: SUMMARY OF FINDINGS TABLE EVALUATION

This study is a part of the same project as Study 2. The research objective of this study is to evaluate the version of the table that we arrived at through the work in the previous study:

Research question 3

What is the effect of including a Summary of Findings Table in a Cochrane Review on user satisfaction, time to find key results, and correct understanding of the main results of a Cochrane Review, compared to a Cochrane Review with no table?

STUDY 4: SUPPORT SUMMARY DEVELOPMENT

The pragmatic objective of this project is to develop a template for short summaries of existing systematic reviews, tailored for use by policy makers in low and middle-income countries. More specifically, the summary should:

- Present evidence from systematic reviews: what is known about benefits and harms, what is not known about benefits and harms, and the quality of the evidence
- Be compatible with the needs of policy makers in low and middle-income countries
- Not oversimplify or incorrectly present the data
- Be acceptable to SUPPORT stakeholders
- Be feasible to produce and publish a series of 100-150 summaries from this template within the SUPPORT project publishing system parameters
- Be short (limited to 5-7 pages)
- Not lose vital functionality or information when reproduced in PDF format or printed out on paper.

The research objective is to answer the follow questions:

Research question 4

How can we improve health policy makers' user experience of SUPPORT summaries of systematic reviews, while at the same time securing stakeholder acceptance and publishing feasibility?

- a) *How do health policy makers experience SUPPORT summaries?*
- b) *What design and content related improvements of the SUPPORT summary template do these findings suggest?*

S C H E M A T I C I L L U S T R A T I O N

On the next page is a figure illustrating the main research questions that this thesis is based on, providing a schematic picture of how the projects, studies and research questions relate to the thesis as a whole.⁶

⁶ Some of the research questions above and in the figure below have a slightly different wording than the objectives in the articles. This is because the articles present both the pragmatic objectives and the research objectives of the studies.

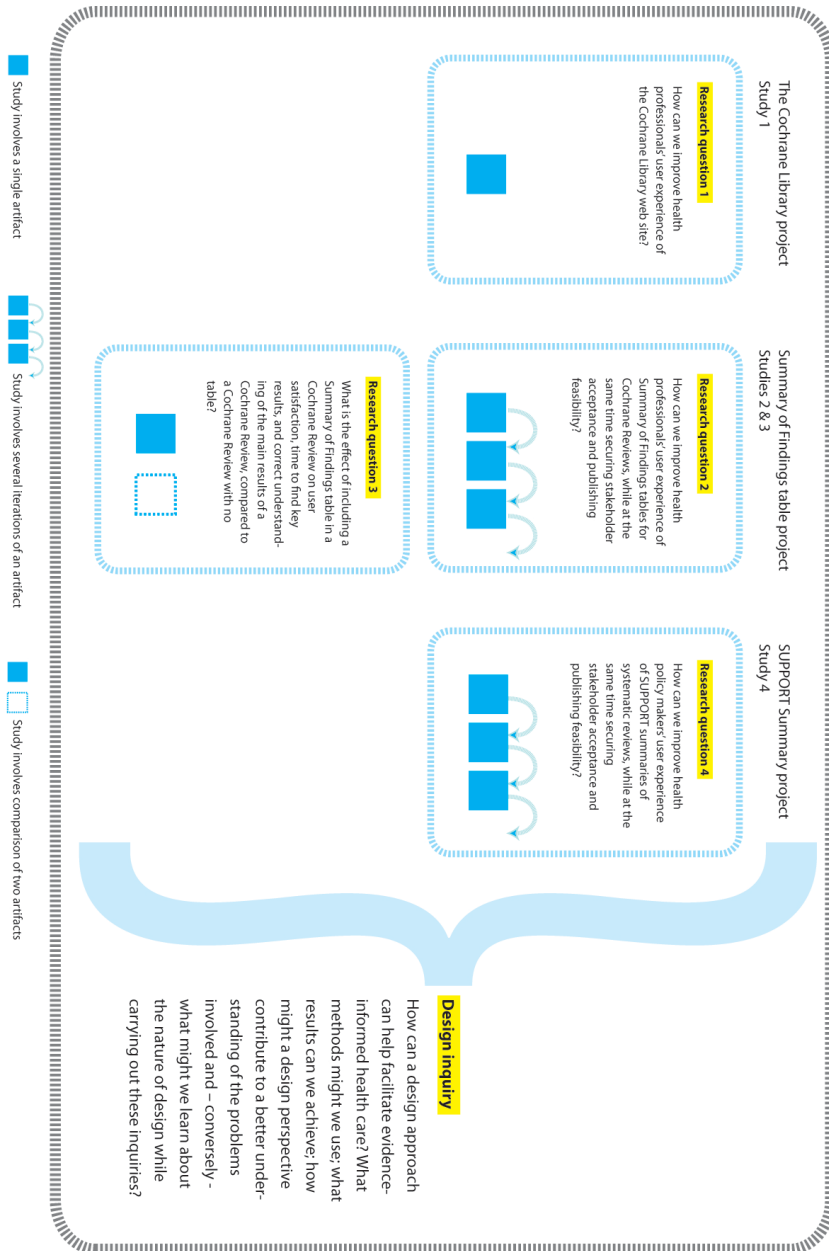


Figure 2. Schematic overview of research questions and inquiries, based on four studies carried out within three projects.

Underlying theories, frameworks and concepts

In this section I present some of the relevant background material that can contribute to a deeper understanding of both the methods used in this body of work and the findings growing out of it. Because of the multidisciplinary nature of the studies, the background information here is taken from several different areas: health care research, cognitive science, education research, and design practice/research.

CONCEPTS AND FRAMEWORKS FROM HEALTH CARE RESEARCH

Much research and practical efforts have addressed the problem of the evidence gap in clinical practice and policy. These initiatives come from a variety of research vantage points and with different aims and scope. Examples of some of these related areas are: evidence-based medicine (EBM), medical decision making, knowledge translation, knowledge transfer, knowledge exchange, research utilization, implementation, research dissemination, diffusion of innovation, continuing education, continuing professional development. There is a plethora of partly overlapping terminology, conceptual frameworks and pragmatic strategies, from which no solid consensus seems to have emerged of how to best describe or frame the work. Criticism in recent years points to the observation that few of the underpinning conceptual frameworks are explicitly theory-based.[95] An in-depth description of this somewhat chaotic landscape is beyond the scope of this thesis; reviews can be found that present them in more detail.[96-98] Those included here illustrate some of the broader developments in this area of work during the past two decades, as well as the areas most relevant to my research and perspective.

Evidence-informed health care

Initially there was an implicit belief in the model of "passive diffusion" - that strong evidence would filter down into practice and policy through journal

publication and medical conferences.[99, 100] Evidence-based medicine (EBM) emerged in the mid 90's, challenging the effectiveness of passive dissemination and promoting a more active approach that quickly gained support.[101] The ideas underlying EBM can be traced back to an a British epidemiologist, Archie Cochrane, who in the early '70's drew attention to disparities of treatment between different health care institutions and emphasized the importance of supporting medical decision making with valid evidence in order to raise the quality and equality of care. The EBM concept emerged from a core group at McMaster University who proposed measures that could be taken by health care professionals to more consciously integrate updated findings from research in their clinical work.[102] Practicing "evidence-based", according to Sackett in 1996, meant:

"The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."[103]

The original EBM model had two components: the integration of "individual clinical expertise" with "the best available external clinical evidence from systematic research." The five-step method (searching, selecting, appraising, integrating and evaluating) is the foundation for training students and professionals in EBM practice.

In recent years, the EBM framework has been extended to include not only the best available evidence and clinicians' expertise, but also patients' values and preferences. This reflects in part the movement in the last decade toward patient centered care and shared decision making⁷. [31] Basic principles from EBM have also been applied to fields in allied health care - e.g. physical therapy - as well as to the broader areas of health system management, guideline development and policy making. Proponents of EBM have increasingly acknowledged the complex nature of health care decisions, whether at the individual patient or national policy level. Decision-making is seen as being influenced by contextual differences such as societal values and priorities as well as local constraints and resources; it can therefore not be "based" merely on evidence.[104] This has given rise to a broader term "evidence-informed health care" and to broader research frameworks, which I discuss briefly below.[105] Though I will occasionally continue to use the abbreviation EBM in this thesis, the phrase "evidence-informed health care" more accurately describes my current understanding of this concept.

⁷ The shared decision making perspective sees the patient as a potentially active co-participant in decisions regarding their own treatment.

EBM has not emerged as an explanatory model of what actually happens, but rather represents an *idealized* version of what should happen for evidence to cross over into practice or policy. The work of the Cochrane Collaboration is aimed at enabling evidence-informed health care, in part by alleviating some of the effort in searching for and appraising evidence through systematically collecting, evaluating and synthesizing the evidence. My work is embedded in the goals of this organization, and is a part of the efforts of researchers with ties to this community to continue to enable evidence-informed health care.

Criticism of EBM

The EBM “movement” and work carried out by Cochrane Collaboration is not uncontroversial.[106-111] For instance, some health professionals feel threatened by what they claim is “cookbook” medicine. This stance seems to be based on a fear that evidence will be used to dictate rather than inform decision-making and a threat to the physician’s autonomy. There are also protests of a philosophical nature, against empiricism in medical practice (the elevation of experimental evidence above traditional clinical thinking which is characterized by a coupling an understanding of basic mechanisms of disease with clinical experience). Critics remind also remind us that philosophers and scientists have considered the basis of empiricism - theory-free objective observation – impossible since the late 19th century. Related to this are concerns that knowledge from randomized control trials will dominate, drowning out evidence from other kinds of studies, such as observational studies, qualitative research or tacit knowledge from practice. Others argue that quantitative data places too much emphasis on effect (which in many reviews is found to be uncertain anyway) and would rather have help understanding *why* an intervention did or didn’t work, or *what* the local cost implications are. Objections are also made that it is problematic to synthesize evidence and assume its relevance across geographical and cultural borders.

Cochrane Collaboration as an organization and EBM as an ideal, on the other hand, don’t represent static, un-reflected positions. Transparency of method, for instance, is an explicit trait of the work of the Collaboration that encourages and invites criticism to begin with, and “encouraging diversity” is part of the ten basic principles.[112] The organization also attends directly or indirectly to criticism about quality or epistemological dominance, for instance by continuous development of their underlying methodology, through systematic dialog with non-research groups (e.g. consumers) and through more recent efforts to include qualitative research in Cochrane Reviews. Cochrane Reviews avoid the cookbook label by publishing only the

findings of the research, not recommendations based on these findings. “Cookbook” criticism might also be responsible for the reframing of “evidence-based” practice to “evidence-informed” practice in EBM literature the last decade.

My own position

My own position on these issues is rather pragmatic – I have approached this work much in the way I would approach other design work. For me, the Cochrane Collaboration represents a “client” with a communication problem, and I am exploring ways in which design can help with this problem. I can, however, see valuable arguments on both sides of this debate. As a non-medically trained citizen, I highly appreciate the kind of knowledge Cochrane Reviews bring to the public discourse and to professional practice and policy that otherwise might be dominated by out-of-date expert opinion or misleading information from drug industries. However, as a designer I understand that information cannot be viewed as objective and context-free. For instance, even the understanding of something as seemingly objective and stable as *numbers* is sensitive to context and subjective point of view, a fact that these studies help illustrate. My position however, does lean toward the evidence-informed standpoint, as I maintain that this kind of scientific evidence should be made widely available. As a patient, I have personally experienced the empowerment of having knowledge about the evidence base of a proposed treatment. I also believe in the general value of the democratization of scientific information. My hope is to help the client move towards this vision, albeit in ways that are informed by end-user perspectives.

Krippendorff suggests that design is a science of second-hand understanding – that a designer’s area of expertise involves a highly focused sensitivity of what artifacts mean to other people, rather than in a direct first-hand understanding of an artifact itself or other subject matter.[15] This is a good way of understanding the role I have had in these projects, trying to accommodate and negotiate the experiences and feedback of others coming from the different worlds of stakeholders, users and researchers. Initially this might look like a neutral position, but as I have been working within mandates from within the Cochrane Collaboration – seeking to improve the use of their artifacts - my position has not been completely neutral. Some implicit concepts have not been questioned by my work, such as whether or not research results from Cochrane Reviews are desirable in the first place. I *assume* an implicit motivation for these artifacts in clinical and policy settings – indeed much of the EBM literature takes the inherent value of evidence from randomized control trials for granted – whereas in real life, that view is not uncontested (as discussed above).

In the final section of this thesis, I reflect in much more detail on my role as designer in these projects: how I experienced participating in this work and what that might mean for ways we think about and talk about design as a profession.

Need for a broader approach

Despite the developments EBM has undergone since its inception, the underlying concept (as it is taught in EBM courses and workshops) remains largely focused on improving the skills of the individual practitioner, an approach that has perhaps not demonstrated huge success on its own. Behavior of the individual is not easily isolated from the social and organizational structures in which work is situated. Broader frameworks were needed for bridging the research-practice gap than those focused only on individual behavior. In 2004 Grol and Wensing outlined a number of theories/models that illuminated factors needing to be considered in improving diabetes care, relating to individuals, to social context, and to organizational and economic context.[113]

Diffusion of innovation - an updated model

Also in 2004, Greenhalgh et al executed a comprehensive literature review of a wide range of literature on the spread and sustainability of innovations in health service delivery and organization.[114] It was based on the same conviction that the research-practice gap needed a broader, multi-level framework, and also that there were lessons to be learned in many other areas of research, particularly those studying diffusion of innovation. Diffusion of innovation is a theory developed by Everett Rogers that seeks to explain the spread of new ideas, growing out of studies from the 50's in rural sociology where researchers were trying to understand the spreading of agricultural innovations among farmers (such as improved corn seeds).[115] Rogers' work identified four main components that influenced spreading (the innovation, the communication channels, time, and a social system) and five stages of progression (awareness, persuasion, decision, implementation, and adoption.)

Greenahlgh et al. identified and summarized evidence from 11 major research traditions that had, largely independently of each other, addressed the issue of diffusion, dissemination or sustainability of innovations. Most but not all were related to health care.⁸ The findings of the empirical studies they

⁸ These included: rural sociology, medical sociology, communication studies, marketing and economics, development studies, health promotion, evidence-based medicine and guidelines

retrieved were grouped under seven categories or components, replacing Rogers four original themes: 1) the innovation itself, 2) the adoption process, 3) communication and influence (including social networks, opinion, leadership, and change agents), 4) the inner (organizational) context, 5) the outer (inter-organizational) context, 6) the implementation/sustainability process, and 7) the linkage between these components. The review authors identified attributes for each of the seven categories that appeared to influence the uptake of new innovations in an organizational context. For instance, the *characteristics of the innovation* found to be important to uptake are:

- Relative advantage
- Compatibility
- Low complexity
- Trialability
- Observability
- Potential for reinvention
- Risk
- Task issues
- Nature of knowledge required (tacit/explicit)
- Technical support

Recommendations for future research from this review maintained that work should not be restricted to a single level of analysis, but rather address important *interactions* between different levels or components, for instance interactions between the innovation and the adopter. The work in this thesis is an example of such research as it involves investigating the interaction between different components of the model: the innovation and the (potential) adopter, and the linkage between producers and users.

Other knowledge translation frameworks

The diffusion of innovation framework is quite broad, covering a whole range of new "things" - such as technological devices - as well as "new knowledge" emerging from research. However, other frameworks deal more specifically on the latter - knowledge from research - and its movement into practice or policy. Not all of these other frameworks are based on an explicit overarching theory, but nonetheless attempt to simplify and conceptualize a complex landscape. See Table 1 for examples, though this is by no means an exhaustive list. Targets differ; some are aimed at changing knowledge and understanding, while others focus on the actual change of behavior.

implementation, organizational studies, knowledge-based approaches to innovation in organizations, narrative organizational studies, and complexity and general systems theory.

Table 1: Examples of conceptual frameworks of evidence production and use

Title	Description
CIHR Model of Knowledge Translation[116]	The Canadian Institute of Health Research Knowledge Translation proposed this model that is based on a cyclical concept of research. Six points are identified within the cycle where interactions, communications and partnerships that could facilitate knowledge translation may occur (KT 1-6). These include defining research questions/methods, conducting research, publishing research, placing findings in context of other knowledge, making decisions/taking action, influencing subsequent research.
Knowledge to Action process Framework (KTA)[98]	This framework by Graham et al is made up of two components - 1) knowledge creation and 2) action - that each contains several phases. Knowledge creation consists of knowledge inquiry, knowledge synthesis and knowledge tools/products. The authors indicate that the process is complex and dynamic, with no absolute boundaries between the components or their phases. The strength of this model is, as in the CIHR model, the inclusion of the knowledge creation process, not just focus on the use.
Lomas' policy decision-making framework.[33]	Lomas has sketched a framework for understanding the context of policy decision-making In this model, he has divided the decision-making world into three domains: 1) the institutional structure for decision-making, 2) values, and 3) information. The information domain contains not only research based evidence, but other things such as anecdotes, experiences, propaganda. There are also many purveyors in addition to researchers (e.g. think-tanks, interest groups, media). The strength of this model is the recognition that "evidence" may hold many meanings for different people and that decisions are not only influenced by many different forms of information, but also by ideologies, beliefs and interests.
Diffusion of Innovation Framework (updated by Greenhalgh et al, 2004) [114]	This update of Roger's original Diffusion of Innovation Framework was based on a meta-review of empirical research from 11 different research domains. It is a whole-systems model, comprised of seven components: 1) the innovation itself, 2) the adoption process, 3) communication and influence (including social networks, opinion, leadership, and change agents), 4) the inner (organizational) context, 5) the outer (inter-organizational) context, 6) the implementation/sustainability process, and 7) the linkage between these components. It is described in more detail above.

Not useful for design work

None of these frameworks provided sets of specific concepts that I found useful for planning or carrying out design work. They were not fine-grained enough for supporting a detailed exploration of user experience of a particular information-based artifact in order to improve its design. Although the Diffusion of Innovation Framework came closest, with its list of characteristics of the innovation that promote diffusion, most of these characteristics were still too general to be useful to me. Since this framework is covering all kinds of innovations (from “things” like corn seeds to “knowledge” like health care evidence and guidelines), it also seems to suffer from a lack of division between an innovation (e.g. a new treatment) and the *representation* of this innovation (e.g. information about this treatment), making it less suitable for my work.⁹

These frameworks, however, do share one important feature that is relevant: they all indicate a gap or at least a distinction between two worlds - production and use - and they all represent attempts to describe the conceptual space where these two areas meet, link, interact or overlap. It is in this conceptual space my work takes place.

A schematic illustration of my work

The description of these meeting points is for the most part rather vague in the frameworks I examined above; concepts such as “transformation” (KTA) or “contextualization” (CIHR) of knowledge are employed. The revised Diffusion of Innovation theory again goes the furthest by including “linkage” between the areas of production and use as an independent component with its own set of characteristics, not just a kind of magical movement represented by arrows. Linkage in the development stage of an innovation is characterized in the following way:

“Linkage at development stage: If the innovation is centrally developed (e.g. in a research centre), it is more likely to be widely and successfully adopted if the developers or their agents are linked with potential users at the development stage in order to capture and incorporate the user perspective (strong indirect evidence[117]). Such linkage should aim not merely for ‘specification’ but for a shared and organic (developing, adaptive) understanding of the meaning and value of the innovation-in-use, and should also work

⁹ Ancker found a similar weakness in risk communication and decision models underlying patient information. (Ancker, J. Interactive graphics for communicating health risks. 2009. PhD thesis. Department of Biomedical Informatics, Columbia University. New York. 2009)

I started with this triangle to represent the “use” side of health care evidence (left half of Figure 4), and added a second triangle to represent the “production” world, with the information artifact as a shared object in both worlds. Although the “production” triangle (the research world) is not based on a review of literature, it was a logical extension of Ancker and Kaufman’s findings and provided me with a conceptual illustration of the actual pragmatic structure of the projects I was working on.

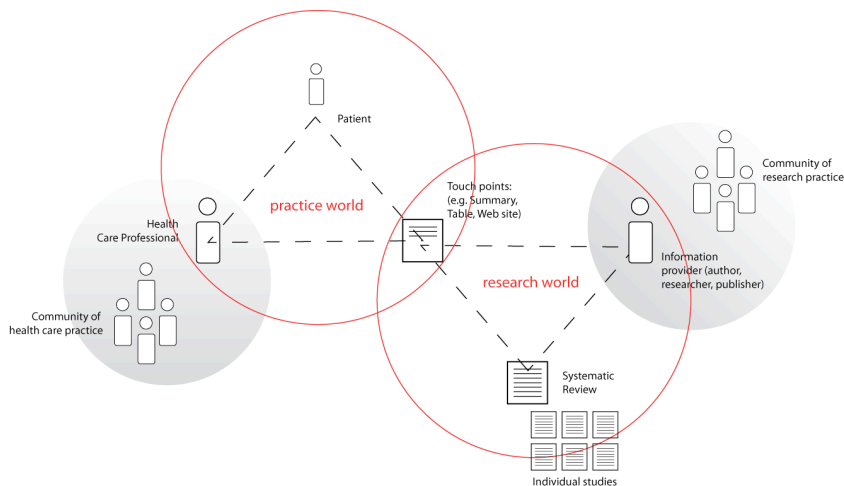


Figure 4: Information artifact as a point of contact between the world of practice and the world of research production.

This figure illustrates the dual role that information artifacts may play as they straddle both worlds, in effect acting as a link or point of contact or collaboration arena between the world of use and the world of production. It also illustrates the importance of collecting feedback from actors from both sides to "get it right" – to create information artifacts that are feasible, acceptable and carry meaning in both contexts. This feedback collection process may also serve to create shared understandings and language between these two worlds, as called for in Greenhalgh’s linkage description above. By studying this micro space, it may be possible - through collaborative efforts to improve shared artifacts - to develop both solutions and shareable knowledge that serve to negotiate and resolve conflicting sets of needs. This work might help to “bridge” a part of the gap or at least contribute to a somewhat better understanding of aspects of this gap.

THEORIES AND CONCEPTS FROM COGNITIVE SCIENCE

If I were reporting on the design of a hammer, I might include a section on the ergonomics of the human body as related to holding a heavy object, or the compatibility of wood or plastic to being modeled to fit the grip of the human hand. But evidence is a knowledge tool for use in cognitive rather than physical activity. Therefore I am including a section here that describes some of the cognitive issues that I find important in understanding people's use of such knowledge tools.

In many studies of the translation of research into practice or policy in health care focus, the focus is exclusively on one type of cognitive activity: decision making. However, my work is looking at the multiple stages of cognitive activity that occur *before* decision-making: acquiring and assessing the evidence (the "searching" and "appraisal" phases according to the EBM five-step model). Concepts from cognitive science can help illuminate the nature of knowledge and knowledge acquisition that will be relevant to understanding my work.

Distributed cognition

Cognitive science is a multidisciplinary domain of study of both the basic aspects of cognition (e.g. attention, memory, comprehension, language acquisition) and the development and evaluation of cognitive artifacts (human made materials, devices and systems designed to extend people's knowledge and cognitive skills). For this reason, it has close ties to both educational research and the Human Computer Interaction domain (described below).[118] Research of cognitive phenomena seems to spring from two vantage points that lie back to back, like two sides of a coin. On the one hand, researchers are looking inward to the processes and structures of the mind and the knowledge and skills of the individual; on the other hand studies focus outward towards the nature of tools and contexts of the external world. The richness of this field comes from the basic recognition that these two domains are not operating in isolation but are closely entwined and interacting with each other.[118] This perspective grows out of the theory of distributed cognition.

The concept that cognition is *distributed* is relatively new. Study of cognition began in the late 1950's with a focus on thought and mental states, growing out of recognition of the limitations of behaviorism, where the analysis of observable behavior was thought to be empirically superior to invisible mental processes such as thinking. The "computer as a metaphor" had an

important influence on the development of the theory of symbolic information processing, which has been the dominant conceptualization of cognition from the late 60's until recently. Early research included laboratory experiments examining the nature of basic mental processes such as perception, attention, memory and problem solving. Newell and Simon pioneered in the area of problem solving and introduced the methods of think-aloud protocol and protocol analysis to capture inner mental processes not otherwise revealed to an observer.[119, 120] In the 70's there was a shift of approach from basic research in the lab to more complex real-life studies, as scientists began to acknowledge the important role of *context*. [121, 122] This move was led by an investigation into the nature of problem-solving expertise in rich real-world settings such as workplaces and classrooms. In recent years, the theory of symbolic information processing – which views cognition as a property of the single individual – has increasingly given way to the concept of cognition as being distributed across groups, tools and cultures: *distributed cognition*. [123] This concept was pioneered by Hutchins, who provided a rich description of a navy navigation team in *Cognition in the Wild*. [124] The finely tuned interaction between socially organized groups of individuals, tools and information in that navigational environment - working seamlessly to continuously update the location of the ship and adjust its course - illustrated how inseparable the properties of the minds, cognitive artifacts, and the environment were in performing complex tasks.

Properties of knowledge

Knowledge is not the same as kind of thing as a package of seeds or a new drug – it has unique characteristics that influence its diffusion, uptake, storage and use. The most basic difference between knowledge and innovations such as corn seeds is that knowledge is not a 'thing'.¹⁰ Earlier concepts of learning conceptualized knowledge as facts to be transported with from teachers to students. [125] However, it is now widely accepted that knowledge cannot be transported in this packaged fashion. The individual learner, based on what he already knows or believes, reconstructs new knowledge. [125, 126] Research about how people learn has shown that learning is accomplished when one is able to connect the new information with what one already knows, building on previously stored background knowledge and concepts. Research also shows that learning is enhanced

¹⁰ The 'diffusion of innovation' framework mentioned earlier didn't make this differentiation. It built on studies that were looking at the uptake of all kinds of innovations, from new corn seeds to technical devices to medical guidelines.

when educators pay attention to the incomplete understanding that learners bring with them and start from there.[127]

Knowledge types and structures

Representations reflect states of knowledge. A cognitive representation “refers to a correspondence between a mental state and objects out in the physical world”.[118] Representations can also be external, for instance texts, numbers, pictures, diagrams, maps, globes, etc. Some examples of different representations of the same content are: 10 per 1000, 10%, 0.10, “one out of ten”, 1/10, or a picture with 10 squares, the first of which is colored green and the remaining nine colored black, (or this text you are reading that is describing that picture).

Complex tasks can often be made simpler by altering external representations, which in turn changes the nature of the task and alleviates cognitive load. Cognitive load refers to the mental effort a person needs to exert to complete a cognitive task. For instance, adding up multiple-digit numbers that are written horizontally in a row is a cognitively taxing task. Altering the representation by lining up the digits vertically changes the task so that the numbers are easier to add. Representations can be changed to be more compatible with the cognitive challenges related to a particular type of user, task or context. A representation that is suitable for one context (such as reading a sign as you enter a subway station) may not be suitable for a similar task in a different context (reading the same sign as you drive at high speed down a highway). Representations may have effects, for instance how different representations of the same content can influence efficiency, task difficulty and behavior.

Propositions are a form of cognitive representation that captures the gist of an idea – the meaning – as opposed to the exact details.[118] Propositions are a central concept in theories of comprehension. Conceptual structures such as schemata can be thought of as having “slots” for propositions. This helps explain why experts in a domain may appropriate new knowledge more readily than non-experts: they already have a ‘slot’ in place for the new information in a well-developed conceptual structure.

Humans are thought to be organizing knowledge in mental structures called *schemata*. [118] These are structures for representing concepts in categories (such as animals, types of transportation or font families). Other types of conceptual structures are mental models, such as an explanatory set of ideas about how the solar system works. Information may be placed in the wrong schema, for instance placing the sun in the category of "planets" instead of

"stars". Or the conceptual structures themselves may have faults, such as the idea that the earth is in the center of the solar system and that the sun revolves around it. The organization of information into conceptual frameworks allows for greater "transfer" - helping a person apply the acquired knowledge to new situations.[127]

Novices and experts - differences that make a difference

In cognitive science, a better understanding of how we use and acquire knowledge is often gained through studies exploring the differences between experts and novices in different domains. Through this work, several general characteristics of experts (independent of domain) have been established. To begin with, an expert typically has a both a larger factual base and a better understanding of these facts and ideas through well-developed conceptual frameworks within his field, such as schemata or mental models. Experts' command of these frameworks also facilitates their understanding and appropriation of new information - they have 'slots' in which to place new propositions (information) and a better ability to see patterns, relationships or discrepancies not apparent to novices. This helps them more quickly identify and remember relevant information. Hence, complex events cause less cognitive load and strain on their attention. Experts are also good at retrieving knowledge relevant to a particular task, through an understanding of the conditions in which the knowledge is useful. Experts have also been shown to monitor their own understanding carefully, taking notice of whether new information was consistent with what they already know, and what analogies could be drawn that would help their understanding. These are meta-cognitive skills. Additionally, experts have a gist (vs. verbatim) processing preference.[125, 127-129] Experts and non-experts have therefore very different perceptions of and relations to the same information. For this reason, experts are not likely to be the optimal source of input on how to design a knowledge-related product for use by non-experts.

In the discussion I draw on some of these concepts to illuminate the findings from the four studies.

THEORY AND CONCEPTS FROM DESIGN

A third area providing useful background knowledge about underlying concepts in this work is the domain of design practice. I have already discussed the nature of design inquiry in the beginning of this thesis. Below is a description of the characteristics of design processes and an introduction to the concept of user experience.

What characterizes design processes?

As mentioned earlier, one basic characteristic of design is that it involves creating things that don't yet exist. This differs from working with things that already exist, although design may also involve doing that to a greater or lesser degree. As pointed out earlier, several scholars - Lawson, Schön, Cross, Gedenryd, - have described design practice as a process of interactive (or reflective) inquiry: creating a solution based on what we know (and what we think the problems are), trying out that solution (either as a sketch for the designers own use or as a prototype to be tried out by typical users), learning from responses to these trials about what works or what doesn't, adjusting the solution (or re-evaluating the problem) and repeating the process.[7-10]

Design inquiry differs from a linear think-plan-act model of problem solving where a problem is stated, an analysis undertaken and a solution executed and evaluated. In design inquiry, a problem may not be understood properly to begin and might therefore be stated tentatively or vaguely. In trying out solutions, the nature of the problem becomes clearer, and can be refined. Many partial solutions may be tried out, and incrementally one arrives at a result. This can entail adjustment underway as to the definition of the problem. Schön stresses the activity of problem setting as characteristic to the nature of design - designers may reframe a problem to fit a solution they know they can produce. Cross points out that designers seem to start out by challenging “givens” or restraints in an assignment, even if these seem relatively straightforward. Goel and Pirolli argue that design problem spaces are unique from non-design problem solving tasks.[6]

The principles of reflective or interactive inquiry are typical of pragmatic design methods. These designerly processes can be useful, not only for developing solutions, but to check underway that our understanding of the problems is correct. We can make sure we are aiming at the right targets to begin with, not just testing the quality of the arrows. I have introduced cyclical reflective methods into these projects as a designerly way of improving the information artifacts we were concerned with.

User-centered design and participatory design

As stated earlier, the underlying values and methodology of this work is grounded in a user-centered design approach. Preece et al suggest that this approach builds on five principles: 1) User's tasks and goals are the driving force behind the development, 2) User's behaviour and context of use are studied and the system is designed to support them, 3) Users' characteristics are captured and designed for, 4) Users are consulted throughout development from earliest phases to the latest and their input is seriously

taken into account, and 5) All design decisions are taken within the context of the users, their work, and their environment.[130]

These projects all involved collecting feedback and input from a large group of people, both end-users and stakeholders, to improve information systems. This extensive collection of feedback and the degree to which it informed the design results can be seen as bordering on participatory design - an approach to systems design that actively involves users or other stakeholders in the making processes. In participatory design, however, users and stakeholders normally partake more extensively in the decision making, rather than just providing input for designers. Nevertheless, there are comparable underlying motivations. In Scandinavia participatory design has its roots from workplace democracy projects in the 70's.[131] The socio-technical approach stresses that all stakeholders share a common interest in a well-functioning workplace, and aims to balance both social/cultural and technical systems in consideration of the organization as a whole. The critical approach, however, emphasizes the uneven distribution of power among differing groups of stakeholders and collaborators, giving the work a more political slant where researchers often have the role of supporting the weaker factions.[132] Bratteteig writes:

“User participation is a characteristic of the Scandinavian approach. Moreover, user participation and participatory design seem to be conceived differently in Scandinavia than many other places, emphasizing that users are co-designers and that systems development is an organizational, technical, and human change process. User participation will improve the knowledge upon which systems are built and enable people to develop realistic expectations, thus reducing resistance to change. In addition, user participation aims to increase workplace democracy by giving the members of an organization the right to participate in decisions that are likely to affect their work – a view that is shared by the critical and socio-technical traditions.”[132]

This work has philosophical traces of both the socio-technical and critical approaches, but is not explicitly defined within either of these traditions. As we were looking at small parts of larger information systems in laboratory-like test settings, exploration of the social cultures and contexts of the work affected by the artifacts we developed were not possible within the boundaries of this work. And, although we were dedicated to bringing in the voices of end-users into development as in the critical tradition, this view is – in principle - not in conflict with explicit values in the Cochrane

Collaboration. What has been previously lacking in the Collaboration is a well-informed and collectively accepted understanding of the consequences of taking the end-user perspective seriously (although under recent changes in leadership this issue has been given increased priority).

Usability

End-user orientation has become almost synonymous with emphasis on “usability”, a concept that has grown out of the fields of human factors and human-computer interaction. Human factors concentrates on ergonomics - the science of understanding the properties of human capabilities, both physical and cognitive, largely in human-machine contexts. Human-computer interaction (HCI) is a related field, but with an exclusive focus on people working with computers. HCI has traditionally focused on studying the instrumental uses of products as they are related to work-tasks, and has been traditionally dominated by the study of usability.

An ISO definition of usability is:

"Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."[133]

Jacob Nielsen defines usability through five quality components:[134]

- **Learnability:** *How easy is it for users to accomplish basic tasks the first time they encounter the design?*
- **Efficiency:** *Once users have learned the design, how quickly can they perform tasks?*
- **Memorability:** *When users return to the design after a period of not using it, how easily can they reestablish proficiency?*
- **Errors:** *How many errors do users make, how severe are these errors, and how easily can they recover from the errors?*
- **Satisfaction:** *How pleasant is it to use the design?*

Nielsen maintains that there are many other key attributes, such as “utility” which he regards as equally important as usability.

Usability is a powerful concept, particularly because there are aspects of it that can be precisely measured: number of tasks successfully completed, time spent and number of errors. It is well suited to evaluation of products or systems created for well-defined and easily measurable tasks, such as using a new electronic device successfully or making an online purchase.

However, testing for usability may place a rather narrow focus on instrumental functionality and doesn't capture as easily other aspects of using an artifact, such as being aware of it or trusting it.

User experience

User experience is a relatively recent concept growing out of the research and practice domains of HCI and interaction design. It has emerged largely as a reaction to the narrow focus on "mere" usability in these domains. I chose to focus on user experience for the same reason – because it potentially covers a broader range of issues than usability alone. There is no current consensus on the definition of user experience, what its components are, or how to best measure it,[135] although some maintain that it is more about what a system or product *feels* like.[130] There is also disagreement about how it relates to usability - whether they are parallel constructs or whether usability is a sub-component of user experience. However, there seems to be agreement that user experience encompasses at least some of the non-instrumental uses and emotional dimensions of humans' interaction with a product.

Here is a selection of some of the definitions that have been proposed:

User experience may be:

“All the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it.”[136]

“... all aspects of the end-user's interaction with the company, its services, and its products.”[137]

“...a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.)”[138]

As a designer with a background in corporate identity, I find much value in the second definition (from Nielsen Norman Group). It acknowledges that user experience is influenced through multiple points of contact with an organization, not just through interaction with one product.

There are also qualities in the third definition (from Hassenzahl & Tractinsky) that are important in the context of this work. They call particular attention to the interrelation between user, product/system and context/environment, reminiscent of the basic concept described earlier that cognition is distributed, not just an internally determined process. In my work, however, I could not begin to separately test all three of these dimensions (user, artifact, and context). I needed a highly pragmatic framework that could lend itself to uncovering aspects of user experience related specifically to improving the design of the artifact, and that could be combined with (relatively rapid) prototyping.

User experience frameworks

Forlizzi and Batterbee group theories and models of experience into three basic categories: product-centered, user-centered and interaction-centered.[139] The first tend to often be checklists or guidelines about how to develop products in order to support desired experiences. The second are approaches that aim to help designers and developers better understand the users of their products. The third has a focus on the interactions between product and user. Their framework defines three different kinds of individual interactions with a product – fluent (automatic, well-learned), cognitive (focus on product at hand), and expressive (forming a relationship to a product) – that result in three different kinds of experiences: “experience”, “an experience” and “co-experience”. Co-experience is about user experience in social contexts, created or shared with other people.

Another dimension of user experience is time. Karapanos et al describe a framework based on their empirical research following new users’ experiences of acquiring an iPhone.[140] They collected data over four weeks, resulting in a conceptualization of three distinct phases: orientation to the product, incorporation of the product into daily routines and identification with the product. They also found that users’ concerns shifted across these phases, for instance from effective use to meaningful use, driven largely by three main forces: familiarity, functional dependency and emotional attachment. They also promoted three directions for practice: designing for meaningful mediation, designing for daily rituals and designing for the self.

Although this work is interesting in drawing up a larger macro perspective of user experience, I didn’t find it suitable for the kind of micro-studies we were undertaking. What I was looking for was a set of concepts, like Nielsen’s list of usability components, but one that covering a wider scope of issues and

that was relevant for exploring and evaluating users' experiences of specific information artifacts.

From my earlier work in a design company, I was familiar a framework by Peter Morville, a tool for exploring user experience in an information design context that he calls the 'honeycomb'. [141] This concept, emerging from his own practice-based work with users, consists of seven different facets of user experience: findability, accessibility, usability, usefulness, credibility, desirability, and value. Since the honeycomb has emerged from design practice rather than formal research, there is no literature describing it (aside from the definitions on Morville's website, accompanied by comments by many readers). I wrote to Morville to ask about how the framework evolved, and received this brief description in reply:

"As far as I can recall, the origin of my thinking on this topic runs back to a talk by Terry Swack in 2000...

http://argus-acia.com/strange_connections/strange007.html

...in which she talked about the importance of usefulness and usability.

Soon after, I started hearing the phrase "useful, usable, and desirable" coming from the experience design community.

Then, around 2002, I realized that "findability" described my passion...

<http://semanticstudios.com/publications/semantics/000007.php>

...and in 2004, I really started to feel that the term "usability" was being used too broadly, that there was a need to define usability as just one important quality of "user experience" and the honeycomb was a way to highlight several other qualities".

I chose Morville's model as my starting point for exploring user experience in these studies, since it provided a conceptual framework that was suitable to the design work I was carrying out: it combined the main facets of usability with some of the more emotional aspects of user experience and the key issue of usefulness. My earlier experience with the concepts in this framework was that they lend themselves well to discovery of compatibility issues between user and product that feed easily into the ensuing practical work of improving the design. I elaborate more on how I used the honeycomb model in the Methods section.

Defining user experience – a work in progress

Since my studies were carried out, further discussion about how to define user experience has taken place. In a paper summing up the relations between usability and user experience to new ISO standards, Bevan writes that facets

of user experience dealing with usability, such as efficiency, effectiveness, or safety, are typical of those that concern organizations because they often impact profitability.[142, 143] User experience goals, on the other hand, are the product seen from the inside out – from the users perspective. They focus typically on dimensions of satisfaction, fun and pleasure. Bevan summarizes these two different categories of goals (pragmatic and hedonic):

“The pragmatic user goals are:

- Acceptable perceived experience of use (pragmatic aspects including efficiency).
- Acceptable perceived results of use (including effectiveness).
- Acceptable perceived consequences of use (including safety).

Hassenzahl identifies three hedonic goals[144]

- Stimulation (i.e. personal growth, an increase of knowledge and skills).
- Identification (i.e. self-expression, interaction with relevant others).
- Evocation (i.e. self-maintenance, memories).

To these I would add:

- Pleasurable emotional reactions to the product (Norman’s visceral category[145]).”

SUMMING UP BACKGROUND

This thesis draws on several different domains of research, practice and theory: EBM provided the motivational background for initiating the projects and the context for implementing results; methods from design practice and inquiry influenced the design of the studies as well as the framing of questions asked; some areas of cognitive science were drawn upon to help analyze and understand the findings.

Methods

One of the central underlying aims of this work as a whole was to explore how a design approach could help facilitate evidence-informed practice/policy making, both through development of practical useful “things”¹¹ and through the knowledge that emerged while developing these artifacts. Therefore, this work had both a pragmatic and a research dimension.

The pragmatic goal was to design or improve information artifacts that might more effectively support the dissemination of summarized evidence from systematic reviews. Our team wanted not only to improve the specific artifacts in a reliable user-centered way. We also needed to produce transparent documentation of the design proposals that would help convince stakeholders, who are largely researchers, of the quality of our results. This provided an added incentive for organizing the pragmatic design work as a set of research studies.

The research-oriented goal was to produce knowledge about these things – their characteristics, the processes used to develop them, knowledge about both users and producers - that was relevant to the designs and the design processes. Our team wanted to extract and communicate findings that could be relevant to others developing similar information artifacts, such authors or publishers of tables, summaries, other documents or web sites with evidence-related content. I also wanted to communicate my reflections of participating as designer in this work back into the design research community.

The four research studies were therefore designed to support both the pragmatic and research goals – to design particular artifacts and to gain design knowledge through that work which could be transferable to those classes of artifacts.

¹¹ I use the words “artifact”, “thing” and “product” intermittently throughout this text.

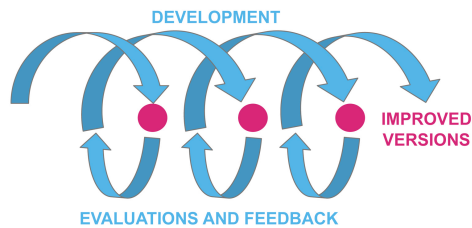


Figure 5. Alternative cycles of development and evaluation/feedback.

A large part of what is meant by “design approach” in this work involved choice of methods. The progression of both the Summary of Findings and SUPPORT projects (articles 2, 3 and 4) was reminiscent of a typical of design project, where alternating cycles of prototyping and evaluating were repeated until an acceptable solution was reached (see Figure 5).

OVERVIEW OF METHODS USED IN EACH STUDY

This overview is to give the reader a snapshot what methods were used in each study. More information can be found in the individual articles concerning issues such as recruitment, choice of outcome measures, how methods were combined, etc.

1) The Cochrane Library study

Pragmatic goal: to evaluate the Cochrane Library web site and make recommendations of how to improve the design of that particular site.

Research goal: to generate findings about barriers and facilitators of use that were transferable to other sites with similar kinds of content.

To achieve both these ends, we explored professionals’ user experience of the Cochrane Library (how they found, used and perceived the Library and its content) through a series of user tests.

Method	Nr of participants	Participant description
User testing Video recorded	32	Health professionals

2 Summary of Findings Development study

Pragmatic goal: to design a template for a table that presented the main outcomes from Cochrane Reviews (Summary of Findings Table - SoF) for use by health professionals.

Research goal: to explore the challenges involved in creating presentations of condensed statistical data from systematic reviews for use by health professionals.

To achieve this we carried out several cycles of the following work: collected and analyzing structured feedback from both users and stakeholders, held workshops to discuss interpretations of problems uncovered through feedback, generated ideas and designed new prototypes.

Method	Nr of participants	Participant description
Brainstorming Workshop	4	Core team of researchers (w/backgrounds in epidemiology, anthropology, journalism, design)
User testing Audio recorded	21	Health professionals (from Norway and England)
Stakeholder Feedback	52	Stakeholders in the Cochrane Collaboration
Graphic design work	1	Carried out by myself

3 Summary of Findings Evaluation study

Research goal: The aim of this study was to measure whether the SoF Table we developed actually facilitated evidence-informed health care by creating easier access to the main findings within a Cochrane Review. We evaluated this by carrying out two small randomized control trials, comparing a systematic review with a SoF Table to a systematic review without a SoF Table, measuring user satisfaction, correct comprehension and time spent finding key messages.

Method	Nr of participants	Participant description
Randomized controlled trial	Trial I: 72	Health professionals attending EBM workshop
	Trial II: 33	Cochrane Entities Meeting participants

4 SUPPORT Summary Development

Pragmatic goal: to develop a template for short summaries of existing systematic reviews, tailored for use by policy makers in low and middle-income countries.

Research goal: to explore the challenges involved in tailored summaries of systematic reviews for use by health policy makers in low and middle-income countries.

To achieve this we designed a prototype, collected and analyzing structured feedback from both users and stakeholders, held workshops and phone conferences to discuss interpretations of problems uncovered through feedback, generated ideas and designed a new prototype. We triangulated results with follow-up questions to original user participants.

Method	Nr of participants	Participant description
Brainstorming Workshop	3	Core team of researchers (w/backgrounds in epidemiology, anthropology and design)
User testing Audio recorded	18 + 3 pilot tests	Policy makers from low and middle-income countries (Argentina, Colombia, Uganda, South Africa, China) Pilot tests: health care development aid advisors (Norway)
Follow-up feedback	13	User test participants (Argentina, Colombia, South Africa, China and Norway)
Graphic design work	1	Carried out by myself

DESCRIPTIONS OF METHODS USED FOR DATA COLLECTION AND ANALYSIS

Below is a general description of each type of data collection and analysis in more depth than was appropriate in the article texts. For information about selection and recruitment of participants, see each individual article.

User testing and think-aloud protocol

User testing (or usability testing) is a method of formative evaluation “with the goal of leaning about the design to improve its next iteration” where products are tested with actual users, as opposed to by developers or experts.[146] It can be described as “an applied form of experimentation used by developers to test whether the product they develop is usable by the intended user population to achieve their tasks”. [130] User testing is carried

out in a controlled laboratory-like setting and is usually scenario based. Common outcome measures are performance (efficiency), errors (effectiveness), recall and user satisfaction. Although often designed to provide both quantifiable findings (e.g. time to complete task, number of errors) as well as qualitative data, user tests are time consuming and usually too underpowered to provide robust statistical results. User testing is commonly employed during the development of web sites and interactive systems to help answer the question: “Does this particular design function as intended?” as well as to indicate what areas need improvement.[130, 147]

The most common techniques for carrying out user testing are eye-tracking¹² and think-aloud technique. As we used the latter, I describe it here in more detail. Think-aloud technique is a form of observation that includes not just watching what the test person does, but encouraging them to articulate their thoughts - what they are looking at, thinking, doing, and feeling - while they are performing a task and capturing that speech through audio (and sometimes video) recording.[120, 147] When observing test participants using web sites, there is a rich wealth of other information that can be captured in combination with speech from thinking aloud: the actions or behavior of the user (such as use of mouse and keyboard), the reactions of the computer hardware and systems, and the “behaviour” of the website. However, when test participants are using paper prototypes (as in two of our studies) the data is not as rich. The paper artifact is static and not talking back, and the users’ external, visible behavior involves mostly only page turning and eye movement. This placed an extra challenge on us to try to capture the internal activity accurately through the think-aloud verbalization.

Qualitative test design

Our test method was closely modeled after usability testing, i.e. controlled and recorded test setting, individual test persons, test moderators and observers, based on a semi-structured interview guide. However, our tests differed somewhat from traditional usability testing in that we placed most emphasis on the qualitative data that was captured through the think-aloud technique for reasons described in more detail below.

First, although we made important notes regarding quantifiable things such as time and errors, we not as concerned with counting/measuring these events as

¹² Eye tracking is a technology that can record eye-movement, so that observers can watch and analyze it. But this is an expensive technology not widely available, and in these studies this was prohibitive.

we were in noting that they occurred and exploring the participants' subjective experience of them. The primary focus of the user tests (reflected in the research questions) was not just to see if the design of the artifacts "worked"; we were equally or more interested in the problem of why or why not, and the participant's own perceptions that would otherwise be hidden from us. For this reason, we emphasized the qualitative dimensions of user testing in our test design and employed methods of qualitative data collection and analysis.[148] (For an understanding of some of the quantitative measurement challenges regarding user experience, see Proceedings of the International Workshop on Meaningful Measures 2008.)[149]

Second, although we had a preconception of how we thought health professionals and policy makers might use these artifacts or why they might find them useful, we weren't completely certain that they shared our point of view. For that reason, an open, semi-structured, exploratory and qualitative test design was more appropriate rather than a rigid predefined set of highly structured tasks with measurable quantifiable outcomes. In this way the tests were designed so that the participant's idea of what was important and what entailed "use" could more readily come to the foreground.

Finally, with the exception of the Cochrane Library study, our artifacts were static paper documents. For this reason, "efficient use" or "number of errors" could not be captured by observation in the way that is possible when testing a web site where participants' mistakes are often clearly visible on the screen. For the Summary of Findings Table, we instead conducted randomized trials to measure efficacy of the paper artifacts, which is a study design much more suitable for measuring effect.

User experience framework

Our method differed in another way from traditional usability testing, in that we wanted to explore a broader (or possibly a more fine-grained) set of constructs than those that are commonly used to define usability. For this reason I chose to employ a user experience framework.



Figure 6: Morville's honeycomb of user experience

As explained in the background, there is no consensus on one preferred framework of user experience. I chose Morville's seven-faceted honeycomb¹³ diagram as a starting point, because it combined usability with usefulness and additional dimensions of user experience that were relevant to use of evidence (such as "credibility"):

Findability: can users locate what they are looking for?

Accessibility: are there physical barriers to actually gaining access, also for people with handicaps?

Usability: how easy and satisfying is this product to use?

Usefulness: does this product have practical value for this user?

Credibility: is it trustworthy?

Desirability: is it something the user wants? Has a positive emotional response to?

Value: does this product advance the mission of the organization behind it?

In these studies I chose not to address 'accessibility', as this is a specialized type of inquiry more readily captured using other methods than user

¹³ The term honeycomb indicates that the set of terms are loosely connected but without any hierarchical order

testing.[150] I also reframed ‘Value’ as meaning ‘value for the user’, not for the producer of the information.

These facets functioned as a thematic starting point for the data collection. I used the six facets as a basis for developing the interview guides for user testing, creating questions and tasks designed to address each individual facet. (We also used these six facets a basis for coding the transcriptions; see the analysis description below for a more detailed description and discussion of this.)

Interview guide

The interview guides were divided into three sections: 1) background information of the participant, 2) task scenarios (both pre-defined and open-ended) and 3) overall evaluation. Tests ended with a few questions about the testing and how it might be improved. An example of one of the interview guides is included in the appendix.

Task and content relevance is a critical factor in achieving reliable feedback, so we made efforts in each study to tailor the task content to the individual participant. In the Cochrane Library study, we gathered information about participant’s particular professional interests by telephone interview prior to testing, and framed tasks around those topics. In the SoF study, the table we tested was taken from a systematic review of relatively universal interest (the effect of compression stocking on long airplane flights). In the SUPPORT studies the participants chose one summary from a set of four that covered different topics that could be of general interest to health policy making in low and middle-income countries. Some examples of questions and tasks from the interview guides follow below.

Examples from the Cochrane Library study:

“Find the Cochrane Library from your home page”.

“Explore this web site and see if you find information here you feel is relevant for you. Use as much time as you normally would. Remember to tell us what you’re looking at, what you’re looking for, what you are thinking.”

“You want to keep updated on a particular topic – interventions for preventing drug abuse – so you want to look through all the *new* Cochrane Reviews relevant to this. How do you find out if there are any new reviews on this topic?” (task tailored to psychologist with specialization in substance abuse).

Examples from the Summary of Findings study:

“Now I’d like you to go through each part of the table, every element, and describe what your understanding of it is. (Show explanation sheet, that they can use it if they want.) Start up here at the top, read aloud and just tell me if things are clear to you or unclear, or if there is anything missing you might be looking for....”

“To the degree you can ”like” a table, did you like this table or not like it? If you could change it in any way (content, language, or formatting) what would you change?”

Examples from the SUPPORT summary study:

“You’ve just had a brief look at this report. Based on this, could you say anything about your impression of the credibility of this report? Do you think you would trust this information? Why, why not?”

“Do you think this report was generally easy or generally difficult to understand? Explain...”

“Would this report would be useful for you if you were going to make a decision about health care policy on this topic?”

Observation

Observation can take place at different levels of participation: complete participation, more marginal participation, observers who also participate and people who observe from the outside without participation.[130] In our user tests we had two kinds of observers: test facilitators who participated somewhat and note takers who participated marginally (passively) or not at all, depending on where they were sitting. The test facilitator had a complex set of responsibilities - guiding the participant through the tasks, posing questions, prompting in periods of silence, confusion, or reluctance to verbalize, watching the time – in addition to listening and observing. Since it is difficult to observe accurately while carrying out these other tasks, we always had at least one observer who’s only task was to take notes, either in a neighboring room watching the video recording or in the same room when sessions were not video taped. Test facilitators did not take any notes during the test, as we felt this helped them concentrate more on the task at hand, as well as helping the participant to relax more and be less self-conscious. Instead, test facilitators made notes about their observations immediately following the end of the test session.

Data collection and transcription

User tests lasted approximately one hour, and were carried out individually. The participant sat in a closed office together with the test facilitator who followed a semi-structured test guide. In the Library study we recorded all movement on the computer desktop through use of Morae usability test software[151] and video-filmed the participant, who was prompted to think out loud during the whole session. We projected the filming of the desktop and the participant as well as the sound track, to another room where two observers transcribed, discussed, and took notes. One observer took notes by placing annotated markers directly in the video recording, which facilitated tracing back to problem areas later on in the analysis. In the SoF and SUPPORT studies, the testing involved a prototype on paper (no computer), and the interview was captured only using an audio recorder. An observer sat in the same room with the participant and the test leader, taking notes. Following the interview, the observer or the test facilitator transcribed the test sessions, usually the same day, and erased the recording.

Ethical considerations

We filed the Cochrane Library study for approval with the Privacy Ombudsman for research (Personvernombudet) (ref: <http://www.nsd.uib.no/personvern/>) at the Norwegian Social Sciences Data Services (NSD), since the study involved registration and processing of video/audio data. A copy of the study consent is included in the Appendix.

The other studies were not registered as they did not meet the requirements for mandatory registration in Norway (i.e. we did not save, process or use any sensitive or personal information that could identify the participants in the data collection). Audio recordings of the interviews were not saved or processed digitally, or used for any other purpose than transcriptions. Transcribing took place for the most part the same day of the interviews, after which audio tapes were erased. We gathered names, emails or telephone numbers of participants only for the administrative purposes of sending information about the study and making interview appointments. This information was not connected in any way to the data after the interviews and was deleted from the project folder at the end of the data collection.

All participants in all studies were provided with both oral and written information according to the guidelines of the Research Ethics Committees.[152] This included, among other things, the purpose of the study, how we intended to use the data, and their right to withdraw their agreement to participate at any time, both during the interview or afterwards. Everyone signed written agreements of consent prior to testing. Some

received either modest gift certificates or removable memory sticks for their participation. All data was rendered anonymous, so that the name of the participant was not associated in any way with the interview files or transcripts. The taped interviews were deleted after use in the study, according to the terms of the agreements made with participants.

Coding

We used three sets of pre-determined categories to code the data: severity ratings, product-specific features and the honeycomb facets of user experience.

1. Severity-ratings: Indicates a problem and the level of difficulty a user experiences with it. We operated with three scores:[153]

- High ('show-stopper', critical errors or things that hindered task completion)
- Medium (much frustration or things that slowed user down)
- Low (minor or cosmetic problems).¹⁴
- Explicitly positive statements and suggestions from the participants were also noted.

Observers assigned severity-rating scores during testing or during re-reading of the transcripts in the analysis phase. This rating indicates the seriousness of problems that were *observed by* us through the user's performance and think-aloud verbalization (as opposed to relying on the users self-reported account of where and what caused problems). This assessment was also based on our in-depth knowledge of the product, and observations of things that went wrong or were misunderstood, even when they didn't appear to be experienced as such by the participant. Example of this kind of 'critical error' is when a participant completely misunderstood the meaning of the data in the Risk columns of the SoF Table, and did not pick up on this mistake himself, thinking he had understood them correctly. Each observer marked severity-ratings individually, usually during the user tests. Then we discussed collectively afterwards and marked them up in the master transcript.

2. Product-specific features: These are a set of categories corresponding to the specific features of the artifact we tested (e.g. the web site menu, query results page, home page) or functionality (e.g. navigation, search). An

¹⁴ Jakob Nielsen recommends these 3-point ratings for development work. When his team is carrying out large-scale research projects, they use 100-point rating scales and mathematically calculate severity based on frequency, impact and persistence ratings. Since our studies were small scale and based on pragmatic development projects, we used the simpler rating model.

example of a list of these kinds of categories for the SUPPORT summaries is: front page including title, logos key messages; background section; characteristics of the review table; summary of findings text; Summary of Findings Table; applicability section; back page). The reason for grouping features also according to functionality is that sub-sets of tasks (e.g. executing a search query and understanding the search results) and the problems arising within these sub-sets (e.g. trouble finding a relevant hit in a search results list) are often interrelated, even though they may occur in different places on the site and in different phases of the user test.[154]

3. Facets of user experience in honeycomb model (minus accessibility): We also coded the data for the themes from the honeycomb framework: findability, usability, usefulness, credibility, desirability, value, as well as new themes that emerged from the data. The reasons for choosing this particular set of themes from the honeycomb framework have already been discussed.

Since we used these pre-established categories as a starting point for the analysis, our approach somewhat resembled a framework analysis,[155] where the objectives are set out in advance. Since we built the interview guide partly around the six themes from the honeycomb framework, it is not surprising that we found them in the data. However, we did not always conform to the structure of the guide. During interviews, we were sensitive to new strands of feedback not covered by the guide and deviated from it when necessary. As we had many participants in each study, we also reached a point of data saturation on some topics before all tests had been carried out. This allowed us more freedom with the later participants, to go in more depth or follow up on uncommon issues.¹⁵ We were also sensitive to new themes that emerged, across interviews when analyzing the transcripts, and accordingly made changes to interview guide manuscripts later on in the work to incorporate some of these themes. One example of a new theme is “understanding”, which we had not anticipated to be a problem when designing the interview guide for the Summary of Findings Table study. In this way our approach shared characteristics with thematic analysis, where

¹⁵ For instance, in one of the Cochrane Library tests, the participant (a doctor who was familiar with Cochrane Reviews) came across a single trial in another database on the Library web site that he mistook it for a Cochrane Review. We decided to let him continue reading the document at his own pace, to see how long it would take him to discover his mistake, instead of moving him on to the next question. (It took him 17 minutes.)

themes may both exist from the start, but can also be reflexively developed from the ground up.[155]

Analysis of user test data

We carried out analyses twice for each study, resulting in two related but separate types of findings: an analysis that could support the pragmatic design goals of improving these *particular artifacts*, and one that could support the research goals that involved writing articles relevant for these *classes of artifacts*. So although the underlying data was the same, we used it in two different ways.

In the first analysis, we needed to understand where 1) there were serious user/artifact problems, 2) what the cause of that problem might be (e.g. reasons leading back to the user or reasons leading back to the artifact), and 3) how we might fix the artifact to reduce or eliminate the problem.

In the second analysis we needed to communicate these same topics in a way that could apply to *classes of artifacts* like ours, not just our specific artifacts. This little detail makes a big difference. For instance, when writing up the article about the Cochrane Library web site, tracing problems back to specific faulty design details for this particular site wouldn't necessarily be relevant to other sites, because they would have a different design. However, problems that pointed back to a set of concepts could be relevant to other sites, particularly other online libraries or sites with evidence-based content.

Therefore we needed to re-sort our findings so that they pointed to *more general concepts* rather than just our site-specific details. In that way we could generate transferable knowledge about our "unique particulars" (e.g. Cochrane Library) to other unique particulars in that class of artifacts (e.g. online libraries with evidence-based content). As a starting point for the second analysis, we used built on the first analysis, but added the theme codes from the honeycomb framework plus the new themes that had emerged. Below I describe how we carried out both of these kinds of analyses, for improvement of our particular artifacts and for writing up the articles. See also flow chart below (Figure 6).

Description of the two analyses processes

We prepared for analysis by rearranging the individual transcripts into one large document, grouping all of the participants' responses together under each interview task or question. At least one other researcher in addition to myself carried out the following analysis work.

For the first analysis, we marked and sorted user issues/problems according to three types of categories:

1. Severity-ratings
2. Product-specific features

Through discussion, we made interpretations of what might be causing these problems based on:

- *Our knowledge of the participant* (based on their answers to intro questions and their display of competence and background knowledge during the test)
- *Our specific knowledge of the interface and the system underlying this site* (both technical and editorial)
- *Our general knowledge of usability and user experience issues* (such as what we know about how people search and typical problems in designing search interfaces).
- *The context of the problem in the individual user test.*

Based on our hypotheses of what might be causing the problem, possible changes to the product were discussed. We either agreed on a change, or agreed to try out multiple solutions in the design phase. (See Brainstorming Workshops and Designing New Prototypes below.)

The second analysis was carried out for the purposes of writing up the findings in article form. This involved going back and adding thematic codes:

1. The honeycomb facets of user experience: findability, usability, usefulness, credibility, desirability, and value
2. Additional themes that emerged from the data *with moderate or high severity-ratings*

This move enabled me to describe users' problems as examples related to a *set of general themes* when I was writing the articles, rather than as local problems embedded in our particular artifacts, making the article texts potentially more relevant for others.

When preparing to write the articles, I looked at the material in light of all three categories of codes together, interpreting the themes in light of how and where critical user problems were associated with this theme. In this way, the analysis involved more than merely "classifying, categorizing coding or collating of the data".[156] By drawing connections through all three sets of categories, I gained an enriched understanding of the *relative importance of each theme* (based on the severity-ratings) and of *where and how critical issues related to these themes were to manifest* in that particular class of

artifacts. I could also supplement these insights with our experiences of *what design or content related moves had been helpful* in resolving the most critical issues.

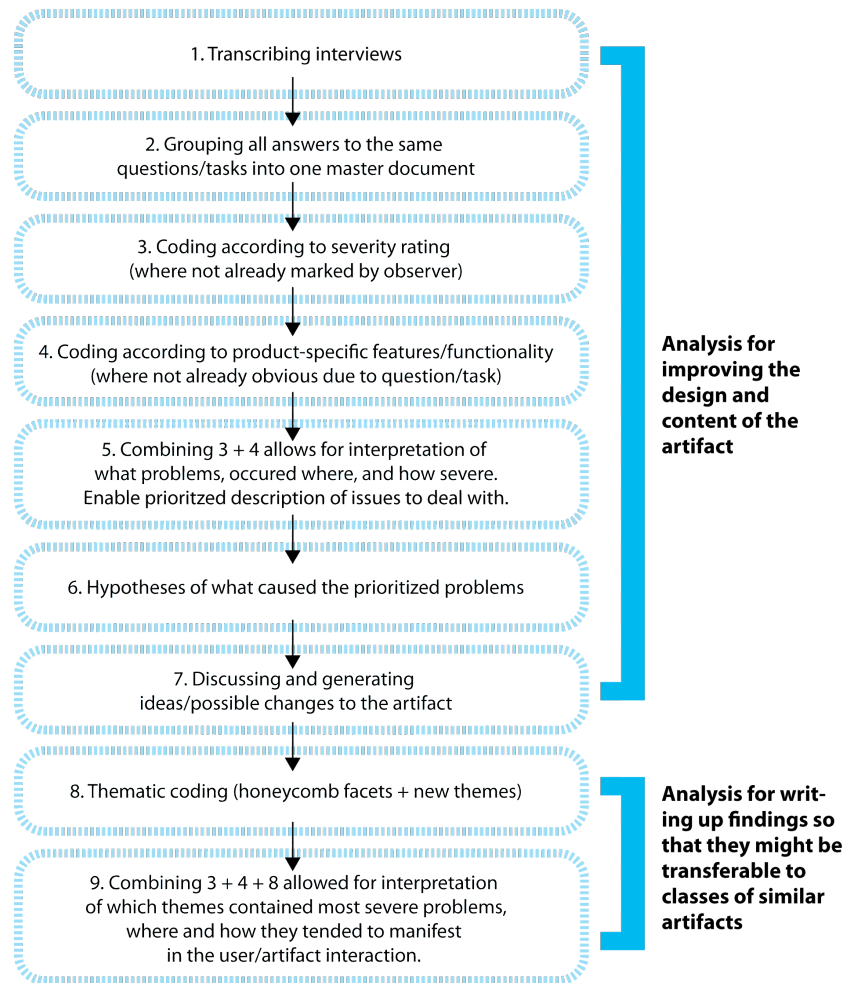


Figure 7. Analysis flow chart for data from user testing

Stakeholder Feedback

Collecting stakeholder feedback

The purpose of stakeholder feedback (utilized in the SoF and SUPPORT projects) was to provide input from a stakeholder perspective. This feedback was necessary in order to make sure we were not misrepresenting the data in our efforts to condense it into tables and summaries. It was also necessary

from a feasibility point of view - could our proposed changes actually be implemented within the existing parameters (such as publishing systems) and if not, were changes to these parameters possible? In the SoF Table study, the stakeholders were organized as an Advisory Group of over 50 people with a range of different roles in the Cochrane Collaboration, including statisticians and other methodologists, review authors, coordinating editors, consumer representatives, publishers, and members of the Steering Group. In the SUPPORT project, the stakeholders were both the authors of the summaries and the interviewing researchers, so feedback served the double purpose of checking our interpretations of the interviews as well as the feasibility of producing the suggested changes. Stakeholder feedback was elicited by email (SoF and SUPPORT) and telephone conference (SUPPORT).

Analyzing stakeholder feedback

The working group carried out analysis of stakeholder feedback. We collected all feedback into one document and looked for issues with:

- high level of agreement,
- high level of disagreement,
- issues not previously considered
- issues of high severity (such as incorrect presentation of data or formatting that was not technically feasible.)

The issues that were of high severity where there was a high level of agreement (that the issue was a problem) received most attention, and were brought these into the brainstorming workshops (see below). In these workshops we also attended to issues not previously considered, as well as feedback that was not controversial and easy to resolve. Issues that triggered much disagreement were fed back to the stakeholder group for in the next round feedback further comments and discussion.

Brainstorming Workshop

After identifying problem areas from user testing and stakeholder feedback, we needed to both determine the underlying cause of the problem and possible ways of generating a solution. “Brainstorming workshop” is the term I use to describe the part of the design process that occurred within the working group in the SoF and SUPPORT studies, and that happened after findings were both scored according to severity and sorted according to product feature. In addition to myself, the two other researchers with much knowledge about Cochrane Review methodology and production participated in the working groups: Claire Glenton (background in social anthropologist), Andy Oxman (background in epidemiology). Hilde Kari Nylund (background in journalism) joined us in the SoF working group. The idea generation work

was carried out for the most part in meetings. We would review a problem already identified in the analysis phase, agree on an interpretation of what was causing it, and throw out many different ideas to resolve it. For the most part, those ideas dealt with altering the representation: e.g. changes in language, numerical formatting and visual form (such as layout, color, typography).

Designing new prototypes

In the SoF Tables and SUPPORT Summaries projects, the goal of the design phase of the work was to develop new versions of the artifact, trying to find more appropriate visual representation to the given content based on feedback. I tried ideas out visually, usually after brainstorming workshops but also sometimes during these sessions. I also carried out individual designing sessions on my own after workshops. This entailed trying out many different solutions, without the added input of the rest of the group, and presenting them for discussion at the next meeting. The core of this work was based on the interpretations we arrived at in the data analysis about what was causing the user problems, but often I tried out other ideas as well that emerged through working with the material and problems at hand.

Although designing for the most part entailed working with *given* content, in several instances we concluded that content needed to be changed in order to produce a good result (i.e. changing language used, changing the way the statistics were formatted, or adding supplementary content).

Designing was a critical part of the protocol in developing SoF Tables and SUPPORT summaries. However to describe the designing as an isolated research method is more problematic. To begin with, design work is highly subjective. Other designers with similar expertise and working from the same starting point might come up with completely different results. There is no effort to reduce subjectivity in design; rather the opposite is often the case. Like qualitative research, it is addressing the specifics of a context, but does not need to be systematic or transparent to be of high quality. Also, although the *results* of designing here are potentially transferable to other artifacts, the designing activity itself is not describable in a way that makes it easy to transfer to other work. Documenting design activity in detail while carrying it out can get in the way of doing it.

The nature of design expertise is elaborated on in the final section of this thesis, and I do argue at the end that a legitimate output of design can be knowledge, not just “things”. But I do not feel there is value in trying to define it as a sub-category of research methodology. I have included the

designing as a part of the methods description here so that the reader can understand the different roles the design inquiry and the research inquiry played in this work, and how they interacted with each other.

Randomized Control Trial

Two randomized controlled trials (RCTs) were executed in the SoF project in order to measure the effect of an SoF Table in the context of reading a Cochrane Review. We carried out the trials at different stages of the table development, comparing a Cochrane Review that included the SoF Table (two different variations of it) with a Cochrane Review not including the table. The objective of the first RCT was to assess users' *satisfaction* with the table; the second RCT aimed to assess the effect of the table on users' *understanding* of the reviews and *time spent* to find answers. See table below for a PICO description of the trials. Description of recruitment, participants, comparisons, randomization, data analysis and outcome measurements are covered in detail in the article "*Summary of Findings Tables improved understanding and rapid retrieval of key information in Cochrane Reviews*".

	RCT 1	RCT 2
Population	Health professionals	Attendees at a Cochrane Entities meeting
Intervention	Cochrane Review with SoF Table	Cochrane Review with SoF Table
Control	Cochrane Review without SoF Table	Cochrane Review without SoF Table
Outcome(s)	User satisfaction	Correct understanding, time spent to find key messages, user satisfaction

Table 2. PICO (population, intervention, control, outcome) for RTC 1 and 2.

Follow-up feedback - SUPPORT study

In order to test our analysis of the findings from SUPPORT Summary user testing and changes we made based on these, we sent new and old versions of the summaries to the user test participants, asking for their feedback over email or telephone about which version they preferred. We also asked them for comments on our analysis of the main problems and solutions implemented in the new version.

SUMMING UP METHODS

In the section above, I have given a short overview of the different methods that were used in each study, and provided a more in-depth description of some of them than space allowed for in the articles.

Results

THREE SETS OF RESULTS

The results of this work take on different forms, as illustrated below.

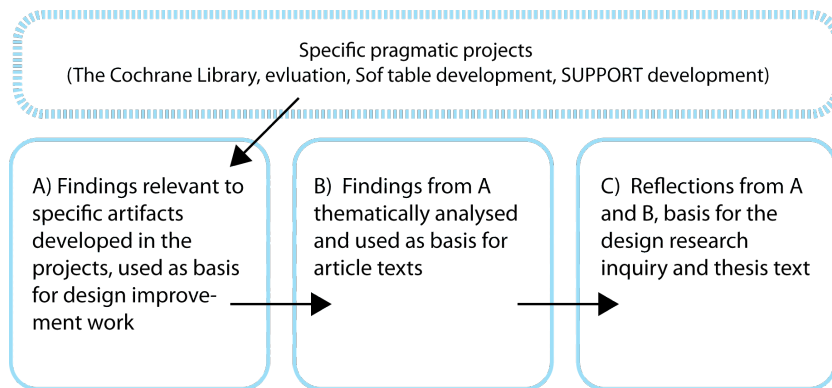


Figure 8: Three types of results

Results A) These are very product-specific results, typical of the type of output coming from a design project.

- The reports we created for the Cochrane Collaboration steering group summarizing user testing and recommending site-specific changes
- The final Summary of Findings Table template
- The final SUPPORT Summary template

The knowledge embedded in these artifacts is mostly only of local value: the templates for SoF Tables and SUPPORT Summaries that inform production of a large number of individual tables and summaries. The Cochrane Library Reports - sent to the Library editors and publishers - have played a part as reference documents for site improvement efforts. Although it could be argued that these artifacts are a part of the research results, I have chosen not

to categorize them as such. This is because, although the main aim of the *projects* was to produce artifacts, the aim of this *thesis* is to produce accessible transferable knowledge. From this perspective, the artifacts and their development were a means to producing design-relevant knowledge, but do not provide much transferable knowledge other than as single case examples. Chow, however, does make a good argument for the value of case studies to design work.[157] The artifacts can also function as helpful illustrations to this text. Therefore I have included examples of a SoF Table and a SUPPORT summary in the Appendix.

Results B) The second type of results builds on the first set and the process of arriving at them. These results are articulated primarily in relation to sets of general themes rather than as findings tied up with the specific details of our artifacts. These formulations are aimed at a wider audience and make up the basis for the articles.[158-161] A summary of these findings is written up below.

Results C) These are a broader set of finding, written up in the final three thesis sections: *Discussion*, *End of Journey* and *Summary and Conclusions*, emerged when I took more of a meta-perspective of the whole body of work.

THE COCHRANE LIBRARY STUDY

We performed a total of 32 user tests in Norway and the UK with health professionals from nursing/midwifery, medicine, dentistry, physiotherapy, social sciences, psychology, and occupational therapy. Twenty-one of the 32 participants were non-native English speakers accustomed to reading in English. Twenty-six participants said they had previously visited the library site, and 25 of 32 could give a basic description of a systematic review. The highest severity ratings were concentrated in three of the facets of the user-experience model: *findability*, *usability* and *credibility*.

User test results

Findability

Participants had much difficulty locating both the site and its contents. Searches were performed to solve most tasks, but were largely unsuccessful due to multiple problems: misspelling and poor recall of terms, misunderstanding search results, misunderstanding 'no hits' and few results, misunderstanding of keyword search results. A few of these problems were due to faults in the system. However most were due to a lack of compatibility between the knowledge of the user (e.g. non-English speaking people making spelling mistakes) and the functionality of the system (e.g. no spell-check

built in). Participants were often at a loss to understanding where the problem originated (system problem? content problem? own problem?) Many ended up blaming their own perceived lack of competence, expressing feelings of ineptitude and alienation. Participants also had problems that they were unaware of, such as making incorrect distinctions between different information types (e.g. confusing a single clinical trial with Cochrane Review). This kind of error occurred even with participants who demonstrated a high level of background knowledge of systematic reviews and the Cochrane Collaboration.

Usability

Participants reacted to the use of jargon throughout the site, which gave the impression that the site was for academic use only and effectively discouraged many from using several of the site's functions. Few were able to describe how the content was structured or point to navigational features.

Credibility

Although most initially expressed a high regard for the site's credibility, some later displayed a mistrust of the independence of the information, largely due to the presence of the unfamiliar logo of the publisher. Others were overconfident, thinking everything on *The Cochrane Library* site shared the same level of quality approval.

Usefulness, Desirability and Value

The Library was perceived by some participants as primarily an academic site and not useful for health professionals. Several indicated that the Plain Language Summaries were useful for understanding the research results. The site was generally seen as being messy and difficult to use compared to other sites. There was much praise for quality of content, but high frustration concerning functionality and ease of use. Many indicated that the site was too frustrating and time-consuming to be of value in their time-restricted professional contexts.

SUMMARY OF FINDINGS TABLE DEVELOPMENT

Brainstorming workshop results

As we collected input from both the advisory group and the user tests, the main focus in these workshops became more apparent: to address the tension between achieving table precision and table simplicity. Tables that included enough information to meet the precision goals of the advisory group tended to be too complicated for user test participants to understand or want to read. There was therefore a continuous re-evaluation about what information was

most critical to include and much effort in these workshops was spent trying to find solutions that accommodated both perspectives.

Advisory group feedback results

We received 58 responses from a total of 52 individuals or groups. Comments fell mostly into two categories: precision of the data representation (e.g. missing, inaccurate or potentially misleading elements) and feasibility of producing and publishing the tables within the current Cochrane system (e.g. limitations of formatting tables in the publishing system).

User test results

A total of 21 people from Norway and UK with a variety of health care related backgrounds took part in the two sets of user tests. The findings that were rated most severe and that led to most changes in the table were concentrated in two of the seven facets of the user-experience model: *usability* and *usefulness*.

Usability

A major finding from the user testing was that data was easily misunderstood, particularly in the first set of user tests. Participants misunderstood or were uncertain about a range of elements: dichotomous outcomes, continuous outcomes, number of studies, meaning of “no data available” or empty cells, terms used in column headings and abbreviations. For instance, five of 13 test participants dramatically misunderstood “9 fewer per 1000” in the column for “Absolute difference”, stating that it meant “9” or “9 or fewer”. This mistake was made by some even when they correctly read the effect statement out loud. Abbreviations such as “RR” (relative risk) and “CI” (confidence interval) also caused some confusion regarding both what the abbreviation stood for and the understanding of the concept it referred to. Participants did not have critical problems related to understanding the GRADE quality of evidence ratings, despite most not having prior knowledge of GRADE.

Usefulness

Participants offered suggestions for changes that would make the tables even more useful in a clinical setting. These included specifying the population/setting/ intervention/control descriptions at the top of the table, describing the intervention in more detail, adding inclusion criteria for high and low risk populations, and including a clear recommendation.

Credibility, findability, desirability and value.

Eighteen of 21 test persons indicated that their perception of the credibility of the table was directly related to the GRADE ratings. Most indicated that they would expect or want a SoF Table to be near the front of the review. Fourteen of 21 explicitly said they liked the table. Although many felt the table contained a large amount of information, this was not necessarily negative. Several participants perceived tables as containing valuable content (worth the effort). Some expected a learning curve for this kind of information and felt tables would be easier to read upon repeated exposure.

Many changes were made to the table based on user test findings prior to the evaluation trials. These are described in more detail in the article.

SUMMARY OF FINDINGS TABLE EVALUATION

RCT I - assessing user satisfaction with the table

Of approximately 90 people present, 72 completed the questionnaire (25 received Cochrane Reviews *without* the SoF Table, 47 received *with* the table divided in two groups with different versions). Participants reading reviews that included the Summary of Findings Table were more likely to respond that:

- It was easy to find results for important outcomes
- It was easy to find quality of the evidence for important outcomes
- The overall accessibility was high (6-7 on a scale of 0-7)

Structured discussion

A structured discussion following the RCT revealed that despite satisfaction with experienced accessibility of the information in the table, many people were misunderstanding the data in one of the columns presenting absolute differences ('9 fewer per 100'). This supported our findings from the user tests carried out the day before and resulted in changes about how absolute effect was presented in the table before the next trial.

RCT II - assessing the effect of the table on users' understanding of reviews and time spent finding answers

Thirty-three completed questionnaires were returned (18 without and 15 with an SoF Table). Multiple choice questions concerning the main findings in the review were answered correctly much more often in the group with the SoF Table:

- 93% versus 44% (p=0.003) identified the correct value for control-group risk

- 87% versus 11%, ($p < 0.001$) identified the correct value for intervention-group risk

The group with the table also spent much less time finding answers (1.5 minutes as compared to 4 minutes in the control group). A larger proportion (73% versus 24%, $p = 0.005$, Pearson's Chi-square) agreed or strongly agreed that it was easy to find information about the quality of evidence for the main outcomes.

SUPPORT SUMMARY DEVELOPMENT

We carried out 18 user tests with health policy makers in Argentina, Colombia, Uganda, South Africa and China, and 3 tests with advisors in aid organizations in Norway.

User test results

Usefulness

Although sixteen of 21 reported that the summary would be useful for them if they were making a decision on the topic, many still expected or wanted content outside the realm of the systematic review: answers to broader kinds of questions, recommendations, measurements of other outcomes, or more detailed information about local applicability and cost. This seemed in part to stem from a poor understanding of what a systematic review was. The graded-entry format (one page of key messages followed by short summary and access to full review) was perceived as useful.

Usability

Some participants still felt that the summary (5-7 pages) was too long and complex. They wanted an even shorter, clearer presentation, some saying they wouldn't have time to read more than one page. Despite the summary authors attempts to write in plain language, some felt the tables and text contained language and statistical/scientific concepts that were still too difficult.

Credibility

Participants seemed willing to trust the credibility of the summary due to recognizing known sources and knowledge that systematic reviews should be of high quality. However, not everybody understood the summary stemmed from a systematic review. Some also expressed doubts about the information when seeing "low quality" GRADE scores, when there was no evidence for important outcomes, or when the studies were too old.

Value and desirability

Seventeen participants felt the summaries were valuable to policy makers like themselves. Fourteen liked the summaries, and several emphasized both the front page with key messages and the 'Relevance for LMIC' section.

Changes made to Summary before follow-up interviews

A number of the findings pointed to obvious solutions, which we adopted, for instance simplifying language and tables, eliminating abbreviations, being more consistent in use of terms and corresponding data in text/tables, etc. But some larger issues were more challenging to address: poor conceptual understanding of what a systematic review is, expectations of information outside the realm of the summary, and the desire for shorter/clearer summaries. We dealt with these problems in different ways:

- Introduced boxes of 'information about the information' in a right column throughout the summary (e.g. "Who is this summary for?" on the front page).
- Broadened the scope of the review by changing the 'References' section to 'Additional information'. Here we added information that helped to understand the problem, provided details about the intervention, or helped place the results of the review in a broader context.
- Made graphic adjustments to the text instead of shortening it further, so that it would be easier to scan and pick out the important parts more quickly:
 - Broke longer texts up into bullet lists
 - Added colored arrows to mark findings
 - Reorganized text in the "Relevance for LMIC" section in a tabular format.

Stakeholder feedback

The advisory group agreed both on our interpretations of the user-test findings and the subsequent changes we made to the summary format.

Follow-up interviews with test participants

Twelve out of 19 participants responded to the follow-up questions. All clearly preferred the new format, explaining that they found it easier to read and more understandable. Reasons for this were mainly the new front page and the addition of the boxes on the front and subsequent pages. There was general agreement that our analysis of the problems was precise and that the new summary resolved the main issues.

SUMMING UP RESULTS

In each of these studies we uncovered not only local findings that fed into the design of these particular artifacts, but also results that can potentially transfer to other similar artifacts. Some of the main findings are:

- Resolving the tension between stakeholders' concern for precision in data presentation and users' needs for simplicity is a major challenge in designing evidence summaries.
- Inclusion of SoF Table in a Cochrane Review improved understanding and rapid retrieval of key findings compared with review with no SoF Table.
- Health professionals and policy makers may not know what a systematic review is or may confuse it with another type of report (e.g. single trial). This can lead to unrealistic expectations or critical misunderstanding of content.
- Unfamiliar language (both foreign language and jargon, abbreviations and unfamiliar terms) may lead to frustration and alienation. They may cause critical barriers not only while reading but also while searching. Non-native English speakers are at an extra disadvantage.
- Outcome effects may be misunderstood when presented as absolute differences in a table; the use of absolute risks can solve this problem.
- Outcome effects presented as continuous outcomes are often difficult to understand when the scales that are used are unfamiliar.
- GRADE scores appear to be easy for people to understand even when they are not familiar with GRADE to begin with.
- Credibility can easily be swayed in a negative direction by the presence of an unfamiliar logo. But users may also uncritically trust the quality of all the content because of a familiar logo such as the Cochrane logo.
- Summaries should not only be designed to be easy to read, but quick to scan.
- Graded entry format is well suited for presenting evidence to policy makers.
- Making review summaries useful for policy makers includes broadening the scope of information, such as adding author's comments about applicability or a wider range of references, as well as information about the information.
- Making summaries useful for clinicians includes better descriptions of interventions and risk groups.

Detailed reports of these findings and discussions of them are written up in the individual articles. In the next section I will take this a step further and discuss the findings that emerged from the projects seen as a whole: important characteristics user experiences with summarized evidence, user-stakeholder "gaps" and the suitability of the honeycomb framework.

Discussion

OVERVIEW BASED ON RESEARCH CATEGORIES

Before moving on to the discussion, it might be useful to map out the remaining content. As the output of this body of work is complex, I have sought ways to better sort it into more understandable categories, both for myself and for the reader. Buchanan has suggested that design research might be seen as falling into three categories based on the type of problem addressed: clinical, applied and basic.[162] Clinical research is directed towards an individual case; applied research towards problems across a class of problems or situations; basic research is directed towards understanding the principles which govern and explain phenomenon. Using these concepts, I have constructed an overview of the different output from this body of work. It should be noted, however, that each type of output builds on knowledge from the previous levels and that the x's indicated here are only for illustrative purposes.

Buchanan categories of design research		'Clinical' individual cases	'Applied': knowledge about classes of problems	'Basic': more general reflection regarding principles
1	Evaluation of Cochrane Library user test reports (2005 and 2006)	X		
2	Design of Summary of Findings template	X		
3	Design of SUPPORT Summary template	X		
4	Characteristics of user experiences, characteristics of stakeholder feedback, characteristics of user-producer gaps, effect of tables		X	

5	Consequences of nr. 4 for design	X
6	Method consequences of nr. 4: Revision of honeycomb framework for use in developing and evaluating 'summarized evidence' related artifacts	X
7	Reflections about cross-platform publishing, templates and general-specific translation	X
8	Reflection about the nature of design practice	(X)

Table 3. Overview based on Buchanan's three categories of design research

Content 1-5 has already been presented: 1, 2 and 3 have resulted in specific artifacts; 4 and 5 have resulted in the four articles.

The topics 4, 5, 6 and 7 have also resulted in broader cross-article discussions presented below, following the strengths and weaknesses section. Topic nr. 8 is the final section, "End of the Journey".

At the very end of the thesis, there is a summary of all of this material, organized as answers to the research questions.

The categories in the table above help illuminate the differences and connections between design practice and design research. In design practice, one is firmly and comfortably situated in the 'clinical' category, concerned primarily with producing knowledge that leads to good solutions of single problems or artifacts. An experienced designer acquires a rich supply of knowledge at the next level ('applied' knowledge relevant for broader classes of problems). Expert knowledge shares many common attributes, for instance it is built on a large factual base and a well-developed set of conceptual frameworks. But personal expert design knowledge is largely tacit. Unless the designer teaches or engages in writing, it is not articulated and does not transfer very far.

It took me time to understand this. Trained as a designer, I was most dedicated to uncovering knowledge that was relevant for practical design work. I thought that rigorous research carried out in this 'clinical' category was sufficient to achieve this goal – I would carry out research to develop artifacts, and write about that. In the studies in this thesis, all the projects were planned and carried out were at this first clinical 'practical' level. However, when I began to write the articles, I was unexpectedly forced up to the 'applied' level, as it gradually dawned on me that findings too deeply embedded in the particulars of single artifacts are not of much use to others.

This movement revealed to me the value of the thematic framework I was already using – enabling me to analyze and thereby lift findings from the ‘clinical’ to the ‘applied’ level. And, although the article texts were of interest to a community of medical publishers and researchers, my advisors reminded me that they alone did not have much relevance to the design research discourse. In this way I was constantly challenged to expand the scope of my writing and the level of investigation I operated on. Although my writing about the ‘design journey’ (at the end of the thesis) does not establish any new basic principles, it touches on larger design-relevant topics that breach the boundaries of ‘clinical’ or ‘applied’ knowledge about singular artifacts or classes of problems. In this way, these categories also make visible my own progression as a design researcher.

STRENGTHS AND WEAKNESSES

Strengths

Although one of the studies exclusively employed quantitative method, the majority of the work in this thesis rests heavily on approaches from a qualitative research tradition. Marshall and Rossman have proposed four categories for evaluating the strength of qualitative research¹⁶: credibility, transferability, dependability and confirmability.[163] Credibility involves providing an in-depth description of the complexities and nuances of the data that is convincing to the readers. Dependability involves attempting to account for changing conditions in the setting. Confirmability asks whether the logical inferences and interpretations of the researcher make sense to someone else. Transferability has to do with applicability and usefulness of the results to other settings. Although Marshall and Rossman argue that the burden of transferability rests with the researcher who would make the transfer, triangulation can enhance the study’s transferability and usefulness for other settings.

Triangulation is the main strength of this body of work, including all four types of triangulation identified by Patton (citing Denzin): methodological triangulation (we carried out both qualitative and quantitative inquiries); data triangulation (we systematically gathered feedback from a wide range of both users and stakeholders); investigator triangulation (more than one researcher collected and analyzed the data); and theoretical triangulation (I drew on

¹⁶ In quantitative reseach, the value or trustworthiness is determined by examining internal and external validity, reliability and objectivity. Marshall and Rossman’s categories provide alternative corresponding concepts for qualitative research.

different theoretical lenses in both framing the questions and discussing the findings.) [164] These forms of triangulation provided corrective and complementary perspectives to three separate but related artifacts, as well as a rich multifaceted description of the work as a whole.

Although studying three different artifacts might be considered a weakness as it limits more in-depth studies of single artifacts, it enabled me to compare findings and see patterns across studies. This arguably makes the results more robust and transferable to artifacts outside of these particular projects, and arguably enhanced the *dependability* of the findings across changing settings.

The level of user testing carried out in these studies far exceeds that for most similar artifacts, at least within the area of evidence-informed health care. Additionally the use of multiple methods of data collection adds depth. The exploration of user experience also likely led to exposure of a more nuanced set of problems than if the work had only focused on measuring usability. Together these factors combine to enhance the *credibility* of the work.

Confirmability is supported by the involvement of many researchers who participated in the data collection and co-authored the articles. Co-authorship means they explicitly agree with the analysis interpretation of the data that I built up in these texts. All of the articles have either been published or are in the process of being published, confirming that informed readers (journal editors and peer reviewers) outside the boundaries of the projects also accept the logic of argumentation and level of transparency in those texts.

A final strength is the carrying out of these studies in actual design projects. The framing of the work from a design-practice perspective increases the usefulness and *transferability* of this work for future design projects, thereby helping move this research into (design) practice. Additionally, by observing real life interactions between producers and users of evidence, we were able to capture important findings that would not occur in artificially constructed development contexts, such as the tension between users and real producers of evidence.

Limitations

The path from evidence-production to evidence-use is complex, involving a large range of factors that may be social, technical, cultural, economical, educational, cognitive, or just plain practical in nature. These sets of studies did not address most of these issues – we did not explicitly examine the motivations of clinicians or policy makers to use or not use research evidence, nor did we measure large populations of users' access, attitudes to,

or awareness of existing evidence. We did not look at the availability of evidence on topics considered relevant by different user groups, or the economic feasibility of implementing treatments supported by evidence. We did not study users in real life decision-making contexts, or examine how communities of practice might influence use of evidence. Our very narrow aim was to examine a few points of contact between summarized evidence from Cochrane Reviews and potential individual users (clinicians and low and middle income policy makers), looking for ways to facilitate better user experiences - and through this, help facilitate evidence informed health care - by improving the design of the information artifacts.

Elaborating on this list, the main weakness of the work is likely the laboratory setting. User test participants were stand-ins for real users, and their testing took place far removed from real-life tasks and use contexts. We also narrowed the problem space down to the interaction between a single individual user and the information, losing any social dimension of use that might affect the user experience; this may have had significant consequences. Additionally, the time frame in user interviews was probably much greater than what users would likely invest in actual working contexts. (On the other hand, the depth of feedback we received through user test interviews would have been difficult to achieve in actual workplaces, where our presence would be disruptive and time available much more limited. Also, the work *was* embedded in real-life projects of improving evidence presentation - with non-fictional limitations regarding content, resources and technology - and included the involvement of the actual stakeholders in those projects, not surrogates.)

In both SoF Table studies, some participants have been included who did not accurately represent our target group.

The individual facets of the user experience honeycomb could have been tested out separately and more precisely, using individually tailored methods for each facet. For instance, ‘usability’ could have been measured more by explicitly measuring exact time to complete tasks in the user test sessions. Accessibility issues related to the Cochrane Library web site or online versions of the tables and summaries could have been evaluated using separate established methods (as pointed out earlier in the Methods section).

The RCT’s were small, and not directly comparable with each other. It is also not clear that the effects observed in these trials would be the same for another similar table that differed from ours, as it is likely that the effect of a table is at least partially contingent on its design. Also, the design of the SoF

Table we tested has not been successfully reproduced in the Cochrane Library for technical reasons. It is not clear either how much a difference that makes for the user experience, and future studies should investigate this issue.

Additionally, for all studies except the Library study, the developers carried out the tests, and participants were aware of this, potentially affecting their responses.

Possible conflict between design goals and research goals?

The pragmatic design goals in these studies strongly influenced the choice of method. This meant that our team worked in a different way than if we were only carrying out academic research unrelated to design activity. For instance, we executed several iterations and short spurts of feedback at multiple points of development instead of recruiting larger sets of participants in more focused, comprehensive studies. One of the peer reviewers of the SoF evaluation article resonated objections one might have to this process seen from a research point of view:

I see no reason why the authors could not have done this study in far larger numbers - why not use 10 workshops ie 10 x 90 people? Not only this, the SoF Table was modified between the first and second RCT - this suggests that the development and user testing of the SoF Table in the first manuscript was perhaps inadequate.

This comment highlights a potential weakness of the rapid-iteration approach characteristic of a designerly way of working (as well as possibly poor understanding on the part of the reviewer as to the nature of our work). Although our design-based method of alternating cycles between development and evaluation worked well for the purposes of designing the specific product we were working on (and produced results about designing these kinds of products that are likely transferable to other publishing projects), it might - as the reviewer indicates - have affected the validity of the specific solutions emerging from the work, despite our efforts to carry out our studies rigorously. It has been argued that measures of user experience should be *meaningful, useful* and *valid*. [165] Achieving a balance between these is a challenge for any research involving or reflecting on practice.

Below I present the cross-article discussions.

IMPORTANT CHARACTERISTICS OF USER EXPERIENCE AND CONSEQUENCES FOR DESIGN

What characterized the user experiences of summarized evidence in these studies? What areas of improvements does our work suggest? In the following text there are three underlying main messages concerning users' experiences with summarized evidence and consequences for design:

1. Clinicians' and policy makers' experiences of summarized evidence can be seriously compromised on a number of levels even by small details of the information artefact or the web site where it is published.
2. Paying careful attention to these details in the design and editing processes has the potential to significantly improve these users' experiences of the evidence, including avoiding misunderstanding that could lead to misuse.
3. Clinicians' and policy makers' experiences of summarized evidence may differ (possibly in a fundamental way) from researchers' own experiences of evidence. Conflicts may therefore arise in attending to both the needs of these groups of users and the needs of the researchers at the same time and in the same information artefact.

First, a discussion of points 1 and 2 are combined in the text below, sorted according to some of the most critical aspects of the user experience model that emerged from our work: searching for and finding evidence, understanding evidence, appraising evidence and the experienced usefulness of evidence. Next is a discussion of point 3, including a reiteration of the main approaches for dealing with these conflicts.

Searching and finding evidence

Research indicates that some of the main obstacles to answering clinical questions with evidence have to do with search.[42] Searching online is a complicated process, involving many sub-steps, such as formulating a search, understanding and evaluating the results, and knowing when to stop. Seeking evidence from a central source of quality appraised summarized evidence such as the database of Cochrane Reviews could potentially eliminate some of these obstacles, by making it easier to locate trustworthy evidence.

However, the Cochrane Library study revealed many obstacles springing out of the design of the web site that hindered successful searching. User tests of the library site carried out by a commercial company in 2006 supported our findings.[166] Rather than helping users find relevant evidence, presentation

of content and both technical and visual design of this particular web site at the time of our study seemed to contribute to the difficulty of this task.

Supporting search

Finding relevant information through the Library's search engine was a major problem for practically all participants. Though displaying no visible trouble reading English text, foreign-language participants were at an extra disadvantage when trying to search. Their problems were related primarily to difficulty recalling and spelling query terms that resulted in relevant hits. Spell check, query translation, or automatic query expansion with synonyms would be helpful to this group of users. International collections of evidence like the Cochrane Library need to be particularly attentive to the needs of foreign-language users. Adding functionality to support them in the stages of search - where they are particularly vulnerable - would require far less use of resources than performing full translations of the site's content¹⁷. These functions would be of help to other users as well.

Special search problems relevant to online collections of evidence

Many of the search problems users experienced are common ones that crop up in all kinds of web sites.[153, 154] However, health care related research collections have particular problems that need special attention. The most serious of these is the confusion between different document types. We observed one reader (who was very familiar with Cochrane) mistaking a single clinical trial for a systematic review, and not discovering the problem for many minutes. Others overlooked the linked labels on the top of the search results page that indicated what type of document followed on the list. Inexperienced users may not have a clear idea of the difference between Cochrane Reviews and other reviews or clinical trials. Misinterpretation of a clinical trial for a systematic review could have serious consequences, whether it is due to misunderstanding of the signposting on the site or to low levels of scientific literacy. Mistakes of this kind are particularly apt to happen in a Library containing many different types of documents, and these need to be clearly distinguished from each other, both in the search results and in the individual document. A more difficult challenge is to communicate

17 Since these studies were carried out, Google has launched a translation feature allowing web pages to be automatically translated into the reader's own language. If the user does not have this feature installed on their browser tool bar, a single web site can provide the same service by installing a simple translation toolbar on their pages, providing a dropdown menu of language choice. Although possibly crude to begin with, the translation power of this service will only increase in time as Google refines their translational dictionaries and begins to refine them according to context of use.

the concept of evidence hierarchy to non-researchers at critical junctions, for instance at the search interface.

An additional problem, also related to the concept of evidence hierarchy, is that a precise query will result in only one or a few hits, as the underlying Cochrane concept is "one review per topic". However, our participants' mental models of how search should function are based on searching Google and PubMed, where simple queries produce a great number of results. The idea that a successful search may result in only a few hits on a site where evidence is collected and synthesized is clearly novel to many users, and ways in which this can be made clearer need to be explored.

eHealth literacy

Difficulties in searching, misunderstandings about document types and poor grasp of the concept of evidence hierarchy can be framed as problems related to literacy. According to a model for eHealth literacy,[167] six literacy skills are needed to navigate the health care information found online: *traditional literacy*, *health literacy*, *information literacy*, *scientific literacy*, *media literacy*, and *computer literacy*. Traditional literacy encompasses text and numerical comprehension, whereas health literacy is "the degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions". A person that is information-literate "knows what potential resources to consult to find information on a specific topic, can develop appropriate search strategies, and can filter results to extract relevant knowledge." Scientific literacy involves "an understanding of the nature, aims, methods, application, limitations, and politics of creating knowledge in a systematic manner".[167] Media literacy involves thinking critically about the media and its messages; computer literacy is the ability to use computers to solve problems.

The umbrella concept of eHealth literacy has grown out of studies of consumer behavior and information needs, and is not often brought into the discussion of health professionals. However, it has clear value also in this context. People who are health-care literate on an expert level do not necessarily have equally high levels of knowledge and skill in other literacy areas. Therefore the term "expert" users may be a misleading. Both documents and web sites targeted at health literate professionals need to be developed so that they are compatible with lower skill levels, for instance in information and scientific literacy.

Is there something wrong with me, or am I just in the wrong place?

Unsuccessful searching led to more than just not finding evidence. Many users lost confidence in their own abilities when they couldn't find what they were looking for. These feelings of ineptitude is perhaps mirrored in an Australian study, where 41% of the participating physicians blamed their own limited search skills as impediments to making better use of research data.[168] The concept of distributed cognition, however, reminds us that users cognitive skills and tools in the external environment are interdependent. Many of the problems our participants encountered were not due to their own lack of skills, but to design flaws that could be solved following usability heuristics[169], research-based guidelines for web design [153, 170, 171] or by implementing better search technology.

Unsuccessful searches increased the suspicion of some participants that the site was for researchers, not for health professionals like themselves. This illustrates how poor search functionality might not only obstruct evidence-informed practice on a practical level, but may also widen the social gap between the world of research and the world of practice by alienating users.

Finding evidence through external search engines

Many users had trouble locating the Cochrane Library website in our tests. Since the studies in 2005-2006, the publisher has allowed crawlers from external search engines to access the databases, one of the recommendations made in our reports. This has greatly improved Cochrane Library ranking on Google search results pages. It also allows users to access specific Cochrane Reviews directly through an external Google search. Although this may help users more readily locate links to Cochrane Reviews, it might increase the chances that users will go elsewhere if they are not quickly satisfied when they arrive on the site.[172] Jakob Nielsen points out that with the development of good search engines, it has become easier for information gatherers to move quickly between different hunting grounds. This places increased pressure on the design of the pages that meets users coming from external search engines, both the front page of the site and any front page of evidence documents.

The value of evidence - related to effort needed to find it

Repeatedly we heard praise for the quality of content of this site. But frustration levels were very high, and several participants said they were ultimately too lazy to bother to use a site that made it so difficult for them. Information foraging theory describes user behavior on the Internet as similar to wild animal's search for food: we want maximum benefit for a minimum of effort.[173]

Understanding evidence

Recognizing document category

Closely intertwined with finding evidence is the problem of user's recognition of document types. Both in the Cochrane Library studies and in the SUPPORT study we observed critical problems due to users not recognizing what kind of documents they were looking at, such as mistaking a single clinical trial for a systematic review. Krippendorff's "theory of artifacts in use" suggests that use of artifacts can be divided into three phases: recognition, exploration and reliance.[15] His recognition phase involves identifying the artifact's location and category; here I am referring to the latter. Designing for successful recognition of category may mean making defining unfamiliar document types more explicitly, for instance through the use of "information about the information" that we added to the SUPPORT summaries. Or it may involve visually distinguishing different types of documents, for instance those published on the same web site and that therefore run a higher risk of being confused with one another.[174] Designing for recognition becomes increasingly important when users demonstrate a low level of information or scientific literacy. These people may not be at all familiar with the document types to begin with, and therefore might not be looking for differences between them.

Comprehension

Once evidence is found and identified correctly, it needs to be read and comprehended in order to be able to be used to inform decisions without misinforming. Studies of evidence comprehension have mostly been carried out in consumer contexts, probably since the Internet dramatically increased access to scientific information for the broader public. However, a few studies have looked specifically at health professionals' comprehension of evidence. A study of first-year medical students found that nearly one-quarter had trouble performing basic numerical tasks and that this seemed to be related to difficulty interpreting medical data.[175] One validation study revealed a considerable misunderstanding of the terms "relative risk reduction," "absolute risk reduction," and "number needed to treat" in a group of 50 general practitioners.[176] Lipkus tested general numeracy performance among highly educated samples in the context of health risk and found that even highly educated participants have trouble with relatively simple numeracy tasks (such as determining which represents the larger risk: 1%, 5%, or 10%).[177] Others document effects of low numeracy and statistical illiteracy on medical decision-making and patient communication.[178-181]

Misunderstandings we documented occurred on many levels that interacted with each other. From a design and development perspective, it is helpful to try to examine these levels separately, as they emerge from different underlying factors. The most prevailing types of misunderstandings we found are summarized below.

Unfamiliar scientific language

The most basic misunderstandings had to do with use of unfamiliar scientific language and jargon. This is a finding that is supported by other studies, particularly those of health policy makers' barriers to using research.[54, 97] Problems with comprehension of jargon create more than just gaps in understanding - they can contribute to participants' impressions that the information is not meant for not for people like themselves, increasing the gap between the two worlds. People who are not fluent in English may have added difficulties with language. Offering explanations for scientific terms or acronyms can help alleviate these problems, but avoiding these issues altogether through consistent use of unabbreviated plain language is an even better solution.

Cognitive load and class reference

The next level of difficulty in our studies had to do with basic misunderstanding of numerical content, particularly in the Summary of Findings Tables. One example was a finding from the SoF study, where "9 fewer per 1000" was consistently mistaken for meaning "9" or "9 or fewer". This may be an isolated finding, related to our particular table, which was complex and filled with many numbers, likely increasing cognitive load. However, correct understanding of the information in this particular case also required a cognitive leap - the class reference of these numbers differed from that of the previous cell in the same row. This was a problem that occurred many places in the table - confusion about what the numbers represented. To avoid these problems, we changed the formatting of the numbers where possible so that class references didn't change from cell to cell. We also added labels in direct proximity to the numbers where class reference was ambiguous instead of just relying on column/row headings. Problems correctly identifying reference class have been uncovered in past work.[182, 183] In a recent review of formats for conveying health risks, Lipkus recommends consistency in use of numerical formats.[184]

Stakeholders maintained that was easier (and more statistically precise) to present the absolute difference (x fewer per 1000) than the alternative we recommended after testing (xx per 1000). Therefore several stakeholders protested in making our proposed change. This case illustrates a characteristic

conflict between the interests of the stakeholders (scientific precision) and the needs of the users (consistent and understandable formatting). Our stakeholders would not have accepted the changes if we had not performed rigorous testing demonstrating that the existing formatting caused problems for users.

The kinds of user problems described above are difficult to pick up on in think-aloud protocols. Since it is the user's own reasoning that is faulty to begin with, he is not aware of the mistake and therefore does not discuss it. If, on closer inspection, a participant understands that he has made an error, he may be reluctant to admit it or describe what he now understands as a mistaken perception, for fear of appearing stupid to the people observing. On the other hand, use of a different study design to test for actual understanding (as in the SoF evaluation study) can document such errors but doesn't provide a deeper understanding of the factors leading up to them.

Numeracy

None of our studies uncovered basic numeracy-related problems (such as comprehending which of two numbers are higher in value), though we know from previous research that these can readily occur even in groups of highly educated subjects.[177] Lack of basic numeracy errors among our test persons may be due to the fact that we designed the first versions of the tables based on previous findings regarding these category of errors: denominators with the base of 10 (e.g. 10, 100, 1000) are easier to comprehend[184], and use of same denominator facilitates comparison.[185]

Statistical and scientific literacy

Participants in all of our studies demonstrated problems related to comprehension of the statistics and scientific method. In the SoF study, participants misunderstood or were uncertain about a range of elements that require some level of background knowledge about statistics and scientific method: dichotomous outcomes, continuous outcomes, confidence intervals, Relative Risk, or the meaning of "no data available". In the SUPPORT study, some participants expressed that the tables and text built on statistical and scientific concepts that were too difficult for them, despite author's attempts to present in simple language.

We have already discussed above how a lack of understanding of scientific method can be framed as one component of eHealth literacy. A related concept - health numeracy - is described by Ancker and Kaufman based on the skills needed to make sense of numbers and statistics related to health care: "the individual-level skills needed to understand and use quantitative

health information, including basic computation skills, ability to use information in documents and non-text formats such as graphs, and ability to communicate orally." [181] As described earlier, researchers and others are beginning to grasp that not only consumers, but also health professionals may lack health numeracy skills and are acknowledging the potentially serious consequences of this. [182]

When lack of conceptual background knowledge (due to low scientific or statistical literacy) leads to misinterpretation or comprehension barriers, two different design approaches might help. The first is to represent the evidence in such a way that *eliminates the need for the missing background knowledge*. An example of this is presenting risks as natural frequencies (e.g. 5 out of 100 people) rather than percentages (5%) or relative risk (0.05). This eliminates the need to know what relative risk means (in addition to the added advantage of providing the baseline numbers, a feature that has been shown to decrease bias in choosing interventions). [186, 187] Earlier studies have shown that risk probability represented as natural frequencies is more intuitively understood than other formats, [182, 188, 189] and doesn't require the same level of background knowledge. The class reference is made clear, and special domain knowledge from statistics is not necessary for correct comprehension. Another example of the same tactic is the use of the GRADE score, included in the SoF and SUPPORT Summary Tables. An earlier study found that one of the main reasons policy makers used systematic reviews was that they helped with the difficult task of evidence appraisal. [69] Including GRADE scores make quality of evidence included in a review even easier to find and understand, by completely eliminating the need for the reader to carry out their own appraisal of the evidence. Participants in our studies did not demonstrate problems understanding GRADE scores, despite most not having prior knowledge of GRADE, and including the scores in tables made this information much easier to find than without tables.

The other approach to dealing with lack of conceptual background knowledge is to *provide this missing background information*. This may help link the new knowledge to what the user already knows. In 2000, The US National Research Council published a summary of main findings from current educational research, hoping to influence the domain of educational practice in the US. [127] One of the main messages in that publication was that for new knowledge to be assimilated, it must build on what the learner already knows, their prior knowledge. An example of this is a tactic we used in developing the SUPPORT summaries. When discovering that participants' demonstrated a weak concept of what a systematic review is and what kind of information it could/couldn't provide, we added explanations in boxes on the

front page. The follow-up participants clearly considered this change helpful. Clinical or policy oriented users of evidence do not necessarily have comprehensive prior knowledge regarding scientific method or statistics. They may need help in constructing appropriate mental models or conceptual structures in order to properly understand evidence resulting from systematic reviews.

Grasping the gist of the information

In the Summary of Findings work we saw that while the user test participants (non-scientists) preferred simplicity in the statistical presentation, the stakeholders group (for the most part scientists) tended to value more details and precision in presentation. This might be seen in light of fuzzy-traces theory as well as what we know about the differences between experts and non-experts ability to acquire knowledge.[125, 190]

Fuzzy-traces theory is a dual-processing theory proposing that we process information along a gist-verbatim continuum the “gist” being the semantic meaning and “verbatim” being the precise details.[190] (An example demonstrating the nature of gist and verbatim representations is the final score of a basketball game. The verbatim representation (exact details) of the final score is: "69 - 68" (Team X versus Team Y). A supporter of Team X would likely extract the gist "We won!" from these numbers, while a supporter from the Team Y would extract a different gist: "We lost...". A referee with no ties to either team might extract the gist "Close game". These three people extract radically different gist based on the same verbatim data, due to their point of perspective. Lack of background information may result in an incorrect "gist", for instance not knowing enough about the game of basketball to understand what the numbers mean. Research has shown that although people process both the verbatim and the gist representations in parallel, they tend towards a gist preference.[129]

Detailed and precise information is important for a scientific audience, since it is precision and transparency of all parts of the work that make it possible for a scientific audience to judge the credibility of results. Stakeholders, who were for the most part statistically literate researchers, emphasized precise numerical detail in the SoF Tables. They had sufficient background knowledge to be able to extract meaning (gist) from this kind of detailed presentation. This group was for the most part experts in statistics. We can assume they had a well-established conceptual structure of quantitative data, with prior knowledge about what details are relevant to seek out in order to understand effect size or quality of the evidence. A non-expert does not readily see these same patterns, and will not as easily be able to glean the

important parts from the extra details. They may also be uncertain if their understanding of the complex information is actually correct. Therefore, even though all the information they need is present (in a verbatim form), they may require extra help to both pick out main parts (the gist) and to be reassured that they have understood these parts correctly. There are several possible ways of supporting non-experts with these tasks.

To begin with, graphical and layout treatment of the information can help emphasize the main messages: graded entry (key messages up front), highlighting (e.g. shading under columns of the SoF Table containing effect sizes or arrows pointing to findings in SUPPORT summaries) and layering of information (e.g. using bold/normal, larger/smaller type or colors of different contrast).[174, 191] The more complex the information, the more this kind of treatment may be needed to help a non-expert reader readily identify important parts and successfully extract the correct gist.

Additionally, text might be added to support correct gist extraction. Examples of this are “statistically significant” or “Favors stockings” texts that were considered in the table development. However, stakeholders felt these small texts were misleading and undermined the neutrality of the evidence. When these kinds of help-texts are not possible, proper graphic treatment to enable easy gist extraction becomes even more important.

Research in fuzzy-traces theory has indicated that preference for gist (over verbatim) processing may actually increase, rather than decrease, with expertise in a domain.[129] In light of this, development of summaries that cater to enabling non-experts to easily extract correct gists might also benefit the expert users of evidence. Perhaps the value of summarizing lies not only in rendering evidence *quicker to access*, but also in supporting *correct gist extraction* of complex data.

Appraising evidence

Appraisal is difficult

Appraising evidence is a more demanding task than merely reading and understanding. This is illustrated by a revised version of Bloom’s taxonomy that includes six cognitive processes associated with learning.[192] These are distributed along a scale that ranges from low to high cognitive complexity:

Remembering > Understanding > Applying > Analyzing > Evaluating > Creating.

Many non-researcher users of evidence may not have acquired the skills at the more advanced end of this scale, such as analyzing and evaluating. As

mentioned previously, earlier research indicates that policy makers find evidence appraisal difficult.[69]

Which elements indicate high quality?

The concept that evidence quality cannot be taken for granted requires a certain level of scientific literacy to begin with; an inexperienced reader may question why anyone would bother to publish weak evidence.[193] However, even if inexperienced appraisers are aware that research findings are likely to vary in quality, they may not know how to make a quality judgment. Some may mix the concept of credibility of source and quality of evidence, ending up relying merely on a familiar logo. In the Cochrane Library study, some participants demonstrated this kind of sweeping overconfidence in the evidence. These users transferred the quality association they had of the methodology behind Cochrane Reviews to the entire content of the library, thinking everything on *The Cochrane Library* site had been quality-approved through an editorial evaluation. This tendency (to base trust on source) has been shown in other research of web sites. While a large study from the Stanford Credibility project showed that consumers placed a lot of emphasis on the *look* of a site to evaluate,[194] a smaller parallel study showed that professional users tended to emphasize the *reputation of the source* when making judgments about the trustworthiness of information found online.[195] While this is great for a commercial company who wants to invoke universal faith in all of their products and services, this kind of brand-based trust is problematic for an evidence collection with individual documents of different levels of quality.

GRADE scores appear to help

The entanglement of source credibility and quality of evidence is difficult to sort out for an inexperienced reader who may not properly understand that poor quality evidence can also come from a highly credible source. However, we did not observe sweeping credibility judgments in the other studies *where GRADE scores were used*. Both in the SoF Table and SUPPORT Summary tests participants said that the GRADE score was the element they looked at when deciding whether or not they trusted the evidence. It is possible that these users were less inclined to make sweeping source-based generalizations about quality of evidence when they were provided with the more fine-grained GRADE scores.

GRADE scores also contributed to more accurate appraisals of the quality of evidence, as we saw in the results of the RCT's, despite users lacking previous knowledge of this system. Our studies have indicated that GRADE scores help readers successfully appraise the evidence.

Usefulness of evidence

While the concept of usability has received an enormous amount of attention the last decade, we hear considerably less about usefulness. However, a user's perception of usefulness may be the more critical construct of the two. The theory of technology acceptance (a theoretical model from Information Systems research) stresses the importance of the users' perception of usefulness to their adoption of new technology.[196] According to this theory, perceived usefulness has been shown to outweigh perceived ease-of-use in predicting technology adoption by users. Also, usefulness has been shown to predicate both intention to use and actual behavior in technology adoption studies. Although it is not clear how transferable these findings are to the realm we are concerned with, usefulness is potentially a very critical facet of the user experience of evidence¹⁸.

But what do we actually mean by usefulness? Morville's description of usefulness in the honeycomb model is framed as a question:

“Usefulness: does this product have practical value for this user?”

In a health care research context, two clinicians proposed a definition of the usefulness of medical literature based on their own experience as physicians:[197]

$$\text{usefulness} = (\text{relevance} \times \text{validity}) / \text{work}^{19}.$$

This formula offers a possible understanding of usefulness as an umbrella concept rather than a single construct, dependent on other characteristics in the context. ‘Work’ is assigned the most sensitive position in this equation, as even a slight increase or decrease in the denominator will result in dramatically different usefulness values. This concept echoes the basic principle from the Information Foraging theory discussed earlier, that the value of information corresponds closely with the amount of work needed to access it. However, this formula seems to exclude multiple perspectives in a rather subtle way, as if there were only single universal measurements of relevance, validity or work, independent of the person doing the measuring or

¹⁸ Although ‘evidence’ as an entity is not necessarily a ‘new technology’, it is mediated through technological artifacts, such as web sites, document files or paper.

¹⁹ The formula is elaborated on in the referenced article: "Relevance" refers to *patient care*, "validity" is technical rigor and "work" can be defined in terms of time, money or effort required to obtain an answer to a clinical question.

the context. Looking back to the Theory of Technology Acceptance that deals with perceived usefulness, we could make a slight adjustment of this formula to bring individual users' experience more explicitly into the equation:

$$\text{Perceived usefulness} = (\text{perceived relevance} \times \text{perceived validity}) / \text{perceived work}^{20}$$

To explore this idea further, I will briefly touch on all three of these concepts, and ask how they might be enhanced when designing/presenting evidence.

Perceived work: We have already covered a number of important ways that 'perceived work' might be reduced (for instance by making it easier for a user to find, understand and appraise evidence).

Perceived validity: This is somewhat trickier - we cannot alter the actual 'scientific validity' of evidence, but we can help users more easily establish a correct perception of that validity through GRADE scores. We can also proceed with caution when other features - such as the presence or absence of a logo - goes beyond affecting a reader's trust of the source and begins to affect their trust of the evidence.

Perceived relevance: The subjective experience of information being relevant seemed to be related to several parts of the content, such as the PICO attributes, the dates of the underlying studies, and the general feeling of whether or not the information was perceived as being for "someone like me". We also found that *adding extra information* might help make evidence more relevant for users by supporting their particular working tasks. For instance, adding explanations of the criteria used to define sub-groups (e.g.

²⁰ By perceived, I do not mean that a subjective perception (for instance of validity) is by itself sufficient – if the evidence is invalid by scientific standards, then a subjective perception of validity would hardly be useful. On the other hand, it does not help for the evidence to be considered valid by scientific standards if the user cannot perceive this quality. Another example of what I mean here can be illustrated by looking at "work": no objective measurement of work (e.g. number of minutes to find evidence) can tell us anything about how different individuals in different contexts experience the work needed to find evidence. For some persons in some contexts, ½ minute will be too much work, for others ½ hour would be considered more than reasonable. In addition, design features can change perceptions. For instance, designers of web sites have found that people are more content to wait for content to download if they have visible feedback about the progress – how much has transpired and how much time left. By using the word "perceived", I mean that the qualities must be accessible to the senses, that perceptions will likely differ, and that design has the potential to change perceptions. This applies to all the terms in the honeycomb model, not just usefulness.

high and low risk populations) helps a clinician to understand how to tell which part of the evidence relates to his particular patient at hand. This may seem like a very small detail that, when missing, actually renders evidence meaningless in a specific clinical context. More detailed information about the intervention (e.g. length of compression stocking) can also increase the relevance of information to a clinical encounter by enabling the user to carry out the “next step” in an intervention based on the information provided, without needing to seek out other sources.[198] This kind of information is not always available in the original research and might need to be collected by the systematic review or summary author.

Additionally, participants in the SUPPORT Summary study participants reported that the relevance section (“Applicability to low and middle income countries”) was highly useful. This was arguably one of the parts they found most valuable after the front page of key messages. What I found most interesting about this discovery was that these sections often did not provide much actual hard data, both because it is not readily available or difficult to provide in a general document that might be used in any number of specific settings. Instead, these texts tended to indicate what factors might need to be considered in order to gather this information for a local setting. Here is an example of a typical applicability text from one of the summaries used in the study:

“Economic Considerations

The studies included in this review did not provide sufficient data to determine the costs of using nurse practitioners, what, if any, savings can be achieved by substituting nurse practitioners for doctors, or to evaluate the sustainability of using nurse practitioners. The potential for scaling up the use of nurse practitioners depends on the availability of nurses and the availability and costs of additional training for them to become nurse practitioners, supervision and continuing education.”

One might argue that, in a user test, participants may have only skimmed this kind of text and only read the heading. This may have been the reason that they thought this section would be very useful for them, wrongly assuming there would be more concrete information (for instance about actual costs) than was written there. On the other hand, we might speculate that although this kind of text does not provide any relevant hard data, it helps the reader define the boundaries of the problem at hand, indicating what factors will likely need to be considered in order to work out economic considerations in their own setting. This could be considered a non-instrumental form of

usefulness, where the document helps the reader delineate and define the problem at hand, rather than directly supporting problem solving or arriving at a decision then and there.

This idea has occurred to me because I see the likeness between some basic aspects of designing and policymaking. Both fields are striving to create new things that do not yet exist. Although this may be a vast oversimplification, it is possible that some of the basic characteristics of the practices may resemble each other. As pointed out in the background, a large part of designing is not just problem solving but problem setting, and these two parts of the work are highly intertwined. This might also hold true for policymaking processes. The kind of information (research-based) that would be useful for problem setting would possibly differ from what would be considered useful from a traditional decision making perspective modeled after a choice between two given alternatives (control and intervention). This could be explored in future research.

Usefulness depends on the definition of use

According to Innvær et al, “the question of what is meant by the concept of the ‘use’ of evidence is the most commonly discussed theoretical issue in the literature on knowledge utilization”. [53] Davies et al maintain that there is little evidence indicating that research is commonly utilized in a rational and linear fashion in policymaking [25] and there is much support for this idea. [53, 66] In the late 70’s Weiss outlined six different models of research utilization: knowledge driven, problem-solving, interactive, political, tactical, and enlightenment. [199] Innvær’s more recent article describes the three most frequent categorizations as direct (instrumental), selective (legitimizing) and enlightening (conceptual) use of evidence. Enlightening or conceptual use, according to Innvær, refers to research that helps establish goals, outline what is attainable, and deepens an understanding of the complexity of the problems as well as possible unintended consequences of action.

Another utilization perspective is provided by Patel, who makes a strong argument for re-framing clinical health care decision according to a problem-solving model. [200] She and her co-authors argue that the importance of conceptual knowledge in decision-making has been underestimated; instead focus has been on providing the factual knowledge of the risk probabilities of alternative treatment choices. They suggest that conceptual knowledge differs in important ways from both factual and procedural knowledge, and that it has a predictable effect on decision processes. Alternative models of utilization are important to consider, because information that is designed specifically for one type of use (such as rational linear decision making

according to an EBM ideal) may be presented in a way that doesn't easily translate to other kinds of uses, and thereby – less useful.

Finally, I would like to point to Buchanan's comments about usefulness related to product experience and design research. He has illustrated internal (users') perspective on products with a simple triangle pointing to three constructs: *Usable*, *Desirable* and *Useful*. [162] He maintains that - while investigation of the *usability* of products points us in the direction of human and cultural factors, and investigation of *desirability* points us towards aesthetics and identification - investigation of *usefulness* points largely in the direction of content and structure. Though designers may meddle with structure, they traditionally avoid responsibility for content. However, these two are intertwined in complex ways that cannot be ignored. Designers who care about user experience need to redefine both content and structure as parts of the artifact that naturally fall in under their sphere of concern.

“GAPS” BETWEEN USERS AND PRODUCERS AND CONSEQUENCES FOR DESIGN

Knowledge gaps

Metaphors of informing medical decision making often refer to new knowledge as though it were some kind of an object or substance to be transmitted from one place (the world of research) to another (the world of practice or policy making). One illustration of this is a quote from Sir Muir Grey; he compared the need for “pure clean knowledge” to water that could be spread through pipes, “ensuring that it reached clinicians and patients at a time of need”. [201]

This is a contestable view, as there are varying assumptions about the nature of knowledge in different research traditions. For instance, a meta-narrative review of the literature on electronic patient record systems found seven main groups of research that built on different philosophical positions and different views about the nature of knowledge. [202] The US National Research Council co-authored a major summary of research on human learning, where knowledge is not presented as fixed objective thing, but as something constructed by the individual in contact with the external world (through own experiences or interacting with social communities), building on top of, replacing or reshaping what the individual already knows. [127] One of the main consequences articulated in that text is that if new information does not link with already acquired knowledge, then learning cannot take place.

From that position, an individual's *prior knowledge base* is an important variable that can cause information to be experienced differently by different people, such as experts vs. non-experts (or producers vs. users) of evidence. A researcher familiar with quantitative data will have a better conceptual platform to build on than a non-researcher. He most likely be scientifically and statistically more literate than non-researchers, enabling him to better understand what type of information he is reading and the language used, more readily extract a meaningful gist from it, better assess the credibility of what he is reading or even just be more skilful at finding the evidence to begin with. Acronyms, for instance, may make perfect sense to people familiar with them, but can render a text meaningless to those who haven't been exposed to them before. A search engine without spell check may provide successful results for English speaking users who easily recall search terms and spell them correctly, but create insurmountable barriers between people who are not native English speakers and the information they seek. Those who haven't heard of the concept "systematic review" before may not readily recognize these document as particular types of evidence syntheses. These examples show how different sets of prior knowledge lead to different experiences of artifacts.

Design can help bridge the gaps between people with different levels of different types of literacy, by presenting and representing evidence in a way that compensates for non-experts' lack of knowledge. The addition of a Summary of Findings Table, for instance, was shown to help readers of systematic reviews perform better on measures of understanding and time spent to find key results. The SUPPORT Summary was improved by adding a description of what kind of document a systematic review is. GRADE scores appear to help non-statistical experts understand the quality of evidence.

Different uses lead to different experiences

Another variable causing artifacts to be experienced differently is that they are used in different contexts and for different tasks. Tools (including texts) that are helpful for certain tasks in one context may create difficulties for other tasks or other contexts. For instance, the narrative review of research on electronic patient records mentioned above found that while EPJ systems apparently created advantages for administration tasks and quality monitoring in health care settings, they seemed to be experienced by doctors and nurses working on the patient wards as less flexible and more time consuming to attend to than paper journals.[202]

I have been referring to clinicians and policy makers as “users” in this entire thesis. However “user” is actually a misleading term because it disguises the fact that researchers also ‘use’ research evidence, albeit for very different tasks than clinicians or policy makers. Researchers use published evidence directly and indirectly for a number of tasks that are central to their work, e.g. establishing themselves as scientists, building their reputation among peers, strengthening their career opportunities and creating possibilities for funding of future work. The published research of others is also used in many ways, for instance to create a context for ones own work, to support or illuminate the findings from ones own studies, or to identify new research questions. Therefore, artifacts emerging from research activity (such as research publications) can take on completely separate meanings for a scientific community than a clinical or policy community because of the different nature of tasks they support – the researchers’ participation in a scientific discourse.

One might even venture to ask whether, in light of the central role research publication plays in this discourse, there might be very strong (possibly unconscious) tendencies to use these published artifacts for the task of drawing up or defending the discourse boundaries of a scientific community, identifying the members from the non-members.[15, 203] Some of the comments emerging from stakeholder feedback suggested this, i.e. an inherent hesitance of some to making evidence accessible to people who didn’t understand all the complexities of it, or the rather persistent insistence on use of scientific lingo. At any rate, if scientists’ use of published evidence to participate in, build (and possibly defend the boundaries of) their discourse community is not openly reflected on, these issues may retreat into the background and continue to manipulate dissemination in ways that are hard to see or change. For instance, needs of researcher-users may appear more like non-negotiable givens - inherent features of their products - rather than being treated as features that have rather developed over time to serve one of many types of users, the scientific community. And it may be precisely some of these features that scientists adamantly insist on, such as use of language or level of detail, that cause problems for other groups of users.

One reader of a draft of this thesis has suggested that the power implications of knowledge flow and translation (for instance how knowledge transfer can change relations between service users and health care providers) possibly merits some more discussion in my work. I will attempt to deal with this point in brief; however it is likely worthy of a more comprehensive treatment than I provide here.

Evidence dissemination efforts seem largely driven by a democratisation ideal – making information transparent and widely available for those whose lives and practices are potentially affected by on it, with the overreaching aim of helping improve quality of care across settings. Making the connection between evidence dissemination and redistribution of power seems relatively obvious in the context of patients’ rights to information. Acknowledgement of patients’ right to be informed has gone hand in hand with a move away from a patriarchal model of medical practice, towards a newer model where medical decision making is seen as potentially a shared process.

However, issues of power that might exist between researchers and health professional ‘users’ of evidence are possibly less obvious to an outsider. One of the central issues has been the everlasting debate about what kind of knowledge - coming from what kind of studies - constitutes robust evidence of effect. But in the studies in this thesis, subtler issues cropped up that had to do with how simple (or how complex) the information can be presented before it loses its scientific validity. There are no easy answers to this question. But researchers need to be sensitive to the fact that these issues are not insignificant. Ignoring them may lead to presentations of evidence that are completely misunderstood by health professionals. One way of levelling out the playing field between scientists and other user groups is for scientists to directly engage with users and observe with their own eyes how their products are being perceived. This could be a surprisingly enlightening experience that encourages them to work towards ways of enabling these users, rather than defending their own scientific territory. Participatory design is a practice springing out of the same ideals – involving users in creating the artifacts that surround them and affect their lives, in order to redistribute the power that is inherent in designing. My ambition with this body of work has been, together with my colleagues, to help move the field of evidence dissemination in that direction.

Helpful concepts for design: Boundary objects and layers

Star describes boundary objects as things that can link disparate social worlds.[204] Because the object tolerates multiple meanings to different groups, coordination of action between these groups is possible. Artifacts presenting evidence have a potential to serve as boundary objects linking the research world of health care with the practice communities. However, to do this successfully they must function sufficiently well in both worlds. They must also be robust enough to resist being changed in the future in ways that might lead them to again become meaningless to one of the groups.

Unstable boundary objects?

Although the SoF Table was found to have a positive affect on finding and understanding evidence in a systematic review, this finding might not be very stable. These positive effects are dependent on the design of the table, which is subject to constant change. The SoF template we developed may be extra vulnerable to change for several reasons. To begin, it is part of a dynamic publishing system in a large democratic organization. The alterations we made during development were quite small, seemingly innocuous. These same small details might be easily changed back again in future SoF Tables, either intentionally or inadvertently by any number of stakeholders in the Cochrane Collaboration who did not participate in our project, and who may be unconvinced of the importance of sustaining the current design and formatting that includes details critical to clinical users.²¹ Unintentional changes may also occur, for instance during translation into publishing technologies that transform the graphical representation in ways that negatively affect user experience.²² Our table template is also potentially inherently unstable because the final result does not seem to be a completely optimal solution for either group. (I have a vision of us having stretched the design of the table between these two user/stakeholder groups like a rubber band, to the point of it snapping.) Because of this built-in tension, it is not unlikely that searches for more other, more optimal solutions will continue.

Layered boundary objects

In the search for increased robustness, it might be useful to conceptualize a systematic review as a boundary object with multiple semantic layers. Krippendorff uses the photocopier machine to exemplify his concept of semantic layers. It has one layer for regular users who make copies – loading documents, choosing options and removing the copies. Then there is a layer that is accessed when something goes wrong – opening the cover, following directions, removing jammed paper. Finally there is a layer for repair people, who enter a technical world that is shielded from the user.

‘Layers’ is a concept that is helpful on many levels of design. At the typographic level, it means pulling some things into the foreground and

²¹ Today there is a group of other researchers in Canada developing a new version of a SoF table, for similar but slightly different set of users (guideline developers). They are adding even more scientific method details to their table, claiming that their users “need” this information, though without user testing. This table will be even more complex than our solution, and probably not suitable for other audiences such as practicing clinicians.

²² So far, the SoF table has not weathered these translations very well, rendering the graphic representation more chaotic and confusing, particularly on the web.

letting other details rest in the background, aiding recognition and gist extraction. At the layout level it can mean using a graded entry format in designing of a document, or adding a table of findings to a systematic review. At the document level it might involve developing increasingly condensed information artifacts, such as those types illustrated by the Haynes 5-S pyramid. Layers can aid understanding and rapid gist extraction, or they can function as complementary interfaces that support different users and different types of tasks.

The SoF Table itself is a layer, lying on top of the more complex information layer - the unabridged systematic review. However, because our SoF Table may be relatively unstable as a boundary object between researcher and practice worlds, more layers might evolve. Possibly new artifacts that are increasingly condensed and simplified will develop grow on top of the SoF Table to more specifically support clinicians, in the same way that SUPPORT summaries are developed to target policy makers. Or possibly less research-oriented clinicians will resort to simpler tables in Cochrane Plain Language Summaries that are being developed for lay audiences, leaving the original more complex SoF Table to researcher-savvy users. Time will tell. In any case, layering appears to be an important feature of artifacts linking the research and practice/policy making worlds.

EVALUATION OF THE USER EXPERIENCE FRAMEWORK

In this section, I discuss the usefulness of the honeycomb framework for this work, and suggest modifications of its user experience facets based on our findings. I also reflect on how the framework might be improved through use of concepts from Krippendorff's "theory of meaning for artifacts in use".[15]

How suitable was the honeycomb diagram for this work?



Morville, with a background in librarianship and information architecture, developed this framework from his own experiences designing web-based interfaces. (I have already described it in more detail in the background and method sections.) Although it is difficult to say how widespread the honeycomb model

is, I am aware of many designers in Norway who find this framework to be a useful practical tool. It has also received a lot of reader attention on his blog in the USA.[141] I have searched for other research work building on or evaluating Morville's honeycomb, and have contacted Morville directly. To our knowledge, there are no published research studies about this diagram.

Honeycomb strengths

Several advantages of using Morville's set of user experience concepts have become apparent during the course of this work. To begin with, the diagram has a useful level of abstraction that lends itself well both to aiding design inquiry (improving specific artifacts in a specific context) but also research inquiry (extracting a certain level of transferable knowledge out of the work). There are difficulties in bringing findings from design research (such as prescriptions of how to carry out work) into design practice,[38] and it has been suggested that part of the problem might lie in a poor understanding of the nature of design practice among researchers.[39] This framework has originated in the practice domain, so essentially my move is a reversal: bringing a practice framework into research. Possibly that is a more fruitful approach for producing practice-relevant knowledge.

The facets of the honeycomb model are also well-suited to examining the *distributed* nature of the influencing factors leading up to a user experience, as each topic can be used to illuminate any or all of the distributed parts of the resulting experience (the user, the product or the environment). Poor findability might be traced back to characteristics of the user (e.g. spelling skills or knowledge of how or where to search); or it might be reflect characteristics of the product (e.g. lack of meta-data or visible labels marking the product); or it might be attributed to characteristics of the environment (e.g. the presence of many similar and competing pieces of information in the same location). If we use the term 'findable' instead, the meaning of the term shifts ever so slightly to only encompassing the attributes of the product, a subtle but critical difference that hides the other two locations from view. However, by using the term 'findability', the conceptual space opens up to including all three - the users, the product and the environment - in our understanding of barriers and possibilities. The product may be changed to compensate for obstacles elsewhere, but is not sole cause of problems – it is always in interaction with attributes of the users and the context.

Areas for improvement

Adding "Understandability"

In using this model over several projects, I became aware an important facet was missing for our particular use, one I have called '*understandability*'. I

had originally interpreted this as a sub-section of ‘usability’, (effective use) but came to the conclusion that in the context of evidence presentation it needed to be given much more explicit emphasis. ‘Understandability’ should cover two different kinds of comprehension: understanding (or recognizing) the document category and understanding the document content. Both of these involve again two separate dimensions: the user’s subjective perception of her own understanding, as well as an objective measure of actual (correct) understanding that can (and should) be tested separately. By correct understanding of the evidence, I mean one that is aligned with the scientist’s understanding, or at least not contradicting it in a serious way.²³ Extra emphasis on this construct is also needed because health professionals’ understanding of evidence seems so far to have been taken too much for granted. There are few studies addressing this topic; our work adds to growing suspicions that problems with comprehension may be underestimated and that they deserve extra attention.

Adding “Affiliation”

Though we observed many language problems across studies that contributed to the basic cognitive tasks of finding and understanding, participants also indicated that scientific language created problems of another nature, signaling that the information was not for “someone like me”. These kinds of reactions also surfaced when users had problems finding what they were looking for. The honeycomb model didn’t have suitable facets to facilitate capture of this part of the user experience problem, which we called “alienation”. We placed these kinds of findings under *desirability* (lack of), but that requires a leap of reasoning that feeling alienated will automatically lead to not desiring the product, a conclusion we did not have grounds to draw. I have considered using the word “alienation” in the framework, but because it is a negative attribute, it does not harmonized intuitively with the other facets that have a positive framing. Considering different terms, such as identification, belonging, membership, I decided to choose affiliation, although I feel that it is less than optimal because it does not adequately associate to the alienating experiences we observed when this dimension is lacking.

²³ I am aware that the concept of “correct understanding” could possibly spur a whole new thesis on the nature of knowledge or diverging views of scientific truths, but here it will suffice to remind the reader that our goal has been to aid alignment of the understandings of the scientists and the practitioners, not to create arguments about the virtues or validities of either groups’ standpoints.

“Identification” has been proposed as a component of another model of user experience, based on a study of new iPhone users.[140] In that study ‘identification’ was found to have two perspectives – personal and social. The former dealt with personalization and incorporation of an artifact into daily routines; the latter was associated with enabling self-expression and creating a sense of membership with a community with shared values and interests.

As mentioned previously, Buchanan described a simplified triangle of user experience with three facets: usability, usefulness and desirability.[162] He maintained that while ‘usability’ points to the concerns of human factors, ‘usefulness’ points more towards content, and ‘desirability’ towards aesthetics but also to the concept of identification.

So, although it seems clear that identification is an important dimension, I find that the term ‘identification’ associates too heavily with a personal sphere. For the purposes of identification within a professional knowledge context, the term affiliation seems to be more appropriate.

Value - for whom?

In the original Morville framework, the facet *value* actually means "value for the client" as opposed to value for user. This seemed like a complicated construct to address in a model for user experience of scientific evidence. We were not testing in the interests of literally selling something, but rather bringing research results all the way into the decision-making doorstep. What would "value" be for publishers of systematic review summaries, other than an optimal user experience that could lead, further down the line, to informed use of the evidence? In our studies we re-interpreted this facet to mean "overall value for the user". However, since this was already covered by the sum of the other facets, participants' responses to the interviewer's questions about value were often redundant. My conclusion is therefore that the facet *value* could be removed for the purposes of exploring users experiences of evidence-based artifacts, or used to double-check answers already elicited.



Figur 9: First revision of honeycomb framework, based on results from these studies.

Lacking dimension of user experience over time

The honeycomb structure is deliberately non-linear and non-hierarchical. There is no attempt to organize the facets in any one chronological order or hierarchy of importance, as this will vary across users, artifacts and contexts. However, some of the facets suggest that there are aspects of chronological order or contingency of the type “if not A, then stop here”. “Accessibility” and “findability” have some of this quality – if information is not accessible or cannot be found, the rest of the framework is not useful.

I have tried several ways of organizing the facets in some kind of chronological order that illustrates this contingency on accessibility and findability, but have always gotten stuck. This is because other factors still may interfere - if a user is motivated enough, he will likely find ways of overcoming even considerable obstacles related to accessibility or findability. The interdependence of the different facets, and their different meanings in various use contexts make it impossible to arrange them in any kind of single order. However, it cannot be denied that there is a dimension of time in a user experience. First impressions may be quite different than experience after 10 minutes or 10 days, as a user moves from novice to expert in handling an artifact.

Other theories of user experience over time

Krippendorff's "theory of the meaning of artifacts in use" addresses both the dimension of time and motivation.[15] As explained earlier, he differentiates three (consecutive) modes of use:

Recognition: This means is "cognizing again, identifying something by its kind (name) and in view of the use to which it could be put". It concerns for the most part the location and category of artifacts, not how they work. Categorization research has shown that people recognize objects by how typical they are in relation to an ideal type of a category. Deviating too far from the ideal type can render an object unidentifiable. Unknown or new types of objects can also take on a metaphorical form, reminding us of something else and helping assign an unknown object a familiar identity. Krippendorff also places the quality of *attractiveness* in the recognition mode, including features such as "newness", "simplicity", "unity" and "intentionality", arguing that these are some of the conditions that attract potential users before acquisition, exploration and use.

Exploration: While users' gain a sense of what an artifact is in the recognition mode, they *explore* an artifact to understand how they can interface with it, for instance by trial and error. There are two entry points to the exploration mode: acquisition and disruption. Users often bring with them conceptual models of how an artifact works, which may or may not be correct. Built-in constraints can help hinder dangerous misuse of an artifact, as affordances can suggest possible correct uses. Designing artifacts in a way that affords multiple understandings through semantic layers of meaning can aid exploration from very different user perspectives.

Reliance: While users gain understanding of the ways of using an artifact in the exploration mode, they no longer ask "how-to" types of questions in the *reliance* mode. At this stage, the artifact falls into the background and is merely "relied" upon. The artifact becomes invisible, as the user attends to the primary task at hand. Heidegger refers to this state as "ready-at-hand". If disruptions occur, one is thrown back into the exploration mode. Krippendorff includes motivation in his description of the mode of reliance: extrinsic motivations stemming from goals to be reached or tasks to be completed, and intrinsic motivation that comes from the pleasure of being immersed in a process. Intrinsic motivation is a key feature of a meaningful artifact, entailing among other things unproblematic interaction, user autonomy, and confidence handling the artifact.

While these categories have an intuitive logic, they are not based on explicit empirical study. Karapanos et al followed people who purchased a new iPhone over four weeks and, based on their findings, proposed an initial framework for user experience over time.[140] This framework also includes three phases: *Orientation*, *Incorporation*, and *Identification*:

“Orientation refers to users’ initial experiences that are pervaded by a feeling of excitement as well as frustration as we experience novel features and encounter learnability flaws. In Incorporation we reflect on how the product becomes meaningful in our daily lives. Here, long-term usability becomes even more important than the initial learnability and the product’s usefulness becomes the major factor impacting our overall evaluative judgments. Finally, as we accept the product in our lives, it participates in our social interactions, communicating parts of our self-identity that serve to either differentiate us from others or connect us to others by creating a sense of community. This phase we call Identification.”

One-size-fits all theory of user experience?

These two frameworks both have three distinct phases of user experience with artifacts that have some degree of correspondence with each other. Karapanos’ is based on empirical study and maps out what characteristics of user experience appear to dominate in each phase. However, this work is carried out in a very different context – consumer products for personal communication. It is not at all clear how well all aspects of this framework would translate to professionals’ use of health care evidence. It is *possible* that, like in use of iPhones, ‘identification’ might be the dominant final phase of user experience of evidence over time. My work has not explored use over time and cannot inform that hypothesis. It could be just as possible that use of research evidence in a professional context differs so much from personal use of a new iPhone that the same categories do not apply, or at least not with the same weight or distribution. For instance, in the Cochrane Library study we saw users who felt alienated already on the home page – a critical kind of user experience related to ‘identification’ that happened in the beginning phases of their meeting with the product. User experience of scientific information measured over time might be more likely affected by attributes such as usefulness and credibility.

Instead of pitting different frameworks up against each other, I maintain that it is difficult to establish a one-size-fits all model. User experiences of different designed things – signs and symbols, products, services,

environments - made of different materials and for very different purposes vary so widely that any framework that would fit all of these would likely be too general to be useful as a conceptual practical design tool.

My intention is to suggest a change to the honeycomb model that adds a dimension of time, but where the weight and distribution of the attributes is flexible, not fixed in relation to the phases of use. This is because different artifacts create different kinds of challenges for users, depending on what problems are embedded in their design, and these are problems that can crop up anywhere in the timeline of a user experience. For instance, a product can change its design slightly and create new problems for expert users that are usually associated with newcomers (e.g. orientation and learnability). Equally, a sense of identification (or the opposite – alienation) is not only important in the later phases, but can draw a user in or drive them away in the very first meeting with a product.

Therefore I have chosen to adapt the time phases described by Krippendorff, as these are more neutral and lend themselves to flexible mapping with the different facets in the honeycomb according to the particular product and context at hand. However future research should evaluate the framework more explicitly, including this temporal dimension.

How can these three modes be used together with the honeycomb framework?

First of all, they can help more firmly establish a broader conceptual model of user experience. There is a tendency in design practice to focus on the exploration mode without much awareness of or attention to recognition or reliance phases of use.

Second, these generalized modes of use can provide a better structure for data collection and analysis in user testing. Questions or tasks could focus more explicitly on one mode at time. Potentially all of the facets of the honeycomb diagram might be meaningful to explore in all three modes of artifact use, but in very different ways. For instance, in the recognition mode, accessibility and findability may dominate. But all of the facets of user experience are also present in this first mode, in the form of the users' *anticipated experience*, their expectations prior to actual use. (Does this artifact look like it will be useable/credible/useful? Is it worth my while to continue, based on what I perceive at first glance?) Visitors to web sites may make these kinds of judgments in a fraction of a second.[205] Additionally the meaning of a facet may change according to which phase it is being used. "Understandability" in the initial recognition phase likely involves *understanding the category* of the

document – what type of information it is. In the second phase, however, it may increasingly involve *comprehension of the actual content*. Likewise for credibility – a user in the initial mode will likely pass credibility judgments on the *source* of information, whereas in later phases the actual content might be more open for scrutiny.

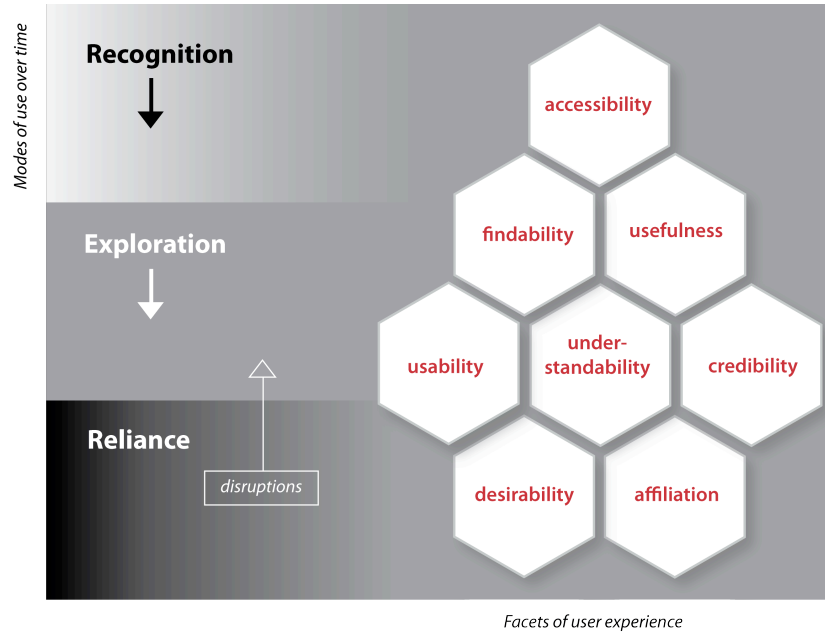


Figure 10. Second revision of honeycomb framework, with proposed modes of use over time.

New framework for user experience of summarized evidence

Above is a revised version of the honeycomb diagram together with the Krippendorff's theory of artifacts in use. I've avoided a direct mapping of the honeycomb facets onto three user modes, but rather placed them beside each other in a way that suggests that all nine facets may be relevant in each of the three modes. Although this combined and adjusted framework may be useful for testing and developing broader set of artifacts than tools and summaries for dissemination of research evidence, an evaluation of that is beyond the scope of this work. Future research should include testing both the revised set of user experience attributes as well as the modes of time.

DESIGNING FOR CROSS-PLATFORM AND
AUTOMATED TEMPLATE-BASED PUBLISHING**Limitations of transferability**

It seems that the closer a design inquiry moves toward generating output tailored for very particular contexts, the less transferable the design of that output is for use in other context or for other similar content. Even translating a design (of identical content) from one program or publishing platform to another can be problematic. For instance, the SoF Table designs we manually created in Word have not translated well to the computer-based publishing systems they were meant for despite comprehensive planning and communication with technology teams. The technology was too different for the design of in one specific context – Microsoft Word – to translate seamlessly to other specific contexts – the Cochrane Review/Cochrane Library PDF and html publishing systems.

Flattened formatting

Additionally, using a single rigid information structure (a template) to form a multitude of individual instantiations based on content which may be similar but not identical does not always provide a smooth translation of design intentions either. Template-based (rule-based) publishing – which treats all content elements in the same way – cannot always produce results that are optimal for every single input variation. Sometimes (I would venture to say ‘often’), manual intervention or tweaking is necessary. In the SoF Tables, for instance, there may be a need to redistribute column widths in a non-standard fashion to accommodate for more content in a particular cell, particularly if the overall intention is to retain a one-page-presentation. Or new subdivisions of cells may be needed that are non-standard (i.e. dividing single cells into several different risk groups.) We also wanted shading of cells and variation of font size and weight within cells (to enable easy gist extraction) that proved to be impossible to design a rule for that could apply for every single table. The result was a flattened typographical result that left out some of the details that helped create clarity and emphasis. Typographical design entails creating a fit between very specific content and visual representation²⁴; that fit is jeopardized or flattened when a machine is doing the work and no adjustments for individual variation in content is possible.

²⁴ This is possibly the reason it is often difficult to establish anything but basic typographical guidelines. Textual artifacts are unique contexts that need individual typographical treatment if the outcome is to be an optimal elegant design that “fits”.

Flattened content

The limitations of the transferability of a template to all kinds of content will be evident to anyone who has published content in web site templates. If you are publishing in a system with standardized templates - for instance where body text is preceded by an ingress - you need to comply with this content structure to obtain a proper result. If you want to introduce a different type of text, such as a text that has no ingress (i.e. a conference schedule, a poem, or just the insertion of a quotation before the ingress) you will either need to have a template accommodating this new type of content structure or you will have to adjust the output manually, for instance by manipulating the html code yourself. When you can do neither of these things, the only choice you have (if you want a proper presentation) is to change your own content, making it comply with the template. The result is a homogenization of content, the eliminations of variations and idiosyncrasies in the published texts.

From general to specific without manual adjustments = poor fit

This is the disadvantage of rigid publishing systems – everything gets treated the same and starts looking the same, even when it isn't. Tufte has complained about this tendency in a scathing criticism of power point templates, arguing that it flattens and homogenizes discourse, even generates errors resulting from camouflaged significance.[206] In our studies we saw that one-size-fits-all publishing templates can set limits on the amount of care that can be given to creating orderly typographical solutions of complex information that help bring the gist of the information forward and prevent cognitive overload. Manual adjustment, or particular attention to the singular details of a particular context, is often needed for obtaining a good typographical fit between content and template in publishing.

The first example – from Word to html/PDF – illustrates a problem of translating from one specific context to another. The second example – from a table template to multiple instantiations of a table – illustrates the kind of problem that can occur when translating from the general to the specific.

I believe there is a parallel to many other general-specific problems in other areas. Off-the-shelf electronic patient record systems (EPR), for instance, have trouble “fitting” the specific workflow contexts in different particular health care settings; there are indications that homegrown solutions may work (fit) better.[207] Along the same lines, structured template-based EPR data structures may not “fit” the myriad of different contexts they are used for such as individual patient histories[208] or particular clinical work flow

(although they may better “fit” more standardize-able tasks such as gathering data for auditing and billing)[202]

Another example of an area characterized by the general-specific dilemma is the research-practice/policy gap that is the topic of this thesis. One of main challenges of research is how to move generalized findings from scientific studies into multiple particular contexts that will always vary in degrees of significance from the original setting, if for no other reason than not being the identical event. Some differences between study contexts and use contexts may not vary much at all, such as the difference between the participants in the compression stockings study and my daughter who flies regularly between Norway and the USA. In other cases, differences may be much greater and have much more potential significance, such as the setting of a study of lay health workers in Africa working in an AIDS-ridden village and a group of health care community workers in Brazil trying to deal with the same disease. In order for generalized results of research, such as those from a systematic review, to “fit” a particular practice or policy context, they may need manual translation, for instance by knowledge brokers [209] or people who are sensitive to both the context of production (research), the context of use and the significance of the differences between them.

How much and what kinds of specific information (of local settings or populations) could be put back into generalized research results, in order for the knowledge to be useful in local settings? In a current project we are conducting in Africa, a pilot study participant suggested we could replace (or supplement) the baseline risk in the SoF Table with local baseline data. This might be one way of adapting generalized results from systematic reviews to more specific local contexts.

The end of the journey

TROUBLESOME EXPERIENCES

I recently attended a party where I fell into conversation with an architect I didn't know well, trying to describe what I do. This is a problem I have had a lot in recent years, because I have moved so far away from the areas of work people can intuitively connect with design (i.e. making publicly visible things like book covers, postage stamps, Olympic pictograms). Usually I can just offer a simple explanation ("I help make medical research easier for doctors to read and understand") and graciously allow them to change the topic of conversation to something they can more easily relate to. But this person, also a designer, was interested in the nitty-gritty of what my work in these research-based projects entailed. What precisely was the design contribution? How could I explain to him, in words an architect would understand, the more exacting nature of this work? In trying to provide him with a satisfactory explanation, I ran into several significant problems.

Designing objects that are out of sight (and outside shared language)

The first problem was to explain the subject matter, the object of my design work. Not only was it out of public sight, it was out of sight, period. Where does a Summary of Findings Table exist? On a web site this architect never heard of, in a pdf document he will never see? These things that I work with are embedded deep into the crevices of medical research publishing systems, where even health personnel have difficulty finding them. It doesn't help the matter any that they also are not part of any easily identifiable tasks that we can associate with doctors' or nurses' work, like measuring blood pressure or ordering medicine. "Checking the research" might be something health professionals do sporadically, but it is on the periphery of more pressing care-giving and decision making tasks that we associate with their work. (Have you ever seen anyone checking the research on "ER"?) Additionally, tables and summaries of systematic reviews do not belong to any established category of objects that non-experts have language for, like "Patient Information". There isn't a parallel category in our common everyday language for "Doctor Information" (not yet). Some of my colleagues call it

“Summarized research about the effect of medical treatments”. I’ve tried this, and I often get a blank face before I even get on to explaining the concept of reviews or condensed summaries of these.²⁵

Working with objects of design that are hidden from public view is certainly not new in the field of design. Information designers often work with designs that are not intended to be noticed, surfacing to people’s consciousness only when they doesn’t work properly (i.e. a book set with too small a font, a grocery store label with a bar code but no price tag, an Internet form that insists you fill in a 5-digit zip code even though you have indicated that your country is not the USA). The difference is that when you tell someone you design books or grocery store labels, they understand what kind of objects you are talking about. My area of design is a type of information so invisible to anybody outside a specific tiny scientific circle that even health professionals (in Norway) often don’t understand what I am talking about, much less architects.

This particular part of my explanation problem is slowly changing, as the kind of information I’ve been working with is gradually bubbling up to the surface of public view in Norway. Going through the newspaper recently I found an article citing the results of a Cochrane Review, which indicated that treating heart attack patients with oxygen might be increasing their risk of dying. The paper called the document “a comprehensive review of available research carried out by the Cochrane Institute”²⁶. I wondered if the architect read that and if he made any connection to our conversation.

This issue may sound like a small problem, but it has professional identity ramifications that should not be underestimated. Although the field of design has expanded enormously in the last 20 years in terms of the type of products being designed, the kind of organizations purchasing design services or the roles designers are creating for themselves,[211] the way we identify designers and talk about what they do does not seem to have followed suit. Designers are by and large still identified through the types of objects they make or material they work with: furniture design, web design, graphic design, interior design, textile design, information design, software design, fashion design. I have often tried to introduce myself as just a “designer” but

²⁵ It’s even worse in Norwegian – “oppsummert forskning om effekten av medisinsk behandling”. And to make it more complicated, these reviews are also about non-medical interventions concerning policy problems, like the effect of interventions to get health personnel to settle and practice in rural areas.

²⁶ The Cochrane Institute, by the way, is the wrong term. It’s called the Cochrane Collaboration (“Cochrane-samarbeidet” in Norwegian).

people invariably ask me to qualify this statement: “designer of what?” If you are designing objects that are invisible from sight, both for the general public and perhaps even for a large part of your intended target group, for which there are exist no readily available linguistic categories, and which are made of “materials without qualities”[212] (digital material) how do you explain to an architect (or anybody else for that matter) what kind of designer you are?

When you can't hang your design (or a picture of it) on the wall

There is another aspect of working with the type of projects I have been focusing on the past few years that make them difficult to explain, perhaps especially to an architect. Even if you dug up one of the hidden-from-sight products I helped design and put it on the table (or screen), it would be hard for anyone (even with visual training) to see any obvious traces of the work of a professional designer.

Let us use Summary of Findings Tables as an example. The final paper prototype of this table is not a pretty object - it is crammed, constructed in Word using an ugly font (arial narrow), full of too much information. It is not elegant or inviting, merely better than the original version of the table that we started with. Word managed to do weird things with the cell background colors and the strokes in the table that didn't correspond to my design intentions and that we weren't able to correct. The table itself is not creatively innovative in any manner, but a static, non-interactive, typographical representation of statistical data, designed within the very finite limitations of an antiquated publishing system. One would be hard pressed to discover much redeeming visual aesthetic quality in this table prototype. At best, it exudes a visual order and structure that helps a reader get through the possibly daunting task of understanding its complex content. The web and pdf versions of these tables look even worse than the prototype, since the intended design did not translate smoothly into these two different technological systems. So the final results in these two different media are even more void of any aesthetic surface quality than our prototype, as well as - much more regrettably - having lost much of what makes them easily readable by people not familiar with their content. What they do retain are basic choices we made regarding the informational structure of the table (number of columns, order of the topics in both columns and rows), choices about how much and what part of the information to include (or discard), choices about which information should be left in the table or relegated to footnotes, choices concerning statistical formatting and its numerical and textual representation, choice of terms and phrases used for labeling columns, and to some degree the typographical choices employed to bring the important parts to the foreground. But some of the decisions about these

elements, such as which words to choose as column headings or what type of statistical representations should be chosen, are not normally included in the repertoire of decisions made by a (visual) designer – they are relegated to those responsible for content. As mentioned earlier, form and content are traditionally separated in graphic design. Designers are still largely evaluated by based on the surface qualities – the aesthetics - of the work they produce. The fact that “meaning” might fall into the cracks between separated form and content does not seem to have grabbed the attention of the design community in any kind of significant way²⁷.

Löwgren and Stolterman write about the challenges of working with digital material, which they call material without qualities.[212] They suggest that this kind of design must be evaluated in the context of a given situation, according to whether it meets users needs. In 2010 this seems so obvious; user needs have been firmly placed on the design practice agenda for many years. Jacob Nielsen placed them firmly in the center of web designing more than a decade ago with his emphasis and arguments for usability, demonstratively illustrating his standpoint by creating a home page stripped of conventional graphic aesthetics.[213] Krippendorff has made a powerful argument for focusing on the meaning of the artifact to the user rather than the form of the artifact, explaining that this began with a concern for product semantics that gained momentum in the 90’s.[15] Both Frascara and Papanek wrote in the late 80’s about how excessive emphasis on aesthetics and “high-tech functionalism” has distracted designers from looking at their work in terms of changes produced in the audience.[214, 215] The move away from

²⁷ Perhaps some of the clearest examples of how text and visual design are inseparable from the users point of view can be found in web site menus. Menus indicate how the whole site is organized – what category of content can be found and where. Single words in a menu, such as “cat”, explain part of the “what”, representing a wealth of underlying content. These words must be chosen judiciously and in relation to each other, as they are not perceived in isolation but in context with each other (e.g. a menu with the terms “cat, dog, hamster, goldfish, turtle, canary” indicates a different kind of web site than a menu with the terms “cat, cougar, cheetah, lynx, tiger, lion, jaguar, leopard”). The visual formatting, on the other hand, provides a different kind of information – it tells us something about the pattern relationships between these words, such as “connected to”, “inside”, “outside” or “part of”. A menu including the terms “cat, food, health, mating, dog, hamster, goldfish, turtle, canary” would be confusing without some kind of visual structure or formatting to indicate that ‘food’, ‘health’, and ‘mating’ are sub-sections – a part of - “cat”. Alternatively, visual menu structure on its own without text would not tell us anything meaningful about the site’s content. The symbolic representation of the content (the words) and the visual representation (the design) in combination with each other form the input that makes menus meaningful

the object and turn towards the user's interpretation of the object is not a new idea.

But although the rhetoric in design literature has promoted user interpretation as a core principle for a long time, this much discussed change is still very slow in actually manifesting itself in design practice or in designers own identities of who they are and what they do, at least in the area I am familiar with – graphic design. For one thing, the move away from focus on the object (and it's qualities, particularly the visually aesthetic ones) and towards the meaning of the artifact for the user is not very apparent in the reward systems in designers' culture. During the same time period that the literature was beginning to challenge designers' concern with the aesthetics, designers' coffee table literature blossomed as never before, and designers began to become "famous" through showcasing their work in juried annuals such as Graphis or glossy magazines such as *Novum Gebrauchsgraphik*, *Idea*, *Eye*, etc. In Norway annual competitions such as *Visuelt*, *The Golden Pencil* (*Gullblyanten*) and *Norwegian Award for Design Excellence* (*Merke for God Design*) have been handing out awards for the best designs since the early 80's, as have their European, American and Asian counterparts. In these competitions, "invisible" work rarely receives an award.

I've experienced the power of these institutions during the course of my career as participator at all levels: submitter of work, winner (and loser) of awards, and jury member/jury chairman. In the capacity of jury chair, I became involved in revitalizing the criteria for the *Award for Design Excellence* some years ago, where we attempted to strengthen the user-advantage aspect of the evaluation process. But despite these efforts, the nature of such competitions cannot side step the fact that user advantages need to be *visible* to the jurists (and hence the audience who applaud them). This is very difficult to do when working with expert-based systems where small "invisible" changes – often made incrementally instead of in one big radical bang - can make enormous differences to the user. When there is no discernable "bang!", it is hard to explain to an outsider what all the fuss is about.

In the end, the most captivating work that wins awards in these competitions tends to be that which you can hang on a wall or print evocative pictures of in a catalog, work that entails visible changes that untrained eyes can clearly see and comprehend. Even when a jury is committed to rewarding work that might be more hidden, it is difficult to hand out public design awards to outcomes of design work that the public just can't see. The Norwegian Design Council has begun to address this issue, by launching a new

competition called “Design Effect”, where the measured and documented effect on the user is the basis for juried evaluation, rather than the jury members’ personal judgment about the aesthetic and functional design. This is a step in an interesting direction, though it presents methodological challenges both in documenting effects and proving that these were the results of design efforts alone.

While writing this text, a very typical example of this precise topic landed in my mailbox. It is an email from “Nora”, working in a design company in Berlin. This company is publishing a “Who’s who” book on graphic designers. The text reads:

*“Dear Mrs. Rosenbaum
My name is Nora. I work in the graphic design studio hesign by Jianping He. We are planning a Who’s Who book of graphic artists publishing. Our desire is to develop a comprehensive publication that provides an overview on the design situation today. This time we want to show graphic designers from around the world and in every age class in it. The target is a kind of graphic artists dictionary to create, so we hope to find about 1000 graphic artists. So you can get an overview of how it looks on the market and see the differences between the various countries and age. You are a graphic artist with appealing and beautiful works. We would like to invite you to present you in our book.”*

I receive similar invitations at least once a year, because I used to make much more visible work that won prizes and secured me membership in an international design organization that lends its mailing lists to publishing projects like these. Who reads these books, and what does that tell them about successful design? What do I tell Nora? How do I explain that I am now working in a less visible area of design where the challenges are very different than creating an eye-catching works that are “appealing and beautiful”, work that would be difficult to describe through pictures in her book?

The media seems for the most part to support Nora’s view of design. In the same issue of the newspaper where I found a clipping of the Cochrane Review, there was a full-page article reporting on the assignment of a new director for Norsk Form²⁸ Throughout this entire article, design and

²⁸ Established in 1992 at the initiative of the Norwegian Ministry of Culture. Norsk Form aims to improve people's quality of life and everyday situation through the use of design and

architecture were referred in the same sentence, as activities that would “improve people’s every day life quality”; Norsk Form would contribute to this goal by increasing “understanding of aesthetic quality in Norway”; the article was entitled “Will make Norway more beautiful”. Is it any wonder that as a designer I have an identity problem when I’m working on things that lack prominent aesthetic qualities? There may be a lot of design literature on the expanding role of designers, but there is little recognition of this tendency in designer’s own reward systems or in the public sphere. The architect is skeptical too; he’s still knitting his brow and shaking his head. He’s looking for the aesthetic qualities and innovation that is visible to him in my explanation of what I do, and he can’t find it.

Interaction design is one field of design where the meaning according to the user has gained actual foothold, where the explicit focus is not on designing “things” but on designing for human communication, action and experiences.[130] Although certainly all areas of design – such as graphic design, product design or architecture –share this human-centered dimension, the end-user perspective in interaction design is particularly strong because the outcomes of designing don’t have an independent material existence way that actual products do. Graphic designs, products or buildings can still be seen, touched, experienced more generally as objects of form outside of their particular use contexts (although much of the original meaning may be lost). Interaction designs, however, are constructed with invisible digital material, “stuff” that can’t be seen and valued except through use. You can’t hang an interaction design on a wall, but you can win a prize for it, and the field definitely cultivates its public heroes who create lots of invisible work – or write about others work – and through this make a mark for themselves (e.g. Jared Spool, Steve Krug, Peter Morville, Jakob Nielsen, Donald Norman). These people are possibly more visible on blogs and at conferences than in coffee table books with colored photographs. The related field of service design is delivering an even more abstract product than interaction design: services that are created at point of delivery rather than things that are produced and distributed. I asked Lavrans Løvlie, partner of Live|Work, an international service design consultancy with a branch in Oslo,[216] how they explained their very invisible product to people such as potential clients or fellow designers. He said that reference cases were his best selling points, and that often the clients coming to him did not have trouble understanding what he was doing, as they recognized their own business and organizational challenges in his casework. However, he said that he had some odd problems

architecture, by initiating and participating in chosen projects and through teaching, events, competitions and exhibitions.

recruiting new staff – it was particularly difficult to find graphic designers that understood his business concept and way of working, unfamiliar as their were on concentrating on problems of use rather than problems of visual expression. What would the architect make of service design? Would he see any kinds of connection to his own way of working and thinking, or would he think of it as foreign design territory?

Who did the designing?

The third thing that makes my contribution in this work difficult to explain to the architect is the distributed nature of the part of the work traditionally relegated to professional designers. To my knowledge, Andy Oxman (the senior researcher who initiated and led the Summary of Findings and SUPPORT projects) had never heard of user-centered or participatory design; yet this was his guiding principle when he designed the overall protocols for this work. In the Summary of Findings Table project, for instance, over 50 stakeholders contributed with input to our work. Andy is a trusted member of the stakeholder communities we queried and understood the politics and pragmatics of making proper space for the feedback of multiple experts who had a scientific interest in the way these tables were designed. 20 users also gave us feedback about what parts of the table gave them trouble. The core working group shared the tasks of collecting and analyzing data and making decisions about which of all these voices to pay most attention to. Compared to other projects I have been involved in, a large part of what is traditionally the designer role was widely distributed among non-designers. Who did the designing? It was definitely “we”, not “I”.

This leads to confusion of professional identity and unclear sense of personal contribution. With the working group at the core and all the people offering feedback at the periphery - where was my role *as designer* (rather than researcher) in this myriad of contributing test participants, opinionated scientists and skilful, user-dedicated core team members? What on earth was left for me when even the possibility of adding aesthetics surface qualities to these products was at a complete minimum?

Some might interrupt here and suggest I should have revolutionized this work by coming up with a brilliant angle no one had thought of before. I should have questioned the limitations of making a static table, and thought about the problem in new ways, as designers are known to do. I did consider this approach at the very beginning, but through conversations with the core team, I quickly understood that this was avoiding the task at hand. Basically, they were going to make tables and they needed help with that. Possibly there were other ways of communicating the information, and, if the output had

been solely web based, there would have certainly been many was of exploring interactive solutions beyond traditional static tables. However, these tables needed to be transferable from the Cochrane Review authoring system to both the web publishing system for the Cochrane Library and the system that generated pdf files of Cochrane Reviews, all of which functioned automatically with no manual intervention. Pdf's needed to be able to be printed out on paper in black and white and therefore had to perform without any interactivity. That meant that one obvious solution - hiding the more detailed information of individual table cells and allowing them to expand on demand – was not an option. Based on all of this, I accepted the very limiting parameters of the project and got to work making proposals of plain old tables.

Another interruption at this point in my explanation might involve the relatively recent emphasis in design research on the importance of emotions.[145] If visual aesthetics are not attended to, doesn't the emotional experience of the artifact suffer? I would argue that in these projects, the critical emotional reactions did not have to do with the aesthetic quality of the artifact. Users clearly felt ignorant because of unfamiliar language, complicated statistical representations, or lack of sufficient prior knowledge to understand document types and content. The strongest emotional reactions we observed were directly tied to these issues. It is unlikely that aesthetically pleasing design would counteract the emotion of feeling like an idiot.

What's left for the designer to contribute?

So, when the design task is stripped of two of its most salient components (adding visual or physical aesthetic quality and coming up with new, unexpected, “big bang” ideas), when dedication to end-user experience is already shared by other team members, and when scores of other people within multiple areas of expertise are also contributing input and suggestions of how to improve the product, what is left that can be called the designer's particular contribution?

I made a list based on my own experience in these projects:

- My participation helped establish (and sustain) a collective awareness that *the artifact matters*, that the numerous choices we made about a myriad of seemingly very small details and the relation of these details to the whole (the character of the thing)[217] can make critical differences to the users.
- My participation also drew attention to the *wider systems and context of use* (e.g. not just looking at use of a document, but also to first find the web site where it is published, and then to find the document on that site).

An obvious next step in these projects emerges from this perspective: to observe their use in natural settings, with an eye for how to both improve the artifacts themselves and the systems that help connect these artifacts to the people who might use them.

- I have contributed by successfully introducing designerly methods for acquiring knowledge to feed into the process of making new things - iterative rounds of solutions and qualitative feedback, including carrying out user tests - into a research community that has other kinds of inquiry traditions and ideas about what valid knowledge is.
- My designer skill set is well suited to working out solutions between conflicting interests (such as conflicting needs of users and stakeholders), and I thrive on these kinds of challenges. My experience has taught me that contradictions or conflicting interests (articulated through language) can often be successfully resolved when dealt with through visual (or physical) form. Working out solutions in sketches also leads to a better understanding of the problem, which can pave the way to redefining it so that conflicts are minimized or disappear altogether.
- Design experience is also about managing uncertainty, keeping the possibility for new solutions or new angles open for as long as necessary, steering comfortably towards unknown solutions. (Although I see this as a typical designerly trait, I think my working group was already relatively comfortable with uncertainty. Perhaps it was because they had lots of previous experience being involved in design projects, or perhaps it was just a reflection of their scientific background?)
- I've helped draw attention to the importance of listening and attending to end-users' perspectives.
- I have brought in other skills based on my design expertise:
 - Seeing: recognizing design possibilities in the feedback
 - Judgment: recognizing core issues in feedback that are relevant for the design, evaluating new sketches in light of these
 - Sketching: trial and error, working towards new solutions, seeking synthesis and unity
 - Thinking: using conceptual skills related to design, such as abstracting problems up a level or reframing problems

Designing takes place at the intersection of so many different dimensions of organizational activity that it can't help but be a strategic tool, if for no other reason than to visualize otherwise hidden consequences of choices. Where there is poor awareness of what the underlying strategic issues are (or might be), design processes can help crystallize consciousness about these. This is what makes design such a potentially powerful resource – its value can extend beyond the production or marketing sectors and far into the

boardroom. In the projects in this thesis, designing was a strategic tool to drive home the end-user perspective in the evaluation and improvement of individual artifacts. But how much responsibility I could claim for that varied from project to project. The earliest study I initiated and organized myself, in an effort to bring users' difficulties with the Cochrane Library to the attention of the Collaboration board and the library's publishers. However, in the other two projects it was not necessary for me to assume the sole role of advocate for the end-user since the working group was already convinced of the importance of this from start. It is possible that this work has influenced attitudes about and consideration of user perspectives, both at the individual and organizational level in the Cochrane Collaboration. Possibly some small ripple effect from the articles, casting a bit of light on the nature of the problem when translating information between disparate worlds of knowledge and practice, might make its way into the wider scientific circles involved in similar kinds of work. But that sort of thing is very hard to assess.

Experiences summed up

This sums up my own experiences while involved in these projects. I've described some of the doubts I've had about my designer identity, through examples of how design tends to be described and evaluated by others. I have also written about the problems of incrementally developing and evaluating products that remain invisible on several levels and possibly not understood as legitimate objects of design. In addition, I have tried to explain what role I took when the traditional roles - sole innovator and generator of ideas, maker of aesthetically pleasing things and advocate for the end user - were shared with non-designers or not priorities.

Maybe the problem I was having with the architect is that while the profession is changing - illustrated through my experiences above - the way we describe and define it, at least in public, has not changed too much. Possibly what is needed is to redraw the map a bit, so that it might better represent the expanding professional landscape.

DESIGN THEORY REFLECTION

What is designing and what are its outcomes?

Traditionally, design has been seen as an activity concerning the planning, making, production or improvement of *things*. "Things" cover a wide range of man-made works: from humans' early tools and images, crafted hand made goods, plans for industrial products and built environments, to highly sophisticated pieces of engineering like space craft. Cross writes: "Everything we have around us - our environments, clothes, furniture,

machines, communication systems, even much of our food – has been designed... Pragmatically, the most essential thing that any designer does is to provide for those who will make a new artifact, a description of what that artifact should be like”.[8]

Newer forms of design, such as software design, interaction design or service design, have challenged this concept. In these areas, a designed “thing” might rather be a *process or an action or a service* rather than a product. The outcome of a software design is not the code but the process inside a machine that the code instigates. Results of an interaction design might be seen as the action of the end user, such as the clicking on a mouse or choosing from a menu. Service design outcomes are defined in terms of end-user’s experiences mediated through multiple designed contact points, some of which are completely immaterial such as help offered over the telephone. Seeing designed “things” in isolation from their meanings, what they mean to users, becomes increasingly difficult when the object of design has no physical independent presence. Hence, the focus of the design outcome is necessarily altered: from creating or planning a physical/visual/symbolic artifact that has a (relatively) stable physical existence in the world, to the processes, behaviors, experiences or actions of machines, people or systems that - like live music - exist only in the moment, in addition to the meanings these designs produce.

Buchanan sensed that old categories of defining design through product or material type was too restrictive, and instead emphasized four “locations of invention”: signs and images, physical products, activities and services, and systems or environments.”[4] This freed the description of design from being bound up by defined by the tangible categories of its results, and rather saw it as it actually was: an indeterminate subject matter that could be framed, constrained or viewed through different conceptual *placements*.

There are some theorists that have pushed the understanding of what design is even further by suggesting new categories of legitimate outcomes (or new areas of invention). Henrik Gedenryd, in his description of designing as an example of cognition as interactive inquiry,[10] made an argument for design as having two types of outcomes. The first is a conventional outcome for productive purposes, similar to those described above (a “product”). A second type of outcome is for inquiry or cognitive purposes, the activity of “knowing” that is a direct result of testing and using an artifact such as a sketch or a prototype. He stated that knowing might be a bi-product of a design process, or it might be the most important result. One central kind of knowing that occurs in designing, according to Gedenryd, is a new and better

understanding of the problem. He, along with many others, argued that this is one of the most typical features of designing: that problem setting is not an analysis phase that can be separated and performed prior to design work, but that designers use iterative imagining, testing and evaluating of trial solutions (sketching) to gain knowledge about the nature of the problem.

The view that knowledge might also be an important outcome of a design project, is echoed by Zimmerman et al,[218] who make a case for many different kinds of knowing occurring as a result of a design process, for use by other designers in other projects or for researchers. Examples of these kinds of transferable knowledge are: *methods* used in a project (e.g. methods for successfully communicate observed user needs to others, such as technical, marketing or business staff), *raw data* from fieldwork (e.g. ethnographical video of air travelers passing through check in areas that can be useful for many different kinds of projects), *important gaps between users and manufacturers* (e.g. how families actually use mini vans contrasted to how manufacturers imagine that their mini vans are used), *characteristics of specific solutions* that could be generalized to similar types of products or use contexts, or *designers' reflections* at the end of a project concerning processes, users feedback or ideas for new concepts or design directions. Zimmerman and Forlizzi call this “research through design”.[219]

This thesis is an example in case of this point of view. The work described here has resulted not only in specific “products”, but also in several knowledge-based outcomes that were the direct result of making those new products. Some examples of these are: knowledge about the method (e.g. use of the honeycomb model and suggestions for improvement), knowledge about the nature of the gap between user and producer (e.g. the need of the scientists to produce detailed and precise accounts of the data contrasted with the need of the users for easily extractable gists), characteristics of the specific products that might be transferred to other projects (e.g. the concept of layering representations, successful use of the 1:3:25 model), characteristics of user feedback (e.g. problems users had recognizing document types), and general reflections about the nature of the work (this text I am writing now).

Another example of knowing that might occur in design projects is when participants or stakeholders gain new insights (e.g. about themselves, their product, their organization, or about others). In our projects, we did not explicitly look for these kinds of outcomes, but saw traces of them through comments from some of the participating researchers who found the process thought provoking. My own experience from earlier projects is that clients or

stakeholders can benefit in surprising ways when they are directly involved in or observe user tests²⁹. In future work this kind of outcome might be studied more explicitly.

Summing up, designing is not an activity that is only concerned with the making of things or processes, material or ethereal. It is also very much an active, interactive inquiry that can result in knowledge. Some of this knowledge has only local relevance to the particular project at hand, whereas other kinds can be relevant to a broader scope of projects and context. (The validity of this latter type of knowledge depends on the way it is collected, processed and documented.) Designers are not the only people who may acquire new knowledge through a design inquiry - stakeholders, clients or users participating in the work may also be learning.

Designing – reflective conversations with the situation

In the early days of design methods study, designing was described as a form of rational problem solving, characterized by a linear stage-by-stage set of methods that began with the problem and ended with the solution. However, empirical studies of what designers actually did documented quite different behavior. Donald Schön studied the practice of architects, as well as several other professions (e.g. psychotherapists, engineers, planners, managers) and found similarities in all of these fields that contrasted with what he referred to as ‘technical rationality’.[7]

“A designer makes things. Sometimes he makes the final product; more often, he makes a representation – a plan, program, or image – of an artifact to be constructed by others. He works in particular situations, uses particular materials, and employs a distinctive medium and language. Typically, his making process is complex. There are more variables - kinds of possible moves, norms, and interrelationships of these – than can be represented in a finite model. Because of this complexity, the designer’s moves tend, happily or unhappily, to produce consequences other than those intended. When this happens, the designer may take account of the intended changes he has made in the situation by forming new appreciations and understandings and by making new moves. He shapes the situation, in accordance with his initial appreciation of

²⁹ Observing user tests (for instance from a remote projection room) is a type of participation that has potential to change the way a client views his business and its offerings. It is conceivably a far more effective and believable experience than reading a report or an article about what other people have observed.

it, the situation “talks back”, and he responds to the situation’s back-talk. In a good process of design, this conversation with the situation is reflective. In answer to the situation’s back-talk, the designer reflects-in-action on the construction of the problem, the strategies of action, or the model of the phenomena, which have been implicit in his moves.

This quote from “The Reflective Practitioner” captures many of the elements that have come to be considered typical characteristics of designerly ways of working: the focus on making things, problem complexity, many variables, trial and error strategy, tight link between problem solution (making new moves) and problem redefinition (new appreciations and understandings of the problem), reflection-in-action, conversation with the situation, the situation’s back-talk.[220] Schön’s description also resonates deeply with my own previously unarticulated experience of practicing design. Each of these issues could be discussed in depth, but I will concentrate on one topic that has most relevance here.

Conversation with the situation

In one of Schön’s chapters, a professor of architecture is helping a student who is “stuck” – her attempts to solve a studio assignment are not leading anywhere. His help, which he verbalizes so that she can follow his train of thought, entails questioning a basic assumption the student has about the nature of problem, reframing her approach based on a new guiding principle and testing this idea out through a series of small sketches to see if it will work. His conversation with the student is a wonderful audible version of the internal conversation he was having with the situation at hand – in this case a “screwy site”, the student’s assignment to design a school building, and paper sketches drawn in pencil. The chapter title is “*Design as a Reflective Conversation with the Situation*”. However, in the same chapter Schön has formulated this concept slightly differently: “I shall consider designing as a *conversation with the materials of a situation*”, (my italics). He also later wrote an article entitled: “*Design as reflective conversation with the materials of a design situation*”.[221] He is often quoted in the literature - possibly incorrectly - with an abbreviated version of these last two phrases: “*a conversation with the materials*”.³⁰

³⁰ One example reference is: The Design Challenge, Pelle Ehn & Lone Malmberg, Scandinavian Journal of Information Systems, 1998. But many similar examples can be found by searching for: “*conversation with the materials*’ Schön”, on Google Scholar.

Conversations with materials - alternating forms of sketching using things like paper, computer, models, or experience prototypes - and the reflection over the speak-back from these materials is arguably the most central and characteristic feature of designing.[8-10] Bratteteig refers to a similar dialog between “materials and ideas” and shows how “ideas” need not only come from a designer in participatory processes; they may also come from other participants when they learn a bit about the materials.[132] Nonetheless, conversations with materials are what designers are particularly good at and trained to do: engaging themselves with the “stuff” in front of them through reflective action-in-practice.

But when Schön’s original concept - “conversation with the situation” - is recast as a “conversation with materials”, something both subtle and critical is lost. The latter is but a sub-set of the former. Possibly we don’t think too much about it, because Schön’s example took place in a school setting where the real life situation (the “wild”) was not visibly present, placing conversations with materials very much in the foreground. In traditional schoolwork, many kinds of design-relevant conversations are missing, due to the artificiality of the environment, rather like a laboratory. Perhaps this is a good thing, as students need to start by mastering the central task of intentionally manipulating and listening to the material, and build skills of reflecting during in this process.

In real life, however, whole symphonies of conversations take place during designing, not just those involving materials. Some are real dialogs that take place between two people, such as client and designer, or between designers and users, or between architects and public planners. Others are internal conversations of the type Schön documented, but they may be with other types of things than materials. For instance a designer may hold an inner dialog with an imaginary design community and their opinions of the work he is doing - this can be a powerful conversation. Even more abstract kinds of “conversations” can take place, such as speak-back from history that converses with a sketch or an idea. The 1992 Winter Olympic pictograms that I designed (inspired by hieroglyphics)[222] are an example of an dialog between a design concept and Norway’s cultural history. This dialog seems to be clearly audible to any Norwegian who sees these symbols.

Real life design processes are not just conversations with materials, because they typically do not take place in isolation. Like Gedenryd’s argument that cognition is not exclusively intra-mental (inside the mind), designing is not exclusively a process that takes places inside the mind of the designer, or even inside the walls of a design studio. Although conversations with

materials are usually very central, it is not just about that. Designing involves lots of reflective conversations, real and abstract, external and internal, between many parts of the situation – clients, users, stakeholders, producers, materials, technologies, systems, “the environment”, ideas, histories, competing products, media, colleagues, regulating bodies ... the list is potentially very long. Some important “conversations” may not involve the designer at all, such as when clients observe users directly and gain new knowledge about their business, or when participators engage with materials and get new ideas that contribute to the design. In complex projects, it is necessary to open up for a wide set of conversations simply because one person can no longer have all the know-how needed to create a good solution based solely on his own expertise. Many conversations – real or imagined – do not just happen on their own but need to be intentionally initiated, such as dialogs with groups of users with special needs or with “the environment”, based on ethical positions that those particular conversations need to be brought forward. Different design processes involve very different sets of conversations. An important part of the design job is to initiate appropriate conversations, ask questions and... listen.

What does a designer do with all these dialogs?

I hope by now that the notion that all designers are typically solo artists that create solution through isolated conversations with their materials is weakened. Though this might be a fitting description for some areas of design more resembling craft or art, many – arguably most – projects in this century are far too complex for designers to work out solutions without engaging in multiple types of conversations. However, the problem remains of dealing with all of this input. How does a designer synthesize the cacophony of voices singing out from both external and internal dialog with people, materials, systems, ideas, cultural history etc, and bring these to bear on the problem at hand? Designing is not just about listening to everything at equal volume and then trying to cater to all registered needs or notions. How are these conversations attended to, how are they used in a design process? Or, to frame the question in a way related to the work in this thesis, what happens when the user tests have been held, the stakeholder feedback collected, analysis of the problems carried out, and potential solutions discussed?

Judging and making

Ultimately selections are necessary. Not all conversations are equally relevant, useful or important. Deciding on which conversations or which parts of a conversation are most significant involves reflective *judgments* about what to listen to, where to focus attention, and what doesn't matter so

much. Sometimes a keenly focused attention in the right spot can short circuit directly to a new idea. Alternatively it could lead to important understandings about the core issues of the situation, to be drawn upon as the work progresses. But as discussed earlier, reflection (including judgment) and making are not easily separable activities. Design is not characterized by pure isolated phases of analysis followed by pure phases of creative activity. The processes of judging and making are interrelated.

An example of judging and making – testing constraints

One of the first rounds of judging that occur in a design project is a design conversation with constraints. An example of this is my description earlier of trying to think beyond the limitations of making a static table at the beginning of the SoF Table project. This is a well-known form of typical design behavior, where designers are known to treat even simple tasks as “ill-defined” problems by challenging the givens in a brief.[8] I don’t think designers do this because they particularly love complex problems, as Stolterman alludes to in his article on the nature of design.[39] Based on my experience, this is a way of locating the true boundaries of the project within which reframing can possibly occur. When I tried to think beyond the limits of making a table, I quickly ran into reasons why this would not work. This kind of inquiry – challenging givens – defines the parameters of the work within which the solution will lie.

Buchanan has said that design has no real subject matter, that a client’s brief “doesn’t present a definition of the subject to be designed, but a problem and a set of issues to be considered in resolving that problem.”[4] If the brief does go so far as to begin to define the subject, that is just the client’s attempt to move from problem to design. Designers learn to politely ignore these client attempts, to not take them literally. This is because one learns that this is merely one of many possible ways of framing a problematic reality – the client’s way – and that it is likely springing out of conventional thinking from earlier solutions. The way to new solutions is often to move backwards and consider reframing of the problem.

But in order to do any reframing, a designer first needs to know where the nonnegotiable boundaries of the project lie. Boundaries, or constraints, are built into a brief often in the form of limitations or requirements. Gedenryd described ‘sources of constraints’[10] (a concept from Lawson) as rigid (i.e. legislative constraints), somewhat flexible (client-imposed constraints) or completely flexible (designer-imposed constraints). In the beginning of a project, a designer needs to understand the constraints embedded in the brief – are they rigid (from a source far removed from the project’s control) or are

they local? If they are local, they can often be changed. Sometimes clients predetermined constraints are not well thought through (or too well thought through), therefore most designers start a project by questioning the framing presented by the brief. One way to check the flexibility of constraints is to envision a new idea that defies them, as good ideas are excellent arguments to convince a client to alter his predetermined constraints.³¹ When a designer has determined where the rigid constraints lie – the ones with distant sources that for all practical purposes cannot be negotiated – and understands which of the client’s constraints are most inflexible, the space remaining inside these boundaries is where the solution will lie.

Sometimes that space is much larger, or covers a somewhat different conceptual territory, than the space defined by the initial brief. This is one of the ways “making” (creating ideas that defy constraints) and “judging” (which constraints are nonnegotiable) interact in design. In the case of the SoF Tables project, the constraints created a rather narrow solution space rather than a broader one. Tables needed to be generated both as web presentation but also pdf for print. This meant parking the idea of making something radically new or possibly interactive, and finding static, printable solutions that could work within these boundaries.

Who is doing the making and who is doing the judging?

When it comes to making and judging of design solutions, who is carrying out this activity and based on what knowledge, experience or values? These are the elements that are increasingly up for grabs in newer forms of design work. When I began my design career at the beginning of the 80’s, nearly all projects I worked on had a clear division of these roles. As designer, I carried out mini-cycles of making and judging that resulted in ideas and sketches, which in turn were ultimately judged by the client, based more or less on personal preference (or sometimes through market testing, for instance of

³¹ Norwegian designer/illustrator Trond Nordahl provided me with a good example of this kind of deviation from an original brief: a client came to him with a very specific brief for the cover of a telephone catalog. They wanted him literally to “draw an illustration of a man and a woman sitting on a couch in their living room. Each of them should have a child on their lap, one boy and one girl. All four should be browsing through one telephone catalog (that was also on their lap) and at the same time be eating cookies”. Nordahl’s idea (and final solution) for this assignment was a dramatic reframing of this brief: a 5-meter tall kinetic abstract sculpture that could be associated with a human figure. A kind of catalog shape was mounted at the end of one of the “arms”. This sculpture, created in steel, was installed outside the company headquarters and photographed for use on the phone catalog cover that year.

packaging designs). Users or other stakeholders were not often visibly present except in such test situations.

Today such a division of labor is not so clear-cut, as many more types of people and conversations are brought to bear on both making and judging processes all through the design work. The work in this thesis is an example, where large groups of stakeholders were consulted several times. Part of the reason for this development is that projects are more complex. Too much separation of judging and making does not lead to successful results in complex projects, and yet no single group has enough knowledge to make good judgments about all aspects of the work or the situation. The movement towards participatory design during the last few decades has also drawn some attention to the political question of who is doing the judging, i.e. who is in control and which other voices should have a right to be heard. They can successfully contribute with both ideas and other aspects of the “making” phases, possibly due to their tacit knowledge about the situation at hand that is difficult to make entirely explicit.[223, 224] In these kinds of cases, the designers’ role might be more about initiating and guiding others’ participation, helping identify and cultivate promising ideas or directions, translating input into final working design solutions, attending to the unity or the character of thing to avoid fragmentary design-by-committee results, as well as keeping an eye on the fit into a larger technological or use environments and systems.

Involving non-professional participants (such as users) beyond the level of informing the project can be very time- and resource consuming, and not all projects are suited to this kind of organization. In addition, not all relevant conversations involve the concerns or perspectives of actual people, like users or stakeholders, who can actively participate in the making processes. Some conversations involve dialogs with more abstract entities, such as “the environment” or the current zeitgeist. Given this, how are the many relevant dialogs brought to bear on both the judging and making in designing?

Active vision

Recent research in the neurology of perception casts some interesting light on this question. In the book “Visual Thinking for Design”, Colin Ware presents a new way of understanding visual perception called “active vision”. [174] Active vision means that seeing is a very dynamic process - not only is the retina sending information for processing to the brain (bottom-up information) but the brain is in parallel sending information in the other direction to steer what is being seen to begin with (top-down information).

Ware describes several concepts that are central to active vision: cognitive threads, visual queries and perceptual critique.

A cognitive thread is defined by intention, the “what” that you are doing at that moment. It is the linked set of concepts forming a narrative in your mind, shaped by the individual task of the moment – searching for oranges in a grocery store or examining a patient for signs of disease. Visual queries are the searching involved in seeing, where eye movement is steered by the cognitive thread, looking for particular information to serve a particular cognitive task. Together the cognitive thread and the visual queries make up the top-down information that affects what you see, or focus attention on. Ware claims that designers typically test their designs by visual queries – making marks on a paper, then evaluating rapidly whether those marks are appropriate to the task in mind. He calls this a “perceptual critique” of the sketch. Researchers such as Lawson, Schön and Cross have already documented this as typical design activity.[7-9] What Ware adds to these descriptions are neurological explanations about what drives this kind of behavior. He also suggests that part of the expertise of a designer is about developing advanced skill in this area of seeing and critiquing - that designers become increasingly adept at *setting their own visual queries aside* for the moment and *imitating the sets of visual queries that other people might have*. Doing so depends on an understanding of others’ cognitive threads – including understanding their cognitive tasks and what motivates (or would motivate) attention to this particular artifact. Ware also explains how the combination of sketching (physically producing external imagery) and imagining new possible moves (mentally producing internal imagery) interacts so rapidly and effectively with this process of visual query.

Evaluating sketches and new ideas

Designing is partially characterized by processes of making and judging. When representatives of the multiple conversations involved in design projects are not directly available to participate in the judging and the making, designers can instead create sets of visual queries and cognitive threads of others to use in this process. These can then be used while sketching, to formulate multiple evaluations – or different ways of seeing the design – to carry out many perceptual critiques from different points of view. Ware writes that designers do this automatically, comparing it to a kind of informal internal cognitive walkthrough. This doesn’t mean the design needn’t be evaluated formally, such as through user testing, but that in between these formal sessions there is an enormous amount of mini-evaluation going on (e.g. during the sketching of new ideas). It is during this phase that an understanding of others’ visual queries is of great value. It also

doesn't mean that designer has no visual queries of her own, just that she becomes increasingly expert at being able to supplement them by setting them momentarily aside and imitating the queries of others.

Such queries might not just be carried out only to evaluate the suitability of a design for a typical user; they might also be used to pass a rapid judgment on possible impact on surroundings, feasibility in production, adherence to key issues in the original brief, etc. By creating a mental store room of large sets of potential visual queries (through conversations with the situation), a designer can move rapidly between them during the sketching activity or when evaluating new ideas. A design involving invisible materials might be evaluated with a similar technique, but one that involves experiential queries rather than merely visual ones. Finally, understanding which queries might deserve most attention is also a skill that increases with expertise, as does the confidence needed to be able to open up to the queries and concerns far removed from one's own.

This accurately describes the activity I carried out after user tests were held, stakeholder feedback collected, analysis of the problems carried out, and potential solutions discussed in working group meetings. After all of this activity, I needed to close the door and sort through all of this input on my own, carrying out phases of rapid sketching, or visual queries. This is the part of the work that most resembles what I used to call designing. What strikes me as most curious is that in a complex project like this, it makes up such a little part of the whole work.

Placements and personas

Ware's description of how designers might imitate others cognitive threads, visual queries and perceptual inquiries bears a strong likeness to two other concepts from design literature that support flexible points of view: placements and personas.

As discussed earlier, Buchanan proposed that the concept of placement was central to the nature of design.[4] Placement, as I understand his concept, is about shifting conceptual point of view so as to look at the problem situation from new locations, new places of thought. For instance, placement might involve shifting to the perspective of another person, another location of invention (e.g. from the single artifact to the whole system), another position in time (e.g. from production phase to whole life cycle), or any move that entails a reframing.

Personas are hypothetical archetypes of real users, precise descriptions of imaginary people and what they wish to accomplish.[225] The strength of personas is that they can help externalize and make explicit mental states or mental activity of users (such as visual queries). They also allow a whole team to engage over time with a stable and shared concept of a user.[226] Designers' thinking about users can be externalized and carried out together with others. The potential weakness of personas is that they are often a product of the imagination of the designer or project team, rather than based on actual input from the real world. My experience with bringing in actual voices from the outside is that these often contribute through viewpoints that nobody involved with the project could ever have imagined or anticipated. An example of this from these studies is our finding that correct understanding of the evidence was a potential problem for professional users.

Seeing design possibilities

A logical extension of the concept of active vision is a better understanding of how a designer brings her own set of visual queries to the table when engaging with the situation – for instance while discussing the brief with the client, observing user tests and reading results of stakeholder feedback. In these parts of the work, a designer is not just listening with a blank mind; she is using active vision, searching for information to suit her own cognitive task at hand of finding a solution to the design problem.

How is a designer's information seeking in these settings any different from any other participant from the outside of the situation? In all of the projects in this thesis, social anthropologist Claire Glenton co-participated in both data collection and data analysis of the user testing and stakeholder feedback. I discussed this issue with Claire, wondering whether she felt that what I – the designer – observed test session and searched the data differently than she did. She replied immediately: “Solutions. You were seeing solutions in the data,” she said. “I was just seeing patterns”.

Designers see solutions probably because we are *looking for* solutions; that is our main cognitive thread. We are not primarily looking for patterns of power (unless that has relevance to the design of the artifact); we are not looking for linguistic patterns (unless these have relevance to the design of the artifact); we are not looking for social network patterns (unless these have relevance to the design of the artifact). As designers, our main cognitive thread is *how to make or improve the artificial world – the artifact, the process, the system or environment* – looking for information that can tell us more about what things are, what they means, what they could be. Through experience we become skilled at recognizing solution-like signals, while observing or interacting

with users or in other relevant conversations. For designers, solutions are like oranges in the grocery store – they send beeping signals to the designer brain, steering our attention to that part of the situation.

Perhaps this is why I have never understood how design processes could be effectively fueled by condensed data that is processed and analyzed by somebody else. I think useful observations leading to new ideas emerge most readily through emersion in unedited data, engaging first hand in the reflective conversations of the situation. This is because *everyone* practices active vision, and one person's edited and processed version of reality might be obscuring or eliminating the details that trigger another person's valuable insights and new ideas. This is likely why real users might be well situated to seeing design possibilities, provided they understand what the materials can do and can be encouraged let their imagination run freely beyond that which already exists. They are already completely immersed in the situation at hand, and have more access to richer information about some parts of it than anyone coming in from the outside.

Summing up, we might say that designing involves potentially many kinds of reflective conversations with the situation, not just with the materials. Part of designer's toolbox of skills is the ability to initiate and attend to these conversations (both real and abstract) and draw core issues emerging from them into the designing process. One way of doing this is for designers to actively search for solutions while directly engaging in these dialogs. Another alternative is to invite actual conversational partners (e.g. stakeholders, users) to contribute to the making and judging parts of the design process. Yet another path maximizes on the concept of active vision: designers use understanding gained through the many kinds conversations to build a repertoire of cognitive threads and visual queries *representing core issues or concerns of others*. These are used in the micro-processes of sketching and perceptual critique that are so characteristic of design work.

Two approaches to design?

The traditional view of the designer is an individual who is conversing in a limited space with his material. The view I am describing differs, in part because it is based on a much wider set of conversations. Gedenryd compared two models of cognition with each other (the old "inside-the-mind" model and the new interactive inquiry model where the outside world plays a much larger role) and suggested they differed from each other in the same way as writing and conversation differ.[10] Possibly our understanding of these two views of designing can benefit from a similar comparison.

Writing is something someone usually does alone. In writing for a particular audience, an author must try to predict what the audience might think, but without their direct feedback. Writing is therefore based on guesswork. (Maybe like a design competition, a stab in the dark without the advantage of conversing with the client and letting their speak-back feed into the process.) A conversation, however involves more than one person. Also you don't have to imagine feedback; you can ask directly for it, and explore what it means. Conversations are exploratory, collaborative process, where many voices can contribute; many people can both listen, and speak and ask questions, learning from each other.

The "old" model of designing, involving one or few people in a limited dialog between themselves and their material is like writing. It produces a stable piece of work, "things" often with a clear independent material existence and definable authorship. There is much room for personal vision and expression, but this work form is limited as to how complex a problem it can address with a reasonable hope of success. The "new" model does not by any means exclude this mode of work, but *expands* the concept of designing so that it also includes work forms more similar to conversation than to writing. Compared to the output of writing, the output of a conversation is harder draw a line around. It is more malleable, more diversified, it is performed and "owned" by more people, and is more a work in progress than a finished piece. This way of working might involve a great number of real and imagined dialogs including many more participating persons, ideas and interests. The result may not be a visible "big bang" but more subtle output that is harder to freeze in time and claim ownership to. Wikipedia says: "Conversations are the ideal form of communication in some respects, since they allow people with different views on a topic to learn from each other." The designer role is not necessarily diminished, just changed – less visible, more distributed, attending to a broader set of concerns. The designer also needs to develop sensitivity to the value of other outputs of the inquiry than merely the visual or technical solutions.

Greenhalgh et al. carried out a meta-narrative review of the literature on electronic patient record systems, in order to map, interpret and critique the research coming from multiple traditions.[202] In this work, they identified five different underlying sets of philosophical assumptions in EPR research: Positivist, Interpretivist, Critical, Recursive (Integrative), and Design. The Design category was divided into two columns and is repeated below (see Table 4 below).

Table 4. Excerpt from table entitled “Philosophical Basis of Different Approaches to EPR Research”[202]

ASSUMPTIONS AND VALUES	DESIGN	
	Conventional <i>(Roots in positivism)</i>	Participatory <i>(Roots in Interpretivism)</i>
Ontology <i>(Assumptions about the nature of reality)</i>	Multiple, contextually situated, alternative world states, socio-technically enabled	Co-creation of socially useful artifact through negotiation and sense making.
Epistemology <i>(Assumptions about the nature of knowledge)</i>	Knowledge emerges through making and testing-in-reality. Design is objectively constrained construction within a context. Meaning is revealed through iterative circumspection.	Knowledge is subjective and value laden and emerges through making, which is a social process that requires shared vision and understanding.
Role and reflexivity <i>(Assumptions about the role of the researcher)</i>	Researcher is creative, precise, technically adept, a seeker of elegance and usefulness in an artifact.	Researcher is a team member and co-producer of a useful artifact.
Methodology <i>(Assumptions about what methods will generate “best evidence”)</i>	Developmental, with a focus on the technical. Measures artifactual impacts on the composite system.	Developmental, with a focus on the social. Measures shared vision, values, and collaborative outputs.
Axiology <i>(What is of value)</i>	Control, creation, progress, improvement, understanding	Fitness for purpose, ownership, engagement, dialogue.

Although this table reflects perspectives found in design research (not practice), it was looking at studies with an emphasis on implementation or use of EPR systems. Greenhalgh commented in an email: “We weren’t really very systematic about the design literature. Indeed we kept trying to exclude papers on design but they kept creeping back in!”[227] Systematic or not, this review echoes the idea that there may be two different approaches to design, a traditional view and an emerging view that I am expressing here.

And that design knowledge, not just design practice, might be increasingly “creeping in”.

The emergence of a new approach to designing has its parallels in other fields, for instance medical practice. As I described in the introduction, the paternalistic model of “doctor knows best” is challenged by a new model of “shared decision making”, an integrative approach that has more room for patients’ own values and preferences as well as input from research. What has been a space largely controlled by the medical professionals’ individual expertise is slowly opening up, not to exclude this expertise or render it obsolete, but to add additional conversations to the locations of health care. This move is not tension-free, especially for those who are schooled and settled in old ways. But a new way of practicing medicine is slowly evolving, probably driven in part by the widespread availability of health related information online and increasingly informed patients. This new doctor role includes improved communication skills and the ability to seek out and synthesize information from a broader set of sources than just his personal expertise to bear on the task at hand, through differing models and degrees of co-operation with the patient. Seen in the light of the work I have been engaging in the past years, I think that these changes in the medical profession resemble more and more the current transformation of my own field.

At this point I return once more to the architect. I think if I had the conversation with him again, I would be able to better explain myself this time. I would tell him I have been exploring a new way of being a designer, one that isn’t so intra-mental but more interactive. It involves the feedback and participation of many people and much reflective conversation with situation far beyond merely a dialog with materials. I would tell him that my role was not necessarily to dominate the arena of ideas or solutions, but to open up the process to including a broader range of relevant conversations, and to attend to the speak-back from these people, constraints, systems, materials, ideas, research studies, histories, perspectives etc. so that these voices made their way into the design work. Sometimes this meant engaging directly in these dialogs while searching for ideas. Sometimes it meant inviting others (non-designers) into either the making or judging processes of design. Alternatively it might mean improving my own skill of identifying and imitating other points of view when carrying out internal reflective conversations with the material. The outputs of this work were not only the “things” we made, but also knowledge we gained underway. Some of that knowledge was primarily of local value and fed back into the artifacts. Some involved methods, findings, placements or visions with potential value for

other situations, which is why we organized the work as research and published it. Some of the knowledge also led me to understand what I did in a different way than when I made more aesthetically oriented work that that I could point to and hang on a wall, or that I could claim individual authorship for. Not better or worse, just different. And I have a feeling that he would understand what I was talking about. This expanded view of designing and design outcomes is not likely unfamiliar to most people in design professions, even those who produce highly expressive personal work. It's just the way we talk about it that hasn't quite caught up yet.

Summary and conclusion

The work in this thesis has resulted in output on several levels: evaluation and design of specific artifacts, knowledge about the characteristics of users' experiences and user-producer gaps, subsequent consequences of knowledge for design, and reflections about the nature of designing based on my own experiences participating in this work. Below is a summary of answers to the research questions and underlying design inquiry.

RESEARCH QUESTION 1

How can we improve health professionals' user experience of the Cochrane Library web site?

- a) How do health professionals experience the Cochrane Library?*
- b) What design and content related improvements of the site do these findings suggest?*

Summarized answer, question 1

In 2005 and 2006 when these studies were carried out, health professional participants had much difficulty locating both the Cochrane Library and its contents. Non-native English speakers were at an extra disadvantage when retrieving relevant documents despite high levels of English-language skills. Many participants displayed feelings of ineptitude, alienation and frustration. Some made serious mistakes in correctly distinguishing between different information types, for instance reviews, review protocols, and individual studies. Although most expressed a high regard for the site's credibility, some later displayed a mistrust of the independence of the information. Others were overconfident, thinking everything on *The Cochrane Library* site shared the same level of quality approval.

Health professionals' user experience of evidence-based online resources can be improved by applying the following principles:

- Follow existing usability heuristics and web usability guidelines, designing especially for findability through search engines, as well as for speed of use particularly important to health professionals.
- If resources are limited, focus on improving simple (non-advanced) search functionality, including technology that will help non-native English speakers.
- Drop “researcher” language and jargon avoid alienating health professionals.
- Don’t assume users possess good mental models of evidence hierarchies. Make document types evident where possible – through information architecture, labeling, and search results design.
- Clearly mark the difference between quality-approved content and not quality-approved content.
- Ownership and authoring must be clear at all levels of the site for supporting and maintaining credibility.

For a more detailed answer to question 1, see:

- The Cochrane Library article (first article, next section)
- The discussion section of this overview under the subheading *Searching and finding evidence, Understanding evidence, Appraising evidence, Other concepts of usefulness – looking at the non-instrumental uses of evidence, Knowledge gaps, Different uses lead to different experiences, Layered boundary objects*

The two evaluation reports can be useful as case study material but are prohibitively long to be included here (23 and 24 pages). They are available upon request to the author. (sarah@rosenbaum.no).

RESEARCH QUESTION 2

How can we improve health professionals' user experience of Summary of Findings Tables for Cochrane Reviews, while at the same time securing stakeholder acceptance and publishing feasibility?

- a) How do health professionals experience the Summary of Findings Tables?*
- b) What characterizes the stakeholder feedback?*
- c) What design and content related improvements of the table do these findings suggest?*

Summarized answer, question 2

User test participants demonstrated unexpected comprehension problems, mainly confusion about what the different numbers referred to (class reference), which in some cases resulted in total misunderstandings about the meaning of the data. They also had difficulties with continuous outcomes, meaning of “no data available”/empty cells, and unfamiliarity with language and concepts such as risk column headings and abbreviations. Many found the table to be complex, though some expected a learning curve for this kind of information. Stakeholders, on the other hand, were primarily concerned about precision in the data representation in addition to production feasibility. Resolving the tension between achieving stakeholders' concern for table precision and users' need for table simplicity was the main focus of group responsible for table design in the SoF project.

It is clear that table design (including details about numerical representation, language use and visual formatting) has a strong influence on users' perceptions, especially regarding correct understanding. However, in order to design a table that is tailored for users, stakeholders' concerns about precision must be addressed. Fuzzy trace theory can be a helpful conceptual tool for resolving what is traditionally thought of as a tug-of-war between complexity and simplicity, by reframing the problem along a verbatim – gist continuum. Other design guidelines for similar tables:

- Avoid class confusion:
 - Use same class reference, especially in number sets that are to be compared

- Support correct class interpretation by adding class labels (e.g. “studies”)
- Describe scales for continuous outcomes in close proximity to the results
- Avoid unfamiliar abbreviations wherever possible, even if they have been introduced in the text.
- Explain empty cells to make uncertainty or lack of data explicit
- Help the reader quickly form the correct gist of the numbers:
 - Use text cues where applicable
 - Align type to make comparison of numbers easier
 - Layer the information visually so that the most important parts “pop out” at the reader

To make tables more useful for clinicians, include:

- Information about the population and setting
- Inclusion criteria for the high/low risk populations
- Description of the intervention

For a more detailed answer to question 2, see:

- The Summary of Findings development article (second article, next section)
- The discussion section of this overview under the subheadings: *Understanding evidence, Appraising evidence, Usefulness of evidence, and “Gaps” between users and producers and consequences for design.*

An example of the final version of our Summary of Findings Table is included in the article as well as in the Appendix.

RESEARCH QUESTION 3

What is the effect of including a Summary of Findings Table in a Cochrane Review on user satisfaction, time to find key results, and correct understanding of the main results of a Cochrane Review, compared to a Cochrane Review with no table?

Summarized answer, question 3

In the first RCT, participants with the table were more likely to “Agree” or “Strongly agree” that it was easy to find results for important outcomes than participants without the table. In the second RCT, participants with the table were more likely to correctly answer two questions regarding results than participants without the table. Participants with the table spent an average of 90 seconds to find key information compared to four minutes for participants without the table ($p=0.002$).

In two small trials we found that inclusion of a SoF Table in a review improves understanding and rapid retrieval of key findings compared to reviews with no table. This study, besides demonstrating that tables “work”, strengthens the position that representation format of evidence (the way it is designed) affects how it is used and understood.

For a more detailed answer to question 3, see:

- The Summary of Findings evaluation article (third article, next section)
- The discussion section of this overview under the subheadings: *Understanding evidence, Appraising evidence, Knowledge gaps, and Unstable boundary objects.*

RESEARCH QUESTION 4

How can we improve health policy makers' user experience of SUPPORT summaries of systematic reviews, while at the same time securing stakeholder acceptance and publishing feasibility?

- a) How do health policy makers experience SUPPORT summaries?*
- b) What design and content related improvements of the SUPPORT Summary template do these findings suggest?*

Summarized answer, question 4

16 out of 21 policymakers found the summary useful. In addition to the 'Key Messages', policymakers particularly valued the section on 'Relevance for LMIC' (low and middle income countries), despite the lack of directly relevant evidence in the systematic reviews that were summarized.

They responded positively to a graded entry format (short summary with key messages upfront), but some struggled with comprehension of text and numbers. The three issues that were the most challenging in redesigning the evidence summaries were policymakers': 1) poor conceptual understanding of what a systematic review was, 2) expectations of information not found in systematic reviews, and 3) desire for shorter, clearer summaries.

These findings suggested the following improvements: Including content about applicability that is specific to policy decisions in LMIC appears to improve the usefulness of evidence presentation, even when it is too general to inform a local decision. Adding boxes of 'information about the information' may also help readers to better understand the nature of information from a systematic review. Although it was difficult for us to further shorten texts without losing scientific credibility, text may be formatted to aid quick scanning. The graded entry format is a good way of presenting information to policy makers, who need to be able to read at a glance the main issues in a document. Text and tables should be as simplified as possible; abbreviations should be avoided.

For a more detailed answer to question 4, see:

- The SUPPORT Summary article (fourth article, next section)
- The discussion section of this overview under the subheadings: *E-health literacy, Understanding evidence, Appraising evidence, Usefulness of evidence, Other concepts of usefulness – looking at the non-instrumental uses of evidence, Knowledge gaps, and Layered boundary objects.*

An example of a SUPPORT Summary is included in the Appendix.

DESIGN RESEARCH INQUIRY

The overarching aim of this thesis is to explore how a design approach can help facilitate evidence-informed health care. This is done in two different ways: 1) through the design development of specific artifacts that lead to improved user experience of systematic reviews and 2) through design-relevant knowledge from that emerged while developing these artifacts. What methods might we use; what results can we achieve; how might a design perspective contribute to a better understanding of the problems involved and – conversely – what might we learn about the nature of design while carrying out these inquiries?

Summarized answer, design research inquiry

Design features that were found to improve user experience during iterative development of three specific types of artifacts are already discussed in depth, both above and in the articles. Design effect is presented in Study 3. Here we demonstrated, albeit in quite small trials, that the representational format of evidence (the way it is designed) does have a measurable effect how it is found, understood and experienced by health care professionals. In study 2, the design methods (gathering explicit feedback from both users and stakeholders) led, among other things, to insights about the nature “evidence-practice gap” as it was manifested between these two groups (simplicity vs precision) and provided a foundation for potential reframing of this problem (along a gist – verbatim continuum).

The use of the honeycomb framework across several studies made it possible to evaluate its suitability for this work, resulting in a proposal for a revised honeycomb for developing/evaluating the design of evidence-based information artifacts. I took out the value facets, and added two new facets: understandability and affiliation. I also proposed adding the dimension of time, borrowing on Krippendorff’s theory of artifacts in use. Comparison of this framework to other existing ones suggests that different user experience frameworks may be necessary for different classes of artifacts or user groups. A more detailed discussion of this topic is in the section: *Evaluation of the user experience framework*.

Experiences from the SoF project illustrate potential problems of transferring designs across technological contexts or through automated templates: the result may be flattened formatting (making gist extraction more difficult),

flattened content (undesirable homogenization) or poor fit. Translational problems are typical of general-to-specific moves. Despite these technological difficulties, design as an approach is a suitable method for translation work (i.e. from generalized knowledge emerging from research to specific users and use contexts), because it focuses on the specific. This is illustrated by all four studies in this thesis. A more detailed discussion of this topic is in the section: *Designing for cross-platform and automated template-based publishing*

Earlier I have seen design as an activity similar to writing – personal, potentially expressive, resulting in finished piece of work. Through my experiences in these projects I have come to understand design in a new way. Design is a conversation with the situation, extending far beyond the reflection-in-action involving materials. Designers engage in many reflective conversations with the situation, involving people, constraints, systems, materials, ideas, research studies, histories, perspectives etc. (In a school context, conversations with the material will dominate, but in real life contexts there will be many conversations to attend to.) The making and judging in design is increasingly shared with others, resulting in a new designer role. Among other things, a designer initiates a broad set of conversations, identifies core issues from these dialogs, engages in and/or guide others making and judging processes, produces or identifies good ideas in suitable materials, attends to content as well as form, develops and protects the character of the thing, keeps an eye on system perspectives. Design is particularly suitable to dealing with conflicting perspectives, as this is a driving force in finding solutions that fit. Not all design processes result in finished “things”; knowledge from design is also a legitimate output.

A more detailed discussion of this topic is in the section: *The end of the journey*.

References

1. Simon HA. *Sciences of the Artificial*. Third ed. Cambridge: The MIT Press; 1996.
2. Sless D. *Philosophy as design: a project for our times*. 2002 [cited 2010 July 24]; Available from: <http://communication.org.au/publications/principles—amp—philosophy/Philosophy-as-design—a-project-for-our-ties/47,30.html>.
3. Alexander C. *Notes on the synthesis of form*: Harvard University Press; 1964.
4. Buchanan R. *Wicked Problems in Design Thinking*. In: Margolin V, Buchanan R, editors. *The Idea of Design*. Cambridge Mass.: The MIT Press; 1996.
5. Chow R. *For user study. The implications of Design*. Braunschweig: Von der Hochschule für Bildende Künste Braunschweig; 2005.
6. Goel V, Pirolli P. The Structure of Design Problem Spaces. *Cognitive Science*. 1992;16:395-429.
7. Schön D. *The Reflective Practitioner*: Basic Books; 1983.
8. Cross N. *Designerly Ways of Knowing*. Basel: Birkhäuser Verlag AG; 2007.
9. Lawson B. *How designers think: the design process demystified*. third ed. Oxford UK: Architectural Press; 2005.
10. Gedenryd H. *How designers work. Making sense of authentic cognitive activities*. Lund: Lund University; 1998.
11. Hookway C. *Pragmatism*. 2010; Spring 2010; [Available from: <http://plato.stanford.edu/archives/spr2010/entries/pragmatism/>].
12. Dewey RE. *The philosophy of John Dewey*. The Hague: Mrtinus Nijhoff; 1977.
13. Field R. *Dewey, John*. 2005 [updated July 14, 2005; cited 2010 July 25, 2010]; Available from: <http://www.iep.utm.edu/dewey/>.
14. Nelson HG, Stolterman E. *The Design Way*. Engelwood Cliffs NJ: Educational Technology Publications; 2003.
15. Krippendorff K. *The Semantic Turn: a new foundation for design*: Taylor & Francis Group; 2006.
16. Archer B. The nature of research. *Co-Design*. 1995(January):5-13.
17. Frayling C. *Research in Art and Design*. London: Royal College of Art 1993.
18. Sevaldson B. *Discussions & Movements in Design Research*. FORMakademisk. 2010;3(1):8-35.

19. Abras C, Maloney-Krichmar D, Preece J. User-centered design. 2004; (unpublished draft of book chapter)]. Available from: http://www.google.no/url?sa=t&source=web&cd=12&ved=0CF0QFjAL&url=http%3A%2F%2Fwww.cs.uta.fi%2Ftjsum%2Fmateriaalit%2FAbras-Maloney_krichmar_Preece.pdf&ei=aAFMTI6jHnKZOLeciZYD&usq=AFQjCNEtz77ODsMncfopm75pSHT-eTNUXw&sig2=mS_BIW0WjE-8S3hM0En3zg.
20. Eason K. Information technology and organizational change. London: Taylor & Francis; 1987.
21. Chalmers I. If evidence-informed policy works in practice, does it matter if it doesn't work in theory? *Evidence & Policy*. 2005(1):227-42.
22. Task Force on Health Systems Research. Informed choices for attaining the Millennium Development Goals: towards an international cooperative agenda for health-systems research. 2004;364:997-1003.
23. Lavis J, Posada F, Haines A, Osei E. Use of research to inform public policymaking. *The Lancet*. 2004 Oct 30-Nov 5, 2004;364(9445):1615.
24. Lewin S, Lavis JN, Oxman AD, Bastías G, Chopra M, Ciapponi A, et al. Supporting the delivery of cost-effective interventions in primary health-care systems in low-income and middle-income countries: an overview of systematic reviews. *The Lancet*. 2008 2008/9/19;372(9642):928-39.
25. Davies HTO, Nutley SM, Smith PC, editors. What works?: evidence-based policy and practice in public services: The Policy Press; 2000.
26. Lomas J, Culver T, McCutcheon C, McAuley L, Law S. Conceptualizing and Combining Evidence for Health System Guidance. Ottawa: Canadian Health Services Research Foundation 2005.
27. Eden J, Wheatley B, McNeil B, Sox H, editors. *Knowing What Works in Health Care: A Roadmap for the Nation*. Washington, D.C.: Institute of Medicine of the National Academies, The National Academies Press; 2008.
28. Oxford American dictionary. 1st ed. New York: Avon Books; 1980.
29. Oxman A, Lavis J, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP). 1. What is evidence-informed policymaking. *Health Res Policy Syst*. 2009;7(Suppl 1):S1.
30. Schunemann H, Fretheim A, Oxman A. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Research Policy and Systems*. 2006;4(1):21.
31. Elwyn G, Edwards A, Gwyn R, Grol R. Towards a feasible model for shared decision making: focus group study with general practice registrars. *BMJ*. 1999 September 18, 1999;319(7212):753-6.
32. Jamtvedt G, Bjørndal A. Kunnskapsbasert praksis. Høgskolen i Bergen and Nasjonalt kunnskapscenter for helsetjenesten; [cited 2010 July 25]; Online course for evidence-based health care]. Available from: <http://kunnskapsbasertpraksis.no/index.php?action=static&id=6>.
33. Lomas J. Connecting Research and Policy. *Canadian Journal of Policy Research*. 2000 Spring 2000(Spring):140-4.

34. Haines A, Jones R. Implementing findings of research. *British Medical Journal*. 1994 June 4;308:1488-92.
35. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet*. 2003 Oct 11;362(9391):1225-30.
36. Balas E, Boren S. Managing Clinical Knowledge for Health Care Improvement. *Yearbook of Medical Informatics 2000: Patient-centered Systems*. Stuttgart, Germany: Schattauer; 2000. p. 65-70.
37. Oxman A, Lavis J, Fretheim A. The use of research evidence in WHO recommendations.
38. Rogers Y. New theoretical approaches for HCI. In: Cronin B, editor. *Annual Review of Information Science and Technology*. Medford, NJ: Information Today; 2004. p. 87-143.
39. Stolterman E. The nature of design practice and implications for interaction design research. *International Journal of Design*. 2008;2(1):55-65.
40. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc*. 2006 Jan;94(1):55-60.
41. Ely J, Osheroff J, Chambliss M, Ebell M, Rosenbaum M. Answering Physicians' Clinical Questions: Obstacles and Potential Solutions. *J Am Med Inform Assoc*. 2005 Mar-Apr 2005;12(2):217-24.
42. Ely JW, Osheroff J, Ebell M, Chambliss M, Vinson D, Stevemer J, et al. Obstacles to answering doctors' questions about patient care with evidence: qualitative study *BMJ*. [research (qualitative study)]. 2002 March 23, 2002;324(7339):710.
43. Davies K. The information-seeking behaviour of doctors: a review of the evidence. *Health Information and Libraries Journal*. [review]. 2007;24:78-94.
44. Green ML, Ciampi MA, Ellis PJ. Residents' medical information needs in clinic: are they being met? *Am J Med*. 2000 Aug 15;109(3):218-23.
45. Krahn J, Sauerland S, Rixen D, Gregor S, Bouillon B, Neugebauer EA. Applying evidence-based surgery in daily clinical routine: a feasibility study. *Arch Orthop Trauma Surg*. 2006 Mar;126(2):88-92.
46. D'Alessandro DM, Kreiter CD, Peterson MW. An evaluation of information-seeking behaviors of general pediatricians. *Pediatrics*. 2004 Jan;113(1 Pt 1):64-9.
47. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. *BMJ*. 1999 Aug 7;319(7206):358-61.
48. Gorman P. Information needs in primary care: a survey of rural and nonrural primary care physicians. *Stud Health Technol Inform*. 2001;84(Pt 1):338-42.
49. 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 Oct 20;282(15):1458-65.

50. Gapminder World. [cited 2010 July 26]; World-wide statistics]. Available from: www.gapminder.org.
51. Williams CD, Pitchforth EL, O'Callaghan C. Computers, the Internet and medical education in Africa. *Med Educ*. 2010 May;44(5):485-8.
52. Gosling AS, Westbrook JI, Coiera E. Variation in the use of online clinical evidence: a qualitative analysis. *International Journal of Medical Informatics*. [research]. 2003 2003;69:1-16.
53. Innvær S, Vist GE, Trommald M, Oxman AD. Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of Health Services Research & Policy*. [systematic review]. 2002 October 2002;7(4):239-44.
54. Lavis J, Davies H, Oxman A, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. *J Health Serv Res Policy*. 2005 Jul;10 Suppl 1:35-48.
55. Harries U, Elliott H, Higgins A. How are policy makers using evidence? Models of research utilisation and local NHS policy making. *Journal of Public Health Medicine*. [research article]. 1999;21(1):29-36.
56. Smith R. Closing the digital divide. *BMJ*. [editorial]. 2003;326(238).
57. Edejer TT-T. Disseminating health information in developing countries: the role of the internet. *BMJ*. 2000;321:797-800.
58. Isaakidis P, Swingler GH, Pienaar E, Volmink J, Ioannidis JP. Relation between burden of disease and randomised evidence in sub-Saharan Africa: survey of research. *BMJ*. [review]. 2002;324:1-5.
59. Reinart L, Greenhalgh T. [P018] Working in partnership with people in the developing world2006.
60. Haines A, Kuruvilla S, Borhcert M. Bridging the implementation gap between knowledge and action for health. *Bulletin of the World Health Organization*. [theme paper]. 2004;82(10):724-32.
61. Rochon PA, Mashari A, Cohen A, Misra A, Laxer D, Streiner DL, et al. Relation between randomized controlled trials published in leading general medical journals and the global burden of disease. *CMAJ*. 2004 May 25;170(11):1673-7.
62. Horton R. Medical journals: evidence of bias against the diseases of poverty. *Lancet*. 2003 Mar 1;361(9359):712-3.
63. Siegfried N, Muller M, Deeks J, Volmink J. Male circumcision for prevention of heterosexual acquisition of HIV in men. *Cochrane Database of Systematic Reviews*. 2009(2).
64. Lomas J. Using Research to Inform Healthcare Managers' And Policy Makers' Questions: From Summative to Interpretive Synthesis. *Healthcare Policy*. [invited paper]. 2005;1(1):55-71.
65. Lomas J. Improving Research Dissemination and Uptake in the Health Sector: Beyond the Sound of One Hand Clapping. *Policy Commentary: Mc Master University Centre for Health Economics and Policy Analysis* 1997 November 1997.

66. Petticrew M, Whitehead M, Macintyre SJ, Graham H, Egan M. Evidence for public health policy on inequalities: 1: The reality according to policymakers. *J Epidemiol Community Health*. 2004;58:811-6.
67. Jewell CJ, Bero LA. "Developing Good Taste in Evidence": Facilitators of and Hindrances to Evidence-Informed Health Policymaking in State Government. *The Milbank Quarterly*. 2008;86(2):177-208.
68. Dobbins M, Jack S, Thomas H, Kothari A. Public Health Decision-Makers' Informational Needs and Preferences for Receiving Research Evidence. *Worldviews on Evidence-Based Nursing*. 2007;4(3):156-63.
69. Dobbins M, Cockerill R, Barnsley J. Factors affecting the utilization of systematic reviews. *International Journal of Technology Assessment in Health Care*. [research article]. 2001;17(2):203-14.
70. Hutchinson A, Johnston L. Bridging the divide: a survey of nurses' opinions regarding barriers to, and facilitators of, research utilization in the practice setting. *J Clin Nurs*. [research: survey of nurses]. 2004 March 2004;13(3):304-15.
71. Greenhalgh T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ*. 1997 September 13, 1997;315(7109):672-5.
72. Cochrane. The Cochrane Collaboration: A brief introduction. [cited 2005 21 August 2005]; leaflet describing the collaboration's aims and activities in brief form]. Available from: <http://www.cochrane.org/resources/general.shtml>.
73. Chalmers I. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci*. 1993 Dec 31;703:156-63; discussion 63-5.
74. The Cochrane Collaboration. Available from: <http://www.cochrane.org>.
75. The Cochrane Library. [Internet]: John Wiley & Sons, Inc.; Available from: www.thecochranelibrary.com.
76. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and Reporting Characteristics of Systematic Reviews. *PLoS Med*. 2007 March 27 2007;4(3):e78.
77. Coumou HC, Meijma FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc*. [literature review]. 2006 January 2006;94(1):55-60.
78. Novack L, Jotkowitz A, Knyazer G, Novack V. Evidence-based medicine: assessment of knowledge of basic epidemiological and research methods among medical doctors. *Postgrad Med J*. [research]. 2006;82:817-22.
79. Lavis J, Davies H, Oxman A, Denis J-L, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. *Journal of Health Services Research & Policy*. [original research]. 2005;10(Suppl 1):35-48.
80. Jones N, Walsh C. Policy briefs as a communication tool for development research2008; 2008: Available from: www.odi.org.uk.

81. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More Informative Abstracts Revisited. *Annals of Internal Medicine*. 1990 1 July 1990;113(1):69-76.
82. Mulrow CD, Thacker SB, Pugh JA. A proposal for more informative abstracts of review articles. *Annals of Internal Medicine*. 1988 April 1988;108(4):613-5.
83. Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *J Med Libr Assoc*. 2004 Jul;92(3):364-7.
84. PLoS Medicine Editors. The Impact of Open Access upon Public Health. *PLoS Medicine*. 2006 May 01, 2006;3(5):e252.
85. Novack L, Jotkowitz A, Knyazer B, Novack V. Evidence-based medicine: assessment of knowledge of basic epidemiological and research methods among medical doctors. *Postgrad Med J*. 2006 December 1, 2006;82(974):817-22.
86. New format for BMJ research articles in print. *BMJ*. 2008 December 29, 2008;337(dec29_1):a2946-.
87. Delamothe T, Mullner M, Smith R. Pleasing both authors and readers. Summary of electronic responses. *BMJ*. 1999 Aug 28;319(7209):580.
88. ACP Journal Club. [December 30 2008]; Clinical synopses. Quality-assessed, clinically rated original studies and reviews from over 130 clinical journals]. Available from: <http://www.acpjic.org>.
89. Evidence Based Medicine for Primary Care and Internal Medicine. BMJ Publishing Group Ltd; [December 2008]; Available from: <http://ebm.bmj.com/>.
90. Evidence Based Mental Health. BMJ Publishing Group Ltd; [December 2008]; Available from: <http://ebmh.bmj.com/>.
91. Evidence Based Nursing. BMJ Publishing Group Ltd; [December 2008]; Available from: <http://ebn.bmj.com/>.
92. Dicenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. *Evid Based Nurs*. 2009 Oct;12(4):99-101.
93. Clinical Evidence. BMJ Publishing Group Ltd; [December 2008]; Available from: <http://clinicalevidence.bmj.com/ceweb/index.jsp>.
94. Glenton C, Underland V, Kho M, Pennick V, Oxman AD. Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. *J Clin Epidemiol*. 2006 Aug;59(8):770-8.
95. Nutley SM, Walter I, Davies HTO. From Knowing to Doing: A Framework for Understanding the Evidence-into-Practice Agenda. *Evaluation*. 2003;9(2):125-48.
96. Estabrooks CA, Thompson DS, Lovely JE, Hofmeyer A. A guide to knowledge translation theory. *The Journal of Continuing Education in the Health Professions*. 2006;26:25-36.
97. Mitton C, Adair CE, McKenzie E, Patten SB, Perry BW. Knowledge Transfer and Exchange: Review and Synthesis of the Literature. *The Milbank Quarterly*. 2007;85(4):729-268.

98. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in Knowledge Translation: Time for a Map? *The Journal of Continuing Education in the Health Professions*. 2006;26:13-24.
99. Scott I. The evolving science of translating research evidence into clinical practice. *Evid Based Med*. 2007 Feb;12(1):4-7.
100. Lomas J. Retailing research: increasing the role of evidence in clinical services for childbirth. *Milbank Q*. 1993;71(3):439-75.
101. Guyatt G, Cook D, Haynes B. Evidence based medicine has come a long way. *BMJ* %R 101136/bmj3297473990. 2004 October 30, 2004;329(7473):990-1.
102. Guyatt G. Evidence-based medicine. *ACP J Club (Ann Intern Med)*. 1991;14(suppl 2):A-16.
103. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996 January 13, 1996;312(7023):71-2.
104. Muir Gray JA. Evidence based policy making. *BMJ*. 2004 Oct 30;329(7473):988-9.
105. Glasziou P. Evidence based medicine: does it make a difference? Make it evidence informed practice with a little wisdom. *BMJ*. 2005 Jan 8;330(7482):92; discussion 4.
106. Cohen AM, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of Evidence-Based Medicine. *Int J Med Inform*. 2004 Feb;73(1):35-43.
107. Little M. 'Better than numbers...' A gentle critique of evidence-based medicine. *ANZ J Surg*. 2003 Apr;73(4):177-82; discussion 82.
108. Wilson K. Evidence-based medicine. The good the bad and the ugly. A clinician's perspective. *J Eval Clin Pract*. 2010 Apr;16(2):398-400.
109. Devries RG. The warp of evidence-based medicine: lessons from Dutch maternity care. *Int J Health Serv*. 2004;34(4):595-623.
110. Gerber A, Lauterbach KW. Evidence-based medicine: why do opponents and proponents use the same arguments? *Health Care Anal*. 2005 Mar;13(1):59-71.
111. Mayer J, Piterman L. The attitudes of Australian GPs to evidence-based medicine: a focus group study. *Fam Pract*. 1999 Dec;16(6):627-32.
112. Higgins JP, Green S, (editors). *Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration*2008.
113. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Austr*. 2004 March 15 2004;180(6 Suppl):57-60.
114. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q*. 2004;82(4):581-629.
115. Rogers EM. *Diffusion of innovations*. 5 ed. New York: Free Press; 2003.
116. Sudsawad P. *Knowledge translation: Introduction to models, strategies, and measures*. Austin, TX: Southwest Educational Development Laboratory2007.

117. Rogers EM. Diffusion of innovations. 4 ed. New York: Free Press; 1995.
118. Patel VL, Kaufman DR. Cognitive Science and Biomedical Informatics. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics Computer Applications in Health Care and Biomedicine. 3rd ed. New York: Springer; 2006.
119. Newell A, Simon HA. Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall; 1972.
120. Ericsson K, Simon HA. Verbal reports as data. *Psychological Review*. 1980;87(3):215-51.
121. Suchman L. Plans and situated actions: The Problem of Human-Machine Communication. New York: Cambridge University Press; 1987.
122. Greeno JG. Learning in activity. In: Sawyer RK, editor. *The Cambridge Handbook of The Learning Sciences*. New York: Cambridge University Press; 2006. p. 79-96.
123. Salomon G, editor. *Distributed cognition: psychological and educational considerations*. Cambridge, UK: Cambridge University Press; 1993.
124. Hutchins E. *Cognition in the Wild*. Cambridge: The MIT Press; 1996.
125. Sawyer RK. Introduction: The New Science of Learning. In: Sawyer RK, editor. *The Cambridge Handbook of The Learning Sciences*. New York: Cambridge University Press; 2006. p. 1-16.
126. Patel VL, Yoskowitz NA, Arocha JF, Shortliffe EH. Cognitive and learning sciences in biomedical and health instructional design: A review with lessons for biomedical informatics education. *J Biomed Inform*. 2009 Feb;42(1):176-97.
127. Bransford JD, Brown AL, Cockning RR, editors. *How people learn: Brain, mind, experience, and school*. Expanded ed. Washington, DC: National Academy Press; 2000.
128. Chi MTH, Glaser R, Rees E. Expertise in problem solving. In: Sternberg RJ, editor. *Advances in the psychology of human intelligence*. Hillsdale, NJ: Erlbaum; 1982.
129. Reyna VF, Lloyd FJ. Physician Decision Making and Cardiac Risk: Effects of Knowledge, Risk Perception, risk Tolerance, and Fuzzy Processing. *Journa of Experimental Psychology: Applied*. 2006;12(3):179-95.
130. Preece J, Rogers Y, Sharp H. *Interaction design: beyond human-computer interaction*: John Wiley & Sons, Inc.; 2002.
131. Gregory J. Scandinavian Approaches to Participatory Design. *Int J Engng Ed*. 2003;19(1):62-74.
132. Bratteteig T. *Making change. Dealing with relations between design and use*. Oslo: University of Oslo; 2003.
133. International Organization for Standardization. *Ergonomic requirements for office work with visual display terminals (VDTs)-. Part 11: Guidance on usability*. Geneve, Switzerland: International Organization for Standardization; 1998.

134. Nielsen J. Usability 101: Introduction to Usability. [cited 2010 31 July, 2010]; Available from: <http://www.useit.com/alertbox/20030825.html>.
135. Law E, Roto V, Vermeeren APOS, Kort J, Hassenzahl M, editors. Towards a Shared Definition of User Experience. CHI; 2008 April 5-10, 2008; Florence, Italy.
136. Alben L. Quality of Experience. *Interactions*. 1996;3(3):11-5.
137. Nielsen Norman Group. User Experience - Our Definition. Nielsen Norman Group,; [cited 2010 31 July]; Available from: <http://www.nngroup.com/about/userexperience.html>.
138. Hassenzahl M, Tractinsky N. User experience - a research agenda. *Behaviour & Information Technology*. 2006;25(2):91 - 7.
139. Forlizzi J, Battarbee K. Understanding experience in interactive systems. Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques; Cambridge, MA, USA. 1013152: ACM; 2004. p. 261-8.
140. Karapanos E, Zimmerman J, Forlizzi J, Martens J-B. User experience over time: an initial framework. Proceedings of the 27th international conference on Human factors in computing systems; Boston, MA, USA. 1518814: ACM; 2009. p. 729-38.
141. Morville P. User Experience Design. [website]: Sematic Studios LLC; 2004 [updated June 21, 2004April 09, 2008]; honeycomb model]. Available from: <http://www.semanticstudios.com/publications/semantics/000029.php>.
142. International Organization for Standardization. Ergonomics of human-system interaction. Part 210: Human-centered design for interactive systems: International Organization for Standardization; 2010.
143. Bevan N. UX, Usability and ISO Standards. CHI 2008; April 4-10; Florence, Italy2008.
144. Hassenzahl M. The thing and I: understanding the relationship between user and product. In: Blythe MA, Overbeeke K, Monk AF, Wright PC, editors. *Funology: From Usability to Enjoyment*: Springer Netherlands; 2003. p. 31-42.
145. Norman DA. *Emotional design: why we love (or hate) everyday things*: Basic Books; 2004.
146. Nielsen J. Usability Laboratories: A 1994 Survey. 1994 [cited 2010 31 July]; Available from: <http://www.useit.com/papers/uselabs.html>.
147. Kuniavsky M. *Observing the User Experience: A Practitioner's Guide to User Research*. 1 ed. Kaufmann M, editor: Morgan Kaufmann; 2003.
148. Malterud K. *Kvalitative metoder i medisinsk forskning*. 2nd ed. Oslo: Universitetsforlaget; 2003.
149. . Valid Useful User Experience Measurement (VUUM). Proceedings of the International Workshop on Meaningful Measures; 2008; Reykjavik: Institut of Research in Informatics in Toulouse.
150. W3C (World Wide Web Consortium). *Evaluating Web Sites for Accessibility: Overview*. W3C; 2010 [updated 2010/05/03; cited 2010 31 July]; Available from: <http://www.w3.org/WAI/eval/>.

151. Morae. TechSmith; Usability Testing for Software and Web Sites]. Available from: <http://www.techsmith.com/morae.asp>.
152. De nasjonale forskningsetiske komiteer. God forskningspraksis. Oslo [cited 2010 31 July]; Available from: <http://www.etikkom.no/no/Forskningsetikk/God-forskningspraksis/>.
153. Nielsen J, Loranger H. Prioritizing Web Usability. 1 ed: New Riders Press; 2006.
154. Nielsen J, Molich R, Snyder C, Farrell S. Search2001.
155. Pope C, Ziebland S, Mays N. Qualitative research in health care: Analysing qualitative data. *BMJ*. 2000 January 8, 2000;7227:114-6.
156. Coffey A, Atkinson P. Making sense of qualitative data. Complimentary research strategies. Thousand Oaks, London, New Dehli: Sage Publications; 1996.
157. Chow R. Case Transfer vs Case Study: An Evaluation of Case Study as a Method for Design Research. Swiss Design Network Symposium 2008; 30-31 May 2008; Berne, Switzerland2008.
158. Rosenbaum SE, Glenton C, Cracknell J. User experiences of evidence-based online resources for health professionals: User testing of The Cochrane Library. *BMC Medical Informatics and Decision Making*. 2008;8(1):34.
159. Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. *Journal of Clinical Epidemiology*. 2010 2010(63):607-19.
160. Rosenbaum SE, Glenton C, Oxman A. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *Journal of Clinical Epidemiology*. [research]. 2010(63):620-6.
161. Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Migniini L, Young T, et al. Evidence summaries tailored for health policymakers in low and middle-income countries. *WHO Bulletin*. 2010;(under submission).
162. Buchanan R. Design Research and the New Learning. *Design Issues*. 2001;17(4):3-23.
163. Marshall C, Rossman G. *Designing Qualitative Research*. 4th ed. Thousand Oaks, London, New Dehli: Sage Publications; 2006.
164. Patton MQ. *Qualitative research and evaluation methods*. 3rd ed. Thousand Oaks, London, New Dehli: Sage Publications; 2002.
165. Law E, VUUM Program Committee, editors. Preface. *Meaningful Measures: Valid Useful User Experience Measurement (VUUM) 2008* 18 June, 2008; Reykjavik: Institute of Research in Informatics of Toulouse.
166. Weedon B, Boguslawski G. *The Cochrane Library. Round 2 Usability Evaluations*. Dublin: Serco Usability Services John Wiley and Sons Ltd.2006.
167. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *J Med Internet Res*. 2006;8(2):e9.

168. Scott I HR, Fairweather P. The use of evidence-based medicine in the practice of consultant physicians. Results of a questionnaire survey. *Australian New Zealand Journal of Medicine*. [research article]. 2000 2000 June 30;June 30(3):309-10.
169. Nielsen J. Ten Usability Heuristics. 2005; Available from: http://www.useit.com/papers/heuristic/heuristic_list.html.
170. Department of Health and Human Services. Research-Based Web Design & Usability Guidelines. Enlarged/Expanded ed. Washington, DC: U.S. Dept. of Health and Human Services; 2006.
171. Fogg BJ. Stanford Guidelines for Web Credibility. A Research Summary from the Stanford Persuasive Technology Lab.: Stanford University; 2002 [April 07, 2008]; Available from: www.webcredibility.org/guidelines.
172. Nielsen J. Information foraging: Why google Makes People Leave Your Site Faster. [website] Fremont, CA: Jakob Nielsen; 2003 [updated April 07, 2008; cited 2008 April 07]; Available from: <http://www.useit.com/alertbox/20030630.html>.
173. Pirolli P, Card S. Information Foraging. *Psychological Review*. 1999 October, 1999;106(4):643-75.
174. Ware C. Visual Thinking: for Design. Nielsen J, Grudin J, Card S, editors. Burlington: Morgan Kaufmann; 2008.
175. Sheridan SL, Pignone M. Numeracy and the medical student's ability to interpret data. *Eff Clin Pract*. 2002 Jan-Feb;5(1):35-40.
176. Young JM, Glasziou P, Ward JE. General practitioners' self ratings of skills in evidence based medicine: validation study. *BMJ*. 2002 April 20, 2002;324(7343):950-1.
177. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making*. 2001 Jan-Feb;21(1):37-44.
178. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med*. 1981 Dec;71(6):991-8.
179. Gigerenzer G, Gaissmaier EK-M, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. *Association for Psychological Science*. 2008 2008;8(2):53-96.
180. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull*. 2009 Nov;135(6):943-73.
181. Ancker JS, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. *J Am Med Inform Assoc*. 2007 Nov-Dec;14(6):713-21.
182. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ*. [education and debate]. 2003 September 27 2003;327(7417):741.
183. Reyna VF. A theory of medical decision making and health: fuzzy trace theory. *Med Decis Making*. 2008 Nov-Dec;28(6):850-65.

184. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Making*. 2007 Sep-Oct;27(5):696-713.
185. Paling J. Strategies to help patients understand risks. *BMJ*. 2003 September 27, 2003;327(7417):745-8.
186. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making*. 2007 Sep-Oct;27(5):638-54.
187. Natter HM, Berry DC. Effects of presenting the baseline risk when communicating absolute and relative risk reductions. Abingdon, UK: Taylor & Francis; 2005.
188. Trevena LJ, Davey HM, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. *J Eval Clin Pract*. 2006 Feb;12(1):13-23.
189. Hoffrage U, Gigerenzer G, Krauss S, Martignon L. Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition*. 2002 Jul;84(3):343-52.
190. Brainerd CJ, Reyna VF. Fuzzy-Trace Theory: Dual Processes in Memory, Reasoning, and Cognitive Neuroscience. *Advances in Child development and behavior*. 2001;28:42-100.
191. Tufte ER. *Envisioning Information*. Cheshire, Connecticut: Graphics Press; 1990.
192. Pickard MJ. The new Blooms taxonomy: an overview for family and consumer sciences. *Journa of Family and Consumer Sciences Education*. 2007 Spring/Summer 2007;25(1).
193. Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the Results of Cochrane Systematic Reviews to a Consumer Audience: A Qualitative Study. *Med Decis Making*. 2010 Jul 19.
194. Fogg BJ, Cathy S, David RD, Leslie M, Julianne S, Ellen RT, editors. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. Proceedings of the 2003 conference on Designing for user experiences; 2003; San Francisco, California.
195. Stanford J, Tauber E, Fogg BJ, Marable L. Experts vs. Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites: Consumers Union of U.S. Inc. 2002 October 29, 2002.
196. Lee Y, Kozar KA, Larsen KRT. The technology acceptance model: past, present, and future. *Communications of the Association for Information Systems*. 2003;12(Article 50):752-80.
197. Slawson DC, Shaughnessy AF. Obtaining useful information from expert based sources. *BMJ*. 1997 Mar 29;314(7085):947-9.
198. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ*. 2008 June 28, 2008;336(7659):1472-4.
199. Weiss CH. The Many Meanings of Research Utilization. *Public Administration Review*. 1979 Sep.-Oct., 1979;39(5):426-31.

200. Patel VL, Kaufman DR, Arocha JF. Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics*. 2002 January 2002;35:52-75.
201. Muir Gray JA. Canadian clinicians and patients need clean, clear knowledge. *CMAJ*. 2006 Jul 18;175(2):129, 31.
202. Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D. The electronic patient record in organisations: A systematic literature review using the meta-narrative method. London: University College London Department of Medical Education 2009.
203. Krippendorff K. Redesigning Design; An Invitation to a Responsible Future. In: Tahkokallio P, Vihma S, editors. *Design: Pleasure or Responsibility*. Helsinki: University of Art and Design; 1995. p. 138-62.
204. Star SL, Griesemer JR. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*. 1989;19:387-420.
205. Lindgaard G, Fernandes G, Dudek C, Brown J. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*. 2006 March-April;25(2):115-26.
206. Tufte ER. *The Cognitive Style of PowerPoint: Graphics Press*; 2003.
207. Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D. Tensions and Paradoxes in Electronic Patient Record Research: A Systematic Literature Review Using the Meta-narrative Method. *The Milbank Quarterly*. [systematic review]. 2009;87(4):729-88.
208. Kay S, Purves I. The electronic medical record and the "story stuff": a narrativistic model. In: Greenhalgh T, Hurwitz B, editors. *Narrative Based Medicine*. London: BMJ Books; 1998.
209. Canadian Health Services Research Foundation. *The Theory and Practice of Knowledge Brokering in Canada's Health System* 2003 December 2003.
210. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv Res*. 2007 Aug;42(4):1758-72.
211. Press M, Cooper R. *The Design Experience: The Role of Design and Designers in the Twenty-First Century*: Ashgate; 2003.
212. Löwgren J, Stolterman E. *Thoughtful interaction design: a design perspective on information technology*: The MIT Press; 2004.
213. Nielsen J. useit.com: Jakob Nielsen's Website. [cited 2010 July 24]; Available from: www.useit.com.
214. Frascara J. Graphic Design: Fine Art or Social Science? In: Margolin V, Buchanan R, editors. *The Idea of Design*. Cambridge Mass.: The MIT Press; 1996.
215. Papanek V. The Future Isn't What It Used To Be. In: Margolin V, Buchanan R, editors. *The Idea of Design*. Cambridge Mass.: The MIT Press; 1996.

216. Live|work. [cited 2010 July 24]; A service design company creating Service Equity through Improvement, Innovation and Strategy]. Available from: <http://www.livework.co.uk/>.
217. Janlert L-E, Stolterman E. The character of things. *Design Studies*. 1997 July 1997;18(3):297-314.
218. Zimmerman J, Forlizzi J, editors. Taxonomy for Extracting Design Knowledge from Research Conducted During Design Cases. *Futureground 04 (Conference of Design Research Society)*; 2004 November 2004; Melbourne, Australia.
219. Zimmerman J, Forlizzi J. The Role of Design Artifacts in Design Theory. [article]. In press 2008.
220. Cross N. Discovering Design Ability. In: Buchanan R, Margolin V, editors. *Discovering Design: Explorations in Design Studies*. Chicago and London: The Chicago Press; 1995. p. 105-20.
221. Schön D. Design as reflective conversation with the material. *Research in Engineering Design* 1992;3:131-47.
222. Official Report of The XVII Olympic Winter Games Lillehammer 1994. Lillehammer: LOOC1995.
223. Bjerknes G, Bratteteig T. The memoirs of two survivors: or the evaluation of a computer system for cooperative work. *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*; Portland, Oregon, United States. 62280: ACM; 1988. p. 167-77.
224. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. *J of Biomedical Informatics*. 2008;41(4):624-35.
225. Cooper A. *The Inmates Are Running the Asylum*: Sams Publishing; 2004.
226. Pruitt J, Grudin J. Personas: practice and theory. *Proceedings of the 2003 conference on Designing for user experiences*; San Francisco, California. 997089: ACM; 2003. p. 1-15.
227. Greenhalgh T. RE: Tensions and paradoxes in EPR - question about design research. In: Rosenbaum SE, editor. e-mail ed2010.

Appendix

Article 1:

User experiences of evidence-based online resources for health professionals:
User testing of The Cochrane Library.

Article 2:

User testing and stakeholder feedback contributed to the development of
understandable and useful Summary of Findings Tables for Cochrane
reviews.

Article 3:

Summary-of-findings tables in Cochrane reviews improved understanding
and rapid retrieval of key information.

Article 4:

Evidence summaries tailored for health policymakers in low and middle-
income countries. (Draft version. The final article will be published in the
December 2010 edition of the WHO Bulletin).

Study Consent Form

Co-authorship role description

Interview guide for SUPPORT study

Example of Summary of Findings Table

Example of SUPPORT Summary