

# **(Mis)Trusting health research synthesis studies:**

## **Exploring transformations of ‘evidence’**

**Submitted by Mila Petrova**

to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Philosophy

In June 2013

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

## Abstract

This thesis explores the transformations of evidence in health research synthesis studies – studies that bring together evidence from a number of research reports on the same/ similar topic. It argues that health research synthesis is a broad and intriguing field in a state of pre-formation, in spite of the fact that it may appear well established if equated with its exemplar method – the systematic review inclusive of meta-analysis. Transformations of evidence are processes by which pieces of evidence are modified from what they are in the primary study report into what is needed in the synthesis study while, supposedly, having their integrity fully preserved. Such processes have received no focused attention in the literature. Yet they are key to the validity and reliability of synthesis studies. This work begins to describe them and explore their frequency, scope and drivers. A ‘meta-scientific’ perspective is taken, where ‘meta-scientific’ is understood to include primarily ideas from the philosophy of science and methodological texts in health research, and, to a lesser extent, social studies of science and psychology of science thinking. A range of meta-scientific ideas on evidence and factors that shape it guide the analysis of processes of “data extraction” and “coding” during which much evidence is transformed. The core of the analysis involves the application of an extensive Analysis Framework to 17 highly heterogeneous research papers on cancer. Five non-standard ‘injunctions’ complement the Analysis Framework – for comprehensiveness, extensive multiple coding, extreme transparency, combination of critical appraisal and critique, and for first coding as close as possible to the original and then extending towards larger transformations. Findings suggest even lower credibility of the current overall model of health research synthesis than initially expected. Implications are discussed and a radical vision for the future proposed.

# Table of contents

## Chapter 1: Background and Aims

---

<b>1.</b>	Introduction	12
<b>2.</b>	What is health research synthesis and why does it matter?	17
2.1.	Research/ knowledge synthesis understood broadly	17
2.2.	The systematic review and meta-analysis – outside and within evidence-based medicine	18
2.3.	Newly emerging methods of health research synthesis	26
2.4.	Impact of health research synthesis studies	29
<b>3.</b>	Interest of the ‘meta-scientific’ fields in health research synthesis	30
<b>4.</b>	Transformations of ‘evidence’ during ‘data extraction’ in health research synthesis studies	34
4.1.	Bringing transformations of evidence from the margins into a focus	34
4.2.	‘Evidence’, ‘data’ and ‘data extraction’ – starting points that are about to change	36
4.3.	Transformations of evidence during data extraction – what does it mean exactly?	37
4.3.1.	Routine, minor transformations	37
4.3.2.	‘Re-location’ of evidence	39
4.4.	How trustworthy are those transformations and how much does this matter?	41
4.5.	Another note on vocabulary	44
<b>5.</b>	Aims, scope and approach of the thesis – initial specifications	45
5.1.	Broad aims	45
5.2.	Scope	46

5.3. Approach	49
5.4. Narrower (but still broad) aims	50
<b>6.</b> Personal background to this work	52
<b>7.</b> Structure of the thesis	56

## Chapter 2: Literature Review

---

<b>1.</b> Contents and structure of this chapter	59
<b>2.</b> Scope, approaches and inclusion/ exclusion criteria for the literature review	60
<b>3.</b> Research synthesis from the perspective of the health sciences	63
3.1. Nomenclature	65
3.2. Defining features of research synthesis studies	66
3.2.1. Clearly stated objectives and pre-defined eligibility criteria	67
3.2.2. Comprehensive literature searches	68
3.2.3. Quality assessment	69
3.2.4. Transparency	71
3.3. Classifications, typologies, 'dimensions of difference'	72
3.4. Rationales, goals, outcomes	75
3.4.1. Increasing and stabilising knowledge	76
3.4.2. Solving a pressing practical problem – primarily in clinical practice and policy making	77
3.4.3. Synthesis work as responding to demands for effectiveness, convenience, ethics and accountability	79
3.4.4. Acquiring 'knowledge about knowledge'	79
3.5. New and emerging synthesis methods – examples and characteristics of the field	81
3.5.1. An overview of methods based on four landmark reviews	81

3.5.2. Examples of methods – extended descriptions	101
3.5.2.1. Narrative synthesis	101
3.5.2.2. Bayesian synthesis	102
3.5.2.3. Meta-narrative	103
3.5.2.4. Evidence synthesis through mechanism-explicit causal models – a proposal from philosophy	104
3.6. Some further characteristics of the field suggested by difficulties in identifying (exemplar) methods	106
3.6.1. We do not know what we do not know	106
3.6.2. Issues of nomenclature	107
<b>4. The unit(s) of analysis/ synthesis in research synthesis studies</b>	<b>108</b>
4.1. Data	110
4.1.1. Data in the methodological discussion	110
4.1.2. Data in the meta-scientific discussion	113
4.2. Evidence	116
4.2.1. Evidence in the methodological discussion	116
4.2.2. Evidence in the meta-scientific discussion	122
4.3. Findings	126
4.4. Facts	128
4.5. Claims, sentences, statements	132
4.5.1. Language in the methodological discussion	132
4.5.2. Sentences, vocabulary and language in the meta-scientific discussion	133
4.6. ‘Becoming and feeding the other’	135
4.7. The debate on ‘paradigms’ in health and other applied sciences research	137
4.7.1. Post-positive knowledge claims	140
4.7.2. Socially constructed knowledge claims	140
4.7.3. Advocacy/ participatory knowledge claims	141
<b>5. Transformations</b>	<b>141</b>
5.1. Translations	143
5.2. Coding	144
5.3. “Travelling”	146

## **Chapter 3: Postscript to Literature Review – Prescript to Methods**

---

1. The field of research synthesis – a working definition and understanding of main components 148
2. ‘Extraction-coding’ 152
3. ‘Findings’ as the preferred term for the basic carrier of primarily empirical knowledge in a research report 154

## **Chapter 4: Approach to Case Study on Extraction-Coding from Cancer Research Publications**

---

1. Topic and question for the test synthesis underpinning the extraction-coding 158
2. Process of identifying the sample of papers 161
3. Features of the performance, observation and exploration of the extraction-coding 162
  - 3.1. Standard and non-standard features of the performance of the extraction-coding 162
  - 3.2. Analysis Framework underpinning the extraction-coding 168
  - 3.3. Software used in the extraction-coding 181
4. Further approaches and tools used 183
  - 4.1. Approaches to exploring types of, similarities and differences between, and grounds for (non-)combinability of findings 183
    - 4.1.1. The Vocabulary of Elements of Findings 184
      - 4.1.1.1. Key elements of findings as per the Vocabulary of Elements of Findings 186
      - 4.1.1.2. Rules of coding 189

4.1.1.3. Some conceptual issues associated with the key elements	192
4.1.2. 'Simple' approaches	203
4.2. Relating findings about types of findings to findings about the presence and effects of matrices	205
4.3. Further generic methods used to complement the Analysis Framework and the Vocabulary of Elements of Findings	206

## Chapter 5: Case study findings

---

1. Contents and structure of the chapter	208
2. The sample of papers	209
2.1. Outcomes of the sampling process and key features of papers	209
2.2. Quantitative/ qualitative status of papers	210
3. The extraction-coding process and outcomes	219
3.1. Basic features of the extraction-coding process and outcomes	219
3.1.1. Extraction-coding in NVivo	219
3.1.2. Extraction-coding in Excel	222
3.1.3. Extraction-coding in EPPI-Reviewer	222
4. A conceptual transformation of transformations	222
5. Differences and similarities between findings	224
5.1. Starting points	225
5.1.1. A (conventional) finding	225
5.1.2. A basic finding	226
5.1.3. Boundaries of findings	227
5.2. Similarities and differences between quantitative and qualitative findings	227
5.3. Other formal similarities and differences between findings	233

5.4. Scope of the thematic differences between findings	234
5.5. Similarities and differences in supposedly highly standardised descriptors	235
5.5.1. Similarities and differences in the collection and representation of basic cancer-related information	235
5.5.2. Similarities and differences in the collection and representation of socio-demographic information	237
<b>6. The multiplicity of findings – illustrations of multiple meanings and uses</b>	<b>238</b>
6.1. Multiplicity of contents, meanings and uses in a negative finding of ‘no difference’	240
6.2. Multiplicity of meanings and uses – multiple affordances for interpretation offered by the same data	244
<b>7. Omissions, silences and half-silences</b>	<b>246</b>
<b>8. Tools for enabling and controlling transformations</b>	<b>250</b>
8.1. Expectations and actualities	251
8.2. Observations on the application of the extensive multiple coding	252
8.3. Observations on the application of the injunctions for transparency and double-edged critical analysis	254

## **Chapter 6: Discussion**

---

<b>1. Structure of the chapter</b>	<b>258</b>
<b>2. Summary of the intent, approach and findings of the thesis</b>	<b>258</b>
2.1. Intent and approach	258
2.2. Summary of findings	260
2.2.1. Findings on the feasibility and effectiveness of methods	260



2.2.2. Main substantive findings from the perspective of transformations	262
2.2.3. Findings throwing light on both methods and substantive issues	266
<b>3. Main contributions to knowledge</b>	<b>266</b>
3.1. Opening up questions and linking disconnected debates	266
3.2. 'Empiricising' what tends to be discussed in the abstract; methodological innovation associated with this	271
3.3. Conceptual innovation and refinement	273
3.4. Highlighting and challenging inconsistencies where none have been noted	274
<b>4. Contribution of this study to the 'nesting' debates</b>	<b>275</b>
4.1. The philosophy of evidence-based medicine debate	275
4.2. The Big Data debate	276
4.3. The quantitative-qualitative divide and integration	277
<b>5. A little more on the key concepts of the thesis</b>	<b>280</b>
5.1. The unit of research synthesis	280
5.2. Transformations	283
<b>6. Limitations</b>	<b>289</b>
<b>7. Vision for the future or an 'impossible synthesis'?</b>	<b>294</b>
7.1. Existing visions for the future of research synthesis	294
7.2. An alternative?	295
7.3. Can it work? Will we want it to?	300
<b>Acknowledgements</b>	<b>307</b>

---

## Appendices

---

### Appendix to Chapter 2

1. Strategies for the bibliometric searches in PubMed presented in Table T 2.1 311
2. Further issues with the meta-methodological discussion of units of analysis/ synthesis 314

### Appendix to Chapter 4

Search strategy used to generate the pool from which papers were sampled for the case study 317

### Appendix to Chapter 5

1. Frequency of codes in terms of the number of papers from which they have “references” 320
2. Frequency of codes in terms of the number of references within them 321
3. Sampling of findings for formalisation with the Vocabulary of Elements of Findings 322

**References** 325

---

### List of Tables and Boxes

---

<b>Table T 2.1:</b> Overview of alternative research synthesis methods	83
<b>Table T 4.1:</b> Types of parameters in the Analysis Framework	171
<b>Table T 4.2:</b> Examples of specific parameters from the Analysis Framework	174
<b>Table T 5.1:</b> Key features of sampled papers	212

<b>Table T 5.2:</b> Presence of the 'other' elements of the Vocabulary of Elements of Findings	232
<b>Box B 5.1:</b> Multiplicity of contents, meanings and uses in a finding of 'no difference'	242

---

## Chapter 1: Background and Aims

### **1. Introduction**

This thesis is about transformations of evidence performed in the process of integrating a number of health research studies – “synthesising” the knowledge contained in them. When taken out of its source study, a piece of evidence inevitably undergoes changes. It has to be enabled to stand more easily on its own, without its contextual support. It may need to be made more similar to other evidence with which it is to be combined, while not compromising its integrity. Its capacity to transform into something somewhat different but still true to its nature may need to be tested. Such transformations of evidence are constantly performed by health researchers involved in “research synthesis”. Can I trim you? Can I extend you? Are you a good brick? Are you a square peg? Can I round you off? At the same time, such transformations of evidence remain curiously under the radar of formal representations and theoretical explorations of health research synthesis. To my knowledge, nobody has made them the focus of an in-depth investigation.

Perhaps this is understandably so. These transformations are largely automatic, apparently minor and often mundane. They also seem too idiosyncratic, too local, not a matter to be dealt with in the abstract. When stumbling upon a difficulty with them, one usually sees a difficulty with this piece of evidence, relative to this question, relative to my limited understanding of the topic.

Yet these transformations of evidence are feeding into research outcomes that determine the healthcare received by each and every one of us, no matter how little healthcare we need or how limited its provision. By being part of health research synthesis studies, and as I will argue in this work – a crucial one – they are part of the effects of evidence-based medicine (EBM). If you have needed healthcare in the developed world in the last, for instance, ten years, you have felt the consequences of evidence-based medicine and those transformations, for good or bad. If you have needed healthcare in the

developing world over the same period, you have most likely felt the consequences of their absence.

Those little processes of transformations of evidence in health research synthesis also offer a view to big philosophical questions of how much we can trust evidence and research and whether evidence is better thought of as solid, fixed and about the things it is about *or* as flexible, freely moving and adaptive, and about many other things than the ones it is about. This thesis is about such philosophy of science questions, too.

Finally, those little processes of transformations of evidence in health research synthesis offer a view to daily challenges of the information and consumer age – how to choose the academic papers we read, the plane ticket we book, the cake we get in the patisserie, the enterprise where we invest our money or, back to health, the dietary and exercise pieces of advice we try to adjudicate between. Health research synthesis shares many of the decision making factors involved in such everyday undertakings – the multitude of information, the need to evaluate options against a number of criteria, the direct relevance of outcomes to the well-being of individuals, the concern with costs, and, as it will become clearer in the course of this work, more of the imperfect rationality of such undertakings than apparent at first sight.

But let us look closer into health research synthesis. The first snapshots we will take will be examples from three research topics – cancer, dementia and depression. Publications on cancer will be used in the case study for this thesis. Dementia is the disease which over two-thirds of my academic colleagues have told me they are most afraid of. Depression is the ailment which I would think they most often suffer with.

As of January 2013, PubMed, the leading gateway to medical and life sciences literature, providing access to more than 22 million citations, holds over 2,4 million references on cancer, over 284,000 references on depression and over 137,000 references on dementia. The ones published in 2012 are 76,624 on cancer, 17,667 on depression and 8,475 on dementia.<sup>1</sup> If you have been trying

---

<sup>1</sup> The numbers will increase at least slightly in subsequent months, as there are some delays in indexing papers. Search run 31 Jan 2013. “Dementia”, “depression” and “neoplasms” (the Medical Subject Heading for cancer) were used as search terms, as free text words.

to keep up to date with the literature on cancer, in 2012 you would have needed to read close to 210 publications a day, every day. For depression, the respective number is slightly over 48 publications a day. You stood a better chance with dementia with about 23 papers (but then you may not remember which ones you have read ...).

This extraordinary publication speed has generated new needs for screening, selecting, abstracting and integrating research and numerous responses to those needs. One such response – research synthesis – will be the broad topic of this thesis. In addition to research studies being potentially many, research findings are often conflicting, sample sizes and detected effects often small, and the quality of studies variable. It is thus necessary to appraise critically and combine/synthesise evidence from a number of studies in order to support decisions about healthcare. According to some views of research synthesis, this is its core – the integration of evidence and, preceding this, its critical analysis so that biased and insufficiently comparable evidence can be excluded.

The exemplar health research synthesis method is the “systematic literature review”. A standard schematic representation of it for those outside health research will be something like the following. If we want to know, for instance, if psychotherapy or drugs are more effective for depression, our best bet is to perform a systematic review comparing the effectiveness of both intervention types. To do this, researchers will perform extensive literature searches to identify as many relevant studies as possible. Ideally, these will be randomised controlled trials (RCTs). Researchers will then bring the evidence, for each of the intervention types, together. Ideally, this will take the form of numerical integration, through the statistical technique of meta-analysis. The evidence for the effectiveness of psychotherapy and antidepressants thus obtained will then be compared and a judgement made.

The misleading simplicity of this representation has contributed to philosophers of evidence-based medicine persevering in their long standing interest in the RCT, gesturing at the technique of meta-analysis, but completely ignoring the complexity of the context of which these become a part in a systematic review. Moreover, there is much more to the field of research synthesis than the exemplar method. Four already dating reviews identify over 30 ‘other’ synthesis

methods between them (Dixon-Woods et al., 2004; Ades and Sutton, 2006; Pope, Mays and Popay, 2007; Barnett-Page and Thomas, 2009). In a very general outline, this thesis aims to draw attention to the broad field of health research synthesis as it is fast developing within evidence-based medicine and to suggest that there is much to it that is philosophically intriguing and challenging.

Returning to numbers, 12,218 (15.9%) of the above mentioned 76,624 publications on cancer from 2012 offer some type of research synthesis, in the form of a “review”, “systematic review”, “meta-analysis”, “guideline” or “practice guideline”.<sup>2</sup> For depression and dementia, the respective numbers are 1,752 (9.9%) and 1,272 (15.0%). Reviews and other integrative work are clearly proliferating, with some variability across fields. One interesting question is whether so many reviews are needed and, if so, why this is the case. A ratio of primary to secondary research ranging between 1 to 6 (cancer) and 1 to 10 (depression) is clearly at odds with the fact that literature reviews in the health sciences tend to have hundreds of references. Does this reflect regrettably low levels of research coordination and high frequency of duplication? Or is it, rather, a consequence of the fact that pieces of evidence and research claims can belong to numerous contexts, including numerous integrative projects? If so, are such integrative projects broadly similar, involving a minimal change of context for the primary research evidence and claims? Or are at least some of them largely unrelated and even rival projects, where recontextualisations of pieces of evidence and claims result in substantial changes to their meanings and function? If this is the case, how much can we trust evidence and its syntheses?

This thesis will address precisely questions about the capacity of evidence to change its form and place, meaning and function, and the implications of this capacity for our trust in the outcomes of research synthesis studies.

The *broad context* of this work is thus methods of research synthesis, alternatively called methods of “evidence synthesis”, “knowledge synthesis”, “integrative methods”, “methods of research integration”, “aggregation of

---

<sup>2</sup> Figure obtained by limiting to the above article types on PubMed.

evidence”, “systematic reviews”, etc. This broad area is largely defined in relation to the framework of evidence-based medicine, whether as a direct realisation of its principles and practices or as a range of responses to its limitations.

The *immediate context* of the focus of the thesis is the initial processing of material from primary studies selected for inclusion in a synthesis study (that is, after initial judgements of relevance have been made). This initial processing includes the reading of primary studies, the identification of pieces of information in these that are relevant to the synthesis study, and the “extraction” of these pieces of information (usually referred to as “data extraction”).

The *focus* of the thesis is on processes of ‘transformation’ and ‘re-location’, re-positionings of evidence that are under way during data extraction and whose aim is either to draw out similarities between a piece of evidence from a certain primary study and pieces of evidence from other studies to be included in the synthesis, or to assign a piece of evidence to an appropriate location in the synthesis framework that differs, formally or functionally, from its location in the source study.

These issues will be explored in the *thematic context* of research on behaviours and ‘mental contents’ (cognitions, emotions, attitudes, etc.) that may have an effect on the incidence, trajectory, experience, recovery or mortality from cancer. A case study will be carried out involving data extraction from such research.

Finally, the *perspective* taken will be that of the ‘meta-scientific’ fields, with an emphasis on debates from the philosophy of science and methodological work on research synthesis from the field of health research itself. Less systematically, work from the social studies of science and the psychology of science will be utilised. In this introductory chapter, I begin to outline the background to this work, its scope, aims, starting points and approach.



## **2. What is health research synthesis and why does it matter?**

### **2.1. Research/ knowledge synthesis understood broadly**

In a classic article entitled “Types of Synthesis and Their Criteria”, Strike and Posner suggest that the concept of (knowledge) synthesis is one of “excessive breadth”, “inherently vague” and “with boundaries that can be made sharp only in arbitrary ways” (Strike and Posner, 1983: 346-347). This is unsurprising – the most widely shared element in definitions of *synthesis* appears to be “the putting together of parts or elements so as to make up a complex whole” (as in, for instance, Oxford English Dictionary Online, 2013; Harper, 2013; Google – “define: synthesis”). Such an activity has few rivals in terms of inclusiveness. A broad terrain for the activities and products of health-related knowledge/ research synthesis is marked out.<sup>3</sup> They can be as varied as the summary and interpretation of findings in a primary study report or review article; the bringing together of two or more theories into a supposedly higher order theory; the aggregation of evidence of effectiveness from individual studies into a summary outcome as in meta-analysis; the writing of a textbook; the assemblage of a bibliographic database. More application-focused activities and outcomes can also be added: the generation of a decision making model; the making of a research-based judgement or decision; the process and products of interdisciplinary work; the development of a public health programme; the fusion of knowledge and practices coming from a range of sources and traditions, as in integrative medicine;<sup>4</sup> even the creation of an artefact – from a new piece of equipment to a living cell, as in synthetic biology. The list can continue indefinitely.

---

<sup>3</sup> In this thesis, *research synthesis* rather than *knowledge synthesis* is generally used in a sense very similar to the one ascribed by Strike and Posner to ‘synthesis’ (and which is referred to as “knowledge synthesis” in the editorial remark preceding the chapter).

<sup>4</sup> There are different understandings of what integrative medicine involves. One view emphasises the integration of orthodox medicine and alternative and complementary therapies. Another seems to remain in ‘proper’ research domains but be more inclusive than biomedicine (e.g. to include the behavioural and social sciences). It draws on ideas of complexity theory and explanatory pluralism to justify integration of knowledge from different areas (Michael Cournoyea offered a critical philosophical analysis of the latter at *the Philosophy of Medicine Roundtable*, San Sebastian, Spain, 2-3 Nov 2011).

Such an inclusive understanding of knowledge/research synthesis in a health and medical context will “stretch the concept towards vacuity” (Strike and Posner, 1983: 347). But it is too early to be specific if we want to avoid being arbitrary and contributing to knowledge fragmentation. At least from one point of view – which I will be defending – the field of health research synthesis is in a state of pre-formation, with its boundaries, objects, phenomena, key issues, terminology, etc. yet to be defined. It appears defined only if we equate research synthesis with its core exemplar, the systematic review inclusive of meta-analysis, which is a frequent move because of the immense popularity and impact of such reviews. But once we broaden our vision towards related methods, even if only towards those defined with reference to the core exemplar, we realise that we do not know what defines and distinguishes the variety of methods of research synthesis and where they fall in a field of logical possibilities (itself non-described) of integrating health research evidence. In Chapter 2 I discuss terminology, proposals for defining features of research synthesis, particular synthesis methods and a broad map of the field. Here I outline some of the most visible, to insiders and informed outsiders, elements of the field – namely its exemplar method, the systematic review inclusive of meta-analysis, and the central role of synthesis studies in clinical practice, health policy and academic research. I also discuss briefly the recent proliferation of alternative methods – another clearly identifiable tendency, though primarily by researchers.

## **2.2. The systematic review and meta-analysis – outside and within evidence-based medicine**

The exemplar research synthesis method and product in the medical and health sciences is the systematic review inclusive of a meta-analysis as associated with the evidence-based medicine movement (e.g. Evidence-Based Medicine Working Group, 1992; Sackett et al., 1996; Petticrew, 2001; Higgins and Green, 2008; Straus et al., 2011). Both respond to a concern that “the conception of research review and integration that prevails in the social and behavioral sciences is one in which the activity is viewed as a matter of largely private judgement, individual creativity, and personal style. Indeed, it is and ought to be

all of these to some degree; but if it is nothing but these, it is curiously inconsistent with the activity (viz., scientific research) it purports to illuminate” (Glass et al., 1981:14). Systematic reviews inclusive of meta-analysis aim to reduce bias by a variety of means – by identifying ‘all relevant literature’, by following a transparent process, and by policing the methodological quality of included studies. They also use statistical means to reduce the imprecision of individual study findings.

Reviews aiming to reduce bias and methods aiming to reduce statistical imprecision by bringing evidence together have certainly been available long before the inception of evidence-based medicine in the 1990s and have been developing outside of medicine, too. In a historical overview of research synthesis, Chalmers, Hedges and Cooper (2002) suggest that the need to synthesise research evidence has been recognised for well over two centuries (*op. cit.*: 12). This takes the form of work that reviews critically what has been published on a subject, with specific efforts made “to reduce the likelihood of being misled by biases and chance” (13). For instance, in the 18<sup>th</sup> century, James Lind, a Scottish naval surgeon, wrote his famous treatise on scurvy in which he also included “a Critical and Chronological View of what has been published on the subject”. That, in Lind’s words, contained “a great deal of rubbish” (Chalmers, Hedges and Cooper, 2002: 13-14). At the end of the 19<sup>th</sup> century, Herbert Nichols published a 76-page review of psychological theories and experiments. As far as numerical synthesis is concerned, in early 19<sup>th</sup> century, a French statistician, Legendre, developed the method of least squares to enable the combination of data from different astronomical observatories (*op. cit.*: 14).

“The science of research synthesis as we know it today” is seen by Chalmers, Hedges and Cooper to begin to emerge in 1904 (*op. cit.*: 14). Karl Pearson, director of the Biometric Laboratory at University College London, published a paper in the *British Medical Journal* where he gathered 11 studies on immunity and mortality from typhoid amongst soldiers. He calculated correlation coefficients for each of the studies (for correlation between immunity and inoculation and mortality and inoculation) and determined their mean values

(*ibid.*). This work is often referred to as the beginnings of meta-analysis.<sup>5</sup> Three years later Joseph Goldberger, of a laboratory that was to give rise to the US National Institutes of Health, performed an analysis of data on bacteriuria in typhoid fever and outlined what are now seen as many of the criteria and steps in research synthesis. These were identifying pertinent studies, applying criteria to select studies for analysis, abstracting and tabulating data, and performing statistical analysis of the abstracted data to obtain the mean rate of bacteriuria (Chalmers, Hedges and Cooper, 2002: 14-15).

In the first half of 20<sup>th</sup> century, methods of research synthesis were also developed in educational research, physics and agriculture (*op.cit.*: 15). Currently, there are systematic reviews in such diverse topics as advertising, agriculture, archaeology, astronomy, biology, chemistry, criminology, ecology, education, entomology, law, manufacturing, parapsychology, psychology, public policy, and zoology (Petticrew, 2001: 99).

In spite of such multidirectional history and spread, the systematic review and meta-analysis have gained their power through the movement for evidence-based medicine. The degree of explicitness of protocol, the strong infrastructure supporting the production of systematic reviews and meta-analyses, and the enormous impact of their findings are new phenomena arising within the context of evidence-based medicine and shaping medicine and healthcare more than any other field. For instance, there are over 5,000 completed Cochrane systematic reviews (with the Cochrane Collaboration being the trend-setter in systematic reviews, discussed below) while the number of Campbell reviews (a parallel organisation that performs systematic reviews on crime and justice, education, international development and social welfare) is slightly over 90 (The Cochrane Collaboration, 2013a; The Campbell Collaboration, 2013).<sup>6</sup> In a review of qualitative research synthesis, Major and Savin-Baden (2011) identified 160 health research synthesis articles vs. 11 coming from education.

---

<sup>5</sup> Accounts differ between sources. Beecher's JAMA study (1955) on placebo effects vs. drug effects is another study referred to as the first meta-analysis of medical research. The term meta-analysis was first used, in the familiar statistical sense, by Glass in his work in educational psychology (Glass, 1976). Bohlin (2012) offers a somewhat different but equally compelling historical account to that of Chalmers, Hedges and Cooper (2002).

<sup>6</sup> There were 93 Campbell Collaboration reviews as of June 13: Crime and justice reviews – 35, Education – 14, International development – 3, Social welfare – 41.

The only ‘proper’ systematic reviews in philosophy<sup>7</sup> are of medical ethics literature, a minor and not exactly high status niche in philosophy. In contrast, a traditional literature review is practically no longer an option for a stand-alone review in health research.

The most robust component of the infrastructure supporting the production of systematic reviews within the context of evidence-based medicine is, undoubtedly, the Cochrane Collaboration. The latter is an “international, independent, not-for-profit organisation of over 28,000 contributors from more than 100 countries, dedicated to making up-to-date, accurate information about the effects of health care readily available worldwide” (The Cochrane Collaboration, 2013b). Its home page features a quote from *The Lancet* representing the Collaboration as rivalling the Human Genome Project in its potential implications for modern medicine (Naylor, 1995).<sup>8</sup> Its history is traced back to Archie Cochrane's book *Effectiveness and Efficiency: random reflections on health services* (Cochrane, 1972). In 1992, the funding for a Cochrane Centre, whose aim was “to facilitate the preparation of systematic reviews of randomised controlled trials of health care”, was approved as a research and development initiative of the British National Health Service (The Cochrane Collaboration, 2013a). As mentioned above, the number of Cochrane Reviews in the *Cochrane Database of Systematic Reviews* exceeds 5,000 (*ibid.*). The most recent (2011)<sup>9</sup> Impact Factor of the *Cochrane Database of Systematic Reviews* was 5.912. This ranked the database in the top 10 worldwide journals in the *Medicine, General & Internal* category (John Wiley & Sons, 2013). In January 2011, The Collaboration was accepted as a “Non-Governmental Organization in Official Relations with the World Health

---

<sup>7</sup> I assume a sense of ‘systematic’ in the tradition of evidence-based medicine, where, for a review to be called ‘systematic’, at least some of the essential features discussed in Chapter 2, Section 3.2. should be clearly expressed. In a looser sense, there are certainly numerous systematic reviews in philosophy.

<sup>8</sup> There is something disingenuous in the way the comparison with the Human Genome Project is presented on the Cochrane website though. Naylor’s article, an authored article rather than a Lancet editorial as might be inferred from the quote’s attribution to *The Lancet*, explores the limits of evidence-based medicine of which the Cochrane Collaboration is a prime realisation and symbol. It is yet another question whether the Human Genome Project is something to be rivalled.

<sup>9</sup> For 2011, published in 2012.

Organization” (The Cochrane Collaboration, 2013a).<sup>10</sup> Usage data for the Cochrane Library from 2009 show that “[e]very day someone, somewhere searches The Cochrane Library every second, reads an abstract every two seconds and downloads a full-text article every three seconds” (The Cochrane Collaboration, 2013c).

There are probably two most immediate associations for ‘systematic review’ amongst health researchers. One is of an extensive literature searching, in over a dozen databases, through search strategies of hundreds of lines and abstruse syntax, fully decodable only by the information specialist who developed them. The second association is of a tedious or trance-like process of “sifting” where from the over-inclusive first search (often between 5,000 and 20,000 citations) a small number of thematically relevant and methodologically robust studies are identified (e.g. 15-20).

More formally and inclusively, a systematic review seeks to “to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made” (Green et al., 2011: 1.2.2). The following features have been identified as its key characteristics: clearly stated objectives; pre-defined eligibility criteria for studies; an explicit, reproducible methodology; a systematic search aiming to identify all relevant studies; an assessment of the validity of the findings of the identified studies; and a systematic presentation and synthesis of the reviewed studies’ characteristics and findings (*ibid*).

The Cochrane systematic review follows a detailed procedure outlined in an over 600-page handbook, the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green, 2008). In an online version, it is updated regularly to reflect methodological advances and user feedback, with a most recent update from March 2011 (Higgins and Green, 2011). Briefly, the review

---

<sup>10</sup> NGOs in Official Relations have “the right to appoint a representative to participate, without right of vote, in WHO's meetings or in those of the committees and conferences convened under its authority” (Section 6.1. in <http://www.who.int/civilsociety/relations/principles/en/index.html>). Special thanks to Lucie Binder, Lisa Bero and Jeremy Grimshaw of the Cochrane Collaboration for clarifying the status of The Cochrane Collaboration in relation to WHO.

process is initiated through the development of a protocol which states the review question, its background, and the objectives and methods of the review (Green et al., 2011: 2.2). Review questions are expected to follow the “PICO-format” and indicate the type of population of interest, the type of intervention and its comparators, and the types of outcomes of interest (**P**articipants, **I**nterventions, **C**omparisons and **O**utcomes). The review question is stated broadly through the review objectives and specified in detail through the eligibility criteria for studies (O’Connor et al., 2011). Randomised controlled trials are the preferred study type to be included in a systematic review due to the method’s status as the gold standard for evaluating the effectiveness of interventions. The review then proceeds towards the identification of studies that meet the pre-determined eligibility criteria. Attempts are made to identify all relevant studies through extensive searching, such as in electronic bibliographic databases and study registers, through citation tracking and contacting study authors. From the pool of studies produced by the searches, relevant studies are selected in accordance with the eligibility criteria and data of interest to the review are collected. Structured tools are used for assessing the risk of bias (study quality). On the basis of this assessment, studies are included in or excluded from further analysis.

Depending on the availability of methodologically sound studies and their “homogeneity” (roughly, similarity), a meta-analysis can be performed. Meta-analysis is a statistical procedure which integrates the results of several independent studies considered by the analyst to be “combinable” (Egger and Smith, 1997; Huque, 1988). Degree of homogeneity is determined on the basis of statistical tests for heterogeneity of study outcomes, such as the Cochran’s Q or the  $I^2$  test (Egger, Smith and Phillips, 1997). Beyond certain levels (e.g. a value for  $I^2 \geq 0.85$ ), it is considered that the study results do not reflect a single underlying effect and meta-analysis is not performed (Egger, Smith and Phillips, 1997; Leonardi-Bee and Rolfe, 2007). The levels of homogeneity/heterogeneity also inform preferences for the general model within which the meta-analysis is carried out – fixed effects models in cases of homogeneity/low heterogeneity and random effects models in cases of moderate heterogeneity. Statistical tests for heterogeneity, however, are not entirely satisfactory (Egger, Smith and Phillips, 1997).

Meta-analysis methods use weighted averages of study results, with larger trials being given higher weights than smaller ones. As there is no single 'correct' method for meta-analysis and each method draws on certain assumptions, the robustness of findings is tested, through sensitivity analysis, to different assumptions and inclusion criteria. This works through splitting findings by, for instance, method (fixed/random effects), methodological quality (higher/lower), publication bias (more or less likely), and early termination (yes/no) and comparing the estimates for the sets of evidence thus obtained (Egger, Smith and Phillips, 1997). If rigorously conducted, meta-analyses are seen as enabling a more objective appraisal of the evidence than traditional reviews, providing a more precise estimate of treatment effects, and facilitating explanations of heterogeneity between individual study results (Egger and Smith, 1997; Egger, Smith and Phillips, 1997).

Most dramatically (and at least in some cases for good reasons), systematic reviews and meta-analyses are seen as saving human lives and public resources. One of the earliest examples of the fact that evidence accumulated as individual studies, unless explicitly brought together, does not 'fully exist' and fails to influence practice and policy is provided in a paper by Antman et al. published in *The Journal of the American Medical Association* in 1992 (the same year and the same journal in which the paper announcing the evidence-based medicine "paradigm" was published, see Evidence-Based Medicine Working Group, 1992). Antman et al. compared the results of meta-analyses of randomised control trials on the treatment of myocardial infarction to recommendations of clinical experts published over the same period in traditional reviews and textbooks. Recommendations were found to be at odds with the summaries of trials. Ineffective treatments were being recommended. Highly effective treatments were not. There were significant time lags between the publication of studies and changes in the recommendations of experts. "[L]ives that could have been saved were lost, and resources were wasted" (Sutton et al., 2000: 6). Another classic example is of evidence of infant sleeping position and sudden infant death syndrome (Gilbert et al., 2005). Sleeping on the front was recommended in books between 1943 and 1988 on the basis of theoretical extrapolations. If a systematic review were performed as early as the 1970s, there would have been statistically significant findings for an



increased risk of sudden infant death syndrome associated with this sleeping position. Over 10 000 infant deaths could have been prevented in the UK and at least 50 000 in Europe, the USA, and Australasia (*op. cit.*).

To bring us closer to our times, there are still media echoes of a broadly discussed, at the end of 2012, Lancet review on over-diagnosis through breast cancer screening (e.g. Gallagher, 2012). Extrapolation from the results of one of the meta-analyses in the review suggests that about three cases are over-diagnosed and treated for one breast cancer death prevented (Independent UK Panel on Breast Cancer Screening, 2012). Although focus group evidence suggests that women still consider screening worthwhile (*ibid.*), the balance of benefits and harms of breast cancer screening has been adjusted. Again, this has resulted from explicitly bringing together evidence which, although available as individual studies, was not ‘real enough’ before having been reviewed and meta-analysed.

In an interesting example of relativism of perspectives, while critics of EBM and the mainstream systematic review tend to see both of these as oppressive orthodoxy, proponents of EBM may consider themselves ‘rebels’ or ‘anti-establishment’ or at least inconvenient voices. For instance, messages of going against accepted opinion were very strong at the 2011 *Cochrane Colloquium* in Madrid.<sup>11</sup> To use a more formal reference, “research synthesis sometimes yields unwelcome results that challenge strongly held opinions and other vested interests” (Chalmers, Hedges and Cooper, 2002: 25).

In summary, the exemplar methodological dyad of the medical and health research synthesis field – the systematic review and meta-analysis – is geared towards answering questions of effectiveness of interventions. It aims to aggregate very specific information from primary studies into a numerical outcome within a broader textual review. It places a strong emphasis on the formulation of a highly specific question and the meticulous identification of ‘all’ studies that are relevant to it. It prioritises evidence obtained through RCTs. It evaluates studies and sanctions their inclusion in the final synthesis through the use of quality assessment tools. It aims to be maximally transparent. It is seen

---

<sup>11</sup> Such claims were made in a number of keynote presentations – 19<sup>th</sup> *Cochrane Colloquium*, 19-22 Oct 2011, Madrid, Spain.

as having the potential to save lives and resources and/or as providing the basis for more informed decision making.

### **2.3. Newly emerging methods of health research synthesis**

Characteristics such as the above make the systematic review with meta-analysis a credible and widely used methodology. But it has well recognised limitations and specificity of application. Some of the research synthesis opportunities it leaves open are immediately visible. For instance, one of the most frequently raised criticisms of evidence-based medicine is that RCT evidence is only one type of evidence relevant to healthcare practice and policy making and that other types of evidence should also be routinely used (Chapter 2, Section 4.2 discusses this in detail). Within health research, the case has been made most strongly for adding to systematic reviews and/or synthesising qualitative research evidence (Green and Britten, 1998; Popay, Rogers and Williams, 1998; Dixon-Woods, Fitzpatrick and Roberts, 2001; Barbour and Barbour, 2003; Popay, 2006).

The potential of qualitative research to broaden the scope of evidence-based medicine arises from some of its basic orientations (Green and Britten, 1998). An orientation towards *naturalism*, in the sense of placing health behaviours in their everyday context, allows to explore the gap between optimum effects from clinical trials and likely effects shaped by everyday concerns. A commitment to identifying *subjective interpretations* of “objective” variables, for instance objective health problems and findings, helps understand the sometimes apparently irrational strategies patients use to deal with health problems. The interest in *process* can reveal, for example, the dynamics of accommodating symptoms, diagnoses and treatments within a patient’s “biography”. Focus on *interactions* can direct attention towards encounters between doctors and patients, where *conflicting explanatory systems* about health and illness tend to clash and negotiation is needed so that good outcomes can be achieved. Findings resulting from these basic orientations of qualitative research can sensitise practitioners to issues they can then usefully explore with patients (*op. cit.*: 1230-1). Evidence of effectiveness in controlled circumstances can thus be

supplemented with evidence of the appropriateness of interventions (the extent to which they meet self-perceived needs of recipients) and of factors affecting decisions and actions (Popay, Rogers and Williams, 1998; Popay, 2006). Further rationales for adding qualitative research evidence to systematic reviews include identifying and refining the review question; identifying meaningful outcomes to be evaluated by the review; adding weight to, challenging or explaining the results of the quantitative synthesis; assisting in making recommendations about improving interventions and implementing them in “real life”, etc. (Dixon-Woods, Fitzpatrick and Roberts, 2001; Popay, 2006).

Work on integrating qualitative research has been a newcomer to evidence-based medicine not only because of resistance against the interpretiveness, subjectivity, small sample sizes, metaphorical language, etc. of qualitative methods by proponents of the traditional systematic review, but also because of doubts from within. “By its very nature and purposes, qualitative research appears resistant to, and endangered by, efforts to synthesize studies. Just as it goes against the nature of poetry to attempt to summarize even one poem about love, so it seems both epistemologically and ethically inappropriate to attempt to summarize findings from one or more qualitative studies about human experiences of health and illness” (Sandelowski, Docherty and Emden, 1997: 366). “For those who hold that each qualitative study is a unique representation of multiple realities or truths the idea of synthesising several studies is anathema” (Pope and Mays, 2006: 143). Britten et al. (2002) summarise the resistances of qualitative researchers in four main directions. First, synthesis studies cannot but ignore the rich detail which characterises good qualitative research, as primary studies are conducted in highly diverse contexts. As a result, generalisations will fail to do justice to the original studies. Next, attempts to develop functional equivalents of meta-analysis are epistemologically naïve. Third, generalisations, whatever their form, have no real grounding and value.<sup>12</sup> Finally, the language games of evidence-based medicine and qualitative research are incompatible (*op.cit.*: 214). In an impasse where qualitative research “appears endangered both by efforts to synthesize

---

<sup>12</sup> The structure of my sentences in this succession required a more liberal paraphrasing of this one, which I hope captures the unsaid correctly. The original was: “Yet others would reject any form of generalisation at all”.

studies and by failure to do so” (Sandelowski, Docherty and Emden, 1997: 365), work on qualitative research synthesis has been steadily growing. As of March 2013, The Cochrane Qualitative Evidence Synthesis (QES) Register contains over 7,100 references to methodology papers relevant to qualitative evidence synthesis and examples of qualitative evidence synthesis (Cochrane QES Register, 2013). Work to describe and conceptualise this variety is lagging behind. In a seminal 2009 paper for which the Register was not available, Barnett-Page and Thomas identified nine more broadly used and three “fledgling” approaches for the synthesis of qualitative research (Barnett-Page and Thomas, 2009). Features of qualitative research synthesis and specific methods, in their relationship to the mainstream systematic review, are discussed further in Chapter 2, Section 3.

Although qualitative research synthesis appears to be the richest and most intensely developing alternative field within the research synthesis landscape, it is not the only one. Dixon-Woods et al. (2004) identify 14 methods for the integration of qualitative and quantitative evidence. Ades and Sutton (2006) catalogue a range of statistical approaches for multiple-parameter evidence synthesis in epidemiology and medical decision making. The methodology of overviews – reviews of reviews – has also been developing (Becker and Oxman, 2011). A large number of studies have produced syntheses without extending a particular study design into a full-blown method.

Such methodological developments have, however, arisen generally independently and unaware of one another. Overall, they tend to have the systematic review with meta-analysis as their only reference enterprise. The field of research synthesis remains uncharted. Outside of the systematic review and some major alternative methods, such as meta-ethnography (Noblit and Hare, 1988; Britten et al., 2002), realist synthesis (Pawson, 2002a,b; Pawson et al., 2004) and meta-narrative (Greenhalgh et al., 2005), it is largely unknown not only to health researchers most broadly but also to synthesis researchers specialising in particular methods. Attempts at mapping innovation in research synthesis, for instance through literature reviews, have been contained within narrowly defined niches – qualitative research (Barnett-Page and Thomas, 2009; Major and Savin-Baden, 2011), qualitative and quantitative research (Dixon-Woods et al., 2004), statistical syntheses (Ades and Sutton, 2006). Such

reviews of subfields are only a few and dating. Further to that, their primary achievement has been to identify and catalogue scattered synthesis methods rather than carry out in-depth critical analysis of the nature, types, promises and limitations of research synthesis. Only initial steps have been made, for instance, to compare approaches; identify their philosophical underpinning and fundamental assumptions; specify methodological types; survey what types of data have been covered and which are still receiving limited attention; highlight good practices that have emerged in one location that may need to be generalised across the field; explore the variety of processes and products characterising research synthesis work. In Chapter 3, I outline one possible map for the field of research synthesis (which is similarly far from addressing such questions but hopefully improves our understanding of the ‘methodological objects’ implicated in them). In summary, the field of alternative methods is dynamic, fragmented and uncharted.

#### **2.4. Impact of health research synthesis studies**

Health research synthesis work has an immense impact on clinical practice, commissioning, health policy, medical education and academic research, to name the most obvious areas of influence. Although reviews are seen traditionally as a way for a scientific community to take stock, it is probably fair to say that current health research synthesis is primarily a route to informing practice, policy and commissioning rather than a way of organising and reporting on scientific knowledge for research purposes. It is also probably fair to say that a piece of health and medical research needs to become part of a synthesis so as to be able to exert a rationally sanctioned influence on healthcare practice, policy and commissioning.<sup>13</sup> Synthesis projects are becoming increasingly the backbone of practice guidelines and policy papers. In the UK for instance, all practice guidelines currently developed by the National

---

<sup>13</sup> Research can, of course, influence policy in many ways, e.g. mediated through its representations in the media or through the hype generated by its promise. My claim is that if *good process of applying the principles of evidence-based medicine is followed*, then the only way in which a piece of research can influence practice, policy and commissioning is by becoming part of a synthesis.

Institute for Health and Care Excellence (NICE) are underpinned by systematic reviews. Practice guidelines are, in turn, used by commissioners to purchase services and by inspectors in service monitoring and evaluation. Their use is also encouraged by “field teams” involved in guideline dissemination and by professional bodies and patient groups (Macbeth, 2011). At the intersection between the academic and the practical, the *British Medical Journal* published between 26 and 42 systematic reviews and/or meta-analyses per year between 2004 and 2011. The numbers represented between 12% and 22% of all research in it.<sup>14</sup> From a more narrowly academic perspective, Chairs in research synthesis are beginning to appear on the job market. Although the influence of research synthesis has been coming primarily from mainstream systematic reviews and meta-analyses, the demand for alternative, e.g. more narrative and conceptual syntheses, is also growing and they too are being incorporated in health services decision making. The field of research synthesis is thus defining of the healthcare we receive and of the patterns of work in the medical and health academic community. At the same time it is almost invisible to the wider society and most scholarly fields interested in health, medicine and research, including the philosophy of science and philosophy of medicine.

### **3. Interest of the ‘meta-scientific’ fields in health research synthesis**

‘Meta-scientific fields’ (or, more loosely, ‘meta-sciences’) is used here to denote areas of traditional academic disciplines dedicated to exploring the processes and products of science and research (both natural and social sciences/research). Of the established and conceivable intersections, the fields of philosophy and history of science are by far the best developed (2111 references containing *philosophy of science* in title – British Library Main Catalogue, Feb 2013. *History of science* – 1866; *sociology of science* – 183,

---

<sup>14</sup> The data are based on the reporting of study design in a paper’s title or abstract which may not be a fully accurate representation of actual study designs. ‘Research’ includes empirical research publications as well as methodological articles about research methods or issues around reporting research. Other types of articles are published in the *BMJ*, too (e.g. editorials, letters, news). If these are taken into account, the relative presence of systematic reviews and meta-analyses in the *BMJ* will be reduced. Sincere thanks to Sara Schroter of the *BMJ* for providing up-to-date data on types of publications.

*psychology of science* – 44, *social studies of science* – 30; *anthropology of science* – 18).<sup>15</sup> The meta-scientific perspective I will take is constituted primarily by philosophy of science work, to a lesser extent by social studies of science work, and by psychology of science thinking further in the background.

At least *prima facie*, the issue of research/evidence synthesis, understood as bringing together primary research studies into a secondary research study, is practically not discussed, at this level of abstraction and in any of the typical health research synthesis terms (e.g. evidence/research synthesis, systematic review, integrative methods, aggregation of evidence, etc.), in the philosophical literature. Of 25 citations containing “evidence synthesis OR research synthesis” and 36 containing “systematic review” in the Philosopher’s Index<sup>16</sup>, two were directly relevant (Bohlin, 2012 and Vergnes et al. 2010, discussed below). A large proportion of the retrieved citations were to actual research syntheses on ethical issues.<sup>17</sup>

The social studies of science and history of science journals are even more reticent as regards research synthesis. No citations were returned by searches on “research synthesis”, “evidence synthesis” and “systematic review” in the *Social Studies of Science*; *Science, Technology, & Human Values* and *Science Technology & Society*.<sup>18</sup> Searches in *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* returned one highly relevant paper through “evidence synthesis” (Stegenga, 2011, discussed below). In the papers retrieved by “research synthesis” (one) and “systematic review” (four in addition to the Stegenga paper), the concepts were all in the bibliographies with one exception. There,

---

<sup>15</sup> The search is only illustrative. It does not include searches by Subject (citations retrieved by a subject search were fewer though).

<sup>16</sup> Search run Feb 2013. When “evidence synthesis” and “research synthesis” were searched as phrases (rather than both “evidence” and “synthesis” appearing together in a citation), the retrieval was 0.

<sup>17</sup> There were also references to papers discussing syntheses or aggregation of other types of material – e.g. Wylie’s criticism of Miriam Solomon’s arguments in favour of “aggregation” of individual opinions in opposition to group deliberation (Wylie, 2006). In a very broad review, this research will be relevant.

<sup>18</sup> Search run Feb 2013, no year limits placed. Search in title or abstract. Provider – Sage.

systematic reviews were problematised to some extent but this remained peripheral to the paper.<sup>19</sup> Inquiring theoretical and philosophical journals in psychology (due to the lack of a specialised psychology of science journal) was similarly unproductive. *Theory & Psychology* returned two and *New Ideas in Psychology* returned four citations containing “systematic review”. These were bibliography or ‘matter of fact’ references or references to other meanings of the concept.<sup>20</sup> None of the two journals contained a reference to “research synthesis” or “evidence synthesis”. *Philosophical Psychology* and the *Journal of Theoretical and Philosophical Psychology* returned no citations containing “research synthesis”, “evidence synthesis” or “systematic review”.<sup>21</sup>

Leaving aside research syntheses of medical ethics papers (e.g. McCullough, 2007; Strech et al., 2008), the philosophical and other meta-scientific publications identified through these and further searches appear to cover the following minimal set of issues. Stegenga (2011) offers a critical exploration of meta-analysis by attending to the numerous decisions made while performing it. These “allow wide latitude for subjective idiosyncrasies to influence its outcomes” (*op. cit.*: 497). He argues that meta-analysis falls short of being “the platinum standard of evidence” which many believe it to be and suggests that “an older tradition of evidence in medicine – the plurality of reasoning strategies appealed to by the epidemiologist Sir Bradford Hill – is a superior strategy for assessing a large volume and diversity of evidence” (*ibid.*). No mention is made of alternative synthesis methods apart from “social methods” for amalgamating evidence such as consensus conferences. Cartwright and again Stegenga (2011) propose a model for the amalgamation of evidence through building

---

<sup>19</sup> Search run Feb 2013, no year limits, search “all fields”. Provider – Science Direct. The ‘exception’ paper was Knaapen and Weisz (2008). The problematisation of systematic reviews related to their embodiment of changing norms of scientific rigour and objectivity and being a new form of standardisation. The focus of the paper was the standardisation of premenstrual syndrome.

<sup>20</sup> Both searches run Feb 2013, no year limits. Access to *Theory & Psychology* – Sage, through journal website (no library subscription). Concepts searched for in “all content”. Access to *New Ideas in Psychology* – Science Direct. Concepts searched for in “all fields”. Examples of uses: systematic review of the highs and lows of life in life history dialogue, a systematic review included in the references, appeal to conducting systematic reviews on computer supported learning tools, etc.

<sup>21</sup> Both searches run Feb 2013, no year limits. Access to *Philosophical Psychology* – Ebsco. Concepts searched for in “Title, Abstract and Full text”. Access to the *Journal of Theoretical and Philosophical Psychology* – Ebsco, through PsycARTICLES. Concepts searched for in all fields.



mechanism-explicit causal models (discussed further in Chapter 2, Subsection 3.5.2.4). Bohlin (2012) presents a fascinating and thorough historical account of the rise of meta-analysis and systematic reviews. Vergnes et al. (2010) consider the ethics of systematic reviews, particularly in terms of unethical or ethically compromised studies that may be contained in them and the fact that research participants have given informed consent for the original studies and not necessarily for meta-studies of these.

Although there is a change since 2010, when the same searches in the social studies of science, history of science and theoretical and philosophical psychology journals returned uniformly 0 articles, it appears evident that the meta-scientific fields are failing to notice the huge complexity and impact of health research synthesis. There is no easily detectible sign of a change in tendencies either (e.g. shifting of discussions at philosophy of medicine events).<sup>22</sup> At the same time, relevant work is more than abundant. Research synthesis touches on a range of core debates within the philosophy of science and other meta-scientific fields, such as unity and disunity of the sciences, (in)commensurability of research traditions, quality criteria in science, the relationship between theory and data, the theory-ladenness of observations, the indeterminacy of theory by data, the nature of evidence, theory choice, reductionism, the nature of interdisciplinary work, the production of scientific facts, the social and political influences on research, to name but a few. The meta-scientific perspective on research synthesis issues is thus both not there and pervasive. In the literature review (Chapter 2), I try to draw a range of disconnected meta-scientific debates together by virtue of their relevance to issues of research synthesis.

Due to the limitations of the meta-scientific debate, I will also be turning to the 'meta-methodological' debate. There is a type of methodological texts in the sciences, or elements of such texts, which, rather than or along with describing a method, take an analytical and/or critical perspective towards it, its application and outcomes. Such texts may, for instance, critique an established or

---

<sup>22</sup> E.g. *Progress in Medicine Conference*, Bristol, 13-15 April 2010; *Philosophy of Medicine Roundtable*, San Sebastian, Spain, 2-3 Nov 2011; philosophy of medicine sessions at the *3<sup>rd</sup> Biennial Society for Philosophy of Science in Practice Conference*, Exeter, 22-24 June 2011; philosophy of medicine sessions at *Causality in the Sciences Conference*, Canterbury, 5-7 September 2012.

emerging method by demonstrating its limitations and the unintended consequences of its application. They may be laying the foundations of a new method, which typically involves the articulation of its theoretical and, occasionally, philosophical commitments and its contextualisation amidst scientific and pragmatic concerns. Publications of this type thus share many of the characteristics of the literature from the meta-scientific fields. I will call them 'meta-methodological' (and occasionally 'methodological' if the applied methodological and the meta-methodological in them are too hard to separate). The distinction from 'meta-scientific' will be important only if debates arising within and outside medical and health research are explicitly compared. I will use 'meta-scientific' as the generic term.

In contrast to the meta-scientific debate on health research synthesis, the meta-methodological one is rich and fast growing, in parallel with the growing number and variety of actual research synthesis studies and methods. Describing it reliably is, however, a research project in its own right. Meta-methodological considerations are scattered across research synthesis fields which are still to be identified consistently. Terminology varies considerably. With regard to newly emerging methods in particular, meta-methodological arguments tend to be only passing comments in publications reporting on actual synthesis studies. The pool of relevant publications is thus enormous. Quality-wise, the robustness and depth of conclusions in the meta-methodological discussion are still quite low – as a result of both its scatteredness and its tendency to appear as incipient arguments growing out of other work. Again, in the literature review in Chapter 2, I attempt to draw traditionally disconnected elements of the meta-methodological debate together.

#### ***4. Transformations of 'evidence' during 'data extraction' in health research synthesis studies***

##### **4.1. Bringing transformations of evidence from the margins into a focus**

Summary descriptions of the research synthesis process tend to list the following elements: formulating the review question; developing inclusion

criteria for studies; searching for studies; identifying research evidence (performing literature searches); selecting studies; data extraction; quality assessment; analysis and synthesis of data; report writing; dissemination (based on Higgins and Green, 2009; Centre for Reviews and Dissemination, 2009). I suggest that the current health research synthesis discourse marginalises the substantial degree of processing of ‘the parts’ from the primary studies that go into ‘the whole’ of the synthesis study and encourage a perceptual switch: *from* seeing all or most of the evidence taken out of primary studies as 1) known-to-be-relevant and 2) remaining unchanged in its movement between the source study and the synthesis study *to* seeing a large part of that evidence as worked upon to establish both relevance and combinability. On the positive side, a much greater richness, versatility and inter-connectedness become visible. A piece of evidence becomes a repository of many more messages and potential uses. The scope of its potential relevance is broadened. Synthesis studies become much more interconnected and mutually informative. On the negative side, evidence and the findings of synthesis studies come to be seen as much less secure and reliable.

In this thesis, I will be demonstrating the centrality of processes of transformation of evidence in research synthesis. I will be arguing for a stronger complementary perspective on evidence which emphasises its transformations, freedom of movement and, possibly more negatively, multiple loyalties. Such a perspective is highly unusual in health research synthesis debates. The phenomena it brings into focus belong to rather distant realms of health research practice and discourse. Once the perceptual filter is supplied, however, there is nothing easier than seeing indications of processes of transformation and active management of relevance and similarity everywhere in research synthesis studies, methodological texts in the field, and the day-to-day concerns and discussions of synthesis researchers. The recognition is already there and in fact quite banal if left at that level.

To go beyond this, I will aim to specify empirically this general claim about transformations characterising research synthesis studies by undertaking and observing a process of data extraction. I will begin to explore, among other things: what is involved in processes of transformation; what factors enable, constrain and shape them; and how pervasive and deep-running (or otherwise)

the potential of evidence for transformations is. Findings will then be related to our trust in and expectations of research synthesis methods. Directions for their improvement will also be considered.

#### **4.2. 'Evidence', 'data' and 'data extraction' – starting points that are about to change**

Some clarification of word usage may be needed before I introduce the topic, aims and scope of the thesis in greater detail. This concerns the concepts of 'evidence', 'data' and 'data extraction'. All three are central to the vocabulary of health research synthesis but their meanings are far from fixed. To start with, 'evidence' and 'data' are used broadly, to refer to any unit of information that is identified in a primary study as potentially relevant to a synthesis study. In the case of typical systematic reviews, 'evidence' and 'data' have become established as the standard generic terms for units of information of interest. It can be argued, however, that such practice is problematic. The compromise with precision is particularly apparent in qualitative synthesis methods and studies, where many of the relevant units of information clearly amalgamate theory and evidence or represent interpretations rather than data.<sup>23</sup> Some authors of qualitative syntheses prefer to refer to 'findings' instead of 'evidence' or 'data' in recognition of that. Others emphatically refer to 'evidence', mainly as a rhetorical device affirming the quality, value and importance of non-RCT research. In the text below, I temporarily put those difficulties aside. As there has been no in-depth analysis in the literature of the units of information taken out of primary studies for inclusion in a synthesis study and thus no precise and fixed vocabulary, I provisionally resort to the terms that are most typically used. The issue is taken up again in the literature review and in Chapter 3.

'Data extraction' in a research synthesis context is, at its most basic, the process of identifying evidence of interest from primary studies and (partly) de-contextualising it with a view to its contribution to and re-contextualisation into the synthesis study. In terms of mechanics, it covers the processes of reading

---

<sup>23</sup> The separation between 'theory', 'interpretations' and 'evidence' and 'data' made here is seen as a matter of degree, with the separation being sometimes easier, sometimes more difficult. I address the issue repeatedly throughout the thesis.

primary reports, picking information potentially relevant to the synthesis study and recording it on a data extraction/ data collection form. The data collection form is seen as a “bridge” between what is reported by the source study authors and what is ultimately reported in the review (Higgins and Deeks, 2011: 7.5.1). It functions as 1) a summary of the review question and the criteria for eligibility of studies; 2) a historical record of decisions and changes thereof that are made during the review process; and as 3) the source of data to be analysed (Meade and Richardson, 1997; Higgins and Deeks, 2011: 7.5.1).

In the context of synthesis methods for qualitative research, the discourse of ‘coding’ is much more frequent than that of data extraction. Qualitative coding is the process of “defining what the data are about”, of “categorizing segments of data with a short name that simultaneously summarizes and accounts for each piece of data”. Codes show how data have been selected, separated and sorted, which is the first step towards making analytic interpretations (Charmaz, 2006: 43). Further consideration of data extraction is given in Chapter 3 and of coding in Chapters 2 and 3. For the moment, I will use “data extraction” as the generic term.

### **4.3. Transformations of evidence during data extraction – what does it mean exactly?**

It is difficult to respond to this question with even a basic definition. This is not a concept that has been used in relation to evidence in research synthesis studies. I will start unfolding it through offering a list of examples and some initial conceptual and critical comments.

#### *4.3.1. Routine, minor transformations*

In health research synthesis studies, at least some evidence is transformed or re-described in preparation for synthesis in a manner that is generally considered routine and minor. This applies to both numerical and textual evidence. In systematic reviews inclusive of meta-analysis, evidence often needs to be standardised so as to become of a uniform format and amenable to

meta-analysis. “Conversions”, as they are also called (e.g. Higgins and Deeks, 2011: 7.7), are usually necessitated by differences and incompleteness in reporting. For example, the parameters typically used in meta-analysis in the case of continuous outcomes are the number of participants, the mean and standard deviation for each intervention group. Their values, however, may not be available directly but obtainable from other statistics. A missing standard deviation can be calculated or estimated from a standard error, confidence interval, t-test statistic, F-test statistic or a p-value (*op.cit.*: 7.7.3.3). In the following formula, the standard deviation is obtained from the standard error of the mean, multiplied by the square root of the sample size:

$$SD = SE \times \sqrt{N}$$

(*op. cit.*: 7.7.3.2)

Conversions (transformations) like these are formulaic and thus highly transparent, replicable and largely reversible. What is involved in the process is known, although more or less secure assumptions may need to be made.

Verbal claims also undergo such routine, minor transformations. Consider a synthesis study of children’s understanding of healthy eating.<sup>24</sup> Dixey et al. (2001) found that children “have some understanding of the idea that there are not healthy and unhealthy foods, but healthy and unhealthy diets” (*op. cit.*: 76). This and the findings of five other studies were synthesised into evidence that children had a “good understanding of the concept of healthy eating”. Such transformations are unavoidable – if you want a synthesis, you accept them – and tend to be seen by synthesis researchers as minor and generally unproblematic. It becomes more controversial whether they are so minor and unproblematic when some of the original data are noted: *All the things that are bad for you are nice and all the things that are good for you are awful* or *The fat*

---

<sup>24</sup> The task to develop a synthesis from the excerpts of six studies on healthy eating in children was given to students in an online module on Research Synthesis I undertaken between May 2010 and July 2010 with The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) at the Institute of Education, London. It was a scaled-down version of an actual research synthesis performed by members of the teaching team (Rees et al., 2011). The same task is the source of the data on inter-coder differences discussed in 4.4.

*squashes the heart and the arteries get blocked* (*op. cit.*: 74). There is much more and much less in nice bad things, awful good things and squashed hearts than “a good understanding of the concept of healthy eating”.

#### 4.3.2. ‘Re-location’ of evidence

At least some evidence in health research synthesis studies is attributed, in preparation for synthesis, to a class of entities, events, processes, relationships, etc. that is different to the one in the source publication. Such ‘re-location’ of evidence is usually seen as strongly interpretive. For instance, in a pilot study on decision making about weight loss<sup>25</sup> evidence that media images are associated with an increase in the importance of weight as a basis for women’s self-esteem and a decrease in their body satisfaction and eating (Strahan, 2003) and evidence that heavy television viewing is associated with increased purging behaviours in Asian-Pacific adolescents (Pinhey and Okinaka, 2004) were both re-located from their original domains (broadly, research on the effect of media images on health) to a new domain concerning the effect of short-lived, minute environmental stimuli on dynamic perceptions of one’s weight and body. The re-location used roughly the following path. Thin images were seen as a kind of minor, possibly subliminally processed, external stimulus. They were shown by the cited studies to affect body satisfaction, the relative importance of weight to self-esteem, eating and purging behaviours. With some further substitutions and amalgamations of concepts and theoretical derivations<sup>26</sup> the cited studies were interpreted as showing that thin images affected the sampled individuals’ perceptions of their bodies and weight. Combined, the above amounted to vicarious evidence lending support to a hypothesis that the dynamic component in self-perceptions of weight may have been underestimated in research and that minor, possibly subliminally processed, external stimuli may contribute substantially to the dynamics of

---

<sup>25</sup> This study (on which I worked, unpublished) compared the research and broader social discussion on decision making about weight loss.

<sup>26</sup> For instance, *perception of one’s body and weight* was seen to be covering some of the elements of *satisfaction* and *importance of weight*. *Perception of one’s body and weight* was also seen as a mediator between being exposed to thin images and engaging in dysfunctional eating and purging behaviours, etc.

those self-perceptions. *Prima facie* at least, the evidence here has been subjected to much more intense and controversial processing than the evidence from the preceding examples. This level of processing is typical of more 'interpretive' qualitative synthesis studies.

Such interpretation heavy re-locations do not happen in qualitative research only. Meta-analysis is dependent upon similar processes. Evidence from a particular study is attributed to newly formed groups that shift focus away from the defining features of a sample, setting, intervention, etc. as these have been asserted in the original study. The 'things' from the original studies are re-located or perspectively transformed into new kinds of things through a revised category attribution. For instance, a recent Cochrane review of interventions in established atopic eczema (Bath-Hextal et al., 2012) brought together, as studies on dietary supplements, studies of three fish oil supplement diets, two vitamin D diets, and single studies of diets supplemented with zinc, selenium, vitamin E, vitamin B6, sea buckthorn oil, hempseed oil and sunflower oil. Uncontroversially, zinc became dietary supplement for the purposes of this review. More controversially, through the same process, zinc became equivalent to sunflower oil.

The other side of a new category formation is that of exclusion from a class. In the vocabulary adopted here, these are cases where transformations are found to be impossible or insufficiently warranted. For example, in a recent Cochrane review of "behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years" (Furlong et al., 2012) studies were excluded because an intervention also included individual interaction between a professional and a parent, rather than only group sessions; because children had learning difficulties, rather than only conduct problems; because the intervention addressed not only parenting problems but also included the training in social skills of the children; because many of the children in a study were above 12 years of age, etc. (*op. cit.*, see table on Characteristics of excluded studies). Supposedly, both the inclusions and exclusions are made on the basis of features that have been clearly



described in the original study or are “analytic”, entailed by the original study category.<sup>27</sup>

The above examples have hopefully given some substance to the claim that processes of transformations are key to research synthesis work. Some further observations will be added under 4.4 concerning the importance and prevalence of those processes. 4.5 comments on the vocabulary of ‘transformations’, ‘re-locations’ and ‘re-descriptions’.

#### **4.4. How trustworthy are those transformations and how much does this matter?**

As indicated above, my interest in transformations of evidence is ultimately an interest in their trustworthiness, which, in turn, is a guide to the trustworthiness of research synthesis studies and the development of adequate expectations of these. We need starting points in terms of credibility of transformations. Currently, these can be obtained mostly indirectly or through anecdotal evidence.

One sign that transformations of evidence are not highly reliable while also being consequential is the practice of double data extraction in research synthesis (i.e. the practice of two researchers independently identifying and recording the information from the primary study which is to be used in the synthesis study). In the case of traditional systematic reviews, double/independent data extraction is standard practice. When this is not feasible as a full-blown process, a proportion of the material (e.g. 20%) is processed by two researchers. Such relatively relaxed standards are not always an option. For instance, double data extraction is mandatory for Cochrane reviews<sup>28</sup> (Higgins and Churchill, 2011). “It is strongly recommended that more than one person extract data from every report to minimize errors and reduce

---

<sup>27</sup> Criticism of those processes of class attribution and collapsing of differences in the context of meta-analysis are well familiar to philosophers. I am trying to contextualise them amongst a range of processes transforming the evidence and entities of interest in research synthesis.

<sup>28</sup> Meaning that good reasons for its non-observance must be provided rather than that the study will be automatically disqualified from being published.

potential biases being introduced by review authors. As a minimum, information that involves subjective interpretation and information that is critical to the interpretation of results (e.g. outcome data) should be extracted independently by at least two people”<sup>29</sup> (Higgins and Deeks, 2011: 7.6.2). Methodological texts justify the need for double data extraction not only through arguments but through empirical research, too. For instance, the Cochrane Handbook cites Jones et al. (2005) who found high prevalence though low impact of data extraction errors (errors in 20 out of 34 reviews). Gøtzsche et al. (2007) found that a minimum of seven out of 27 reviews had substantial errors in data extraction.

In the mainstream systematic review and other ‘more objectivist’ synthesis methods, differences between researchers are seen as a sign that an error has been made or that at least one of the decisions is of lower quality.

Disagreements are resolved in discussion between the two researchers, potentially with the help of a more experienced arbiter. Errors or lower quality decisions are corrected. Criteria and rules are specified further if necessary. The researchers’ approach may also be additionally “calibrated”. The strong emphasis on such processes of control over data extraction is there to ensure the validity and reliability of the synthesis outcome. But it can also be read as a photographic negative. It addresses the fact that the explicit instructions of the method do not constrain unambiguously the handling of the evidence. A replication of the study is incorporated into the study design because of a high likelihood of error and/or high degree of indeterminacy, both with potentially substantial consequences for the outcome.

In qualitative or mixed research synthesis methods, double (or triple) coding or data extraction is also standard practice (e.g. Thorne and Paterson, 1998; Walter et al., 2004; Oliver et al., 2005; Smith, Pope and Botha, 2005).

Discrepancies in data extraction are expected here, too. But only to an extent they are seen as a matter of quality differences. Discrepancies are more often taken as a consequence of legitimate differences of perspective, as resulting

---

<sup>29</sup> The Handbook gives advice on the background and training of extractors, too: “it is preferable that data extractors are from complementary disciplines, for example a methodologist and a topic area specialist. It is important that everyone involved in data extraction has practice using the form and, if the form was designed by someone else, receives appropriate training” (Higgins and Green, 2011: 7.6.2).

from the ambiguity of data and language, and/or reflecting different researchers' sensitivities to different aspects of the data. In any case, the discussion between the coders supports the generation of higher quality material to be used in the synthesis. At the same time the indeterminacy involved in turning primary study evidence into synthesis study evidence is there, a shadow in the background.

A feature concerning the reporting of double data extraction also raises doubts of its reliability, especially relative to the above cited empirical findings on errors in data extraction. There is no established practice of reporting the degree of initial inter-researcher disagreement in synthesis studies, including that in data extraction. It can be argued that such reporting is superfluous. A rigorous process was followed. Inter-researcher disagreement was superseded or used to good advantage. Its extent is irrelevant to the assessment of the synthesis product. Perhaps it is irrelevant. In contrast, it can also be argued that it is missing at least partly because it belongs to a backstage that does not speak particularly well of the validity and reliability of health research synthesis findings. Most probably, the degree of disagreement in data extraction is highly variable rather than consistently high, depending both on the specifications of the method and the particular data used. In some cases it may be genuinely minimal. In other cases it may be minimal because of very restrictive methodological specifications, which pass over the richness of data and compromise validity for reliability, or at the cost of intense calibration of the practice of researchers. In still other cases, it may be disconcertingly large and best not brought to light. For instance, in coding for a test synthesis in which I participated, one of us, both trained researchers, generated 18 and the other 75 codes for the same source material, working to specifications of the same method as laid out in the same publication and having read the same instructions. This is probably an extreme case and it is a single one. But while it is rational to discount it for those reasons, it is also rational not to dismiss it.

Over the last few pages, I have made loom large transformations in data extraction and the fact that practices ensuring their reliability are, as a photographic negative, indicative of problems inherent in them. In case the

direction of attention has begun to imply that transformations rule data extraction and that they are unreliable, I want to balance this out. It may be the case, it may not be. This is one of the questions I will be asking here. My starting point is that the transformations of evidence and the ways in which they may be unreliable need more attention, but that there are enough literal, non-transformative data extraction and enough minor, totally appropriate, unproblematic transformations in synthesis studies. For instance, if a number of studies find that children prefer fruit to vegetables (as has been the case, see, for instance, Dixey et al., 2001) and if the synthesis question explores children's attitudes to healthy eating, it seems unproblematic to extract the evidence about this preference and unnecessary to subject it to any pre-processing before aggregating it. At this level of abstraction, it is hardly important if children in some schools know that tomatoes are a kind of fruit and in others do not. Similarly, there seem to be numerous transformations that are made easily and intuitively and which are regarded as unproblematic. For instance, evidence on "perceptions of weight" relative to objective measures of weight, on "judgements of weight status" relative to objective measures of weight and on "degree of error in self-reports of weight" relative to objective measures of weight are, for all practical intents and purposes, evidence on the same issue. The processes of transformations and re-locations are thus neither fully defining of research synthesis studies, nor by default fraught with uncertainty (unless we disavow the possibility for any certainty whatsoever, be it only pragmatic). This has two important consequences. First, 'straightforward' cases of no or minimal transformation and no or fully acceptable uncertainty can be used as a yardstick in exploring what is happening in more complex cases. Second, straightforward cases can be seen as the stable basis underpinning synthesis studies, which makes synthesis studies meaningful and feasible. One of the versions of the most generic concern of this thesis is what the parameters of this meaningfulness and feasibility are, and how we can improve on them.

#### **4.5. Another note on vocabulary**

The vocabulary of transformations, re-locations and re-descriptions used so far is provisional but needs some preliminary clarification nonetheless. I have used

'transformation' or 're-description' where the source evidence is indeed 'lost' for the synthesis, where it has been substituted by a *representation* of the source study evidence that claims to be *the* source study evidence. This is the case, for instance, when the technical procedure used is one of reading a primary study report and filling in, in a data extraction table, a new description for a particular evidential statement or a set of statements. A similar situation can be observed when information is coded for in the margins of a paper and it is the code that will be used in the synthesis.

The use of procedures and display formats which preserve the original evidence intact or secure immediate access to it is, however, becoming widespread.<sup>30</sup> This is the case when the coding is performed in specialised electronic software (see Chapter 4, Section 3.3. for a brief discussion of packages). There, the original statement is directly visible or a click away. In such cases a vocabulary of 'transformations' may be seen as at best dramatic and at worst misleading. The evidence is still as it was. What has happened is that a new statement or category has appeared. Discussing such processes and outcomes in terms of re-classification or subsumption under a new hypothesis may appear more appropriate. I prefer 'transformations' as at a later stage the original evidence will nonetheless be superseded. It is the claim under which the original evidence was re-classified or the new hypothesis which it came to support that will continue into the synthesis study, not the evidence from the source study. Refinements of terminology concerning transformations will be one of the concerns of this work.

## ***5. Aims, scope and approach of the thesis – initial specifications***

### **5.1. Broad aims**

Most broadly, I will aim to give a descriptive account of the transformations and re-locations of evidence under way during data extraction for the purposes of research synthesis studies – what they involve, their frequency and range (on a

---

<sup>30</sup> In 2010, when the original version of this chapter was written, a Word data extraction form was still quite frequent, possibly dominating the way in which synthesis studies were carried out. This is no longer the case and certainly changes some of the original meaning of 'transformations'.

continuum of 'small' to 'big' transformations), and their drivers (the factors that enable, constrain, shape, determine them). The outcomes of this exploration will be seen as throwing light on the validity and reliability of research synthesis studies – i.e. the extent to which we can trust their results. They will also be used to underpin proposals for methodological developments to help improve the rigour and heuristic potential of transformations in synthesis studies. More theoretically, findings will inform our thinking on the nature of evidence. In particular, its position on three overlapping continua will be considered: of stability vs. flexibility (capacity to preserve or change its form and contents in cases of re-use), fixedness and adaptability (within a particular framework or to a range of frameworks), and singularity vs. multiplicity of meaning.

## **5.2. Scope**

The processes of transformation and re-location of evidence will be viewed primarily as processes concerned with similarity and relevance. The context in which they will be investigated will be that of initial processing of primary study material (through data extraction). The transformations and re-locations of interest will be ones of textual or textual plus numerical material, to the exclusion of purely numerical, formulae-based transformations. This material will be highly varied, drawn from a range of research papers on cancer informed primarily by behavioural, social sciences and humanities thinking. The process of data extraction will be introspectively observed (observed as performed by the author) and will be in response to a guiding synthesis question on, broadly speaking, changes in behaviours and mental contents and their relationship to outcomes in cancer.

These choices exclude numerous approaches and contexts to exploring processes of transformation and re-location of evidence which are likely to be key to an adequate account. For instance, processes of transformation and re-location can also be explored as a field of negotiation, where researchers negotiate what a valid interpretation of a particular piece of evidence is. Or they can be seen from a cognitive psychological perspective and the mechanisms relating input to output experimentally explored. Data extraction is at best the second stage where data are re-located – judgements of relevance in study

selection are an earlier phase where re-locations of evidence may happen. Much data will also undergo further transformations before the final synthesis is reached. Levels of transformation and re-location in synthesis studies involving highly varied material (to be addressed here) will be markedly higher than in synthesis studies of largely similar source material (with the most popular method, the mainstream systematic review, being of the latter type).

These inclusion-exclusion decisions have been made with the aim of capturing some of the most fundamental, interesting and opaque processes implicated in the processing of information to be included in a synthesis study while also ensuring feasibility of the work. Below are some justifications.

The context of *data extraction* is the first stage in a synthesis study where varied and analytically intense transformations of evidence happen. Transformations may be underway during the earlier stage of abstract sifting – namely as re-locations where studies of a distant theoretical framework are seen as relevant and re-located to the theoretical framework of the synthesis study. However, transformations during abstract sifting involve limited analysis and no change in the form and contents of evidence. Data extraction is thus a natural starting point for exploring processes of transformation most broadly.

I chose the perspective of *similarity and relevance* as it seems to reflect what are, arguably, the primary aims of processes of transformation and re-location in the context of research synthesis studies – of ensuring similarity, of ensuring relevance, of testing for similarity and of testing for relevance. I will question this framing throughout the thesis, while taking as a starting point that all processes of transformation and re-location of evidence serve: 1) to ensure that relevant pieces of evidence become sufficiently similar, along a dimension of interest, so that they can be brought together into an aggregated piece of evidence or summary statement concerning the dimension of interest; 2) to ensure that broadly relevant (to the synthesis study) pieces of evidence are attributed to the ‘right’ object (e.g. group of patients, type of intervention, type of context) or the ‘right’ parameter within the synthesis framework (concept, hypothesis, etc.); 3) to test for lines of relevant similarity between pieces of evidence which are not

immediately recognisable; and 4) to test for relevance of a piece of evidence which does not fit immediately within the emerging synthesis framework.

Only *textual material or textual plus numerical material* will be explored for feasibility reasons. The decision to narrow the scope of the work in this way took into account several factors. Most of the research to be analysed in the case study was likely to be health research informed by the behavioural and social sciences and the humanities. Such research is either fully narrative, or combines narrative and numerical elements with a predominance of the former. The synthesis question to be used to guide the data extraction was rather broad. As such, it seemed to require that a narrative be developed first. Textual transformations currently appear much more opaque than numerical transformations. Finally, numerical transformations seem to become relevant after initial data extraction rather than as part of it. 'Textual or textual plus numerical material' is used to refer to statements that are purely textual; statements that incorporate numbers; and the partly textual – partly numerical contents of tables (which are another way of representing propositional knowledge). Even when numerical information is included, however, the type of transformations and re-locations of interest will be ones which are performed through language and concepts, rather than statistical techniques.<sup>31</sup>

*Highly varied material* was preferred for the case study so that as broad a range of transformations can emerge, 'difficult cases' can be encountered (e.g. it is, presumably, much more difficult to combine scientific and humanities thinking than two RCTs on the same topic) and so that the limits of the processes of transformation can become clearer.

Finally, I chose a largely *introspective methodology* as it would allow me to accommodate the greatest variety of questions. None of the alternatives I was forgoing was of superior rigour, although each had some relative advantages. There seemed to be three first-line approaches to exploring questions of transformations and re-locations of evidence during data extraction: 1) comparing the contents of source studies and that of respective data extraction

---

<sup>31</sup> In some cases such transformations may affect the numerical evidence (e.g. numbers may be categorised into ones that represent small, moderate and large effects); in others the numerical evidence may remain unchanged, with transformations affecting other elements of the proposition.



tables from completed or in-progress syntheses; 2) observing the process of data extraction while performed by researchers and eliciting introspective reports from them, possibly in combination with 1); and 3) performing the process myself, observing its deployment and taking extensive notes on the observations, possibly in combination with 1). A number of practical and methodological reasons led to the third choice. For instance, data extraction tables or coded documents of synthesis studies other than systematic reviews are not typically published and not easy to obtain. Also, published or shared tables are extensively edited and thus some of the unembellished processing is lost. Observing and inquiring about other researchers' work would slow down studies that tend to be required within tight timeframes. The questions asked of the research participants might frame their reports too strongly, while a fully open, non-leading inquiry about a process which is not normally discussed may limit unacceptably the collected data. Most importantly, however, performing the data extraction myself was going to allow me to extend standard practice. I could explore a set of conjectures and opportunities for methodological innovation that would otherwise have remained closed for investigation.

If I need to prioritise the ways in which these choices limit the generalisability of findings, I will suggest that findings from this study are likely to overestimate the extent and non-transparency of transformations during data extraction (by virtue of focusing on highly heterogeneous studies and excluding purely numerical transformations) as well as obscuring the importance of transformations during other stages of the synthesis process (by virtue of focusing on data extraction).

### **5.3. Approach**

The performance, observation and exploration of a process of data extraction from recent publications on cancer will be at the core of the study, with the data extraction guided by the following test synthesis question:

*What services and interventions are likely to be effective in changing behaviours and mental contents (cognitions, emotions, attitudes, etc.) that can affect the incidence, trajectory, experience, recovery or mortality from cancer? In a contextualised version of this question, what is the*

*contribution to improved cancer outcomes of such services and interventions relative to biomedically-based ones (e.g. medication, screening, genetic tests, etc.)?*

I will *not* be aiming to complete a synthesis study. The test synthesis question is so broad that it is unrealistic to seek an answer to it not only within the constraints of this work but also in the context of a standard research project resourced for several years and a small- to medium-sized team. The test synthesis question aims to draw in highly varied, potentially hard-to-combine material and direct the data extraction. The observation of the data extraction will be guided by a detailed Analysis Framework specifically developed for this work, drawing on ideas from the philosophical and other meta-scientific literature as well as ideas arising from my previous engagement with research synthesis studies. The data extraction and observation processes will not be easily separable. For instance, probing into the reasons for certain transformations is a way of observing the process but also a way of extending it. Data extraction for research synthesis purposes naturally involves self-observation, at the very least for bias control and critical analysis. Most consequentially, this difficulty of separating the observation from the extraction will result in the stretching of standard data extraction, as some of the questions guiding the observation will be about unexplored opportunities and limits. For instance, the data extraction performed here will be much more comprehensive (more of the information within a paper will be used) and will aim consistently, rather than occasionally, to elicit multiplicity of meanings (involve the consistent extraction of the same piece of data under several categories). These issues are discussed in detail in Chapter 4.

#### **5.4. Narrower (but still broad) aims**

The following key questions will be addressed with the aim of informing a descriptive account of transformations of textual evidence (including textual evidence incorporating numbers) in the process of data extraction for the purposes of health research synthesis:

- What is involved in transformations and re-locations of textual evidence during its decontextualisation from the source study for the purposes of its recontextualisation into the synthesis study? What steps are involved? What information is used? What judgements are made? What types of transformations are emerging?
- How frequent are those transformations and re-locations of evidence and what is their range? How often do they happen? How 'big' or 'small' are they 'on average'? How much change do the 'big' transformations introduce and how close do the 'small' ones keep to the original evidence?
- What is it that drives transformations? What factors enable, constrain, shape, determine them? In particular, what features of study reports – the main source of evidence for research synthesis – drive them?

The following key questions will be addressed with the aim of providing initial pointers for critical analysis of current approaches and proposals for methodological improvements:

- Does extracting more data, envisioning more uses for it and heightening the transparency of the data extraction process even further – which will be prominent features of the methodology employed in this study – ultimately facilitate and enhance the quality of data extraction and synthesis, notwithstanding the more complex process? For instance, does such a process enable the identification of more relevant and stronger lines of similarity? Or, on the contrary, does it bring forth differences between studies which demonstrate that few pieces of evidence are justifiably combinable?
- Using findings on the effectiveness of the case study approach (roughly of 'more data', 'more uses' and 'greater transparency') as a comparator, what can be said about the ability of current health research synthesis methods to optimise the potential of available evidence and put it to new

uses? Does the field as a whole tend to overestimate the fixedness and incommensurability of evidence coming from different research traditions? I.e. does it see too much difference and fail to transform and re-locate evidence enough? Or, on the contrary, does it go too far, too quick? To what extent do current health research synthesis methods incorporate adequate methodological steps to constrain or enable the transformations and re-locations of evidence? What are the implications for our trust in health research synthesis?

- What further methodological steps can be developed so that transformations and re-locations of evidence become more rigorous and creative? How feasible would it be to introduce such steps?

These narrower but still broad aims will be specified further in Chapter 4.

## ***6. Personal background to this work***

In this final section I will describe how I came to the issue of research synthesis and a philosophical and broader meta-scientific exploration of it. This helps explain some of the methodological choices made (e.g. decisions to test certain hypotheses that were not derived from the literature review or choice of search strategies), examples given so far (many of which on overweight and obesity), as well as some of the presentation style.

I became interested in research synthesis through a project on identifying publications on health-related values in diabetes, obesity, schizophrenia and dementia (understood broadly to include issues such as patients' and other stakeholders' values, perceptions, preferences, beliefs, experiences, satisfaction, quality of life, etc.) and mapping their key messages (Petrova et al., 2012a, b). I also conducted a pilot synthesis which aimed to compare the research and the broader social debate on decision making about weight loss. As the methodology of the mainstream systematic review could offer the most detailed guidance on developing literature searching strategies, I started reading on it. As it was not suited to the development of conceptual syntheses, which was my aim, I was on the lookout for alternative synthesis methods.

This involvement in search strategy development and research synthesis methods came shortly after an MA in which the philosophy of science module had made me as excited and jittery as no other.<sup>32</sup> Some of the ideas presented there felt revelatory and liberating. Others had to be wrong – because I did not like them. The juxtaposition of an engagement which required pragmatic outcomes and a theoretically informed perspective generated a number of concerns. Some of these, as I found later, have been extensively addressed within the critical debate on evidence-based medicine.

Others of my concerns did not seem to be receiving any focused attention. The main of these was that there must be something wrong with our methods or conduct of primary research, our methods or conduct of synthesis studies, and/or our idea of science equalling quality and rigour if so many systematic reviews reach a conclusion that “there is insufficient high quality evidence and more research is needed”. This was the conclusion I had come to expect of systematic reviews. This was the conclusion colleagues would so often deliver or hear and which has the status of a usually trite and sometimes hilarious community joke. I recently checked if this perception was not a matter of bias, where we tend to notice the negative rather than the positive. For lack of published data, I carried out a microstudy of 32 recent Cochrane systematic reviews (most recent in February – March 2012). More than three primary studies were included in 23 of these reviews (i.e. in nine reviews evidence was clearly insufficient in volume, too). Within these 23 reviews, assessments of low risk of bias dominated in only 3 studies (13.0% of 23 and 9.4% of 32).<sup>33</sup> These assessments concerned primary studies which, typologically, are believed to have a low risk of bias (mostly RCTs). If the overwhelming perception of researchers and the ratios from this pilot study are on the right track, most health research, even when it has gone through peer-review, is of low quality. Research synthesis both further polices peer-reviewed studies, including those of the gold standard design, and makes do with this degree of quality. Is there something wrong with the standards of research synthesis studies, with the quality of our primary research, or with our expectations of research? Should we

---

<sup>32</sup> Which I owe to Professor Tim Thornton.

<sup>33</sup> They used the GRADE system, which meant that a number of discrete assessments of risk of bias were performed.

rather say that the majority of studies were of standard (rather than low) quality and, expectedly, that there were few or no exceptional studies?

Logic (or perhaps sensitivity to failures to live up to a self-created image, as shaped by my psychology training) also made me question the 'integrity' of systematic reviews. They are so committed to comprehensiveness, 'extremeness' in literature searching as a way of achieving rigour but do not match this with comprehensiveness, 'extremeness' in data extraction. They make so much of their transparency (and indeed, they are remarkably transparent of what decisions have been taken and why in comparison to any traditional review), yet some of their processes are black boxed or only superficially transparent-ised. For instance, why have studies for subgroup analysis been grouped in this way rather than another and what are the potential consequences? Why not discuss the many ways in which database platforms and indexing terms affect search strategy retrieval and from there findings? Why not give the data for inter-rater disagreement, which rarely starts as a beautiful picture as much as it may end up as one? How truly conducive to rigour is it that we have the input and output of some transformation decisions in a table when these are far from reflective of rigour and only exemplary cases? Why are health research synthesis methods failing to question themselves with the intensity with which they question primary research? Of these 'failures of integrity', what I perceived as limitations of transparency became a key interest for me. The reason, to the extent to which I can reliably identify it, is another consequence of my psychological training. I am acutely aware of the many ways in which we self-deceive and trust in a far greater rationality and self-awareness than we are justified to. I see this pattern as applying to science and research, too, rather than to our everyday lives only.

My third main concern was associated with the creative use of data while avoiding the danger of finding what you want to find. In the pilot synthesis on decision making about weight loss I mentioned, a storyline emerged some of whose elements were heavily researched within the context of weight loss and others had a corresponding blank slate. At the same time there was vicarious evidence to draw in support of those claims, with some transformation. Was this a good use of the potential and multiplicity of evidence or was it a stretching to fit a belief? Could we try to re-interpret data in as many directions as we can

think of and then compare the numerous pictures that emerge? What happens if we apply ideas about pluralism in the philosophy of science to how we analyse data and group phenomena for the purposes of research synthesis? Such concerns and questions explain many of the methodological decisions which will be introduced in Chapter 3 much better than the literature review, where similar debates were not apparent.

To conclude this section on personal starting points and their effect on this work, I will make a comment on ‘belonging’ and its impact on the writing style adopted here. Presumably, connecting methodological debates on research synthesis and debates from the philosophy of evidence-based medicine, philosophy of science and the meta-scientific fields more broadly must have a huge heuristic potential. The merger, however, is not easy and not only because the meta-scientific fields have not addressed research synthesis extensively. My personal impressions from talking to health research colleagues and health professionals and attending philosophy of medicine events where health professionals were present is that, polite comments aside, philosophy of medicine is, at best, considered interesting but complex and detached from actual research and clinical practice; at worst, completely misguided, telling nothing of real use to medical science or even harmful; and somewhere in the middle “as strange or as missing the point or as demeaningly ironic” (Krieger, 1992: xvii cited in Chang, 2011).<sup>34</sup>

In commenting on health research papers I have been writing prior to starting my PhD, clinical colleagues would often suggest cuts to theoretical paragraphs. They would jokingly remark that “it is too complex for doctors”. In hearing what my PhD is on, several health research colleagues and doctors have said it is very interesting but “too complex for my little brain” (Yours?!). A quote from a highly regarded blog on “dubious and dishonest science” which features primarily health and medicine related posts (Colquhoun, 2006-2013) illustrates an extreme of the other typical attitude mentioned above, of a perception of uselessness or harm:

---

<sup>34</sup> Krieger, M., 1992. *Doing physics: How physicists take hold of the world*. Bloomington: Indiana University Press.

*Having long since decided that it was Fisher, rather than philosophers, who had the answers to my questions, why bother to write about philosophers at all? It was precipitated by joining the London Evidence Group. Through that group I became aware that there is a group of philosophers of science who could, if anyone took any notice of them, do real harm to research (Colquhoun, 2011).*

I cannot hope to change such attitudes (and in many respects I do share a view that the philosophy of evidence-based medicine could and should be better) but I have taken them into account. I tried to keep the philosophical and other theoretical and critical elements of this work clear and readable by health researchers and clinicians, too. Stylistically, this thesis shifts much. It tends to take discourse features of the field where an issue it discusses at a particular moment is typically located, whether philosophy, health research, sociology of science or psychology, with the former two dominating. It is, however, probably closest to a health research or 'education and debate' publication for a general medical journal, with a higher frequency of I-forms. Thematically and conceptually rather than stylistically, I hope that it falls in a good middle – between philosophy, the broader meta-scientific field, and the health sciences.

## **7. Structure of the thesis**

This introductory chapter is followed by a literature review of three main parts. The first is on the concept of research synthesis, defining features and types of research synthesis, and specific methods other than the mainstream Cochrane-type systematic review inclusive of meta-analysis. I present both tabular summaries and extended descriptions of exemplars. I discuss issues of research synthesis primarily on the basis of sources from health research due to the limitations of the meta-scientific debate outlined above. The second main part of the literature review covers debates on five supposedly basic carriers of empirical knowledge – data, evidence, findings, facts and claims both from a meta-methodological (health sciences) perspective and a meta-scientific perspective. The third (much shorter) part of the literature review addresses key debates on 'transformations'. The literature review is a dense chapter due to the need to address an unstructured topic – research synthesis, a range of very



broad concepts – such as evidence, data and transformations, and to do so both from a meta-scientific and meta-methodological perspective. It needs its breaks of reading.

Chapter 3 is a brief Postscript to Literature Review – Prescript to Methods chapter. The literature review could not create a coherent picture for many of the key concepts needed for the subsequent exploration. The debates adduced in it were often too disparate, too fragmented. In the Postscript-Prescript chapter I stabilise some shaky grounds concerning research synthesis, data extraction and units of analysis/synthesis so that the investigation can proceed.

In Chapter 4, I present the approach, methods and tools used for the case study of data extraction from research papers on cancer. I justify the choice of cancer as topic for the case study and outline the approach to sampling papers. I describe the Analysis Framework used to guide the data extraction and five key ‘injunctions’ which complement it – for comprehensiveness, extensive multiple coding, extreme transparency, combination of critical appraisal (for methodological rigour) and critique (of fundamental assumptions of a piece of work), and for coding as close as possible to the original, with no or minimal transformations, and then extending towards larger transformations. A key component of the analysis toolkit is the Vocabulary of Elements of Findings, also described in Chapter 4.

In Chapter 5, I present findings from the empirical investigation of the 17 papers from which data were extracted. As the ‘behaviours’ of the tools and the findings obtained through their application were quite different to what I expected, I re-conceptualised ‘transformations’ and organised the emerging findings around this revised view. A substantial part of the chapter is dedicated to findings about similarities and differences between pieces of evidence. These similarities and differences concern composition and structure, thematic contents, and standardised parameters such as cancer-related and socio-demographic parameters. I also present findings about the inherent multiplicity of pieces of evidence and the extent of missing information in pieces of evidence. Finally, I summarise my observations on the deployment of the processes of transformation and on the difficulties and affordances of capturing it.

In Chapter 6, I summarise and re-interpret the case study findings with the aim of throwing light on the credibility of current synthesis methods and opportunities for methodological improvements. I specify the intellectual contributions of this work and its limitations. I also offer an ambitious and perhaps utopian vision of a new model of health research synthesis before concluding with comments on its feasibility and application in minimal ways.

A note on graphical conventions may be needed before closing this chapter. The layering of quotes in a study which looks into studies that represent other studies can be complex. For instance, there are cases when I am quoting a piece of research where a research participant is quoted who herself quotes another person. The picture is further complicated by the fact that the study looks at how the same thing can be represented in this way and in that way and in that other way, as a result of many different transformations. Finally, it is a study in a little explored area and as such requires the introduction of a number of new terms. I have adopted a convention where other scholars' quotes or terms appear in double quotation marks and my rephrases or terms are in single quotation marks. I 'mention' words<sup>35</sup> or indicate a metaphorical, ironic or hesitant usage also with single quotation marks. Italics are used for quoting too, usually with longer quotes. They are also used for placing emphases. The graphical conventions in studies where a wide range of perspectives and meta-commenting are implicated is in itself a topic for exploration. Overall, my aim has been to facilitate reading, the attribution of claims and the prioritisation of messages in a way that leaves the conventions inconspicuous (in some cases this means that slight variations are there from context to context). Please let the text lead you – it is only worth remembering the difference between single and double quotation marks around newly introduced terms. Single quotation marks are around my terms which, at the moment, have no currency. Double quotation marks are around other scholars' terms.

Now onto a chapter where double quotation marks prevail – the literature review.

---

<sup>35</sup> In the sense of the philosophical 'use-mention' distinction, where to 'mention' a word is to draw attention to the signifier. Above, I 'mentioned' *mention*.

## Chapter 2: Literature Review

There is an interesting somewhat hidden, somewhat false paradox at the heart of research synthesis studies. They are often said to be bringing forward and together ‘what is known’ on a topic. But if this is needed, then we do not properly know the known. The known may be known in terms of being there, having been discovered or generated by some, but not properly known in terms of being inaccessible to many. It may be perceived as known, but when joined with other known things to be shown as false. It may be known to many in their seclusion and believed to be precious, unique knowledge, but once its sharedness in isolation becomes known, its value degrades. In any case, there is something truer, more definitive, more useful to be known by extracting further knowledge from what we know or by ascribing credibility and importance to it.

This promise is sometimes met, sometimes not. Sometimes what was vaguely known will come out as all that was to be known, albeit more clearly. We are likely to feel disappointed but may also be reassured. Sometimes we will find out that we know less than we thought we did. We may feel at a loss but excited at the opportunity, too. Sometimes superior knowledge may be generated but be lifeless, a “terminally boring recital or catalogue of previous studies” (Schwandt, 1998: 410). The outcome of a research synthesis and the response to it are unpredictable.

With such a framing, centuries of scholarly work on knowledge and meta-knowledge are beckoning with their relevance. I wish I could buy more Big Issues.

### ***1. Contents and structure of this chapter***

This literature review has four main sections. In Section 2, I describe the variety of approaches used to collect information and make decisions about the coverage of the review. In Section 3, I outline current thinking on health research synthesis. I begin by presenting more abstract claims and debates

concerning the 'nature' and 'defining features' of research synthesis<sup>36</sup> and follow these by brief descriptions of exemplary synthesis methods outside the mainstream systematic review. Finally, I draw on the difficulties I encountered in identifying exemplar synthesis methods to articulate further features of the field.

In Section 4, I present a range of views concerning five main carriers of empirical knowledge – data, evidence, findings, facts and claims. In trying to pin down the unit of analysis/synthesis in research synthesis, I became aware of the immense variety of units suggested. I chose the above partly because they are mentioned frequently in research synthesis texts, partly because they are likely to be amongst the most basic possible units of analysis/ synthesis. This section also includes a brief discussion of widely accepted 'paradigms' in health research which are presumed to be shaping evidence, data, findings, etc. Their apparent incommensurability is often used as an argument against the viability of certain types of research synthesis.

In Section 5, I consider the literature on 'transformations'.

Due to the novelty of questions and approach and the fact that very different literatures were brought together, no stable picture emerges from the literature review. In the brief next chapter, Postscript to Literature Review – Prescript to Methods, I temporarily stabilise uncertain starting points so as to proceed with the investigation.

## ***2. Scope, approaches and inclusion/exclusion criteria for the literature review***

In prefacing a volume on *The Structure of Scientific Theories*, Frederick Suppe suggests that "[i]t is only a slight exaggeration to claim that a philosophy of science is little more than an analysis of theories and their roles in the scientific enterprise" (Suppe, 1977: 3). With concepts like research synthesis, evidence, data and transformations, it is only a slight exaggeration to say that one can include the whole of the philosophy of science, the broader meta-scientific field

---

<sup>36</sup> Nobody who has read and studied with Professor John Dupré can use words like 'nature', 'defining features' and 'essence' without long and impassioned explanations concerning the plurality of things that things can be (or occasional quotation marks in their stead).

and all the methodological literature in health research in this literature review. This section describes the main decisions I took so that the review can be feasible while striving towards systematicity and rigour. Such transparency is certainly out of form in a philosophy of science thesis but feels indispensable in a piece of work discussing research synthesis. I sought a compromise in brevity.

I took two main approaches in identifying relevant literature. The first was the use of what in primary research is known as “maximum variation sampling” (Lincoln and Guba, 1985: 233), aiming to ensure that as much as possible of the variety of circulating debates can be tapped into. The second was the performance of what I will call ‘microstudies’ – pragmatic small scale studies of imperfect methodology which, in spite of their limitations, are still more reliable than impressionistic claims concerning the same issues.

To enable maximum variation sampling, I first mapped issues by consulting a range of major and/or highly relevant databases and collections. I used this approach for topics that were central to the work and well defined at least in some respects. For instance, the search on ‘evidence’ involved searching the British Library catalogue, the Philosopher’s Index, PhilPapers online, MEDLINE, PsycINFO, readers in the Philosophy of Science, the Stanford Encyclopedia of Philosophy, ‘books on evidence’, and my personal collection accumulated over a number of years. The searches on ‘research synthesis’ involved searching the British Library catalogue, MEDLINE, PsycINFO, the Philosopher’s Index, individual journals identified through the list of ZETOC,<sup>37</sup> my personal collection, and ‘passive’ searches through alerts from the Cochrane Qualitative and Implementation Methods Group.

In the case of ‘transformations’ – a topic which ‘does not exist’ in the way proposed in the thesis – I first articulated, as key subthemes corresponding to aspects of the concept, the background knowledge on the basis of which the concept emerged. I then refined those subthemes and articulated new ones through further reading and conceptual analysis. Finally, I identified exemplar texts for all subthemes. With topics which I took to be less central to the

---

<sup>37</sup> A British Library service giving access to over 28,000 journals – <http://zetoc.mimas.ac.uk/>.

formulation of the thesis questions (e.g. data, findings, facts), I sought exemplar texts directly. This was usually through relatively simple searches, citation tracking and advice rather than through microstudies.

I took the meta-methodological literature on research synthesis and the philosophical literature in evidence-based medicine as the primary literatures for the review. However, in a number of cases relevant debates in those fields were limited or lacking (e.g. discussions of data in the philosophy of evidence-based medicine) or derivative of debates in a neighbouring field (e.g. discussions of paradigm differences in the methodological literature on research synthesis seem entirely derivative of discussions of paradigm differences in the methodological literature on primary, mostly mixed methods, research). In such cases I reviewed literature from the broader or closest field.

I performed microstudies primarily to ensure the rigour of selecting issues to present but also to collect information that was lacking in the literature. The latter is in line with the appeal of Onwuegbuzie and Frels (2011) who maintain that a 'literature review' label is restrictive for what should nowadays be included in a literature review. The microstudies included two thematic analysis microstudies performed in QSR NVivo (QSR International, 2013),<sup>38</sup> two linguistic content analysis microstudies, and a number of bibliometric microstudies.

The first of the thematic analysis microstudies aimed to identify types of goals, outcomes and rationales for performing research syntheses. It involved coding, until saturation, of exemplary texts on innovative synthesis methods. I then tested the codes thus generated against texts on the mainstream method and further texts on innovative methods (selection of findings in 2.4). The second thematic analysis microstudy aimed to identify recent tendencies in the philosophy of medicine and philosophy of evidence-based medicine. It involved coding the contents of relevant abstracts retrieved from the Philosopher's Index for the last five years, 2007-2012 (120 abstracts; search string "philosophy of medicine OR evidence based medicine OR EBM"; search run Feb 2012, findings in 3.2.2).

---

<sup>38</sup> One of the popular software packages for the analysis of qualitative data. Versions 8 through to 10 were used.

The linguistic content analysis microstudies were used to identify attributes and labels ascribed to 'data' in two exemplary methodological texts – the *Cochrane Handbook* (Higgins and Green, 2008) on the quantitative side and Green and Thorogood's *Qualitative Methods for Health Research* (2004) on the qualitative side. The identification of phrases containing 'data' was done through the search function of Adobe rather than formal word frequency analysis and software (findings presented briefly in 3.1.1).

Finally, the bibliometric microstudies included: 1) identification of the most cited papers on evidence-based medicine in the Web of Science so that these can be used to structure the presentation of the debate on evidence in the meta-methodological literature (see 3.2.1); and 2) identification of relative interest in new and emerging synthesis methods as measured by number of citations in PubMed, the Web of Science and Google scholar (2.5.1).

Micro as they were, those studies were time consuming. Such an approach could not be applied to all topics. A sizeable proportion of the literature I included thus came from simple searches, opportunistic familiarity, availability and advice. Most notable influences outside of my supervisors' advice and the (less than perfect) holdings of the Exeter University library were events on data sharing hosted at my home institution and one of its lead researchers, Dr Sabina Leonelli,<sup>39</sup> and technical briefings by Professor Barry Barnes.

### **3. Research synthesis from the perspective of the health sciences**

Good succinct definitions of research synthesis understood broadly, but also of specific methods of it, are hard to find. Definitions tend to be:

minimal: *Literature reviews accumulate learning and avoid the pitfalls of relying on single studies* (Harden et al., 2004: 794);

---

<sup>39</sup> Such as the international workshop on *Data-driven research in the biological and biomedical sciences*, Exeter, 15-16 April 2010, and the *Making Sense of Large Datasets Conference*, Exeter, 20-21 June 2011.

narrowly procedural: *Conventional systematic review developed as a specific methodology for searching for, appraising, and synthesising findings of primary studies* (Dixon-Woods et al., 2006: bmc 2<sup>40</sup>);

somewhat repetitive: *All reviews involve combining multiple studies for the purpose of demonstrating their collective relevance for solving some problem, for understanding some issue, for explaining some relationship, and so on. Reviewing is an interpretive undertaking insofar as it is an effort to make sense of those studies and to establish their meaning* (Strike and Posner, 1983a). *A view widely held is that reviews are a means of collecting and organizing the results of previous studies so as to produce a composite of what we have already learned about a particular topic* (Schwandt, 1998: 409);

substantially laden with theory: *From a Bayesian perspective, synthesis of research evidence is a decision-making process, and individuals who make decisions do not approach evidence in isolation* (Roberts et al., 2002: 1596-7);

a bumpy combination of formality and informality: *Systematic reviews or research syntheses can be thought of as a technique for sorting out the bits of the jigsaw, weighing up where they might go, and putting bits together. They also help us see more clearly where there are gaps. Systematic reviews are more cumulative and more critically robust and, along with the related idea of evidence-based policy and practice, have been likened to ice breakers sailing through the pack ice of opinion and assertion* (Sutton et al., 2000: 6);

or beautifully poetic but little instructive: *Qualitative metasynthesis is ... a complex exercise in interpretation: carefully peeling away the surface layers of studies to find their hearts and souls in a way that does the least damage to them* (Sandelowski, Docherty and Emden, 1997: 370).

---

<sup>40</sup> The BMC journals, which are only published electronically, indicate underneath each page number that it is “not for citation purposes” (every article starts from p. 1). Adding *bmc* acknowledges that their numbers are special – but not so special so that they are not used for citation purposes ...



I will take as a starting point the following broad conception: *'health research synthesis' is used here to refer to the process(es) and/or outcome(s) of bringing together information from, primarily or only, health research reports where an attempt is made to articulate the process in some detail.* In the Postscript to Literature Review – Prescript to Methods Chapter, I will provide an extended account of research synthesis as understood in this work.

### 3.1. Nomenclature

Current texts concerning the bringing together of health research reports refer to “systematic reviews”, “reviews”, “overviews”, “structured reviews”, “systematic narrative reviews”, “evidence synthesis”, “best evidence synthesis”, “research synthesis”, “systematic (research) synthesis”, “summaries of research”, “integrative research”, “secondary research”, “aggregation” (of research, evidence, data), “knowledge accumulation”, “meta-analysis”, “synopses”, “meta-aggregation”, to name some of the most frequently applied ones.<sup>41</sup> The ‘process vocabulary’ of the field includes “synthesising”, “pooling”, “integrating”, “aggregating”, “summarising”, “accumulating”, “assembling”, “compressing”, “reducing”, “condensing”, “drawing together”, “combining”, “collating”, etc.

Some of these generic terms are used in narrower and/or value-laden ways in particular contexts. For instance, ‘synthesis’ may be preserved only for ‘wholes that are more than the sum of their parts’ – for outcomes that go beyond the averaging of results or juxtaposition of briefly presented study descriptions. Thus, synthesis can be variably opposed to “mere”, “simple” summaries, reviews, aggregations of findings, etc. or other only “summative”, “cumulative” outcomes (e.g. see Greenhalgh et al., 2005: 428; Paterson et al., 2001: preface p. 2; Noblit and Hare, 1988; Barnett-Page and Thomas, 2009). For others, a study can be a synthesis even if it only “involve[s] the juxtaposition of findings from multiple studies, perhaps with some analysis of common themes or findings across studies” (Popay et al., 2006: 7). Still others tend to preserve ‘synthesis’ for the ‘synthesis proper’ part of a review, which comes after

---

<sup>41</sup> Meta-analysis may seem out of place here since generic terms concerning research synthesis are discussed, but at least in earlier years ‘meta-analysis’ has been used generically too and further qualified, as in “qualitative meta-analysis” (Schreiber, Crooks and Stern, 1997 who credit Stern and Harris, 1985 with the first use of the term).

specifying the review question, literature searching, study quality appraisal, data extraction and before the reporting of findings. They tend to oppose methods for performing reviews to methods for performing (solely) a synthesis (e.g. Popay et al., 2006:10-11, Dixon-Woods et al, 2006: bmc 6).

‘Summaries’ and ‘synopses’, too, can have more specific meanings. For instance Haynes, one of the pioneers of EBM, proposes the “5S model” – of studies, syntheses, synopses, summaries, and systems – for the evolution of information services for evidence-based healthcare decisions (Haynes, 2006). In this model a *synthesis* is a systematic review and *synopses* are “succinct descriptions of an individual study or a systematic review” (*op. cit.*: 162). Together with studies, syntheses and synopses tend to address one aspect of the management of a health condition (e.g. the effectiveness of a specific drug). *Summaries*, in contrast, “provide a full range of evidence concerning management options for a given health problem”. Finally, *systems*, such as an advanced electronic medical record, are “decision support services that match information from individual patients with the best evidence from research that applies” (*op. cit.*: 162-3).

For the moment, I will not be adhering to any of the distinctions proposed. One of the aims of this work is to see what distinctions are worth preserving or creating in the field. Although my central term will be ‘research synthesis’, I will also be using others, in particular ‘review’ and ‘integration’.

### **3.2. Defining features of research synthesis studies**

The Cochrane Handbook posits five features defining of a systematic review: 1) clearly stated objectives, along with pre-defined eligibility criteria for studies; 2) an explicit, reproducible methodology; 3) a systematic search aiming to identify all relevant studies; 4) an assessment of the validity of the findings of the identified studies; and 5) a systematic presentation and synthesis of the reviewed studies’ characteristics and findings (Higgins and Green, 2011: 1.2.2). Various combinations of these have been defended as essential to non-mainstream synthesis studies and research synthesis as a whole. Yet all of them as a set and each of these separately have also been subjected to

extensive criticism (maybe with the exception of the fifth feature, which has received limited attention). A range of positions are discussed below for the first four of the proposed defining features.

### 3.2.1. Clearly stated objectives and pre-defined eligibility criteria

The requirements of clearly stated objectives and pre-defined eligibility criteria for including studies in a review are often interpreted as a requirement for a well formulated research question: “In Cochrane reviews, questions are stated broadly as review ‘Objectives’, and specified in detail as ‘Criteria for considering studies for this review’” (i.e. eligibility criteria)<sup>42</sup> (Higgins and Green 2011: 5.1.1). “A statement of the review’s objectives should begin with a precise statement of the primary objective, ideally in a single sentence. Where possible the style should be of the form ‘To assess the effects of [*intervention or comparison*] for [*health problem*] in [*types of people, disease or problem and setting if specified*]’. This might be followed by one or more secondary objectives” (*ibid.*).

Some new and emerging methods also insist on a clearly formulated research question, such as framework synthesis (Lloyd Jones, 2005) or realist synthesis (Pawson, 2002a, b; Pawson et al., 2004). In the context of other methods, however, the advance specification of a precise review question is perceived as “neither possible nor desirable” (Dixon-Woods et al., 2006: bmc 3). The format of the initial research question may need to be “broad, open-ended” (Greenhalgh et al., 2005: 420). In reviews of complex evidence the number of review questions may be “almost infinite”, to be curtailed by pragmatic criteria (*ibid.*, note 1). The question may not be “finally ... settled until the end of the review” (Dixon-Woods et al., 2006: bmc 3).<sup>43</sup>

---

<sup>42</sup> Eligibility criteria also relate to quality assessment, see below.

<sup>43</sup> Undoubtedly, there are differences of approach, driven by differences of question types, source material, epistemological convictions, etc. Yet at least some of the controversy follows from stipulating different beginnings for a synthesis study and from differences in representational style. In a Cochrane review, for example, the question is developed during the protocol development stage. This is an advance specification of a precise question only if we agree that the review starts after the protocol has been developed.

With eligibility criteria, there is a similar spread of positions. The Cochrane Handbook provides detailed guidance on identifying study characteristics to inform them. The prompts to consider in developing the thematic aspect of eligibility criteria are enlisted on five pages (Higgins and Green, 2008: 85-90). In general, it is useful to consider exactly what intervention is delivered, at what intensity, how often it is delivered, who delivers it, and whether people involved in delivery of the intervention need to be trained” (Higgins and Green, 2008: 86).

Some of the newly emerging methods attempt to emulate this degree of specificity. Texts that problematise it usually do so from a quality assessment perspective: primarily the methodological requirements in eligibility criteria are attended to (see also 3.2.3 on quality assessment and 4.2 on evidence). One of the few aspects of eligibility as thematic relevance to be addressed critically, yet somehow always in passing, concerns “background” or “secondary” papers. This is the broader range of publications that have had an influence on a review but are not formally included in it (Marston and King, 2006). A subissue here is that with conceptual reviews, once you have read a paper, you cannot unread it and fully exclude it from your synthesis, even if you do not use its claims and data directly (Booth, 2011).

### *3.2.2. Comprehensive literature searches*

The performance of a comprehensive literature search is probably the most defining characteristic of conventional research synthesis studies in terms of researchers’ perceptions. “Systematic reviews of interventions require a thorough, objective and reproducible search of a range of sources to identify as many relevant studies as possible (within resource limits)” (Higgins and Green, 2008: 97). This will include searching the major health and medical databases, any subject-specific databases, trial registers, conference proceedings, grey literature sources, screening the contents pages of particular highly relevant journals, hand searching, contacting study authors, etc. Time and budget restraints are nevertheless clearly acknowledged (Higgins and Green, 2008: 97). There is also a move to develop simplified approaches to literature searching, partly capitalising on the improved reporting of studies, the improved

bibliographic indexing, and the accumulation of entries on RCT databases. It has been suggested that exhaustive searches are no longer cost-effective (Royle and Waugh, 2005).

Some alternative approaches and studies appropriate the feature of exhaustive and reproducible literature searches (e.g. Walter et al., 2004; Oliver et al., 2005; Smith, Pope and Botha, 2005). Others question the value of such searches in view of the possible outcomes of subsequent (qualitative) analysis of all this data – namely unacceptably “gross generalisations”, “trite conclusions” (Noblit and Hare, 1988: 27-8). Concepts such as theoretical saturation, conceptual saturation, iterativity are advanced as more appropriate (Pawson et al., 2004; Dixon-Woods et al., 2006; Barnett-Page and Thomas, 2009). In realist synthesis for instance, theoretical saturation is the aim, and neither “encyclopaedic coverage of all possibly relevant literature”, nor a “census” of all evidence is sought (Pawson et al., 2004: 20, v). Arguments against exhaustiveness of searches are often accompanied by expressions of dissatisfaction with reproducible, highly structured searches (Dixon-Woods et al., 2006, Greenhalgh et al., 2005). A “more organic” process is adopted, “fit[ing] better with the emergent and exploratory nature of the review questions” (Dixon-Woods et al., 2006: bmc 3). The paradox of performing a comprehensive search while not knowing what is there in the field is also frequently noted.<sup>44</sup>

### 3.2.3. *Quality assessment*

Quality assessment – and the evidence hierarchies, tools and scales associated with the concept – is probably the most vigorously contested feature of systematic reviews. Some of the intensity of the debate comes from the phrase itself, and practitioners of evidence-based medicine are cautious in using it. Nevertheless, the performance of some appraisal of study validity, risk of bias, confidence in estimates of effect is a cornerstone of the exemplar method. The EBM hierarchies of evidence provide the foundations for risk of bias assessment, through ranking the robustness of very broad methodological

---

<sup>44</sup> I have often heard it raised by researchers (e.g. at the presentation of Booth, A., 2011). A good reference evades me. Noblit and Hare make a related point – that it is difficult to know when one is being exhaustive since not all studies are published or publicly available (Noblit and Hare, 1988: 27).

types. Details of the mainstream approach will be discussed under 4.2.

Evidence.

Alternative methods take again a variety of positions. Some apply “rigorous selection criteria” and “standards for scientific rigour” (Sherwood 1997: 33-34). A commitment to structured evaluation procedures can branch off in a range of choices. Arguments in favour of alternative, for instance qualitative, hierarchies of evidence and general expectations of their contents are put forward (Popay, Rogers and Williams, 1998). EBM-style tools for broad use are developed, such as the CASP qualitative research checklist (Critical Appraisal Skills Programme, 2013; for application see, for instance, Walter et al., 2004). Customised quality assessment tools are designed (e.g. Karunanathan et al., 2009; Oliver et al., 2005) or existing ones adapted (e.g. Thorne and Paterson, 1998). Context-sensitive approaches are suggested, such as the “best evidence synthesis” of Slavin where inclusion/ exclusion is decided on the basis of what is available in the literature on a particular topic (1995).<sup>45</sup>

Many alternative approaches are more radical. They take a stance, for instance, against evaluating quality at the start of a synthesis, against prioritising methodological criteria, and against treating the quality of a study as uniform across all its aspects. For instance, Pawson et al. suggests that “[t]rue quality appraisal comes at the coup de grâce and not as a preliminary pre-qualification exercise” (Pawson et al., 2004: 23; see also Noblit and Hare, 1988: 11, 34-35). The value of a study may not be in its empirical findings but in concepts, ideas and insights that are not affected by methodological flaws. Thus, a choice is made to apply no or minimal criteria of quality to exclude only “fatally flawed” studies at sampling (Dixon-Woods et al., 2006; bmc 4). More stringent judgements of quality may then be made during or after the analysis (e.g. Weed 2005: 8; Kearney, 2001: 274). In addition, it is often the case that excluding studies is not an option – once very specific theories concerning very specific

---

<sup>45</sup> Slavin offers a very clear analogue from law to explain this: “In law, there is a principle that the same evidence that would be essential in one case might be disregarded in another because in the second case there is better evidence available. For example, in a case of disputed authorship, a typed manuscript might be critical evidence if no handwritten copy is available, but if a handwritten copy exists, the typed copy would be inadmissible because it is no longer the best evidence” (*op. cit.*: 11).

aspects of an intervention are identified, the data that can test them can be rather hard to find. The process may be more akin to “scavenging” for information rather than allowing to pick and choose between studies (Pawson et al., 2004: 12, 21).

More radically, appraisals of methodological rigour may be seen as insufficient to uncover flaws and uncertainties in the construction of a study. An examination of the social, historical, ideological, etc. context in which a piece of knowledge was produced is also required (Barnett-Page and Thomas, 2009: bmc 3,7). Dixon-Woods et al. suggest the broader concept of “critique” as opposed to the conventional “critical appraisal” for methodological rigour. Such critique involves challenging “the ways in which the literature constructs its problematics”; the epistemological and normative assumptions it draws on; the traditions that have guided particular research fields; the particular forms of discourse available; the factors influencing the proposed solutions (*op.cit.*: bmc 2, 6, 9).

#### 3.2.4. Transparency

“Transparency” in a systematic review context is taken to stand for the explicitness of methodology, the generation of a clear audit trail, the auditability of the synthesis process. This will include, for instance, the full description of one’s search strategy, the referencing of all included/consulted studies, statements about degree of agreement between researchers performing the same process (inter-rater reliability), the inclusion in the review of data extraction tables, etc. Transparency enables clarity about the sensitivity of results to different assumptions and judgements (Sutton et al., 2000: 8), accountability, the independent assessment of bias, and the reproducibility of the process – in theory for its replication, but in practice primarily for the purpose of updating reviews. Transparency is an aspect of the “scientific and open way” in which systematic reviews are carried out (Sutton et al, 2000: 8). The requirement for transparency is also self-referential – it allows reviewers to appraise the systematicity of their process and correct for digressions in it as they go along.

Generally, alternative synthesis methods concur with the conventional systematic review on the value of transparency. But they also tend to challenge the degree to which it is desirable and achievable. For instance, Dixon-Woods et al. (2006) present Critical Interpretive Synthesis as a method that explicitly recognises the “authorial voice” and the non-transparency of some aspects of the review production, and “does not aim to offer a series of pre-specified procedures for the conduct of review” (*op.cit.*: 6, 10). Pawson et al. (2004) object to the “reproducibility principle”. First, it is on the grounds of the “sheer impossibility of making transparent every single decision involved in research synthesis” (*op.cit.*: 37). Experience, sagacity, intuition are used in making relevant judgements (*ibid*). Secondly, they question “whether objectivity in science has ever stemmed from standardisation of procedure” (38). For these authors, “laying down one’s methodological tracks”, “surfacing one’s reasoning” are there not for reproducibility purposes, but to expose a developing theory to scrutiny and critique, and in this way enable its revision and refinement (*ibid*).

In the context of secondary data analysis, Hammersley (1997) also argues that the attempt to render the research process fully explicit creates hardly recognised ethical issues that relate to researchers, too. He quotes Bond in relation to an anthropologist’s fieldnotes: “[t]hey are personal property, part of a world of private memories and experiences, failures and successes, insecurities and indecisions”.<sup>46</sup> “[T]he work of even the most competent researcher would look poor if documentation of the research process were compared to some idealised, overly rationalistic model”. Transparency thus comes with a danger of “researchers producing documentation that bears only a remote relationship to how the research was done” (Hammersley, 1997: 136).

### **3.3. Classifications, typologies, ‘dimensions of difference’**

As little informative as it may be, the strongest distinction in the field of research synthesis is still between conventional, Cochrane-type systematic reviews and

---

<sup>46</sup> Bond, G. C., 1990. Fieldnotes: Research in Past Occurrences. In: R. Sanjek, ed. *Fieldnotes: The Makings of Anthropology*. Ithaca: Cornell University Press.



'other' reviews. The individuation of review methods tends to go along the lines of comparison and contrast with the systematic review (and, in earlier years, with the traditional literature review), rather than claiming a membership of any of the classes discussed below. The quality of currently circulating classifications is also suboptimal.

Possibly the most widespread positively defined distinction is between "aggregative" and "interpretive" reviews. It was first introduced by Noblit and Hare (1988), in their seminal book on meta-ethnography. Aggregative reviews "assemble" and "pool" data concerning largely comparable phenomena, potentially through the use of techniques such as meta-analysis. The concepts and variables under which data are summarised are seen as generally secure and well specified. Aggregative reviews are used mostly to test theories (Dixon-Woods et al., 2006: bmc 2). Such reviews involve a belief in general laws of prediction and in increase of objectivity through the use of procedural methods (Gough and Elbourne, 2002: 226). In contrast, interpretive reviews, drawing on interpretive social science, look for an empathetic understanding of meaning and are "directed towards generating new conceptual understandings and theoretical explanations" (Pope, Mays and Popay, 2007: 72). "Mixed" reviews have also begun to be added to the dichotomy of aggregative (often quantitative) and interpretive (often qualitative) reviews (Pope, Mays and Popay, 2007; Harden, 2011). Mixed reviews are able to accommodate diverse evidence, including non-research evidence. This is an eclectic group: the sources mixed and approaches used vary widely (Pope, Mays and Popay: 95).

In the context of qualitative research synthesis, the most detailed and streamlined framework for describing synthesis methods seems to be the "dimensions of difference" of Barnett-Page and Thomas (2009). They propose seven such dimensions:

(1) the epistemology underlying the method – with distinctions drawn between "subjective idealism", "objective idealism", "critical realism", "scientific realism" and "naïve realism";

(2) the extent of iteration – for instance, in the literature searching or synthesis proper;

- (3) the approach to quality assessment – for instance, the use of checklists for assessing methodological quality or making unstructured judgements of how a research finding informs a theory;
- (4) the extent of problematising the literature – such as whether attempts are made to place it in its historical and socio-political context or deconstruct theories;
- (5) the degree of heterogeneity, diversity of primary studies;
- (6) the extent to which the synthesis goes “beyond” the primary studies to produce a whole that is more than the sum of its parts;
- and, finally, (7) the “synthetic product” in terms of its utility – whether it can be directly used by policy makers and programme developers or is “more complex and conceptual”, potentially “operating on the symbolic or metaphorical level” (*op. cit.*: 5-9).

For the subfield of qualitative research synthesis, this is a detailed and informative framework, albeit presenting substantial difficulties in positioning studies along dimensions.

Booth, Papaioannou and Sutton (2012) also stress that the classification of review types is multidimensional. They give examples of dimensions such as the purpose of the review, types of included studies, nature of included data, type of question, phenomenon investigated, underlying intent, context and “‘philosophy’ regarding subsequent use” (*op. cit.*: 20). They do not, however, explore those dimensions<sup>47</sup> or apply them to the description of their own review typology. The latter includes 11 review types – critical review, integrative review, literature review, mapping review/systematic map, meta-analysis, mixed studies review/mixed methods review, overview, qualitative systematic review/qualitative evidence synthesis, rapid review, scoping review, state-of-

---

<sup>47</sup> The examples illustrating them are limited and, overall, do not throw sufficient light on the contents of a particular dimension and its distinction from similar dimensions. E.g. purpose of the review is illustrated through the example of “the mapping review”, underlying intent through “meta-ethnography for theory generation or realist synthesis for theory verification”, and philosophy regarding subsequent use is illustrated through “best evidence synthesis”. It is not intuitively clear where the boundaries between purpose, intent and philosophy regarding subsequent use lie and how, for example, best evidence synthesis relates to philosophy regarding subsequent use.

the-art review and systematic search and review. These are characterised by means of brief overall descriptions and more or less heterogeneous comments and labels on type of search, presence or absence of appraisal of studies, type of synthesis and features of the analysis<sup>48</sup> (*op.cit.*: 26-27). The authors add further examples of typologies and dimensions of difference from the literature (21) in what becomes a telling, and maybe not fully intended, illustration of the fact that no broadly accepted typologies of research synthesis studies/ reviews exist, that key features are bandied about but remain unspecified, and that existing typologies are quite loose in terms of descriptions of dimensions, points along those dimensions, and operationalisation criteria for locating studies relative to these. In the Postscript-Prescript chapter, I suggest a further 9-component grouping of research synthesis studies, itself a 'working hypothesis'.

### **3.4. Rationales, goals, outcomes**

Some detailed lists of the multitude of rationales, goals and outcomes of qualitative research synthesis have been produced (e.g. Dixon-Woods et al., 2004, Major and Savin-Baden 2010), but there does not appear to be a systematic collation and presentation of these more broadly, to extend to other types of research synthesis. Seven broad areas emerged through coding, in NVivo, of relevant contents in methodological texts. Namely, research synthesis studies enable:

- Increasing and stabilising knowledge on a particular topic/in a broad field;
- Solving a pressing practical problem – primarily in clinical practice and policy making;
- Responding to demands for effectiveness, convenience, ethics and accountability in the production and utilisation of information;
- Improving the preconditions for knowledge (such as through identifying research directions or domains to underpin the development of measurement instruments and scales);

---

<sup>48</sup> E.g. the labels on types of synthesis are relatively limited and consistent: narrative, conceptual, chronological, graphical, tabular and qualitative, with some more extended descriptions - e.g. "tabular (matrices, graphs, charts, or networks) usually according to a framework", while the descriptions and labels of, for instance, the search and analysis are quite varied.

- Acquiring ‘knowledge about knowledge’: developing a meta-perspective on the work within a research field;
- Enhancing the relative status of a research field;
- Bringing research in one area to bear on research in another area where previous interactions have been limited – i.e. interdisciplinary work done in a systematic way and at the level of review studies.

Below I give illustrations of claims concerning four of the above – the three types of rationale, goals and outcomes that seem to be invoked most often (acquiring more and more stable knowledge, the resolution of pressing problems, and responding to practical and ethical demands of information use) and one which seems to be gaining momentum through recent methodological developments (acquiring knowledge about knowledge).

#### *3.4.1. Increasing and stabilising knowledge*

Historically, research synthesis work has been seen primarily as enabling the accumulation of large amounts of data. This helps reach conclusions of much greater clarity, generality and confidence than available through individual studies. Synthesis studies generate large samples which may be very difficult, or practically impossible, to get in a single trial. The power of estimates is increased. Random error in the assessments of treatments may also be reduced. This is particularly important in widely practicable interventions with small or modest benefits for common and serious conditions. If their benefit is detected, such interventions can save thousands of lives (Sutton et al, 2000: 10-11). Pooling studies for synthesis has “the ability to offset to some degree the limited scope of single reports”, to strengthen generality by the breadth of geographic and cultural settings (Kearney, 2001: 274).

In qualitative and mixed methods syntheses, the emphasis seems to shift towards finding order in a field of knowledge, acquiring a (more) complete picture of it, and towards theory generation. Synthesis studies “can consolidate a body of widely scattered literature into a usable and coherent whole” (Paterson et al., 2001: 13); enable questions and ideas to be “meaningfully grouped and classified” (Greenhalgh et al., 2005: 423); help map out a full(er),

(more) complete picture concerning an issue of interest or a thematic field (Roberts et al., 2002). They also provide an opportunity for bringing forth easy-to-miss phenomena and difficult-to obtain data types, e.g. about sensitive topics or hard-to-reach populations (Dixon-Woods et al., 2004: 4).

Research synthesis can also enable the generation of (middle-range) theories and models (Estabrooks et al., 1994; Schreiber, Crooks and Stern, 1997; Finfgeld, 1999; Sherwood, 1999; Kearney 2001; Britten et al. 2002). These are likely to be of greater breadth, scope, generalisability, credibility and explanatory power than existing ones (Sandelowski, Docherty and Emden, 1997; Sherwood, 1999; Paterson et al., 2001). The breadth of other carriers of knowledge can also be extended. Synthesis work often comes up with higher order analytical categories (Dixon-Woods et al., 2004) and “meanings that extend well beyond those presented in the available body of knowledge” (Paterson et al., 2001).

#### *3.4.2. Solving a pressing practical problem – primarily in clinical practice and policy making*

The more effective and judicious use of research findings to inform clinical practice comes up as the primary rationale for performing research synthesis work.<sup>49</sup> In the words of Rosenberg and Donald, “[f]or decades people have been aware of the gaps between research evidence and clinical practice, and the consequences in terms of expensive, ineffective, or even harmful decision making”. In early days of evidence-based medicine, when clinicians were expected to identify and appraise the evidence themselves (see, for instance, Evidence-Based Medicine Working Group, 1992), the process was supposed to be prompted by occurrences in everyday clinical practice. It was to be carried out “for our patient”, for this particular “77 year old woman living alone ... admitted with non-rheumatic atrial fibrillation and her first bout of mild left ventricular failure” (Rosenberg and Donald, 1995: 1122). The ‘particular patient’ prompting and the individual clinician’s work of appraising are no longer a significant part of the vision of the synthesis process. Nevertheless, the

---

<sup>49</sup> This particular rationale is usually applied to evidence-based medicine and by extension to research synthesis.

rationale of producing evidence to be “harnessed” for everyday clinical practice (*ibid*) is still presented as paramount.

The other main type of problems to require the identification and integration of research are found at the level of policy and organisational decision making. For instance, meta-analysis and the movement towards evidence-based medicine have been seen as a response to “[a]ccumulated public and private frustrations regarding escalating costs and substantial evidence of variation in medical practice” and as offering “immense” opportunities for improving routine practice (Clancy and Kamerow, 1996: 329). The potential of pooling together of evidence for policy purposes has led to the formation of a number of bodies with such a mandate. For instance, the Health Development Agency (HDA) in the UK “was established to, among other things, build the evidence base in public health, with particular reference to reducing inequalities in health” (Kelly and Swann, 2004: iv). It was thus “faced with the job of drawing together the known evidence and putting it in a form that could be usable for practitioners and policy-makers, and in such a way that the evidence could inform questions about effectiveness and inequalities” (*ibid*). In some cases, concerns about how exactly to bring together knowledge – which are highly relevant to the range of goals under the previous section – are seen as being at odds with practical needs and the urgency of human suffering:

*First, and most importantly, inequalities in health and the human pain and misery that flow from them are too great to be ignored on the grounds of philosophical or methodological problems. Certainly there are epistemological and practical differences between research traditions, but this in itself is not a reason for inaction. The HDA cannot detain itself with what, in our view, is a misrepresentation of the philosophical issues. The practical problem to be solved is that of the premature mortality and excess morbidity that disproportionately affect the poor to a far greater degree than the well-to-do* (Kelly and Swann, 2004: v).

*The HDA’s position is that much of the methodological debate is completely unhelpful from a point of view of trying to bring about reductions in inequalities in health* (*ibid.*).

### *3.4.3. Synthesis work as responding to demands for effectiveness, convenience, ethics and accountability*

A closely related and continuously re-echoed rationale – more out of frustration than a need to persuade – concerns the capacity of research synthesis to enhance dramatically the use of primary research. Integrative studies enable practitioners to deal with the information explosion, deluge of information (Rosenberg and Donald, 1995: 1122), tsunami of data (Hawkes, 2011), mountains of evidence (Pope, Mays and Popay, 2007: 3). They condense knowledge for clinicians, policy makers and researchers in a way that makes its use convenient, even possible. Arguments are also frequent for the avoidance of research duplication or disuse – this is seen as thwarting progress, unethical, “wasteful” (e.g. Sandelowski, Docherty and Emden, 1997: 366-7; Roberts et al., 2002: 1596; Sutton, Cooper and Jones, 2009).

### *3.4.4. Acquiring ‘knowledge about knowledge’*

A less debated potential goal of research synthesis, which, however, seems to be growing in importance with the development of methods like meta-narrative (Greenhalgh et al., 2005) and critical interpretive synthesis (Dixon-Woods et al., 2006), concerns its capacity to question the foundations of a research field. Broad-based research synthesis studies can reveal the cultural, historical, social, disciplinary, etc. situatedness of research; the implications of this for conceptualisations of truth and progress in science; and the formative effects of research on practice, society and its own future. Research synthesis can thus be thought of as a method for engaging in the studies of science.

I will only include one telling example from the work of Greenhalgh et al. (2005) to illustrate the type of insights which such work can offer. The aim of the team’s project, funded by the UK Department of Health, was to synthesise the knowledge base on the “[d]iffusion, spread and sustainability of innovations ... with a view to informing the modernisation agenda for UK health services” (*op. cit.*: 418). What the authors gradually came to recognise was the pro-innovation bias of much innovation research. The specific example of such a bias below concerns the effect of a context of food shortages on political ideology and

scientific priorities, and through them – on the interpretation of data. Years later, in a context of agricultural overproduction and conservation concerns, an entirely different interpretation is automatic to any ‘modern’ reader:

*Back in 1954, one of the Iowa farmers that I personally interviewed for my PhD dissertation research rejected all of the chemical innovations that I was then studying: weed sprays, cattle and hog feeds, chemical fertilisers, and a rodenticide. He insisted that his neighbours, who had adopted these chemicals, were killing their songbirds and the earthworms in the soil. I had selected the new farm ideas in my innovativeness scale on the advice of agricultural experts at Iowa State University; I was measuring the best recommended farming practice of that day. The organic farmer in my sample earned the lowest score on my innovativeness scale, and was categorised as a laggard (Rogers, 1995: 425 cited in Greenhalgh et al., 2005: 424-5).<sup>50</sup>*

At a more abstract level and in a Kuhnian vein, Greenhalgh et al. add that “[r]esearchers in different traditions had conceptualised, explained and investigated diffusion of innovations differently and had used different criteria for judging the quality of empirical work. Moreover, they told very different overarching stories of the progress of their research”. One of the outcomes of Paterson et al.’s meta-study methodology is, similarly, the identification of “both overt and subtle ways in which researchers have contributed to the current interpretation of chronic illness experience” and of practical implications of this interpretation (Paterson, B et al. 2001: Introduction, p. 14). So far, such critical, partly deconstructivist positions, have had limited presence in research synthesis work.

---

<sup>50</sup> Rogers, E. M., 1995. *Diffusion of innovations*. 4th ed. New York: Free Press.



### **3.5. New and emerging synthesis methods – examples and characteristics of the field**

#### *3.5.1. An overview of methods based on four landmark reviews*

A 'master' review for the field of research synthesis, understood in some broad sense, is yet to appear. Texts offering an overview of available methods tend to stay within a particular niche, for instance qualitative synthesis (e.g. Barnett-Page and Thomas, 2009; Major and Savin-Baden, 2011). More up-to-date, comprehensive and rigorous reviews of methods within particular niches are also needed. There are certainly advantages to research isolation and independence, such as incubation of minority ideas. But in the case of research synthesis the negative consequences of field fragmentation are beginning to weigh heavily. There is much duplication of effort, misplaced identity and uniqueness claims, self-righteousness, multiplication of vocabularies, and missed opportunities for the spread of methodological innovation and taking challenge and impetus from closely aligned work.

To start representing the broad field of health research synthesis beyond the systematic review, the methods listed in four reviews were pulled together – Dixon-Woods et al. (2004) on integrating quantitative and qualitative research, Ades and Sutton (2006) on statistical methods for evidence synthesis, Pope, Mays and Popay (2007) on quantitative, interpretative and mixed synthesis methods, and Barnett-Page and Thomas (2009) on qualitative synthesis. The outcome is presented in Table T 2.1 below, along with bibliometric data for citations on PubMed (all and last five years), Web of Science and Google scholar. Briefly, it shows that in terms of PubMed citations for all years, the first five most popular synthesis methods are content analysis (403 citations retrieved), narrative synthesis (315), (qualitative) meta-synthesis (274), case survey (169) and thematic analysis/synthesis (165). The most actively developing methods, judged by the number of publications from the last five years relative to all publications on a method on PubMed, are critical interpretive synthesis (92.9% of publications are from the last five years – 13 out of 14), realist synthesis (90.5% – 19/21), narrative synthesis (80.6% – 254/315), meta-ethnography (76.3% – 58/76) and thematic analysis/synthesis

(77.0% – 127/165). The most popular methods as per citations of exemplary publications on Google scholar are qualitative comparative analysis (due to Ragin, 1987 with over 4,345 citations), hierarchical models of (statistical) evidence synthesis (Sutton et al., 2000 – 1050; Spiegelhalter et al., 2004 – 743), meta-ethnography (Noblit and Hare, 1988 – 995 citations), realist synthesis (Pawson, 2006 – 647) and (qualitative) meta-synthesis (Sandelowski et al., 1997 – 378).<sup>51</sup> With regard to Web of Science citations, meta-ethnography is the leading method (with 188 citations for Campbell et al., 2003), followed by (qualitative) meta-synthesis (Sandelowski et al., 1997 with 186 citations), qualitative comparative analysis (Cress and Snow, 2000 – 152), thematic synthesis (Thomas and Harden, 2008 – 84) and the quantitative case survey (Yin and Heald, 1975 – 76). The table also lays out a range of uncertainties about individuating synthesis methods and identifying exemplar publications.

---

<sup>51</sup> Miles and Huberman, 1994 (with over 37,000 citations) came up highest but their work is a 'predecessor' rather than one proposing a synthesis method.

**Table T 2.1. Overview of alternative research synthesis methods – as identified by Dixon-Woods et al. (2004), Ades and Sutton (2006), Pope et al. (2007), and Barnett-Page and Thomas (2009)**

This is a long table of several layers of information. The following notes describe its contents and the principles for inclusion of information in it.

The references accompanying descriptions of synthesis methods in the above reviews were taken to be “exemplar publications” for those synthesis methods. In some cases the referenced work was described in some detail in the reviews. In other cases only a reference was provided. Exemplar texts from outside health research were included only if they were a pioneering publication which has influenced health research too and/or if no examples from health research were given in the reviews (generally up to two publications from outside health research were included). Overall, exemplar texts include an actual synthesis of studies. In some cases important predecessors where no synthesis is offered or discussed are also listed (e.g. Miles and Huberman’s work, see entry 21 in the table). If this is the case, ‘the year of earliest exemplar’ (column 3 in the table) lists the year of the first worked example in a synthesis context.

If a large number of health research publications were referenced in a review (which happens relatively rarely), up to 5 publications were included from that review (generally by order of presentation by the review authors, which tends to be chronological). Discrepancies between reviews in terms of exemplar publications could result in more than 5 publications presented in the table.

\*\*\*\*\* is used to introduce comments on uncertainties about the information presented (including because of discrepancies between reviews in terms of exemplar publications). They thus indicate findings which limit the drawing of clear conclusions and suggest that more in-depth investigation is required.

Four types of bibliometric data are provided. These are 1) retrieval for the method name in PubMed for all years (search strategies can be found in Appendix to Chapter 2, Section 1); 2) retrieval for the method name in PubMed for the last 5 years; 3) citation counts in the Web of Science; and 4) citation counts in Google scholar. In cases where the Web of Science does not hold a publication type (e.g. books or grey literature), ‘not covered’ is entered in the box (‘not found’ is used when the publication could not be retrieved, either because it is not available on the Web of Science or the searching for it was ineffective).

The full references are given in a separate section in the bibliography, unless cited elsewhere in the dissertation.

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
1.	<b>Bayesian approaches to cost-effectiveness analysis</b>  Pope et al. (2007)	1999	<b>Luce B R and Claxton K (1999)</b> Redefining the analytical approach to pharmacoeconomics.  <b>Turner D, Wailoo A, Nicholson K et al. (2003)</b> Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.	<i>Health Economics</i>  <i>Health Technology Assessment</i>	41	17 (41.5%)	8  Not found	18  152
2.	<b>Bayesian meta-analysis/synthesis</b>  Dixon-Woods et al. (2004)  Pope et al. (2007)	2002	<b>Roberts K A, Dixon-Woods M, Fitzpatrick R et al. (2002)</b> Factors affecting uptake of childhood immunisation: a Bayesian synthesis of qualitative and quantitative evidence.  **** Earlier references to Bayesian meta-analysis available on PubMed.	<i>The Lancet</i>	131	78 (59.5%)	50	97
3.	<b>Case survey (quantitative case survey)</b>  Dixon-Woods et al. (2004)  Pope et al. (2007)	1975	<b>Yin R K and Heald K A (1975)</b> Using the case survey method to analyse policy studies.  <b>Larsson R (1993)</b> Case survey methodology: quantitative analysis of patterns across case studies.  <b>Jensen J L and Rodgers R (2001)</b> Cumulating the intellectual gold of case study research.  **** Last exemplar only in Dixon-Woods et al. (2004). Pope et al. (2007) list it under comparative case study.	<i>Administrative Science Quarterly</i>  <i>Academy of Management Journal</i>  <i>Public Administration Review</i>	169	52 (30.8%)	76  74  26	246  290  214

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
4.	<b>Comparative case study</b>  Pope et al. (2007)  ***** Primary research use only but "has potential application for the synthesis of qualitative studies" (op.cit.: 78)	1991	<b>Jensen J L and Rodgers R (2001)</b> Cumulating the intellectual gold of case study research.  <b>Agranoff R and Radin B A (1991)</b> The comparative case study approach in public administration.	See 3. above  <i>Research in Public Administration</i>	3	2	26  Not found	213  Captured only as cited by others
5.	<b>Comprehensive decision modelling</b>  Pope et al. (2007)	2001	<b>Cooper N, Sutton A and Abrams K (2002)</b> Decision analytic economic modelling within a Bayesian framework: application to prophylactic antibiotics' use for caesarean section.  <b>Dowie J (2001)</b> Towards value-based, science-informed public health policy: conceptual framework and practical guidelines.	<i>Statistical Methods in Medical Research</i>  WHO/Department of Health Consultation	24  ***** see search strategy, Appendix to Chapter 2	9 (37.5%)	Not found  Not found	41  14
6.	<b>Confidence profile method</b>  Ades and Sutton (2006)	1989	<b>Eddy D M (1989)</b> The confidence profile method: a Bayesian method for assessing health technologies.  <b>Eddy D M, Hasselblad V and Shachter R (1992)</b> Meta-analysis by the Confidence Profile Method.	<i>Operations Research</i>  Book	13	0	46  Not covered	73  264

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
7.	<p><b>Content analysis/ synthesis</b></p> <p>Dixon-Woods et al. (2004)</p> <p>Pope et al. (2007)</p> <p>Barnett-Page and Thomas (2009)</p> <p>**** "fledgling approach" according to latter</p>	1995	<p><b>Evans D and Fitzgerald M (2002)</b> Reasons for physically restraining patients and residents: a systematic review and content analysis.</p> <p><b>Suikkala A, Leino-Kilpi H (2001)</b> Nursing student-patient relationships: a review of the literature from 1984–1998.</p> <p><b>Miller D and Reilly J (1995)</b> Making an issue of food safety: the media, pressure groups and the public sphere.</p> <p>**** The reviews choose different exemplars. The first is in Dixon-Woods et al. and Barnett-Page and Thomas, the second in Barnett-Page and Thomas, and the third in Pope et al.</p> <p>**** Retrieval is substantial in PubMed, yet seen as a "fledgling approach" by Barnett-Page and Thomas.</p>	<p><i>International Journal of Nursing Studies</i></p> <p><i>Journal of Advanced Nursing</i></p> <p>"Food and Nutrition as Social Problems" edited collection</p>	403	180 (44.7%)	27  7  Not found	59  52  Available only for whole book (54)
8.	<p><b>Critical interpretive synthesis</b></p> <p>Pope et al. (2007)</p> <p>Barnett-Page and Thomas (2009)</p>	2005	<p><b>Dixon-Woods M, Kirk D, Agarwal S et al. (2005)</b> Vulnerable Groups and Access to Health Care: A Critical Interpretative Synthesis.</p> <p><b>Dixon-Woods M, Cavers D, Agarwal S et al. (2006)</b> Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups.</p>	<p>National Co-ordinating Centre for NHS Service Delivery and Organisation R&amp;D</p> <p><i>BMC Medical Research Methodology</i></p>	14	13 (92.9%)	Not found  Not found	62  209

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
9.	<b>Cross-design synthesis (grouped meta- analysis)</b>  Ades and Sutton (2006)  Pope et al. (2007)	1992	<b>No named authors (1992)</b> Cross-design Synthesis: A New Strategy for Medical Effectiveness Research.  <b>Droitcour J, Silberman G and Chelimsky E (1993)</b> Cross-design synthesis: a new form of meta-analysis for combining results from randomised clinical trials and medical-practice databases.	US General Accounting Office  <i>International Journal of Technology Assessment in Health Care</i>	9	4 (44.4%)	Not found  Not found	8  60
10.	<b>Ecological triangulation</b>  Barnett-Page and Thomas (2009)	<b>2003?</b>  ***** likely year , see column to the right	<b>Banning J (undated – 2003 is the year of a similar paper in the author's publication list)</b> Ecological Triangulation: An Approach for Qualitative Meta-Synthesis.  <b>Banning J (undated)</b> Ecological Sentence Synthesis.	Author's webpages	Not found	Not found	Not found  Not found	16  3

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
11.	<b>EPPI-approach</b>  Pope et al. (2007)	2001	<b>Harden A, Oakley A and Oliver S (2001)</b> Peer-delivered health promotion for young people: a systematic review of different study designs.  <b>Oliver S, Oakley L, Lumley J and Waters E (2001)</b> Smoking cessation programmes in pregnancy: systematically addressing development, implementation, women's concerns and effectiveness.	<i>Health Education Journal</i>  <i>Health Education Journal</i>	41	11 (26.8%)	Not found  Not found	73  14
12.	<b>Estabrooks, Field and Morse's aggregation of findings</b>  Dixon-Woods et al. (2004)	1994	<b>Estabrooks C, Field P and Morse J (1994)</b> Aggregating qualitative findings: an approach to theory development.	<i>Qualitative Health Research</i>	18	8 (44.4%)	Not found	149
13.	<b>Framework analysis/ synthesis</b>  Pope et al. (2007)  Barnett-Page and Thomas (2009)	2005	<b>Lloyd Jones, M. (2005)</b> Role development and effective practice in specialist and advanced practice roles in acute hospital settings: systematic review and meta-synthesis.  <b>Brunton G, Oliver S, Oliver K &amp; Lorenc T (2006)</b> A Synthesis of Research Addressing Children's, Young People's and Parents' Views of Walking and Cycling for Transport.  <b>Oliver S, Rees R, Clarke-Jones L et al. (2008)</b> A multidimensional conceptual framework for analysing public involvement in health services research.	<i>Journal of Advanced Nursing</i>  EPPI-Centre report  <i>Health Expectations</i>	24   ***** see search strategy, Appendix to Chapter 2	18 (75%)   ***** see search strategy, Appendix to Chapter 2	61  Not covered  30	21  21  49



	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
13.	<b>Framework analysis/synthesis, cont.</b>		<p>**** The reviews choose different exemplars. The first reference is in Pope et al. and the remaining two in Barnett-Page and Thomas.</p> <p>**** The first example (Lloyd Jones, 2005) is also included under Miles and Huberman's cross-case analysis (method 21) and thematic analysis/synthesis (method 31). It also contains 'meta-synthesis' in the title and can thus be considered an example of meta-synthesis (method 28).</p>					
14.	<b>Grounded (formal) theory</b>  Dixon-Woods et al. (2004)  Pope et al. (2007)  Barnett-Page and Thomas (2009)	<b>1996</b>	<b>Estabrooks C, Field P and Morse J (1994)</b> Aggregating qualitative findings: an approach to theory development.  <b>Dixon D (1996)</b> Unifying concepts in parents' experiences with health care providers.  <b>Kearney M (2001)</b> Enduring love: a grounded formal theory of women's experience of domestic violence.  <b>Eaves Y D (2001)</b> A synthesis technique for grounded theory data analysis.	See 12. above  <i>Journal of Family Nursing</i>  <i>Research in Nursing and Health</i>  <i>Journal of Advanced Nursing</i>	150	37 (24.7%)	See 12. above  Not found  46  39	See 12. above  29  100  193

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
14.	<b>Grounded (formal) theory</b>		<p><b>Finfgeld D (1999)</b> Courage as a process of pushing beyond the struggle.</p> <p>**** The reviews choose different exemplars, apart from Kearney (2001) which is present in all. The first two papers are from Dixon-Woods et al. and the latter two from Barnett-Page and Thomas.</p> <p>**** Eaves (2001) synthesises grounded theory texts to produce a technique for the analysis of primary data. It is likely to be applicable to synthesis studies but I believe the following claim of Barnett-Page and Thomas to be incorrect: "Eaves undertook her own synthesis of the synthesis methods used by these authors to produce her own clear and explicit guide to synthesis in grounded formal theory" (bmc 2).</p>	<i>Qualitative Health Research</i>			14	61
15.	<b>Hierarchical models</b>  Ades and Sutton (2006)	? **** Very rich field, dedicated historical analysis needed.	<p><b>Sutton A J, Abrams K R, Jones D R et al. (2000)</b> Methods for Meta-Analysis in Medical Research.</p> <p><b>Whitehead A (2002)</b> Meta-analysis of Controlled Clinical Trials.</p> <p><b>Spiegelhalter D J, Abrams K R and Myles J (2004)</b> Bayesian Approaches to Clinical Trials and Health-care Evaluation.</p> <p>**** Numerous further exemplar texts given. The books themselves discuss a variety of approaches.</p>	Book  Book  Book	138	48 (34.8%)	Not covered  Not covered  Not covered	1050  442  743

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
16.	<p><b>Meta-ethnography</b></p> <p>Dixon-Woods et al. (2004)</p> <p>Pope et al. (2007)</p> <p>Barnett-Page and Thomas (2009)</p>	1988	<p><b>Noblit G W and Hare R D (1988)</b> Meta-ethnography: synthesising qualitative studies.</p> <p><b>Paterson B L, Thorne S and Dewis M (1998)</b> Adapting to and managing diabetes.</p> <p><b>Nelson A M (2002)</b> A metasynthesis: mothering other-than-normal children.</p> <p><b>Beck C T (2002a)</b> Mothering multiples: a meta-synthesis of qualitative research.</p> <p><b>Beck C T (2002b)</b> Postpartum depression: a metasynthesis.</p> <p><b>Jensen L A and Allen M N (1994)</b> A synthesis of qualitative research in wellness–illness.</p> <p><b>Britten N, Campbell R, Pope C et al. (2002)</b> Using meta ethnography to synthesise qualitative research: a worked example.</p> <p><b>Campbell R, Pound P, Pope C. et al. (2003)</b> Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care.</p> <p><b>Clemmens D (2003)</b> Adolescent motherhood: a meta-synthesis of qualitative studies.</p>	<p>Book</p> <p><i>Image – the Journal of Nursing Scholarship</i></p> <p><i>Qualitative Health Research</i></p> <p><i>MCN, the American Journal of Maternal/Child Nursing</i></p> <p><i>Qualitative Health Research</i></p> <p><i>Qualitative Health Research</i></p> <p><i>Journal of Health Services Research and Policy</i></p> <p><i>Social Science &amp; Medicine</i></p> <p><i>MCN, the American Journal of Maternal/Child Nursing</i></p>	76	58 (76.3%)	<p>Not covered</p> <p>Not found</p> <p>43</p> <p>Not found</p> <p>94</p> <p>Not found</p> <p>Not found</p> <p>188</p> <p>Not found</p>	<p>995</p> <p>153</p> <p>91</p> <p>57</p> <p>206</p> <p>89</p> <p>222</p> <p>369</p> <p>91</p>

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
16.	<b>Meta-ethnography, cont.</b>		<p><b>Sandelowski M and Barroso J (2007)</b> <i>Handbook for Synthesizing Qualitative Research</i></p> <p>**** The reviews choose different exemplars. Of the 10 publications above, 3 are referenced in all three reviews – Noblit and Hare (1988), Britten et al. (2002) and Campbell et al. (2003).</p> <p>**** Becks (2002 a and b), Nelson (2002), Clemmens (2003) and Sandelowski and Barroso (2007) refer to their work as 'meta-synthesis'.</p>	Book			Not covered	265
17.	<p><b>Meta-interpretation</b></p> <p>Barnett-Page and Thomas (2009)</p> <p>**** qualified as "fledgling approach"</p>	2005	<p><b>Weed M (2006)</b> "Meta Interpretation": A Method for the Interpretive Synthesis of Qualitative Research.</p> <p>**** Bibliometric search returns Finfgeld (1999) as another example of meta-interpretation (see 14. on grounded theory).</p>	<i>Forum: Qualitative Social Research (Sozialforschung)</i>	5	3	Not found	69
18.	<p><b>Meta-narrative</b></p> <p>Barnett-Page and Thomas (2009)</p>	2005	<p><b>Greenhalgh T, Robert G, Macfarlane F et al. (2005)</b> Storylines of research in diffusion of innovation: a metanarrative approach to systematic review.</p>	<i>Social Science &amp; Medicine</i>	43	32	75	199

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
19.	<b>Meta-needs assessment</b>  Dixon-Woods et al. (2004)	2000	<b>Gaber J (2000)</b> Meta-needs assessment. **** <i>The extent to which meta-needs assessment can be considered a separate approach is questioned by Dixon- Woods et al.: "Meta-needs assessment provides a general framework combining elements of other approaches rather than a distinctive new approach in its own right" (27).</i>	<i>Evaluation and Program Planning</i>	Not found	Not found	1	14
20.	<b>Meta-study</b>  Dixon-Woods et al. (2004)  **** <i>in Pope et al. (2007) mentioned only as offshoot of meta-ethnography</i>  Barnett-Page and Thomas (2009)	2001	<b>Paterson B L, Thorne, S E, Canam C and Jillings C (2001)</b> Meta-study of qualitative health research.  <b>Zhao S Y (1991)</b> Metatheory, Metamethod, Meta-Data-Analysis – What, Why, and How?	Book  <i>Sociological Perspectives</i>	48	25 (52.1%)	Not covered  35	341  80
21.	<b>Miles and Huberman's cross-case analysis/ synthesis (meta-matrix)</b>  Dixon-Woods et al. (2004); Pope et al. (2007)	2000	<b>Miles M B and Huberman M (1994)</b> Qualitative data analysis: an expanded sourcebook.  <b>McNaughton D B (2000)</b> A synthesis of qualitative home visiting research.  <b>Lloyd Jones, M. (2005)</b> Role development and effective practice in specialist and advanced practice roles in acute hospital settings: systematic review and meta-synthesis.	Book  <i>Public Health Nursing</i>  <i>See 13. on framework analysis</i>	8	3	Not covered  34  <i>See 13. above</i>	<b>37,056</b>  72  <i>See 13. above</i>

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
22.	<b>Multiparameter evidence synthesis</b>  Ades and Sutton (2006)	<b>2002?</b> ***** <i>Unclear how important pre- decessors</i>	<b>Ades A E and Cliffe S (2002)</b> Markov Chain Monte Carlo estimation of a multi- parameter decision model: consistency of evidence and the accurate assessment of uncertainty.	<i>Medical Decision Making</i>	10	7	48	78
23.	<b>Narrative summary</b>  Dixon-Woods et al. (2004)	<b>1999</b>	<b>Harden A, Weston R and Oakley A (1999)</b> A review of the effectiveness and appropriateness of peer-delivered health promotion for young people. (given as an example of a theory-led approach)	EPPI-Centre report	105	59 (56.2%)	Not found	74
24.	<b>Narrative synthesis</b>  Pope et al. (2007)	<b>2006</b>	<b>Popay J, Roberts H, Sowden A et al. (2006)</b> Guidance on the conduct of narrative synthesis in systematic reviews.  <b>Arai L, Britten N, Popay J et al. (2007)</b> Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions.	ESRC methods programme report  <i>Evidence &amp; Policy</i>	315	254 (80.6%)	Not covered  Not found	167  34

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
25.	<p><b>Qualitative comparative analysis</b></p> <p>Dixon-Woods et al. (2004)</p> <p>Pope et al. (2007)</p>	1987	<p><b>Ragin C C (1987)</b> The comparative method: moving beyond qualitative and quantitative strategies.</p> <p><b>Ragin C C (2000)</b> Fuzzy set social science.</p> <p><b>Melinder K A and Andersson R (2001)</b> The impact of structural factors on the injury rate in different European countries.</p> <p><b>Cress D. and Snow D. (2000)</b> The outcomes of homeless mobilization: the influence of organization, disruption, political mediation, and framing.</p> <p>***** <i>Only first publication referenced in both reviews.</i></p>	<p>Book</p> <p>Book</p> <p><i>European Journal of Public Health</i></p> <p><i>American Journal of Sociology</i></p>	36	16 (44.4%)	<p>Not covered</p> <p>Not covered</p> <p>12</p> <p>152</p>	<p>4345</p> <p>1782</p> <p>18</p> <p>370</p>
26.	<p><b>Qualitative meta-analysis</b></p> <p>Dixon-Woods et al. (2004)</p> <p>Pope et al. (2007)</p>	1985? (see right)	<p><b>Schreiber R, Crooks D and Stern P N (1997)</b> Qualitative meta-analysis.</p> <p><b>McCormick J, Rodney P and Varcoe C (2003)</b> Reinterpretation across studies: an approach to meta-analysis.</p> <p>***** <i>Reviews choose different exemplar publications.</i> ***** <i>Schreiber et al. refer to an earlier use of the term (Stern and Harris, 1985).</i></p>	<p>“Completing a qualitative project” edited collection</p> <p><i>Qualitative Health Research</i></p>	45	20 (44.4%)	<p>Not covered</p> <p>30</p>	<p>107</p> <p>73</p>

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
27.	<p><b>Qualitative metasummary</b></p> <p>Barnett-Page and Thomas (2009)</p> <p>**** qualified as "fledgling approach"</p>	2007	<p><b>Sandelowski M and Barroso J (2007)</b> <i>Handbook for Synthesizing Qualitative Research.</i></p> <p>**** Bibliometric searches suggest a further article – Sandelowski M, Barroso J and Voils C I (2007).</p> <p>**** Book also included as illustrative of meta-ethnography, see 16.</p>	See 16. on meta-ethnography	15	9 (60.0%)	See 16. above	See 16. above
28.	<p><b>(Qualitative) meta-synthesis</b></p> <p>Dixon-Woods et al. (2004)</p> <p>Pope et al. (2007)</p>	1994	<p><b>Jensen L A and Allen M N (1994)</b> A synthesis of qualitative research in wellness–illness.</p> <p><b>Sherwood G (1997)</b> Meta-Synthesis of Qualitative Analyses of Caring: Defining a Therapeutic Model of Nursing.</p> <p><b>Sandelowski M, Docherty S and Emden C (1997)</b> Focus on qualitative methods. Qualitative meta-synthesis: issues and techniques.</p> <p><b>Sherwood, G. (1999)</b> Meta-synthesis: merging qualitative studies to develop nursing knowledge.</p>	<p>See 16. on meta-ethnography</p> <p><i>Advanced Practice Nursing Quarterly</i></p> <p><i>Research in Nursing and Health</i></p> <p><i>International Journal for Human Caring</i></p>	274	183 (66.8%)	<p>See 16. above</p> <p>Not found</p> <p>186</p> <p>Not found</p>	<p>See 16. above</p> <p>55</p> <p>378</p> <p>62</p>



	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
28.	(Qualitative) meta-synthesis, <i>cont.</i>		<p><b>Higginson I J, Finlay I, Goodwin D M et al. (2002)</b> Do hospital-based palliative teams improve care for patients or families at the end of life?</p> <p><b>Nelson A M (2002)</b> A meta-synthesis: mothering other than normal children.</p> <p><b>Schreiber R, Crooks D and Stern P N (1997)</b> Qualitative meta-analysis.</p> <p>**** Very strong presence in Dixon-Woods et al. (all but last reference), only minimal in Pope et al. – they discuss it in relation to terminology of interpretive methods and do not individuate as a method.</p> <p>**** Some overlap with references under 16. meta-ethnography. Also, a number of the references classified by Dixon-Woods et al. under 'meta-ethnography' and not appearing here contain 'meta-synthesis' in their title.</p> <p>**** Pope et al. include Schreiber et al. (1997) here, while Dixon-Woods et al. include it under qualitative meta-analysis (where Pope et al. choose a different exemplar).</p>	<p><i>Journal of Pain and Symptom Management</i></p> <p>See 16. on meta-ethnography</p> <p>See 26. on qualitative meta-analysis</p>			130	214
							See 16. above	See 16. above
							See 26. above	See 26. above

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
29.	<b>Realist synthesis</b>  Pope et al. (2007)	2002	<p><b>Pawson R (2002)</b> Evidence-based policy: in search of a method.</p> <p><b>Pawson R (2006)</b> Evidence-based Policy: A Realist Perspective.</p> <p><b>Pawson R and Bellamy J (2006)</b> Realist synthesis: an explanatory focus for systematic review.</p> <p><b>Pawson R, Greenhalgh T, Harvey G and Walshe K (2005)</b> Realist review – a new method of systematic review designed for complex policy interventions.</p>	<p><i>Evaluation</i></p> <p>Book</p> <p>Discussion paper for the Health Development Agency</p> <p><i>Journal of Health Services Research and Policy</i></p>	21	19 (90.5%)	<p>Not found</p> <p>Not covered</p> <p>Not covered</p> <p>Not found</p>	<p>223</p> <p>647</p> <p>10</p> <p>506</p>
30.	<b>Textual narrative synthesis</b>  Barnett-Page and Thomas (2009)	2004? (see right)	<p><b>Lucas P J, Baird J, Arai L et al. (2007)</b> Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews.</p> <p><b>Harden A, Garcia J, Oliver S et al. (2004)</b> Applying systematic review methods to studies of people's views: an example from public health research.</p> <p>***** <i>Harden et al. (2004) speak of thematic analysis and do not mention "textual narrative synthesis". Harden is in the acknowledgements of Barnett-Page and Thomas's paper – there should be good but unarticulated reasons.</i></p>	<p><i>BMC Medical Research Methodology</i></p> <p><i>Journal of Epidemiology and Community Health</i></p>	Not found	Not found	<p>19</p> <p>60</p>	<p>47</p> <p>129</p>

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
31.	<p><b>Thematic analysis/ synthesis</b></p> <p>Dixon-Woods et al. (2004)</p> <p>Pope et al. (2007)</p> <p>Barnett-Page and Thomas (2009)</p>	<p>1999? ***** Very basic primary research method – too late for first synthesis use?</p>	<p><b>Garcia J, Bricker L, Henderson J et al. (2002)</b> Women's views of pregnancy ultrasound: a systematic review.</p> <p><b>Harden A, Weston R and Oakley A (1999)</b> A review of the effectiveness and appropriateness of peer-delivered health promotion for young people.</p> <p><b>Neill S J (2000)</b> Acute childhood illness at home: the parents' perspective.</p> <p><b>Dixon-Woods M, Cavers D, Agarwal S et al. (2006)</b> Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups.</p> <p><b>Lloyd Jones M (2005)</b> Role development and effective practice in specialist and advanced practice roles in acute hospital settings: systematic review and meta-synthesis.</p> <p><b>Thomas J and Harden A (2008)</b> Methods for the thematic synthesis of qualitative research in systematic reviews.</p> <p>***** <i>The reviews choose different exemplars. This can, however, be expected as thematic analysis is a broadly used and flexible method. The review authors choose to exemplify specific aspects of its use, e.g. Pope et al. point to Lloyd Jones (2005) as an example of sophisticated inter-relating of themes rather than juxtaposition.</i></p>	<p><i>Birth</i></p> <p><i>See 23. on narrative summary</i></p> <p><i>Journal of Advanced Nursing</i></p> <p><i>See 8. on critical interpretive synthesis</i></p> <p><i>See 13. on framework analysis</i></p> <p><i>BMC Medical Research Methodology</i></p>	165	127 (77.0%)	64	133
							See 23. above	See 23. above
							10	27
							See 8. Above	See 8. above
							See 13. above	See 13. Above
							84	207

	Approach; review(s) in which mentioned	Year of earliest exemplar	Author (year) Title	Source	PubMed retrieval Apr-May 2013	PubMed, last 5 yrs No (% of all)	Web of Science retrieval Apr-May 2013	Google scholar retrieval Apr-May 2013
31.	<b>Thematic analysis/ synthesis</b>		<p>***** <i>Dixon-Woods et al. refer to Neill (2000) suggesting that that thematic analysis was used, incorrectly labelled as content analysis.</i></p> <p>***** <i>Uncertain phrasing used, e.g. in Dixon-Woods et al.: "Other papers that appear to have adopted thematic approaches to synthesis include Harden et al. (1999)" (15).</i></p> <p>***** <i>Barnett-Page and Thomas relate the thematic synthesis of Thomas and Harden (2008) to meta-ethnography and grounded theory rather than thematic analysis in primary research.</i></p>					

### 3.5.2. Examples of methods – extended descriptions

Below I describe four methods in greater detail, to give a more tangible sense of the nature of integrative projects outside the mainstream systematic review.

Narrative synthesis is a highly flexible synthesis method used to combine heterogeneous studies. Bayesian synthesis (Roberts et al., 2002) is a highly structured method that brings together quantitative and qualitative studies and explicitly incorporates subjective judgement. Meta-narrative (Greenhalgh et al., 2005) pays close attention to the ‘paradigm’ within which a study was generated and the combinability of studies from different paradigms. As a counterpoint in a field dominated by applied sciences proposals, I added a model coming from philosophy (Cartwright and Stegenga, 2011).

#### 3.5.2.1. Narrative synthesis

In the context of the conventional systematic review, narrative synthesis is performed if meta-analysis is “either not feasible or not sensible” (Higgins and Green, 2008: 245). Put simply, narrative synthesis is a formalised version of the expert literature review (Greenhalgh et al., 2005).<sup>52</sup> As per the most detailed guidance on narrative synthesis (Popay et al., 2006), although the latter can incorporate the manipulation of statistical data, it is primarily a textual approach aiming to “tell the story” – a trustworthy story – of the findings from the included studies (*op.cit.*: 5). It can be used in preparation for a specialist synthesis, such as meta-analysis or meta-ethnography; instead of an intended specialist synthesis, when the identified studies are found to be too dissimilar to allow for one; and when the review question demands the inclusion of a broad variety of studies (*op.cit.*: 7, 14). Narrative synthesis tends towards the “simple” juxtaposition of findings, but may involve a greater degree of integration and/or interpretation (7).

Four main elements to a narrative synthesis process are identified in the guidance (11-16): 1) development of a “theory of change” – a theory of how the intervention works, why and for whom (not all reviewers performing a narrative

---

<sup>52</sup> Which is also sometimes referred to as “narrative review” but not “narrative synthesis”.

synthesis choose to do this); 2) development of a preliminary synthesis; 3) exploration of relationships – between characteristics of studies and their findings, and between the findings of different studies; and 4) assessment of the robustness of the synthesis. For each of these elements, the authors suggest a range of tools and techniques (16-22).<sup>53</sup> The preparation of a preliminary synthesis may be enabled by producing brief textual descriptions of studies; generating groupings and clusters of studies; tabulations of results, study characteristics and quality assessment outcomes; transformations of quantitative data into common metrics; “vote counting” to calculate the frequency of different types of results across studies; “translating data” through thematic or content analysis, etc. (16-19). The tools that focus on the synthesis process may include, for example, critical reflections on the synthesis process and contacting primary study authors to test the validity of synthesis interpretations (Popay et al., 2006: 22).

#### 3.5.2.2. Bayesian synthesis

Roberts et al. (2002) develop a Bayesian approach to the synthesis of qualitative and quantitative evidence in the context of exploring factors affecting the uptake of childhood immunisation in developed countries. Within a Bayesian framework, the synthesis of research evidence is seen as a decision making process which individuals approach with pre-existing beliefs, subjective judgements and access to external sources of evidence. These contribute to “a prior probability distribution” (*op. cit.*: 1596-7). In generating prior distributions in this study, the reviewers first stated, independently of one another, their subjective beliefs about the factors likely to affect uptake of childhood immunisation and ranked them. They then read, in a randomised order, the qualitative studies retrieved through the literature searches, extracted relevant factors from them and indicated their relative importance. Each reviewer then produced a revised list of ranked factors. As a result of accessing the qualitative literature, pre-existing subjective beliefs shifted substantially and grew more similar, and some new factors were identified (*op. cit.*: 1597). Finally, common

---

<sup>53</sup> In their worked examples, the authors find that tools identified for one stage are quite similar to tools identified for another stage; do not find all tools helpful; and do not use some of the proposed tools. Also, they have not been able to identify any specific tools and techniques for developing a theory of change.

descriptive categories were generated so as to bring the reviewers' lists together. The rankings for each factor from each reviewer were combined into a communal prior probability, with the probabilities for all factors calibrated to sum to unity (1597).

Data extraction from the quantitative studies followed. The data were coded into the categories so far generated, with new categories also being added. The evidence obtained from the quantitative studies was then combined with the prior probability for a particular factor to form a posterior probability that this factor was important in determining immunisation uptake. For each factor, Bayes factor methods were used to compare two fixed effect meta-regression models for the log odds of uptake, with data from the quantitative studies used in fitting the models (1597-8). The addition of quantitative data resulted in posterior probabilities that differed, at times substantially, from the prior probabilities. For instance, the probability for "child's health" determining immunisation uptake increased, and the factor was also raised from third most important to most important. The probabilities for the other two most important factors decreased (health professionals' advice and structural issues) (1598). Findings were interpreted as showing that "use of either qualitative or quantitative research alone might not identify all relevant factors, or might result in inappropriate judgments about their importance, and could thus lead to inappropriate formulation of evidence-based policy" (1596).

### 3.5.2.3. *Meta-narrative*

Meta-narrative is a method for synthesising evidence across multiple disciplinary fields developed by Greenhalgh et al. (2005), through a project for systematically reviewing the literature on the diffusion, spread and sustainability of innovation (*op. cit.*: 418). It is conceptualised as drawing on Kuhn's notion of paradigm while taking into account criticisms of the idea of incommensurability suggesting "that apparent differences in findings can always be explained at some level of abstraction" (428). Initially, the "over-arching storylines of the rise and fall of diffusion research" within each research tradition of interest were described. Identified were the key conceptual, theoretical, methodological and instrumental elements of the research 'paradigm'; its key actors, events and

discoveries; and its prevailing language and imagery. Thirteen “meta-narratives” were thus developed, for fields as disparate as rural sociology, communication studies, marketing, health promotion, evidence-based medicine and complexity studies. Then, primary studies in each of the traditions were judged according to the tradition’s internal quality criteria, as derived through the analysis of seminal sources. Further to that, seven key dimensions concerning the diffusion of innovation in organisations were identified. The concepts and findings from each tradition concerning each of these dimensions were “distilled” from the meta-narratives and a narrative account of these given. Epistemological, as well as “pragmatic and realistic” explanations were sought for differences in findings and recommendations between traditions. Finally, the key messages from the literature and other relevant evidence (e.g. budget, policy making priorities) were brought together and recommendations for practice, policy and research were developed.

#### *3.5.2.4. Evidence synthesis through mechanism-explicit causal models – a proposal from philosophy*

This last example of emerging synthesis methods is of a model proposed by Cartwright and Stegenga (2011). (The designation above was added by me. Strictly speaking, the paper is not on research synthesis – phrases such as research/evidence synthesis, systematic review, etc. have not been used in it and the suggested scheme for “amalgamating” or combining evidence is one amongst several key concerns.)

Cartwright and Stegenga outline a theory “for the use of evidence in predicting policy effectiveness *in situ*” (*op. cit.*: 292) – for evaluating the probability that, were a proposed policy to be implemented in the way it would in fact be implemented, it will produce the targeted outcome (292, 294). The theory has three central principles.

The first principle is that a reliably successful way of evaluating the likelihood of a policy producing a targeted outcome is to construct a causal model. This is to include a list of the causes at work in the situation independent of the policy action, any changes in these resulting from the policy action, and a rule for calculating the joint effect of the causes (298). Although additivity of effects is a



common assumption in policy making – that adding a good thing can only improve results – things are not so simple (299). Ideally, the rule for calculating the joint effect is specified through an equation (303). On the one hand, such quantitative precision may be quite difficult to achieve in a policy context. On the other, equations too can leave a lot out, and a lot of what is relevant to policy success (303, 319).

The second central principle of the proposed theory is that a cause is, following J. L. Mackie, an INUS condition - an **I**nsufficient but **N**ecessary part of an **U**nnecessary but **S**ufficient condition (304). This philosophical conceptualisation has a good parallel in epidemiology where a sufficient cause is thought of as a constellation of component causes which, together, are sufficient to cause a disease. Diagrammatically, this is often represented through pie charts, where a slice is a component cause (= INUS condition) and a pie is a sufficient cause (*ibid.*).

Finally, the third main principle proposed is that “mechanisms matter”. Quite informally, a mechanism is thought of as an answer to the question of *How would the policy variable bring about the desired effect?* (317). The tracing of the causal pathway from variable to effect helps identify the auxiliary factors required if the policy variable is to operate successfully (*ibid.*).

The above principles are intended as the basis for a guide to help users construct their own causal models and judge their adequacy against available evidence (321). This is in contrast to the predominant tendency in synthesis proposals coming from the empirical field. Evidence synthesis is researchers’ business, albeit it may be user-driven and incorporate substantial user input.

### 3.6. Some further characteristics of the field suggested by difficulties in identifying (exemplar) methods

*Indefatigable I build in the light*  
*Ruthless I destroy in the night*<sup>54</sup>

Table T 2.1 highlighted a range of uncertainties concerning the history and exemplars of alternative synthesis method and, in some cases, my extraction of information from the reviews. In the process of obtaining information for the table, I encountered a number of further difficulties and limitations which are briefly described below. This negative meta-information, whose provision responds to demands for transparency and self-reflexivity, is also positive substantive information about the field and phenomena of interest.

#### 3.6.1. *We do not know what we do not know*

To my knowledge, Table T 2.1 is the most comprehensive formalised list of research synthesis methods (as opposed to reference lists to methodological articles, for instance). Nevertheless, it is ruefully incomplete. Consider the following estimates. The total of PubMed citations identified by searches for the above 31 method names was found to be 2,344. The citation count for systematic reviews inclusive of meta-analysis on PubMed is currently 12,777.<sup>55</sup> The citation count for systematic reviews is 37,012. This means that the titles, abstracts and indexing terms of 24,235 systematic reviews held on PubMed do not include 'meta-analysis'. A number of 2,344 does not go far in accounting for the methods those reviews have used, even when we acknowledge the likelihood that a large a proportion of those 24,235 reviews may not have been properly labelled and in fact include a meta-analysis or offer the default option of a narrative review. It accounts for even less than appearing at first sight as many alternative syntheses are not indexed as 'systematic reviews'.

---

<sup>54</sup> Debelyanov, D., 1910. *Black Song*. Original in Bulgarian.

<sup>55</sup> "Systematic review" AND meta-analysis. This and the following searches carried out 02 Mar 13.

### 3.6.2. Issues of nomenclature

At least some of the above inefficiency of identification of methods is due to issues of nomenclature. A key limitation of the PubMed searches is the assumption that the method names identified in the landmark reviews are relatively well established. This does not seem to be the case. Much more extensive than the simple search strategies used (see Appendix to Chapter 2, Section 1) are required for a rigorous exploration.

For instance, some of the most popular terms in the field are polysemous. The field's fragmentation also leaves authors unaware of relevant polysemy and few, if any, attempts at drawing distinctions are made. The term "meta-synthesis" and its combinations with "qualitative"<sup>56</sup> is one such example. It is used to refer to: 1) a phase within methods for the synthesis of systematic reviews (e.g. within the EPPI-approach, as per Pope, Mays and Popay, 2007, and the approach described by Ryan, Kaufman and Hill, 2009); 2) a number of particular methods for the synthesis of qualitative research, the degree of commonality between which is unknown (e.g. Jensen and Allen, 1996; Sherwood, 1997); 3) at least one method for the qualitative synthesis of *quantitative* research (Higginson et al., 2002); 4) at least one method for the synthesis of systematic reviews (Ryan, Kaufman and Hill, 2009); 5) the synthesis of qualitative research most broadly (e.g. Sandelowski, Docherty and Emden, 1997); 6) the *interpretative* synthesis of qualitative research more narrowly (as documented in Pope, Mays and Popay, 2007: 75); and, 7) not infrequently, for more than one of these in the same broad context with narrower contexts determining particular meanings (Ryan, Kaufman and Hill, 2009). Although meta-synthesis is probably the clearest example of such polysemy, it is by no means an exception.<sup>57</sup> The reverse of this problem – the

---

<sup>56</sup> E.g. "qualitative meta-synthesis" or meta-synthesis of qualitative research, studies, etc.

<sup>57</sup> The terms 'narrative synthesis', 'thematic analysis' (used in a synthesis context), 'thematic synthesis', 'content analysis' (used in a synthesis context) and 'meta-study' exhibit similar behaviours.

application of a number of terms for practically the same methodological entity – can also be observed.<sup>58</sup>

#### **4. The unit(s) of analysis/ synthesis in research synthesis studies**

I will now be turning to the next major section in the literature review – on candidate units for the most basic unit of analysis/synthesis in research synthesis.<sup>59</sup> ‘Evidence’, ‘data’, ‘findings’, ‘research’ and ‘studies’ seem to be the most frequent objects of integrative processes to be referred to in methodological texts on health research synthesis, but an abundance of other concepts is also used. Dixon-Woods et al. write, for instance, of integrating “qualitative and quantitative *forms of evidence*”<sup>60</sup> or “qualitative *approaches with trial designs*” (Dixon-Woods et al., 2004: 1). Pawson occasionally refers to *methods* as that which is to be “blended” (Pawson, 2002a: 179). Greenhalgh et al. take as their initial unit of analysis “*the unfolding ‘storyline’ of a research tradition over time*” (Greenhalgh et al., 2005: 419). The Cochrane Handbook emphasises that systematic reviews have *studies*, as opposed to study reports, as their “primary units of interest and analysis”, yet clarifies that *reports of studies* are currently the most convenient and primary source of information about studies and their results (Higgins and Green, 2008: 95, 97, 167). Pope, Mays and Popay (2007) suggest that the *findings* of interest in an interpretive synthesis may be “the *interpretations* offered by the authors, typically in the form of *analytical concepts, metaphors or themes*”, while in a realist synthesis the focus will be “more on overarching *theories or explanations* which can be

---

<sup>58</sup> Detailed analysis of specific uses is needed, but at least some uses of, for instance, qualitative meta-analysis, thematic analysis and narrative summary overlap with uses of (qualitative) meta-synthesis, thematic synthesis and narrative synthesis.

In addition to the polysemy and polynomy described above, there are also cases of ‘ignorant’ overlap of the meaning of terms – e.g. the familiar from primary research confusion between thematic and content analysis.

<sup>59</sup> Here I am discussing those units of analysis/synthesis that represent and carry knowledge. There is a further typical, maybe even more typical sense, of ‘unit of analysis’ in the context of research synthesis, in terms of the object of which data have been collected – e.g. an individual, hospital, class of intervention programmes, etc.

<sup>60</sup> All italics in this paragraph have been added.

synthesized” (*op. cit.*: 42). Britten draws attention to the layering of interpretations in line with ideas of Alfred Schutz (1962): “the building blocks for the synthesis were the second-order interpretations of the original studies, from which we constructed several (third-order) interpretations” (Britten et al, 2002: 211).

The unit of analysis/synthesis is rarely placed into sharp focus and discussed in the abstract. When done, presentation is relatively brief; a problem is laid bare and only a pragmatic solution developed, e.g. study findings comprise all the text labelled as ‘results’ or ‘findings’ (Thomas and Harden, 2008); discussion is circumscribed as highly contextual (e.g. qualitative synthesis only, or a particular method only) (Paterson et al., 2001: 7-8, 10; Greenhalgh et al., 2005: 419, 423); choices of units and meanings argued for are inconsistently followed (Fingeld, 2003; Estabrooks et al., 1994); the argument has rhetorical strength but, on deeper analysis, runs into difficulties (Paterson et al., 2001 and Greenhalgh et al., 2005; see Appendix to Chapter 2, Section 2 for further details). The main theoretical concern in debates on the units of analysis/synthesis seems to be derivativeness and distance from the data – that a synthesis study performs an “analysis of analyses” (Paterson et al., 2001: 10); that it represents “findings thrice removed” (Sandelowski and Barroso, 2007: xvi).

I will focus on five supposedly basic carriers of empirical knowledge – potential candidates for the most fundamental unit of analysis/synthesis in research synthesis studies. *Evidence*, *data* and *findings* are the most frequently mentioned and problematised information carriers in methodological texts. *Claims* (sentences, statements) is infrequently used there but has been occasionally problematised. It is also of a similarly basic level as the others. *Facts* is found almost exclusively in the meta-scientific literature (the term is very rare in recent health research texts) but again has substantial overlap with the previous four concepts. There is also an up-and-coming research niche in philosophy on the “travelling” and re-use of facts (discussed later in this review: Howlett and Morgan, 2011; Leonelli, 2012a). Many of its concerns are akin to my concern with transformations in research synthesis. I address *observation* only as part of the debates on the other units. It is another concept of differential

importance between the (health) sciences and the meta-scientific fields. It is fundamental to debates in the philosophy of science; not entirely uncommon, but certainly not particularly important, in primary health research; and practically non-existent in the research synthesis discourse.

There is no way of doing justice to any of these concepts apart from at a book series length. I outlined the general selection principles used in the literature review in the Introduction. Further to these, I prioritised discussions that relate data, evidence, findings, etc. to some solid ground, to “the basis for our knowledge, the ultimate evidence” (Glymour, 1980: 10). In this way I sought to anchor arguments about the reliability of evidence, which is one of the main concerns of the thesis. I also focus on the variety and types of the above units and features along which variety and types can be constructed. Presumably, transformations of evidence are easier and more effective when the types of evidence to be brought together are ‘somehow’ more compatible.

#### **4.1. Data**

##### *4.1.1. Data in the methodological discussion*

In the Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2008) ‘data’ has been defined, “for the purposes of this chapter”, as “any information about (or deriving from) a study, including details of methods, participants, setting, context, interventions, outcomes, results, publications and investigators” (*op. cit.*: 156). An accompanying table further specifies the items to consider in data collection or extraction (157)<sup>61</sup>.

The most detailed presentation of formal types of data in the Cochrane Handbook includes the following:

- 1. dichotomous (or binary) data, where each individual’s outcome is one of only two possible categorical responses;*

---

<sup>61</sup> These include: data about the *source* (e.g. study ID, citation and contact details); *eligibility* for review (confirmation thereof or reasons for exclusion); *methods* (e.g. study design, blinding procedures); *participants* (e.g. data on numbers, diagnostic criteria, age, sex); *interventions* (e.g. details of the intervention, sufficient to allow replication if feasible); *outcomes* (e.g. outcomes collected, outcome definitions, unit of measurement); *results* (e.g. sample size, means and standard deviations); *miscellaneous* (e.g. key conclusions of the study authors, miscellaneous comments of the review authors).

2. *continuous data, where each individual's outcome is a measurement of a numerical quantity;*
3. *ordinal data (including measurement scales), where the outcome is one of several ordered categories, or generated by scoring and summing categorical responses;*
4. *counts and rates calculated from counting the number of events that each individual experiences; and*
5. *time-to-event (typically survival) data that analyse the time until an event occurs, but where not all individuals in the study experience the event (censored data) (op. cit.: 249-250).*

In a broader statistical context, there is a fundamental division between nominal, ordinal, interval and ratio scales and associated data. In nominal scales, items are labelled rather than ranked (e.g. male – female). In ordinal scales, items are ordered along a continuum (e.g. questionnaire respondents on a life stress scale). In interval scales, items are ordered along a continuum where equal intervals represent equal differences. Celsius and Fahrenheit temperature scales are typical examples. Ratio scales have the added characteristic of having a true zero point, which allows to make claims including phrases such as “half as much”, “1.222 times as long as X”. Common ratio scales are those of time, weight, length, volume, etc. (Howell, 2011: 18-24). Overall, there are some well established types of numerical/ quantitative/ statistical data.

In the qualitative research literature, much of the ‘data discussion’ is dedicated to the strengths of qualitative data. These include, for instance, their “focus on *naturally occurring, ordinary events in natural settings*” (Miles and Huberman, 1994: 10; italics in the original); their capacity to enable “a strong handle on what ‘real life’ is like”; their “local groundedness”, referring primarily to the fact that the data are collected on specific cases embedded in their local context; their richness, holism and vividness, which have a strong impact on the reader (*ibid*). Frequently highlighted are also the prior conceptual contents embedded in data and the “rapid ramification” of a “descriptive, first-order ‘fact’” into interpretations and explanations, first of the people being studied and then into the overlaid interpretations and explanations of the researcher (*op.cit.*: 9). Such

features are partly associated with the potentially non-representational, in terms of some uniform external reality, nature of data (Green and Thorogood, 2004: 87). The validity of qualitative data as accounts, perceptions, interpretations of the world is often defended against a positivist expectation that data should be a proxy representation of the external world (*op. cit.*: 89).

The typologies of qualitative data are less well established than those of statistical data. The most widely used classifications seem to be in terms of source (e.g. interview, documentary analysis, focus group data, etc.) and modality (e.g. textual, visual, auditory and multimodal data).

In addition to such more or less standard typologies of quantitative and qualitative data, there are numerous other labels with an extremely high frequency of usage in the research literature. The following list is a selection based on phrases appearing in the *Cochrane Handbook* (Higgins and Green, 2008) and Green and Thorogood's *Qualitative Methods for Health Research* (2004).

Most obviously, data in both texts were qualified in terms of their thematic contents or 'object' to which they related (e.g. "survival data", "quality of life data", "data on blood pressure and proteinuria") and/or methodological provenance (e.g. "clinical trials data", "survey data", "observational data"). But many other implicit classifications can be found, too.

A very clear one goes, for instance, along the lines of the scientific virtues or usability of data. The 'troublesome' data or data pools can be "incomplete", "missing", "insufficient", "skewed", "sparse", "limited", "unnecessary", "duplicated", "misleading". The 'good' data or data pools can be "complete", "full", "normally distributed", "trustworthy", "useful", "reliable", "unbiased", simply "good", "valuable", "rich"<sup>62</sup>. The 'reformed' data can be, for instance, "imputed" and "log-transformed".

Data can also differ in terms of their proximity to the data collection procedure and, supposedly, the world – "raw data"; "primary data" (opposed to a number of data types, e.g. "reported data", "transcribed data", "interpreted data"); "data

---

<sup>62</sup> The majority of these descriptors have been taken from the *Cochrane Handbook* – so much so for the value-free language of quantitative research.



‘as extracted’” vs. “consensus data”; “original research data” vs. “summary”, “pooled” “aggregate(d)”, “de-aggregated” data.

Many perspectives have been taken towards data in the methodological literature. One dictum appears again and again. It is found in methodological texts on quantitative and qualitative, primary and secondary research; it is reiterated in undergraduate and postgraduate courses; it is rediscovered by experienced and novice researchers alike. You are asking which one? Well, it is obvious. You need to ask, you need to interrogate data. “Data never speak for themselves.”

#### *4.1.2. Data in the meta-scientific discussion*

To my reading, there are three main aspects of the concept of data in the philosophical and broader meta-scientific literature: 1) data of the senses and/or “sense data”; 2) data as subsumed within the concepts of observation and perception in science and as used to develop and test scientific theories; and 3) data as very material traces and (arte)facts, including data as a commodity.

The debates on data of the senses and/or “sense data” concern that of which we are aware of in perception and its directness. In much of modern epistemology, data of the senses have been regarded as the foundation of all evidence (Hacking, 1975: 32). “Sense data” can have a more technical meaning than that of data of the senses/ perceptual data most broadly. Sense data can be seen as “the alleged mind-dependent objects that we are directly aware of in perception, and that have exactly the properties they appear to have” (Huemer, 2011). The general doctrine here is that “we never see or otherwise perceive (or ‘sense’), or anyhow we never *directly* perceive or sense, material objects (or material things), but only sense-data (or our own ideas, impressions, *sensa*, sense-perceptions, percepts &c.)” (Austin, 1962: 2, italics in the original). Sense-datum theorists seek the “incorrigible” (*op. cit.*: 103-5), even in cases where we go wrong in our observations, as when we call aeroplanes ‘birds’ (Hanson, 1958: 22). Debates on sense-data extend towards deep philosophical worries – such as of the possibility of true knowledge of the external world, solipsism, mind-body dualism, difficulty in locating sense data in physical space

(Huemer, 2011, Feigl, 1958), and similarly deep “hankerings”– after absolute certainty, truth, foundations (Austin, 1962: 104). Topics as grand as objectivity, empiricism and scepticism are not far away either.

More strictly science-related philosophical debates on data are those on observation and perception as grounds for the development and testing of scientific theories. The word ‘data’ is rarely incorporated in the formulation of key problems and arguments in such debates, but is subtly omnipresent – used directly but with secondary functions, used as an alternative label for certain meanings of the concepts of observation and perception, or even absent but being the word which scientists/researchers are more likely to use in similar contexts.

Historically, the richest debates on observation and perception in relation to scientific theories seem to take up the logical positivists’ proposals of a purely observational language – and the idea that only what is expressible in this language can provide an independent test for a theory (discussed in section 4.5.2 below). This view has been discredited in favour of what is still current orthodoxy, of theory-laden, paradigm-dependent observations (Hanson, 1958; Kuhn, 1996). Presently, the liveliest philosophy of science debate where observation and perception remain centre stage is that on the relationship between the observability and reality of theoretical entities (e.g. van Fraassen, 1980; 2002; 2008; Churchland, 1979; 1982; Hacking, 1983). In a language more comprehensible to scientists, this debate concerns the extent to which the data/evidence we have of theoretical entities give us a good enough reason to believe in the existence of those theoretical entities. To give some specifics of the issues addressed, I will use Bogen and Woodward’s classic article on “Saving the Phenomena” (1988).

“According to a widely shared view of science”, they begin, “scientific theories predict and explain facts about ‘observables’” – i.e. objects and properties perceived by the senses or detected through processes that can be regarded as extensions of perception, for instance involving the use of instruments (*op. cit.*: 303, 305). The authors argue against this view. Their argument revolves around a distinction between data and phenomena. “For the most part” data (such as bubble chamber photographs, patterns of discharge in electronic particle

detectors and records of reaction times in psychological experiments) can be “straightforwardly observed” (*op. cit.*: 305, 306), but are not typically predicted or systematically explained by a theory. It is facts about phenomena (such as weak neutral currents, the decay of the proton, and chunking and recency effects in human memory) that scientific theories aim to predict and explain. Although detected through the use of data, phenomena are generally not observable in any interesting sense of the term (306). Both facts about data and facts about phenomena can serve as evidence, but with regard to different ‘targets’ – respectively claims about phenomena and high level general theories (306). Bogen and Woodward suggest that this historically strong but misguided emphasis on observables as the focus of explanation and prediction of scientific theories and the failure to distinguish between data and phenomena have had a number of negative consequences. For instance, it is not appreciated that discrediting the reliability or possibility for observation (e.g. by appealing to its theory-ladenness, expectancy effects and sets, or the impossibility of observing theoretical entities) does not discredit the possibility for having objective criteria for the comparison of theories or claims to reality and scientific legitimacy (305). Similarly importantly, the significance of non-perceptual considerations that have a bearing on reliability (e.g. related to avoiding confounding factors, procedures for data analysis and statistical inference, etc.) has escaped philosophical attention (311, 312, 327).

Finally, the third strand in the meta-scientific debate on data highlighted here concerns the objects available to, generated, handled and shared by scientists – such as specimens, datasets, photographs, recordings, model organisms, simulations; often peculiar, unwieldy and precious. In recent years there has been a rapidly growing meta-scientific interest in data-driven science, data-sharing and the travelling of data, facts and artefacts (Howlett and Morgan, 2011; Leonelli, 2012a). Issues such as standardisation, commodification and meta-data accompanying data are major concerns. I return to this issue in the section on transformations (6.3).

Now let data travel. They can go. In medicine and health, it is all about evidence.

## 4.2. Evidence

### 4.2.1. Evidence in the methodological discussion

In a classic editorial entitled “Evidence based medicine: what it is and what it isn’t”, Sackett et al. (1996) state that “[b]y best available external clinical evidence we mean clinically relevant research, often from the basic sciences of medicine, but especially from patient centred clinical research into the accuracy and precision of diagnostic tests ..., the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens” (*op. cit.*: 71-2). For Guyatt et al. “any empirical observation about the apparent relationship between events constitutes potential evidence” (Guyatt et al., 2000: 1292). They accompany this broad definition with a Hierarchy of Strength of Evidence for Treatment Decisions which ranks, in a descending order of strength, the following sources of evidence:

- 1) N of 1 randomized trial (randomised trials of treatment responses of an individual patient),<sup>63</sup>
- 2) Systematic reviews of randomized trials;
- 3) Single randomized trial;
- 4) Systematic review of observational studies addressing patient-important outcomes;

---

<sup>63</sup> In representations of the limitations of EBM it is often claimed that the hierarchy of evidence is topped by the randomised controlled trial, and, more rarely, by the systematic review of RCTs. Although I cannot claim systematicity, I have not come across a hierarchy in any of the authoritative texts where the RCT, as opposed to the systematic review of RCTs, comes first. In recent texts, N of 1 trials or meta-analyses of these (trials of the treatment response of individual patients) are consistently at the top, but as can be seen from Guyatt et al.’s article above they were there at least as early as in 2000. Of course, an argument can be made as to what types of studies are most often carried out or available to decision makers but, strictly speaking, claims that the hierarchy of evidence in EBM is dominated by the randomised controlled trial are false.

It is a further question of how exactly N of 1 studies are similar and different to randomised trials and whether the designation of ‘N of 1 randomized trial’ is a reasonable typological ascription or sign of the fetishisation of RCTs (this is one of many astute observations that came from my supervisors, in this case Prof. John Dupré. And it is one of those I feel I would never have thought of myself – the term was so natural to me!).

- 5) Single observational study addressing patient-important outcomes;
- 6) Physiologic studies;
- 7) Unsystematic clinical observations.

Hierarchies of evidence are a defining feature of EBM. They differ according to the type of decision that needs to be made (e.g. associated with treatment, prognosis, diagnosis, etc.) and also demonstrate some internal variety. In defence against a recurrent criticism, Sackett et al. (1996) state that “[e]vidence based medicine is not restricted to randomised trials and meta-analyses” (*op. cit.*: 72). For instance, assessments of the accuracy of a diagnostic test would require “proper cross sectional studies of patients clinically suspected of harbouring the relevant disorder”. Sometimes the evidence “will come from the basic sciences such as genetics or immunology”. Even with some questions about therapy an RCT may not be required, as with “successful interventions for otherwise fatal conditions” or if it cannot be waited for (*ibid.*).

The notion of evidence in evidence-based medicine has been vigorously attacked in the methodological literature. The “definition of evidence challenge” is seen as one of four challenges that EBM should meet “to establish its claims to universality and legitimacy” (Upshur, 2004: 197). In her paper “Accounting for EBM: Notions of evidence in medicine”, Lambert (2006) summarises the criticisms of evidence-based medicine raised in over 86 publications in medical and health professional journals into a typology of six alleged limitations (*op.cit.*: 2634, note 1). One major type of concern is that evidence derived through population studies is “incommensurate” with the forms of evidence required for the clinical management of individual patients (2634). A second concern is that EBM has legitimised a bias towards individual focused and simple treatments (e.g. drugs) as opposed to behavioural, psychosocial, community based and multiple component interventions. This is because evidence of effectiveness of the former is much easier to obtain through evidence-based approaches than evidence of effectiveness of the latter (2635-6). A third salient criticisms is that EBM fails to take patients’ views into account: its reliance on strict evidence hierarchies results in predominance of evidence of clinical and cost effectiveness (2635), whereas the narratives of patients and doctors should

also bear on decision making (2642). Finally, a range of concerns has been raised about the very dominance of evidence and scientific knowledge in medicine. EBM is seen as excluding clinical skills – “the diagnostic art”, clinical expertise and judgement, clinician’s intuition and tacit knowledge (2635).<sup>64</sup>

Some of the above criticisms have had an impact. There are numerous re-explifications of evidence-based medicine to the effect that it is not restricted to the randomised controlled trial and that clinical judgement, patient preferences, values and other types of evidence are of ultimate importance (e.g. Sackett et al., 1996; Petticrew, 2001). More systemic changes can also be observed. For instance, the Cochrane collaboration is opening up to qualitative research<sup>65</sup> and “logic models” (roughly, models that represent middle-level theorising as opposed to ‘pure’ evidence) (Anderson et al., 2011). The depth and motivations of such changes, or the possibility for a “truly satisfactory” reformulation of evidence to fit the distinctive epistemologies and standards of different disciplines have, however, been questioned (Lambert, 2006: 2639, 2643).

When it comes to types and dimensions of difference along which bodies and pieces of evidence can be described, the variety of labels is enormous (analogous to that for data illustrated in 4.1.1). Established typologies seem to go primarily along the lines of source study methodology, as represented in a variety of hierarchies of evidence. As these have already been attended to, I will turn to three features of evidence which seem to pervade the methodological literature: its strength, relevance and heterogeneity/homogeneity. I will start with the least technical of these – relevance.

---

<sup>64</sup> The remaining types of criticisms as summarised by Lambert are less ‘evidence-focused’ and concern: the development of formulaic guidelines (e.g. clinical guidelines, protocols, algorithms), seen as potentially eroding clinical autonomy, limiting patient choice and impacting negatively on physician learning; and the problem of translating research evidence into clinical practice, health policy and health services delivery.

<sup>65</sup> Adding templates for systematic reviews of qualitative research in RevMan was discussed as an upcoming development 19<sup>th</sup> *Cochrane Colloquium*, 19-22 Oct 2011, Madrid, Spain.

In the current health policy, advocacy and research discourse, ‘relevance’ of evidence is understood in terms of the likelihood that some body of evidence will 1) be used in clinical practice, self-care, commissioning and policy development and 2) be understood and felt as needed, important and helpful by various stakeholders outside the research and/or policy making community.<sup>66</sup> The concern with relevance is associated with a well defined, emotional and rhetorically dense debate, with frequent opposition between research-oriented and user-oriented evidence and powerful-user-oriented evidence (e.g. for clinicians and policy makers) and ultimate-user-oriented evidence (for patients and carers). The following excerpts are typical illustrations:

*Consumer or public uses of evidence invoke a fundamental question about the nature of the evidence: who decides what evidence to seek and how to seek it? A discourse about citizens invokes the right to be involved in the generation of evidence as well as the planning and review of services. Patient and public involvement in decisions about their own care, or reviewing services, is meaningless if, in an evidence-informed culture, the evidence is irrelevant to either group (Oliver et al., 2008: 73).*

*INVOLVE defines public involvement in research as research being carried out **‘with’** or **‘by’** members of the public rather than **‘to’**, **‘about’** or **‘for’** them (INVOLVE, 2013).*

Next, it may be worth introducing the methodological debate on the strength/quality of evidence by suggesting what it isn’t. It is not exhausted by a concern with evidence hierarchies. There is richness to the methodological debate that critical texts, including many philosophy of evidence-based medicine texts, fail to appreciate. Hierarchies of evidence are only one of the tools, and a crude one, through which evidence is assessed. As for the very concept of “quality of evidence”, many texts and speakers presenting quality assessment endeavours are at pains to explain why the term is inaccurate and

---

<sup>66</sup> This notion of relevance is different to the philosophical notion of relevance of evidence (where relevance of evidence is a function of its relationship to a particular hypothesis). At the same time, in philosophy of science informed texts relating to evidence-based policy, such as Cartwright and Stegenga, 2011, there is a similar concern with making a distinction between the viewpoints of evidence producers and evidence users (in predicting policy effectiveness).

corresponds to stronger claims than those they are prepared to make and why it may be preferable to use, for instance, “risk of bias”, “study validity” (Higgins and Green, 2008: 189-190) or “confidence in estimates of effect”. This does not mean, however, that hierarchies of evidence are particularly flexible; that the crude assessments made on their bases are not amongst the most consequential; that quality is not a fetish in some texts or the fall-back-on word even in texts which disclaim their use of the term (see, for example, Higgins and Green, 2008: 190).

Over 10 years ago, a report of the Agency for Healthcare Research and Quality on *Systems to Rate the Strength of Scientific Evidence* (West et al., 2002) identified over a 100 systems (tools, checklists, scales<sup>67</sup>) to rate the quality of studies and/or grade the strength of bodies of evidence from one or more studies. The differences between the pronouncements of different tools have been an object of investigation (Jüni et al., 1999), as well as the development of “walkovers” from one grading system to another (Ebell et al., 2004: 555). It is more likely that the number has tripled since 2002 than to have remained the same.

In the context of the *Cochrane Handbook*, evidential quality is conceptualised in terms of minimal risk of bias. It is assessed with the “Risk of Bias Tool” (Higgins and Green, 2011: 8.5). Recently, the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) system of rating quality of evidence in systematic reviews and guidelines, and of grading strength of recommendations in guidelines (Guyatt, Oxman, Akl et al., 2011) has been gaining popularity. It is likely to become the orthodoxy in evidence evaluations.<sup>68</sup>

---

<sup>67</sup> The Cochrane Handbook “explicitly discourages” (191) the use of scales in assessing risk of bias and the summary scores they provide. Some of the main concerns relate to the use of difficult to justify weights in calculating a summary score and the lack of transparency of those summary scores.

<sup>68</sup> The rating capacity is one of the aspects of GRADE. It is claimed that it is “much more than a rating system. It offers a transparent and structured process for developing and presenting evidence summaries for systematic reviews and guidelines in health care and for carrying out the steps involved in developing recommendations. GRADE specifies an approach to framing questions, choosing outcomes of interest and rating their importance, evaluating the evidence, and incorporating evidence with considerations of values and preferences of patients and society to arrive at recommendations” (Guyatt et al., 2011: 384).



It is predicated on a starting point of high quality evidence for RCTs and low quality evidence for observational studies. Five quality parameters may be used to downgrade evidence – risk of bias, inconsistency, indirectness of the evidence, imprecision of the estimates and publication bias. Evidence can be upgraded on the grounds of large effect, the presence of a dose-response gradient, and a judgement that all plausible confounding factors would normally decrease an observed effect (Guyatt, Oxman, Sultan et al., 2011). GRADE is “outcome centric” – the quality of evidence for a specific outcome/finding is rated rather than that of the study as a whole (Guyatt, Oxman, Akl et al., 2011: 385).

Across quality of evidence tools, it is stressed that these (may) involve subjectivity, judgement, interpretation and do not eliminate disagreements (Higgins and Green, 2008: 191; Guyatt, Oxman, Akl et al., 2011: 392). Quality is a continuum and there are limitations to using discrete, rigid categories (Guyatt, Oxman, Akl et al., 2011: 389). The difficulty of distinguishing between incomplete reporting and inadequate study conduct is a key obstacle to valid assessments (Higgins and Green, 2008: 191). In brief, the strength of evidence is constituted by numerous building blocks and diminished by numerous missteps whose relative importance and effect may be very difficult to judge objectively, even with structured tools.

Finally, the third characteristic of (bodies of) evidence that seems to dominate the methodological literature concerns its homogeneity/heterogeneity. The narrow statistical meaning of homogeneity/heterogeneity covers the variability of intervention effects across studies (see also Chapter 1, Section 2.2). Methodological texts on the systematic review also discuss clinical heterogeneity, associated with variability in participants, interventions and outcomes of interest, and methodological heterogeneity, associated, for instance, with the use of blinding or allocation concealment or different approaches to defining and measuring outcomes.<sup>69</sup> Admissible heterogeneity is determined primarily by decisions about the scope of the review (Higgins and

---

<sup>69</sup> Statistical heterogeneity results from clinical or methodological diversity/heterogeneity, or both.

Green, 2008: 277). In the context of reviews that use a greater variety of studies, heterogeneity is often associated with a diversity of disciplines, theoretical orientation, worldview assumptions and 'paradigms'. A brief discussion of 'paradigms' in health research is given in 4.7.

#### *4.2.2. Evidence in the meta-scientific discussion*

The nature of evidence in evidence-based medicine is the pre-eminent topic for the philosophy of evidence-based medicine. Most frequently, this is addressed through (often harsh) criticisms of EBM's over-reliance on randomised controlled trials and meta-analyses; through arguments undermining claims about their unbiased nature and methodological supremacy; through criticisms of EBM's strict, categorical, solely study design-based hierarchies of evidence; through insistence on the admission of a broader range of quality of evidence criteria and methodological alternatives (e.g. observational evidence, mechanistic evidence, causal claims, pathophysiologic rationale); through emphasising the distance between RCTs and effectiveness in real life (Worrall, 2002; 2007a, b; La Caze, 2008; 2009; Bagshaw and Bellomo, 2008; Grossman, 2008; Cartwright, 2011; Cartwright and Stegenga, 2011; Howick, Glasziou and Aronson, 2010, Howick, 2011). Further examples of the various other ways in which EBM's concept of evidence has been unsettled in philosophical writings include: challenging the thoroughgoing avoidance of any prior (subjective) knowledge in obtaining evidence in the style of classical as opposed to Bayesian statistics (Dowe, 2008); highlighting the irreducible fuzziness of the boundary between evidence and lack of evidence (Vineis, 2004); drawing attention to the implicit normativity in the production and presentation of facts within evidence-based medicine (Molewijk, 2004, Goldenberg, 2006).

The second main line of debate in the philosophy of EBM extends the concern with the 'other' types of evidence advanced as valuable. The relationship between EBM and its alternatives is explored in terms of the types of knowledge they generate, the practices in which they are used, the sources of their authority and the values associated with them. Specific topics include: clinical judgement, wisdom and reasoning (Wifstad, 2008; Parker, 2002; Upshur, 1999; 2003); tacit knowledge (Thornton, 2006, Henry, 2006, Braude, 2009); the

relationship between EBM and psychiatry and psychotherapy, which are prime examples of fields of meanings, values, irreducible personal individuality and contextual understanding (Fulford, 2004; Fulford, Thornton and Graham, 2007; Gupta, 2007; Ayob, 2008; Bolton, 2008; Falkum, 2008; Thornton, 2008); the relationship between EBM and the complementary and alternative therapies (Ernst, Cohen and Stone, 2004; Parker, 2007; Clark-Grill, 2007; Derkatch, 2008; Dysart-Gale, 2008); between EBM and basic science (Sehon and Stanley, 2003); EBM and clinical medicine (Kulkarni, 2005). In some cases the discussion is critical, sometimes condemnatory of evidence-based medicine. It is seen as a threat to those 'others' or neglectful of them to the detriment of good care. In other cases their role is taken to be to limit, balance EBM. In still other cases some form of integration is sought, usually requiring minor adjustments to claims and standard interpretations of EBM.

Two further issues that appear in the above lines of research have a strong independent presence, too. The first is the relationship between evidence-based medicine and values, ethics and a values-based medicine. The ethical implications of the spread of EBM are explored (Goodman, 2003; Saarni and Gylling, 2004; Goldenberg, 2006). Critical analyses are offered of the explicit and implicit values with which EBM aligns (Molewijk, 2004; Goldenberg, 2006; Zarkovich and Upshur, 2002; Rogers, 2004; Watine, 2011) or of the value commitments it fails to make (Bolton, 2008). Proposals are made as to how to incorporate values, alongside evidence, into clinical decision making (Fulford, 2004; Fulford, Thornton and Graham, 2007; Zarkovich and Upshur, 2002; Thornton, 2006; 2008). The second of those further issues is the tension between the evidence-based as general(ised), objective, population-based and external and the individual, subjective, context-specific, internal to a situation (and supposedly non evidence-based). It is explored, for instance, by Tonelli, 1998; Kulkarni, 2005; Ayob, 2008; Falkum, 2008; Wifstad, 2008.

Overview works on the philosophy of evidence-based medicine or ones that claim to bring the philosophy of science to bear on EBM generally cover the issues outlined above. For instance, Sehon and Stanley in "The challenges of evidence-based medicine: A philosophical perspective" (2003) go back to a closer reading of Kuhn to argue that the advent of EBM is not a paradigm shift,

as claimed by its proponents. Quine's metaphor of "web of belief" is then used to intertwine EBM and the basic sciences (*op. cit.*: bmc 9). Ashcroft (2004) suggests that "further work is needed on the theory of evidence and inference; causation and correlation; clinical judgment and collective knowledge; the structure of medical theory; and the nature of clinical effectiveness" (*op. cit.*: 131). Goldenberg in "On evidence and evidence-based medicine: Lessons from the philosophy of science" (2006) brings ideas from post-positivist, feminist and phenomenological philosophies of science to problematise objectivist claims made about evidence in evidence-based medicine and draw attention to a variety of features of its social nature and social and political effects. Howick in *The Philosophy of Evidence-Based Medicine* (2011), which seems to be the only book-length treatment of the subject, provides detailed critical, yet often sympathetic, analysis on EBM's stance on randomisation, double masking, placebo controls, mechanistic reasoning and expert judgement.

There is no preferred classification scheme for evidence in the philosophy of evidence-based medicine. A broad range of labels, possibly with a predominance of method-based ones, are used. Rather than adding detail in this direction, I will extend the discussion with a typology of *concepts of evidence* in science offered by Peter Achinstein in *The Book of Evidence* (2001). It comes from the general philosophy of science rather than the philosophy of medicine. It seems, however, a particularly useful tool in distinguishing between different concepts of evidence in operation in health research synthesis and in considering how the status of a piece or body of evidence can change as a result of being contextualised within a synthesis study.

Achinstein argues that there are at least four different concepts of evidence or perspectives to the concept of evidence in science (Achinstein, 2001: 18). He distinguishes between "ES-evidence" (with "ES" standing for "epistemic situation"), "subjective evidence", "veridical evidence" and "potential evidence".

In the first case, evidence is evidence that  $h$  relative to a particular epistemic situation.<sup>70</sup> In 1883, Heinrich Hertz was in an epistemic situation – including his experimental setup, results, the techniques for removing gas from cathode tubes at the time, his background assumptions – which completely justified his belief that cathode rays are not charged (*op. cit.*: 18). Later, it was demonstrated that they were in fact negatively charged. The reasons for what, with the benefit of hindsight, was found to be an erroneous conclusion on the part of Hertz was that he had not evacuated sufficiently the gas in his cathode tubes, yet scientific knowledge of his time did not allow him to recognise this (19). In his epistemic situation, he was justified in believing the hypothesis that cathode rays are not charged. The requirements of this view are not as weak as they may appear either – to be justified in believing a hypothesis  $h$  on the basis of  $P_1, \dots, P_n$ , a person must be justified in believing  $P_1, \dots, P_n$  (21).

Achinstein formalises the subjective view, the second concept of evidence in science, as follows:

$E$  is  $X$ 's evidence that  $h$  at time  $t$  (with  $X$  being a person or a group) if and only if at time  $t$

1.  $X$  believes that  $e$  is evidence that  $h$ ;
2.  $X$  believes that  $h$  is true or probable; and
3.  $X$ 's reason for believing that  $h$  is true or probable is that  $e$  is true (23).

Subjective evidence only demands that  $X$  believes that  $e$  is true, not that  $e$  is true (24), as well as that  $X$  has a reason to believe that  $h$  is true or probable on the basis of  $e$ , not that  $X$  be *justified* in believing  $h$  on the basis of  $e$ . Even if it was the case that Hertz's experimental results had not warranted his belief in the neutrality of cathode rays, he could have viewed his results as evidence for that hypothesis (*ibid*).

In the case of "veridical evidence", if  $e$  is evidence that  $h$ , then  $e$  provides a

---

<sup>70</sup> "[A]n epistemic situation is different from what philosophers usually call "background information." The latter consists solely of propositions (assumed to be true). The former is an abstract type of situation in which, among other things, one knows or believes that certain propositions are true, one is not in a position to know or believe that others are, and one knows (or does not know) how to reason from the former to the hypothesis of interest, even if such a situation does not in fact obtain for any person" (Achinstein, 2001: 20).

good reason to believe  $h$  (24). When a belief is justified without the invocation or implicit presence of an epistemic situation, a “good reason to believe” has been provided (25). “Good reason to believe” functions as a “sign” or “symptom”. A certain type of a rash may be a sign or symptom of a disease irrespective of medical experts’ unawareness of the connection and thus lack of justification for believing that the disease is present. Nevertheless, the rash is a good reason to believe that the disease is present (*ibid*). The concept of veridical evidence requires the truth of hypothesis  $h$  (26). Veridical evidence should, however, be distinguished from “conclusive” evidence where not only is  $h$  true, but  $e$  establishes  $h$  with certainty (27).

Finally, the concept of “potential evidence” does not presuppose the truth of the hypothesis (28). A patient may have a rash typically associated with the measles virus but not have the measles virus. If we say that whatever his rash was, it was not a sign or symptom of the measles virus, we are invoking the concept of veridical evidence. If we say that, in spite of his not having the measles virus, he had its signs or symptoms, we are employing the concept of potential evidence. Thus, in the case of potential evidence, some fact  $e$  may be a good reason to believe  $h$  even if  $h$  turns out to be false (27-28).

### **4.3. Findings**

“Seek and you will find”. You will, but little, if you are seeking for a discussion of ‘findings’ in meta-scientific texts. The concept appears only marginally there. I thus discuss ‘findings’ only from a methodological perspective. And while in most methodological texts it is a vague concept, some posit findings as *the* unit.

Originally, this has been done within the context of distinguishing between synthesis studies and the secondary analysis of data. In the latter, primary study datasets are re-analysed as opposed to only drawing on study reports. Some authors sharpen this contrast to suggest that secondary analysis always deals with raw data while synthesis studies always deal with reported and interpreted data referred to as ‘findings’. For instance, Finfgeld describes metasynthesis as “a complete study that involves rigorously examining and interpreting the findings (versus the raw data) of a number of qualitative

research studies” (Finfgeld, 2003: 894). Estabrooks, Field and Morse (1994) state that “data aggregation is appropriate in the single study, whereas aggregation of findings is the appropriate approach when working with multiple studies” (*op. cit.*: 505). “Like a secondary analysis, qualitative synthesis could involve re-interpretation, but unlike secondary analysis it would be based on published findings rather than primary data” (Britten et al, 2002: 209-10).

The most extensive conceptual exploration of findings in the context of research synthesis studies appears to have been performed by Sandelowski and Barroso (2002), who analysed reporting practices and purported findings of 99 studies on women and HIV.

The authors assert that research integration projects require that the findings in primary studies be clearly delineated (Sandelowski and Barroso, 2002: 218), separable from other elements of the study, easy to find. Many theoretical positions associated with qualitative research challenge the assumption of separability of findings from other elements of a study, such as research problem, method, data, analysis, interpretation, researcher (*op. cit.*: 214, 215). Practices of using non-standard representational forms also complicate the identification of qualitative findings (215). As much as the authors acknowledge the grounds for such views and recognise the value of alternative representational formats (215, 219), they insist that synthesis projects necessitate such separations, “albeit imperfect and controversial” (214).

Five main problems in “finding the findings in qualitative studies” are identified. The first is the misrepresentation of data as findings. In reports where this happens, researchers were prone to “descriptive excess” but offered practically no interpretation of their data (Sandelowski and Barroso, 2002: 216, also quoting Lofland & Lofland, 1995: 165).<sup>71</sup> The opposite to this is a tendency towards “analytic excess”. In such studies researchers present their manipulations and rearrangements of data (lists, coding schemes, categories, concepts, frequency counts, etc.) as if they were coherent interpretations of the

---

<sup>71</sup> Lofland, J. and Lofland, L.H. , 1995. *Analyzing social settings: A guide to qualitative observation and analysis*. 3rd ed. Belmont, CA: Wadsworth.

phenomenon of interest (Sandelowski and Barroso, 2002: 216, also quoting Lofland & Lofland, 1995: 164).

A common problem is the use of quotes and examples that do not seem to be fitting the findings they are supposed to illustrate. This leaves the reader wondering whether the researchers had other evidence to support the interpretation and, potentially, how to handle the line of interpretation suggested (Sandelowski and Barroso, 2002: 217). Further difficulties include failures to provide the grounds for designating something as a “theme or pattern” (*op. cit.*: 217) and problems of researchers’ use of existing concepts or theories. For instance, researchers drift from one concept to another in presenting their findings, leaving the reader to decide if to report several different findings. In other cases, the theories used to interpret data do not fit the data well, or the purpose of using a theory is unclear (218).

The authors conclude by insisting on the importance that findings of qualitative research be found: “if research integration is considered an essential means toward that end [generating knowledge for practice], any report of a study in which the findings cannot be discerned might just as well not have been written. Indeed, the study might just as well not have been conducted” (219).

#### **4.4. Facts**

Similarly to findings, ‘facts’ is a one-home term. It does not appear in recent health research texts, while being a key concept in the philosophy and social studies of science. In an age where everything in the wider world is information and data, it was also curious to find that facts, of the five carriers of knowledge discussed here, appeared most often in claims with predicates of ubiquity and exhaustiveness. For Wittgenstein “[t]he world is the totality of facts, not of things” (1922: 1.1). Russell inhabits a similar world where everything is a fact, but things are facts, too: “Everything that there is in the world I call a “fact”. The sun is a fact; Caesar’s crossing of the Rubicon was a fact; if I have a toothache, my toothache is a fact. If I make a statement, my making it is a fact, and if it is true there is a further fact in virtue of which it is true, but not if it is false”



(Russell, 1948: 159). Recently, those same ubiquitous facts are also seen as travelling, circulating, being transmitted and transformed (Howlett and Morgan, 2011: i) – i.e. as going even ‘more everywhere’. Below, I will point to four themes in specifying a network of meanings associated with ‘facts’: facts in relation to laws; facts in the context of the analytic-synthetic statements distinction; facts as firm or infirm, of the world or of us; and facts in opposition to values.

The received view in the philosophy and history of science is that important, interesting, even beautiful, facts can be distinguished from unimportant, uninteresting facts by their recurrence (Poincaré, undated: 17), reliability and validity across contexts and times. Recurring facts make scientific laws, or at least provide the basis from which they can be derived. Outside of these, we have accidental, contingent facts. Scientific laws thus state the (important) facts about the world. This intuitive claim has been discredited in the philosophy of science, most notably in the work of Nancy Cartwright (Cartwright, 1983, Cartwright, 1999). One of her central arguments rests on the idea of “nomological machines” which are seen as the source of our most valued scientific facts. A nomological machine is “a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws” (Cartwright, 1999: 50). The cost of those scientific truths, however, is a huge number of constraints and specifications that need to be met for those facts to obtain. It is thus questionable whether they can be said to obtain in the world, *simpliciter*.

Facts are also party to the debate about the existence or not of “some fundamental cleavage between truths which are analytic, or grounded in meanings independently of matters of fact, and truths which are synthetic, or grounded in fact”, which, in Quine’s famous phrase, is one of the “two dogmas of empiricism” (Quine, 1951: 20).<sup>72</sup> A sharp boundary between the two is rarely

---

<sup>72</sup> As traced back by Quine, Kant’s cleavage “was foreshadowed in Hume’s distinction between relations of ideas and matters of fact, and in Leibnitz’s distinction between truths of reason and truths of fact” (*ibid.*).

The original paper has “truth which are *synthetic*” rather than “truthS”. I take it to be a typo that has not attracted much speculation about deeper meanings.

asserted in present-day philosophy, with different positions as to whether a weaker version of the distinction is justified or not and how it can be drawn.

Facts, or most facts, or proper facts, or at least some facts are again in more traditional views something real, solid about the world. Cognitively, they provide a test for the truth or falsity of beliefs and statements, including those embedded in scientific theories. Affectively, they better be met with acceptance as many are independent of our volitions and even of our existence (Russell, 1948: 160). Conatively, they provides grounds for rational action. Not that it is easy to do any of the above. But facts are foundational and some more so than others – the “ultimate”, “irreducible”, “brute” (Feigl, 1958: 472); “hard”, “stubborn”, “ineluctable” facts (Russell, 1948: 160). Science either reveals those or builds on them to reveal other hard, reliable facts.

This cluster of views – of (some, proper, scientific, proper scientific, etc.) facts being rock solid, providing the grounds from which to build reliable knowledge and science, and of the world outside rather than of human perception and interaction – has been challenged in a number of ways. Some of these involve arguments familiar from the theory-ladenness and paradigm-dependence debates. Polanyi, for instance, questions the stability of facts relative to interpretations: what is “frequently said that the facts of science remain and only the interpretations change” is “not true or ... at least very misleading” (Polanyi, 1946: 74). If we recognise many of the facts of earlier science, it is because their basic interpretation is still preserved. “But while to Kepler in 1596 it appeared as an indubitable fact that the planetary orbits are related to the geometry of perfect solids, we regard this to-day as mere fancy” (*ibid*). This is one of many expressions of the “factual relativity doctrine” – “the doctrine that factual propositions exist only within a given conceptual framework rather than independently of theories and conceptual schemas” (Weimer, 1975: 450).

Cartwright challenges views of reliable knowledge being based on the solid foundations of solid scientific facts through a different approach (Cartwright, 1999). According to her, it is not the exemplary scientific facts – facts “that are legitimately regimented into theoretical schemes”, that are generally “about behaviour in highly structured, manufactured environments” – that should have a privileged status and be seen as grounding objectivity. It is the vague,

imprecise, inexact, “refugee facts” that are to perform this role – knowledge that an oak tree can grow from an acorn but not from a pine cone, that a child will grow more secure if nurtured, or that one can head north if they follow a compass needle (Cartwright, 1999: 23-25).

In the meta-scientific fields more broadly, the stability of (scientific) facts has often been undermined through attending to the process of constructing facts, its contingencies and the social negotiations involved in it. For Latour (1987) for instance, facts are historical, socially constructed and mutable, the result of the collective efforts of a large number of people. They either do not possess a strength of their own, or it is of little importance. Their fate and development depends to a small extent on their primary builder and mostly on the behaviours of others. The strength and consequently the reality of a scientific fact depend on the number of elements tied to it and the extent to which those elements have been glued together into an organic whole. Facts are always dependent on people, even when they have turned into a black box (Latour, 1987, see Chapters 1 and 3 in particular).

Finally, and going back to the broader themes associated with facts, the philosophy of science is showing an increasing interest towards values and the fact-value dichotomy (Longino, 1990; 1996; Kitcher, 2001; Machamer and Wolters, 2004; Fulford, Thornton and Graham, 2007; Kincaid, Dupré and Wylie., 2007; Douglas, 2009). The fact-value distinction is typically associated with the writings of Hume (as the impossibility to derive ‘ought to’ from ‘is’). It is seen as fundamental to the received view of science. Its validity, usefulness, absoluteness have, however, been extensively questioned or its strength relaxed.

## 4.5. Claims, sentences, statements

### 4.5.1. *Language in the methodological discussion*

Claims, sentences, statements – and similar concepts highlighting linguistic form, propositional structure and issues of representation involved in handling and sharing scientific knowledge – are rarely used in health research synthesis texts to denote the basic unit of analysis/synthesis. But the research synthesis debate is concerned with language and expression. For instance, the non-transparency, incompleteness, ambiguity and imperfections of source study texts are frequently commented on informally and occasionally in methodological texts. Problems tend to be attributed to authors' skills and precision and space constraints in journals rather than to something more general about language and representation. The politics, meta-messages, subtle leanings, epistemological and ethical commitments, etc. of language are an object of analysis in more 'deconstructive' synthesis methods, e.g. the meta-study of Paterson et al. (2001), meta-narrative of Greenhalgh et al. (2005) or critical interpretive synthesis of Dixon-Woods et al. (2006).

In the broader health research literature, the treatment of the linguistic and propositional forms in which data, evidence and findings come<sup>73</sup> shows clear differences between fields. In the context of clinical research, which feeds into the mainstream systematic review, language tends to remain unnoticed. If noticed, it is generally seen as a tool that is difficult to master, but once mastered and used well, as providing a reasonably transparent window to research findings and the world. Discussions of language occur most prominently in handbooks on survey design (how to formulate and order questions and prompts so that valid and reliable responses are obtained) and how to write scientific/research texts (e.g. Alley, 1996; Fraser, 1997; Browner, 2006; Hall, 2013).

In the context of methodological literature on qualitative health research, language is seen as having a much more central place (Green and Thorogood,

---

<sup>73</sup> Obviously, evidence, findings and data used in research synthesis studies are represented in other ways, too. Ideally, one should also address how tables, graphs, images, etc. are discussed in the methodological literature.

2004: 81), a house and a view rather than a window only. The form of the majority of data produced by qualitative researchers is linguistic. Many of the methods used in data production are also based on language use (e.g. interviews) (*ibid*). But there is more. Qualitative research uses a range of methods to scrutinise language for non-obvious meanings and states of affairs. Linguistic analysis, for instance, may uncover hints of social acceptability, engagement or detachment by attending to the use of active or passive voice (Kvale and Brinkmann 2009: 220). Conversation analysis may be observing “turn-taking sequences” and “repair of turn-taking errors” so as to trace the creation and maintenance of intersubjective understanding (*op. cit.*: 221). In narrative analysis, socially accepted plots, temporal and social structures, actors’ and objects’ typological characteristics will be elicited (222-225). In discursive and dialectical approaches, contradictions in individuals’ discourses will be laid open and related to historical situations and discursive practices (226-227). The majority of synthesis methods, however, stay with a ‘language as a window’ view, usually labelled as ‘positivist’ by critics. This is not necessarily a justified attribution, as will be shown immediately below.

#### *4.5.2. Sentences, vocabulary and language in the meta-scientific discussion*

The naive empiricist view of the relationship between claims and state of affairs is, in the words of Kuhn, “that truth and falsity are uniquely and unequivocally determined by the confrontation of statement with fact” (Kuhn, 1996: 80). In an interesting way, this facile idea is both advanced and implicitly opposed by the logical positivists. It is advanced through their belief in an observational language. It is implicitly opposed through the very attention they give to language.

Logical positivists were concerned with developing a “logically perfect” language (Wittgenstein, 1922, as per Suppe, 1977: 13), which would guarantee the cognitive significance of whatever was expressed in it. Cognitive significance was ensured by the possibility for empirical verification. This language L was to

be free of all metaphysical terms, which were considered nonsensical (Suppe, 1977: 13-14). It would include three sublanguages (*op. cit.*: 50-52):<sup>74</sup>

- a) The observation language  $L_O$  contains only terms from an observation vocabulary  $V_O$  and a logical calculus  $K_O$ . It contains no theoretical terms and no quantifiers or modalities (50). The terms in the observation vocabulary  $V_O$  refer to directly observable physical objects or attributes of physical objects (45). This is a physicalist vocabulary. There was initial disagreement as to whether  $V_O$  terms should be given a phenomenalist sense-data interpretation or a physicalist interpretation, but the two alternatives came to be seen as equivalent and a physicalist interpretation was settled for (45-46). Carnap suggests that “any quantitative magnitude that can be measured in a relatively simple, direct way” is an observable (*op. cit.*: 47, citing Carnap, 1966: 225-226).<sup>75</sup> The assertions made using  $V_O$  as their only non-logical terms are seen as “intersubjectively unproblematic with regard to truth: any two observers who possess the words from  $V_O$  used in the assertions, regardless of their scientific or theoretical background, will be able to agree upon the truth of such  $V_O$  assertions” (Suppe, 1977: 48). That is, such assertions are scientifically and theoretically neutral (*ibid.*).
- b) The logically extended observation language  $L_O'$  is an augmentation of  $L_O$  which incorporates a more complex logical apparatus and syntax, including, for instance, quantifiers and modal operators (*op. cit.*: 49, 51). The price for this augmentation, however, is that the sentences of  $L_O'$  do not satisfy the verification criterion of cognitive significance, which requires for sentences to be non-analytic, i.e. amenable to complete verification by observational evidence (49).
- c) Finally, there is the theoretical language  $L_T$  which contains only terms from a theoretical vocabulary  $V_T$  and its associated logical calculus,  $K_T$ .

---

<sup>74</sup> Description as per the “final version of the Received View” – the logical positivists’ view of scientific theories.

<sup>75</sup> Carnap, R., 1966. *Philosophical Foundations of Physics*. New York: Basic Books.  
134

In addition to sentences formulated in any of the above sublanguages, L also contains “mixed sentences” which contain both  $V_O$  and  $V_T$  terms. One of the key types of mixed sentences are the correspondence rules C, which offer a partial observational interpretations of theoretical terms (51).<sup>76</sup> In this (positivist) view, data, evidence, facts and findings are likely to be represented through, or potentially just are, statements made in an entirely observational language or as mixed sentences.

This and related positivists view have received harsh criticism. The dichotomy between observation and theory, observational and theoretical statements has been challenged (e.g. Maxwell, 1962; Feyerabend, 1975; Fraassen, 1980). It is current orthodoxy in the philosophy of science, particularly as a result of the work of Hanson (1958) and Kuhn (1996), that observations are theory-laden.

Not all agree with such a state of affairs. Hacking (1983), for instance, sees this as a philosophical fashion which has distorted commonplace facts about observation (Hacking, 1975: 167). There are plenty of pre-theoretical observation statements (*op.cit.*: 168).<sup>77</sup> Experimentally, philosophically minded psychologists have demonstrated that there is a difference between observational and theoretical terms and that statements made using predominantly observational terms may indeed be more intersubjectively shareable (Clark and Paivio, 1989). Current philosophical research on data sharing takes it to be the case that, after all, observations, data, evidence may not be so theory-laden (Leonelli, personal communication).

#### **4.6. ‘Becoming and feeding the other’**

Whatever defining features and distinctions between data, evidence, findings, facts, claims (and observations) may have emerged above, these require much

---

<sup>76</sup> The restriction to “partial” is as per later formulations of the Received View. Earlier formulations required the complete reduction of theoretical terms to observable consequences by means of correspondence rules. Another later development is the idea that C be seen as an interpretative system that relates theoretical entities to their empirical manifestations and thus enables theory-testing as a whole, rather than each theoretical term being related to its observable consequences through a corresponding rule of its own – again as per Suppe, 1977: 25.

<sup>77</sup> More importantly, observation according to Hacking has been over-rated.

further exploration. The boundaries between the six units are very uncertain – the units are ‘in the habit of becoming and feeding the others’.

The above terms are often used interchangeably in the same text, for instance:

*This remark is not, of course, intended as a criticism of **data gathering**: those engaged in the process no doubt try to **select facts** that might prove relevant to future hypotheses (Hempel, 1996: 13).*

*[I]t is the methods used to collect ‘**evidence**’ ... that define whether or not **data** are potentially applicable to evidence-based practice (Lambert, 2006: 2641).*

A number of basic unit terms often appear together in the same sentence. In-depth analysis may or may not help specify the nature of their relationships and identify fine distinctions of roles. For example:

*With respect to their **evidential role** what distinguishes **data** from phenomena is not that only **facts** about **data** may serve as **evidence**, but rather that **facts** about **data** and **facts** about phenomena differ in what they serve as **evidence** for (**claims** about phenomena versus general theories) (Bogen and Woodward, 1988: 306).*

*[Quality criterion 7] Inclusion of sufficient original **data** to mediate between **evidence** and interpretation (Harden et al., 2004: 796).*

In further cases, warnings are given against the tendency for some units to become other units:

*I labor these familiar points because I want to rescue from their normal oblivion three facts which I believe to be highly important: First, **facts are not data**. They are mental artifacts, selected by human concerns and abstracted from experience by filtering through a screen of schemata (Vicker, 1981: 152).*

Further relationships in which the different terms enter can be specified. But the above has probably been sufficient as direct illustration of the porousness of boundaries between units since implicit parallels between the extended



descriptions of evidence, data, findings, facts and claims are highly likely to have been detected, too. Looking back, it seems often a historical accident why certain debates have clustered around one rather than any of the other units (e.g. why not have “theory-ladenness of facts” or “data-based medicine”).

#### **4.7. The debate on ‘paradigms’ in health and other applied sciences research**

As mentioned in Chapter 1, there is a foundational belief in much health research in the ‘purely empirical character’ of high quality evidence. As typically held, this belief stands for two main ideas which may or may not appear together. The first is that high quality evidence is free of bias, often equated with a claim that non-admissible effects of context and agents of production have been effectively controlled. The second is that evidence is direct signal from the world, independent of theory, untestable metaphysical claims/ worldview assumptions and its cultural-historical context of production.

A range of positions at odds with the latter idea was already presented as part of the review on the five basic units. In the health research field, the debate that questions most strongly the idea of the purely empirical character of high quality evidence is that of ‘paradigm differences’ between quantitative and (types of) qualitative research. A background acknowledgement of ‘paradigms’ is formative of health research synthesis debates and methods. Issues of combinability of studies, findings, evidence, etc. are seen almost exclusively as derivative of issues of ‘paradigm’ differences between quantitative and (types of) qualitative research. Such differences are either claimed to determine the impossibility, difficulty or questionable value of combining studies from across the divide or denied such effects. They seem invariably mentioned though, even when negated.

While having its origin in the mainstream philosophy (and history) of science, the ‘paradigms’ debate has developed a life of its own in health research and in other areas where the quantitative-qualitative dichotomy is internal to the field (e.g. educational, social services, evaluation research, etc.). There, it has

become the norm to refer to the constellations of philosophical commitments, methods, techniques, preferred questions, interpretative moves, etc. on which quantitative and qualitative research draw as ‘paradigms’. In a selective breach-adherence to the Kuhnian idea (1996), those constellations are seen as generally incommensurable but not necessarily so. They are also constantly struggling for relative supremacy but definitely existing contemporaneously and expected to continue to do so. While earlier sources are cautious of stretching Kuhnian thinking, this has ceased to be the case. The usage of ‘paradigm’ in methodological texts in health research may thus be seen as grossly uninformed or simply reflecting a concept in need of a name. Below, I will follow convention in the applied sciences and use the concept in this modified sense, too.<sup>78</sup> I will be indicating substantial distance from the original senses with single quotation marks or referring to ‘broad ontological-epistemological-axiological paradigms’ or, briefly, ‘broad epistemological paradigms’.

The description of ‘paradigms’ which follows is based on a classic methods book by John Creswell, *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (2003), which is a typical good quality example of describing social and human sciences research ‘paradigms’. In this text, the quantitative approach is seen as underpinned by post-positivist knowledge claims and the qualitative by social constructivist or advocacy/participatory knowledge claims.<sup>79</sup> I complemented and adjusted Creswell’s description by using Creswell and Plano Clark (2007) and Patton (2002). I also toned down a small number of peripheral claims that would attract professional philosophers’ raised eyebrows. My aim has been to recreate ‘good’ applied research-style descriptions of epistemological positions underpinning quantitative and (types of) qualitative research.

Looking at those descriptions from the point of view of mainstream philosophy and social studies of science, it seems that they combine local versions of

---

<sup>78</sup> Ironically, not using the term in that context is a sign of ignorance of the technical/expert language of the debate rather than ‘informed abstaining’.

<sup>79</sup> Pragmatist knowledge claims, as associated with the mixed methods approach, are not discussed here for feasibility reasons. There is also disagreement whether pragmatism is a separate paradigm. Proponents of the mixed methods approach claim to be starting from philosophical positions that diffuse the paradigm conflicts between quantitative and qualitative research. These are pragmatist positions, or the “pragmatist paradigm”.

debates which have their 'gold standard' counterparts in the mainstream fields. These include, for example, debates on the realism – antirealism of theories and theoretical entities; on the normative presence of non-cognitive values in science; on the nature and generation of scientific theories and scientific knowledge more broadly; on the rhetoric of science and research; on the nature of evidence, etc. What also seems to be the case, however, is that the philosophy and social studies of science internal to health (and applied social sciences) research have not drawn upon those gold standard debates in their recent, and sometimes decades-old, forms. Neither have they contributed to them.

For example, it is probably safe to assert that in present-day philosophy of science the hypothetico-deductive model of scientific research, the belief in universal laws, or the belief in an unproblematic access to a 'world out there' have been irreparably discredited. In the 'paradigms' debate, however, such views and aims are represented as underpinning one of the valid ways of doing science, labelled as the quantitative/post-positivist way.

In the mainstream philosophy of science, it is still unconventional to hold the view that science and research are, legitimately, a space of non-cognitive as well as cognitive values. Many, perhaps most, health researchers and practitioners will acknowledge that some research is legitimately about values and value-laden and that this too is respectable scientific research.

Assumptions that science and research are looking for objectivity and generalisations (regardless of whether we can achieve them or not) rather than subjectivity and idiosyncrasy are much weaker in health and applied social sciences research than in the philosophy of science. Both researchers and practitioners have grown to see some research as aiming to preserve subjectivity and idiosyncrasy.

There are many issues to untangle here. Careful distinctions need to be made between credible ways of doing science and credible ways of describing the doing of science. Some of my descriptions of predominant views may be imprecise. My point for the moment is that the mainstream and 'internal' studies of science each have a trajectory of their own, and the intersections are fewer than one would expect.

After this long introduction, I finally present the illustrations of ‘good’ applied research-style descriptions of epistemological positions underpinning quantitative and (types of) qualitative research.

#### *4.7.1. Post-positive knowledge claims*

In Creswell’s rendering, post-positivism is a label that is generally interchangeable with the “scientific method”, empirical science, quantitative research, positivist/postpositivist research, etc. (Creswell, 2003: 6-7). The ‘post’ in post-positivism is explained as denoting a move away from the idea of the possibility for “absolute truth”. Evidence is always imperfect and fallible. Research begins with theories, hypotheses and claims. These are then refuted, supported or refined on the basis of acquired data. Most quantitative research involves theory- or hypothesis-testing. It is also reductionist, here perceived in the sense of reducing complex questions to small sets of discrete, highly specified ideas and variables to be tested. Post-positivist research is seen as primarily concerned with causes and their probabilistic relationship to outcomes, and as a result favours the experimental approach. The world is taken to exist “out there” and to be governed by laws. Observations, measurements and experiments, with their strong numerical slant, are means of acquiring objective knowledge of an external reality. Validity, reliability and the control of bias are major concerns. In researching human beings, post-positivists are seen as prioritising the study of behaviour and the development of numerical measures of it.

#### *4.7.2. Socially constructed knowledge claims*

Social constructivism, frequently combined with interpretivism, is proposed by Creswell as one of the two main philosophical positions underpinning qualitative research. Within the social constructivist view, individuals are seen as motivated by a search for understanding of the world in which they live and work. The meanings they develop are highly subjective, yet social. They are forged through engaging with a world open to interpretation, in interactions and discussions with other people, and in very particular historical, cultural and

social settings. The variety and complexity of the meanings thus emerging are huge. The aim of research is to capture and preserve this subjectivity, sociality, variety and complexity rather than narrow meanings into a limited number of categories and ideas. The questions of such research are thus typically broad, general and open-ended so that participants are unconstrained in constructing their own meanings of a situation. Constructivist researchers are particularly interested in processes and interactions in and through which meaning is constructed and in the life and work contexts which people inhabit. Qualitative research tends to be inductive, with data collected in the field used to develop theory or identify patterns of meaning. Researchers openly “position themselves” in the research and acknowledge how their interpretations follow from their own personal, cultural and historical background and experiences.

#### *4.7.3. Advocacy/ participatory knowledge claims*

The third position presented by Creswell endorses knowledge claims obtained through an advocacy/ participatory approach. Knowledge is political. Inquiry should be driven by an agenda to help marginalised and disenfranchised people. Issues such as empowerment, inequality, oppression and alienation are paramount to this position. The research process should be collaborative so that the participants'/collaborators' voice is truly heard – participants may help formulate questions, engage in data collection and analysis, and be rewarded for their participation. The aim of research is to bring about change – in the lives of participants, the institutions they are part of and engage with, and in the researcher's own life. Advocacy/participatory forms of inquiry are often informed by theoretical-political perspectives such as feminist thinking, critical theory, queer theory, anti-racism and disability movements, etc.

### **5. Transformations**

This is the final section of this chapter. As highlighted at the outset, there does not seem to be any published work that falls naturally and fills up much theoretical space under a heading of 'transformations of evidence in health research synthesis studies'. It could hardly be otherwise if, as asserted here,

the very field of research synthesis is in a state of pre-formation. Theoretical discussions have not had the chance to develop extensively. Overlapping concepts from disconnected subfields are yet to be linked together. Of course, in a broadening circle of relevance, debates abound. Some need to be prioritised.

If we look at particular processes of transformation in particular niches of the field of health research synthesis, there seem to be four major practices associated with sound articulation of processes and/or concepts, or at least the presentation of exemplars to follow: numerical conversions (e.g. *Cochrane Handbook*, 2011: 7.7); the practice of “translation” as conceptualised by Noblit and Hare in the context of meta-ethnography (1988); the practice of coding as underpinning most qualitative research syntheses; and visual re-representations (e.g. mapping and charting, widely used in synthesis studies).

If we go to the broader methodological literature, we find the somewhat vague and maybe now outdated concept of “data reduction” (e.g. see Miles and Huberman, 1994: 10-11) which may be formulated both as encompassing processes of transformation of data (*ibid.*) and being an instance of them.

In the meta-scientific literature, the concept of “travelling of facts” is gaining momentum, as discussed in 3.1.2 (Howlett and Morgan, 2011).

No matter where we look, we will be finding mentions of ‘interpretation’ which is, undoubtedly, a major subtype and/or element of processes of transformation of evidence in health research synthesis.

No matter where I looked (in the methodological and meta-scientific literature), I did not find discussions of three further processes alluded to in initial descriptions of transformations (Chapter 1): reformulations/rephrasing involving a supposedly no or minimal interpretation; alternative “chunking” of information where, for instance, sentences are broken down in different ways; and switching focus towards alternative features of entities and phenomena. In an ever widening circle of relevance, I could have found discussions of these in general psychology but some limits had to be placed.

For reasons of space constraints, I will only address briefly the literature on translations, coding, and travelling of facts. These were chosen in preference to

others of the above debates as they had a well defined theoretical and conceptual component and as each of them was likely to be unfamiliar to at least one specified group of readers. The presentation below adds key ideas to the background from which subsequent explorations of transformations were undertaken. It also illustrates the affinity of the concept of transformation to big concepts, big debates and rich metaphors, with all the associated consequences of vagueness, complexity and controversy – and with all the potential for some useful disentanglement.

### 5.1. Translations

The idea of research synthesis (of qualitative research) as involving “translations” of studies, accounts, metaphors, interpretations, concepts, themes, etc. comes from Noblit and Hare’s work on meta-ethnography (1988). It is based on Turner’s (1980)<sup>80</sup> argument of sociological explanations as translations (Noblit and Hare, 1988: 29). According to Turner, in Noblit and Hare’s reading, all social explanation is comparative, whether implicitly or explicitly. The breakdown of an expectation of “same practices” across groups and cultures yields an explanatory puzzle. Such puzzles are solved on the basis of “translations” of one case into another (*op.cit.*: 30-31). Further, the basic form of a translation is of an analogy: “[o]ne program is like another except...” (28-29). At the same time, a translation is “more involved than an analogy” (28). “An adequate translation maintains the central metaphors and/or concepts of each account *in their relation to other key metaphors or concepts* in that account. It also compares both the metaphors or concepts and their interactions in one account with the metaphors or concepts and their interactions in the other accounts.” (*ibid*, italics in the original). Translations in a meta-ethnography can be “reciprocal”, when accounts are directly comparable, or “refutational”, when they are in opposition to one another. Studies may also come together into a line of argument rather than represent a reciprocal or refutational translation

---

<sup>80</sup> Turner, S., 1980. *Sociological Explanation as Translation*. New York: Cambridge University Press.

(36). When translations, which are a form of synthesis, become too many, a second level of synthesis/ translation is possible – translations are compared and potentially translated into one another (28).

## 5.2. Coding

In the words of Miles and Huberman, “coding is analysis” (56). “Codes are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study”. Codes are ascribed to segments of information of varying size – words, phrases, sentences, paragraphs. The meaning seen in a segment of text is a choice out of a range of possibilities. The choice derives from a particular logic or conceptual framework of which the researcher may or may not be aware (*op. cit.*: 56-57). From a perspective emphasising ‘ordering’ rather than meaning, “[c]odes are shorthand devices used to label, separate, compile, and organize data” (Charmaz, 1983 cited in Eaves, 2001).<sup>81</sup>

Eaves (2001), who brought together a range of authoritative understandings of coding to construct her synthetic grounded theory analysis technique, articulates nine coding stages used in her substantive work on care-giving in rural African-American families. First, in “line-by-line *in vivo* coding” the researcher uses key phrases in the informants’ own words (*op. cit.*: 658). Then, “shorter code phrases” are developed on the basis of the *in vivo* codes. Third, code phrases are further reduced by grouping together similar code phrases. Fourth, these groupings are brought together into ‘clusters’. Clusters are further reduced into “meta-clusters with labels”. Fifth, the labels for the meta-clusters became ‘concepts’. Sixth, similar concepts are grouped together to develop ‘categories’. “Categories are classifications of concepts, and are discovered when codes are compared against one another ... Categories, then, are of a higher, more abstract order than are codes”. ‘Subcategories’ are then identified, which represent “characteristics and properties of categories along a continuum or dimensional range”. Linkages are made amongst categories and, finally, core

---

<sup>81</sup> Charmaz, K., 1983. The grounded theory method: An explication and interpretation. In: R. M. Emerson, ed. 1983. *Contemporary Field Research: A Collection of Readings*. Prospect Heights, IL: Waveland Press, pp. 109-26.



categories are identified (*ibid.*).<sup>82</sup> This take on coding illustrates clearly the iterativity of the process and the multiple levels at which it happens. It also offers a particular (and maybe particularly confusing) interpretation of the relationships between codes, concepts, categories, classes, properties, characteristics, etc. Interpretations of these relationships tend to vary substantially across authors writing on coding.

One can create codes starting from a provisional “start list” – using “the conceptual framework, list of research questions, hypotheses, problem areas and/or key variables that the researcher brings to the study” (Miles and Huberman, 1994: 58). In more “grounded approaches”, as originally developed by Glaser and Strauss (1967), no pre-coding is done (Miles and Huberman, 1994: 58). Partway alternatives are also available, of starting from very general, formal coding schemes (*op.cit.*: 61). Some codes are descriptive. Others are interpretive, at varying level of interpretiveness, e.g. Miles and Huberman see “pattern codes” as more inferential and explanatory than typical interpretive codes (57).

Codes undergo revisions. “[S]ome codes do not work; others decay. No field material fits them, or the way they slice up the phenomenon is not the way the phenomenon appears empirically.” Such codes are discarded or their level is changed. “Other codes flourish, sometimes too much so”. Such codes need breaking down into subcodes (61). The coding and recoding is completed when “all of the incidents can be readily classified, categories are “saturated”, and sufficient numbers of “regularities” emerge (62).

In addition to codes, a researcher generates “marginal remarks” (or “annotations”) in a coding process – ideas and reactions to the material and the coding process as they come in the coding process. They may suggest new interpretations, capture leads, draw connections to other parts of the data, point to issues to look into in further data collection. They may point to problems with codes and provide the grounds for revisions of the coding scheme (66-69). More extensive “memos” are also produced in the coding process – “the

---

<sup>82</sup> The process continues, but already too far away from codes and their groupings – into theories and explanatory frameworks

theorizing write-up of ideas about codes and their relationships” (Glaser, 1978: 83-4, cited in Miles and Huberman, 1994: 72).<sup>83</sup>

### 5.3. “Travelling”

While ‘coding’ is an old and ubiquitous concept in qualitative research, the “travelling of facts” is a new and possibly still niche concept in the studies of science. It is associated with a research project which ran at the London School of Economics between 2004 and 2009 and a subsequent publication on *How Well Do Facts Travel?* (Howlett and Morgan, 2011). It appears to be the concept from the meta-scientific literature that is closest to that of ‘transformations of evidence for the purposes of health research synthesis studies’.

In Howlett and Morgan’s edited collection, facts are seen as having the tendency to “travel relatively independently to other users, without much reference to their producing context” (*op.cit.*: 26). Facts are also seen as “travelling well” when two conditions are met. First, a certain integrity of the fact is preserved, even if some change in shape, kind or form may have occurred (25). Facts retain their original content even if wearing down, rounding off, enriching and sharpening occur, with some [facts] “becom[ing] simpler and los[ing] information while others add information and become more complex as they travel” (17-18). The second condition for taking facts to have “travelled well” is their fruitful usage – users have had them “fulfil various other functions than those of their production and intended use” (*ibid.*). “At its most fruitful, the use of travelling facts creates a new pattern, a new coherence, a new narrative or fulfils a new role” (20). As much as they were described as travelling independently, facts need travelling companions. These “range from the mundane level of labels and packaging to the more material vehicles of

---

<sup>83</sup> Glaser, B. G., 1978. *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press. Miles and Huberman also discuss “reflective remarks” that are added to field notes in the process of their writing up. These may include, for instance, remarks on the relationship with the informants, now the researcher is off the site; second thoughts on the meaning of what the informants were saying; doubts about the quality of some of the data; new hypotheses, issues to pursue; cross- allusions to other material; personal reactions, etc. They also refer to Bogdan and Biklen (1992) as dividing reflective remarks into remarks on analysis, method, ethical dilemmas, researcher’s frame of mind and points of clarification (*op. cit.*: 66).

transportation, as well as to the people involved in chaperoning, and from the various kinds of institutional structures that support travelling knowledge to the technical standards that carry facts with them” (27). Good companions are seen as supporting the travels of a fact, but not being part of the latter. When the destination is reached, they can be discarded (*ibid*).

This chapter will conclude here, itself travelling towards becoming a much different collection of concepts, quotes and claims. It has provided a glimpse into a broad range of issues that demand many more examples of instantiations in the literature and extensive conceptual and theoretical analysis. The picture that emerges is hardly coherent. In the next brief chapter, I tighten some of the concepts that will be central to subsequent chapters.

## Chapter 3: Postscript to Literature Review – Prescript to Methods

The literature review has shown that we are far from having a network of established concepts for entities and processes concerning research synthesis. In this brief chapter, I lay out the ‘working understandings’ and vocabulary choices I have made in relation to research synthesis, data extraction and the basic unit of analysis/synthesis in research synthesis. It is becoming of little relevance what research synthesis is most broadly since from the next chapter I will be turning to a specific case study. Yet something essential will be missing if, for a concept from the thesis title, nothing more structured is offered than the collection of issues discussed in the review. Most of the vocabulary conventions related to data extraction I use are largely intuitive, but articulations of intuitions reduce ambiguity. Finally, I have settled for ‘findings’ as the main term for unit of analysis/synthesis in research synthesis for pragmatic rather than theoretical reasons (it appears to be the vaguest and thus most open of concepts).

### ***1. The field of research synthesis – a working definition and understanding of main components***

I suggest that there is heuristic value in seeing the following two (composite) features as ones by virtue of which a process or outcome could be classified as a research synthesis process or outcome, with each of the two features being sufficient by itself and the co-presence of both possible but not necessary. These two features have come together in the exemplar research synthesis method – the systematic review inclusive of meta-analysis – but the latter is both much more than them and much less than the range of their potential realisations:

- (1) The primary or only source material that is used in the ‘bringing together’ is health research reports. There is thus a strong element of a literature review, yet unlike the case of traditional literature reviews an attempt is made to reduce bias by making the process more comprehensive than usual and articulating its steps in some detail.

(2) There is a high level of integration of the material used, where the input cannot be recovered from the output, but the transformation process is either fully transparent or an attempt has been made to articulate it in some detail.

I also suggest that there is heuristic value in construing, as a starting point, the health research synthesis field as comprising the nine subfields outlined below. As discussed in Chapter 2 (Section 3.3), the most widespread distinctions made are between traditional systematic reviews and ‘other’ reviews as well as between aggregative (largely quantitative), interpretive (largely qualitative) and mixed reviews. One of the most extensive typologies presented identified eleven review types: critical review, integrative review, literature review, mapping review/systematic map, meta-analysis, mixed studies review/mixed methods review, overview, qualitative systematic review/qualitative evidence synthesis, rapid review, scoping review, state-of-the-art review, and systematic search and review (Booth, Papaioannou and Sutton, 2012). The parameters around which the subfields below have been formed are source material, question type and main approaches used.

- (1) **Exemplar “evidence synthesis” studies.** These address questions of intervention effectiveness, take the form of systematic reviews, incorporate meta-analyses and have the RCT as their gold standard method;
- (2) **Procedurally typical systematic reviews on untypical topics.** Some examples are systematic reviews of methodological issues and approaches (Eaves, 2001; Harris et al., 2008, see also work of the 16 Cochrane methods groups (The Cochrane Collaboration, 2013d); appraisal tools (Vlayen et al., 2005); predictors (Davey et al., 2009); arguments-based clinical ethics literature (McCullough, 2007). It is unclear whether systematic reviews on untypical topics tend to apply the practice of systematic searching while analysing and presenting their material in quite traditional narrative ways (meta-analysis is not applicable in most cases), or whether they tend to propose innovative integrative processes and products.

(3) **Qualitative research synthesis.** An open access review by Barnett-Page and Thomas (2009) brought together dispersed and little known methods of qualitative research synthesis and seems to have channelled work in this subfield. The latter is quite active, with a respective Cochrane subgroup (Cochrane Qualitative and Implementation Methods Group), voluminous register of relevant studies (Cochrane QES Register, 2013, see also Chapter 1, Section 2.3), training courses, and a community for whose members qualitative research synthesis is an important professional identity-defining commitment.

(4) **Integration of qualitative and quantitative research.** Some of the subfield's apparent logic of development – of a large number of methods arising independently then brought together in an influential review (Dixon-Woods et al., 2004) – is similar to that of qualitative research synthesis. Albeit older in terms of its landmark review, the subfield seems less consolidated than that of qualitative research synthesis. Studies of this type are potentially called mixed methods research syntheses, which I preserve for (6).

(5) **Integration of diverse data types on broad research questions.**

These are cases where the quantitative/ qualitative distinction is insufficient to represent the diversity of material used.

An example of a very broad systematic review, designated a “multidisciplinary systematic literature review”, can be seen in the work of the Canadian Initiative for Frailty and Aging (Karunanathan et al., 2009). The investigators reviewed the literature in ten aspects of frailty – biological basis, social basis, prevalence, risk factors, impact, identification, prevention and management, environment and technology, health services, and health and social policy. Although the outcome itself was not a comprehensive framework of frailty (i.e. a ‘proper’ knowledge synthesis), the study was seen as contributing towards the future development of such a framework (Karunanathan et al., 2009: bmc 10).

Realist synthesis (Pawson 2002a, b; Pawson et al., 2004) is an approach of a different order – usually employed in intervention evaluations – which also uses a wide variety of data types, including “action research, documentary analysis, administrative records, surveys, legislative analysis, conceptual critique, personal testimony, thought pieces”, etc. It may also go across policy domains (Pawson et al., 2004: 11).

Institutionally, The Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre) at the Institute of Education, London, is well known for its work on integrating diverse types of evidence on broad research questions (e.g. Harden, Weston and Oakley, 1999; Oliver et al., 2001; Harden et al., 2004; Brunton et al., 2006; Oliver et al., 2008). Some speak of the “EPPI-approach” to research synthesis (Pope, Mays and Popay, 2007).<sup>84</sup> Reviews of this type are also sometimes referred to as “mixed methods reviews” (e.g. Harden, 2011), the idea being that they mix a variety of largely single-method studies.<sup>85</sup>

Research synthesis work integrating diverse data types on broad research questions is strongly policy driven, tends to involve stakeholders quite intensely and is thus amongst the most influential and influenced type of health research synthesis.

- (6) **Reviews of mixed methods studies.** Here, these are conceptualised as reviews of studies that, at the primary level, mix quantitative and qualitative approaches, or research and non-research material. A good mixed methods study report will already have fused findings from the different methods it applied into a well integrated whole. At present, this type of mixed methods review is a rarity. To complicate matters further, there are mixed methods reviews which are reviews interested in the methods used by mixed methods studies (and if they are systematic, can

---

<sup>84</sup> In an informal conversation at the 2011 Cochrane Colloquium a researcher working at EPPI stated quite forcefully that “there is no EPPI-approach!”

<sup>85</sup> More precisely – single-method studies, studies that use a standard combination of methods, or studies that use a combination of methods where the contribution of each method is clearly delineated.

be thought of as systematic reviews on an unusual topic, the first category discussed above).

- (7) **Overviews of reviews.** These are reviews that bring together findings from a number of systematic reviews (Becker and Oxman, 2009). They thus provide a meta-level to what is already a meta-level.
- (8) **Narrowly statistical evidence synthesis methods.** There is much more to statistical evidence synthesis of health and medical evidence than meta-analysis. As mentioned on several occasions, Ades and Sutton's review of multiparameter evidence synthesis in epidemiology and medical decision making (2006) provides a glimpse of this.
- (9) **"Social methods" of evidence synthesis.** These include consensus conferences and approaches such as the Delphi technique.<sup>86</sup>

Finally, I would suggest that whatever features become established as defining of a field of health research synthesis, amongst its most important boundaries will be with secondary data analysis/ data re-use, mixed methods research, interdisciplinary research, the traditional literature reviews, Campbell systematic reviews, integrative medicine, health decision modelling and the intervention mapping approach for developing behaviour change interventions.

## **2. 'Extraction-coding'**

I use 'extraction-coding' as a shortcut to represent at least four inter-related processes which are under way while primary studies are processed so as to identify and 'take out' information that may be used in a synthesis study. These are *picking/highlighting*, *data extraction*, *coding* and *'tagging'*.

In both data extraction and coding there is first, or simultaneously, a process of **picking/highlighting** of information of interest. We can extend this with

---

<sup>86</sup> This is a subfield of which I would not have thought if I have not read Stegenga (2011), after having spoken to the author at a conference.



**constructing** if we want to emphasise the fact that the unit of information is not or need not necessarily be simply there, clearly individuated.

**Data extraction** was discussed in Chapter 1 (4.2) and **coding** in the Literature Review (5.2). Only a note on typical associations will be added.

The term 'extraction' has developed a strong connotation of directness, lack of change – with the extracted information being identical with the source study information. Extraction is thus associated with objectivity and reliability. In contrast, coding tends to be associated with interpretiveness and subjectivity.

Data extraction, however, cannot avoid interpretativeness entirely. The process of transfer to a new place requires at least re-classifying on at least some occasions (see, for instance, the formation of new groups under 4.3.2 in Chapter 1). Similarly, some coding may be quite 'non-interpretive' / 'non-transformative': the code repeats what is in the strip of text of interest and the coding acts as highlighting/picking rather than 'proper' coding involving a (re-) conceptualisation. This type of coding is very similar to data extraction as associated with 'lack of change'. As a result, how transformative a gesture of data extraction or coding is is a matter of judgement in particular cases only.

I use **tagging** for the process of adding comments and annotations of the following kinds: 1) critical appraisal comments addressing the methodological rigour of studies; 2) 'critiquing comments' challenging a study's foundations as supplied by the research tradition in which it is embedded; 3) meta-critical comments on 1) and 2), which question my grounds for making the comments under 1) and 2); 4) 'research process' tags – showing the development of the ideas and conclusions of the study – preliminary observations, operationalisations, change of response to the material, pragmatic issues, e.g. behaviour of the software, etc.; 5) 'self-monitoring for rigour' tags – notes serving to observe and potentially revise problematic processes and to provide an 'audit trail' for problematic decisions; and 6) 'self-monitoring for tendencies of thought' tags – notes that aimed to capture my broader interests, concerns and proclivities as triggered by the material analysed. This tagging took to extremes, in a broadly rationalist framework, two of the main virtues of research synthesis – transparency and critical analysis.

In view of the above, the typical processing of source material for research synthesis can be labelled as **picking-extraction-coding-tagging**. A brief and precise re-labelling of the picking-extraction-coding-tagging would be **picking-processing**. To make it easier to trigger associations for those familiar with the field of health research synthesis, I use **extraction-coding**.

### ***3. 'Findings' as the preferred term for the basic carrier of primarily empirical knowledge in a research report***

In Chapter 1, I used 'evidence' and 'data' as the standard generic terms for units of information of interest to a synthesis study. I anticipated revisiting the terminology as I wanted to avoid the strong connotations of 'pure', 'uninterpreted', 'theory-free' knowledge which these terms have in the context of evidence-based medicine. I was inclined to think that 'findings', which in qualitative research synthesis has come to have connotations of interpreted, processed data, may be a better term for the basic unit of analysis/synthesis in research synthesis. It seemed more likely to trigger associations with 'transformations' as well as with the frameworks, factors, causes, etc. that shape pieces of empirical knowledge over and above 'the way the world is' and make it more or less similar to other pieces of empirical knowledge. The literature review showed, however, that each of the basic carriers of empirical knowledge considered (data, evidence, findings, claims, facts and observations) is associated with influential debates on the complex shaping of empirical knowledge. Claiming that research synthesis is a synthesis of findings, rather than evidence or data, was not going to trigger unique associations.

I nevertheless chose 'findings' as the preferred term for what, as a start, could be seen as the basic unit of analysis/synthesis in research synthesis as it seemed to be the vaguest and thus most inclusive concept. Very small fragments of empirical knowledge (data points), rich clusters of information (extended interpretations), 'normal sentences' which represent them, etc. are comfortably referred to as 'findings'. For the purposes of the subsequent analysis, I will thus take as (a representation of) a finding any sentence within the text or any proposition represented elliptically in a table or box which has

been reported in the Findings/Results section of the sampled papers. This operational definition is reconsidered in Chapter 5.

I will only add a supporting concept to be used around the concept of findings, primarily in the Methods Chapter. 'Matrix' will refer, in a most generic way, to the factors that shape findings other than how the world is (if it is a certain way). It is a looser concept than the Kuhnian "paradigm". First-line examples of matrices, corresponding to easily identifiable debates in the meta-scientific and meta-methodological literature of what shapes evidence, are: paradigms, theories, methods, instruments, frameworks of aims and objectives, fundamental assumptions about the world and knowledge, concepts, representational conventions, models, networks of auxiliary assumptions, and cultural-historical contexts of production.

In earlier versions of the thesis, 'matrix' was a key concept. Currently, it does not carry any weight in representing the findings about transformations and their discussion. These were more naturally framed around other concepts. However, it provided an organising role in the Analysis Framework for the case study and will occasionally appear in the next (Methods) chapter.

## Chapter 4: Approach to Case Study on Extraction-Coding from Cancer Research Publications

There was little methodological work that could foreground directly the case study for this thesis, with the untypical questions it set out to explore and its aim to bring closer together the meta-scientific, methodological and empirical. Some broad similarities can be drawn with the work of Sandelowski and Barroso (2002) who sought to develop a research protocol for conducting qualitative metasyntheses in any health-related field and used qualitative research on women with HIV as their “method case”. The authors were not interested in the metasynthesis per se but in creating an audit trail of the process and communicating the challenges arising in it (Sandelowski and Barroso, 2002: 213). Similarly, Popay et al. (2006) developed their guidance on narrative synthesis by first performing an extensive literature review on appropriate tools and techniques and then applying it to two bodies of evidence while taking detailed notes of all major decisions and the reasoning behind them. That is, they prospectively documented the synthesis process and urged reviewers adopting a narrative approach to do the same (Popay et al., 2006: 67). Some parallels can also be found with the work of Hasok Chang (2011) who argues in favour of a new framework for the description and analysis of scientific practice and makes initial proposals in the direction of capturing the physical, mental and paper-and-pencil “epistemic activities” performed by scientists, the “doings” and “happenings” in science (Chang, 2011: 208).<sup>87</sup>

These were, nonetheless, minor and somewhat latecoming influences. The multi-method approach I used is not easily relatable to any single methodological tradition or study design. It combines some very basic, fundamental and conventional health research (synthesis) and broader scholarship practices with practices that depart radically and uncomfortably from what can be found in ‘normal’ health research (synthesis) studies. It is a combination of the highly conventional and highly unconventional.

---

<sup>87</sup> In distinguishing physical, mental, and paper-and-pencil operations Chang follows Percy Bridgman (1959: 3). “Doings” and “happenings” are also Bridgman’s words.

Described minimally, this case study was *an empirical investigation of similarities and differences between types of evidence and their transformations in the context of an 'extreme' data extraction from academic journal publications on cancer.*

A more extended but still condensed description can be as follows:

*This case study involved the performance, observation and exploration of a process of 'extraction-coding' from research papers on cancer and exploration of the input and output of this process.*

The *process of extraction-coding* was brought to life by carrying out a test synthesis on how to introduce changes in behaviours and mental contents that may relate to cancer outcomes.

The main *input* to the extraction-coding (source studies for the test synthesis) were highly varied research publications on cancer.

The main *output* of the extraction-coding were preliminary groupings of findings, claims and concepts. A complete synthesis was not aimed at.

The *performance of the extraction-coding* involved 'standard' extraction-coding, as done for the purposes of most synthesis studies, but was also driven by five non-standard 'injunctions'.

The *non-standard injunctions* driving the extraction-coding took a virtue of synthesis studies to an extreme or extended its scope (e.g. transparency became extreme transparency or the degree of comprehensiveness of literature searching was also applied to the extraction-coding).

The *observation and exploration* of the process of extraction-coding and the exploration of the input/output material focused on the deployment of the processes of transformation, their frequency and scope, and their drivers. The driver of transformations that received particular emphasis, by virtue of being embedded in the preliminary conceptualisation of transformations, concerned the nature and dynamics of similarities and differences between pieces of evidence. The observation and exploration were both highly structured and receptive to the unexpected: an extensive Analysis Framework was prepared in advance and tools and methods were designed as the study evolved.

I unpack those descriptions below. In Section 1, I justify the choice of topic and test synthesis question. In Section 2, I describe how the material subjected to extraction-coding was obtained – namely the search strategies used and the approach to randomly sampling from the initial retrieval of over 18,000 papers. In Section 3, I specify features of the performance, observation and exploration of the extraction-coding. I first address the five untypical injunctions driving the extraction-coding (3.1). I then describe the Analysis Framework (3.2), starting from its backbone of five broad parameters and moving towards an extensive tabular illustration of specific parameters. Software is covered briefly in 3.3. In Section 4, I outline a range of supporting tools and approaches. These were developed largely to complement the Analysis Framework or probe further into unexpected findings obtained through it. Central amongst these was the Vocabulary of Elements of Findings.

Further justifications of methodological choices were already given in Chapter 1. In Section 5.2, I explained the attention to data extraction rather than some other phase in the research synthesis process; the preference for a strongly introspective and personal approach rather than more externalised and community-based one; and the narrowing to textual and textual-numerical transformations. In Section 5.3, I discussed briefly the difficulty of separating the performance of data extraction from its observation and exploration. I will not be revisiting these justifications here but will attend to them again, in a more critical fashion, in the Discussion.

### ***1. Topic and question for the test synthesis underpinning the extraction-coding***

The extraction-coding was performed with a view of a test synthesis addressing the following broad question on behaviour and ‘mind’ change in cancer:

*What services and interventions are likely to be effective in changing behaviours and mental contents (cognitions, emotions, attitudes, etc.) that can affect the incidence, trajectory, experience, recovery or mortality from cancer? In a contextualised version of this question, what is the*

*contribution to improved cancer outcomes of such services and interventions relative to biomedically-based ones (e.g. medication, screening, genetic tests, etc.)?*

I formulated such a broad question to ensure as wide variety of findings as possible. As discussed in Chapter 1, Section 5.3, it is unrealistic to seek an answer to this question not only within the constraints of the thesis but also in the context of a standard research project. Only a huge research programme of many years and researchers stands some chance of performing a high quality review of the available evidence. The breadth of the question was seen, however, as enabling the natural coming together of health services research, epidemiological, psychological, psychoneuroimmunological research, behavioural medicine, social sciences research on health, organisational research, health economics, philosophy of medicine, medical humanities, etc. – variably individuated health-related fields outside the biomedical. These would bring in their potentially incongruous fundamental assumptions and concerns; theoretical, methodological and pragmatic frameworks; concepts, vocabulary, values, etc. A variety of source studies would allow for a broad range of transformations and ‘difficult cases’ to emerge during the extraction-coding (e.g. it is, presumably, more difficult to combine scientific and humanities thinking than two RCTs on the same topic). The variety, opportunities and limits of processes of transformation could thus be better explored.

I chose cancer as the thematic focus for the same reasons – maximum variability and potential incongruity of types of findings and their frameworks. Other topics considered were obesity, dementia, psychosomatics, genetic testing, minor environmental cues in making health-related decisions, etc.

Cancer is a highly heterogeneous, prevalent and potentially serious condition. As per one view, there are over 200 types of cancer, reflecting the variety of cells in the human body (Cancer Research, 2011). For the US population, the lifetime risk for developing cancer has been estimated as 44.81% for men and 38.17% for women. The risk of dying from cancer has been estimated as, respectively, 23.08% and 19.39% (American Cancer Society, 2012). Cancer is thus, unsurprisingly, an intensely researched condition. Close to 11% of the

publications on PubMed (over 2,4 million out of 22 million) are indexed for “Neoplasms”, the controlled vocabulary term for cancer (search run Mar 13).

Cancer is also associated with a highly complex psychological, social, cultural, ethical, etc. background that seems to have a tangible effect on how it is studied. A telling illustration of the effect on cancer research of such factors (in this case a dramatic socially shared image and psychological poignancy) can be found in the relationship between its volume on the one hand and a rationally assessed seriousness of the condition and burden relative to other serious conditions on the other hand. Consider the statistics below relative to the fact that 11% of PubMed publications are on cancer.

Amongst the ten leading causes of death throughout the world (accounting for 52% of all causes of death), there is one cancer-related: trachea, bronchus and lung cancers are estimated to cause 2.4% of the deaths around the world (World Health Organization, 2008). The three leading causes are ischaemic heart disease (12.8% of all deaths), stroke and other cerebrovascular disease (10.8%) and lower respiratory infections (6.1%). The ten leading causes of death in low- and middle-income countries do not include any type of cancer. In high-income countries, there are three types of cancer in the top 10 – trachea, bronchus and lung cancers (5.9% of all deaths), colon and rectum cancers (3.3%) and breast cancer (1.9%) (*ibid.*). It was assumed that if psychological, social, cultural and suchlike factors are affecting the volume of research on cancer, they may also be affecting its contents and presentation more strongly than in the case of other health topics. As a result, a greater degree of critical analysis of the primary research may be demanded, which is likely to result in ‘more and greater’ transformations.

Crucially, I chose cancer in preference to other burdensome, intensely researched, and psychologically, socially, ethically, etc. complex conditions because of its strong association with alternative practices premised on mind-body interactions and a growing scientific interest in the same direction. Claims of mind-body interaction may concern, for instance, the effect of prevalent emotional states, stress, personality type, degree of optimism, etc. on the likelihood for developing and recovering from cancer. These are controversial issues, especially when it comes to non-orthodox therapeutic modalities



entering into conflict with biomedical models of aetiology and treatment. But they are also ones from the leading edge of research (as is the case with psychoneuroimmunology). As a result of both the controversy and innovativeness, a vagueness of models and a substantial degree of conflict and discrepancy between them can be envisaged. Again, I saw these characteristics of the field as likely to be associated with substantial difficulties in the combinability and transformations of respective findings. Why have it easy when it can be difficult.

## ***2. Process of identifying the sample of papers***

To capture studies from the variety of fields targeted (e.g. health services research, epidemiological, psychological, social sciences research on health, organisational, health economics, medical humanities, etc. research), I combined a previously developed “brief and precise filter” for identifying publications on “health-related values” in MEDLINE (Petrova et al., 2012) and a Mind-Body search strategy. The conceptualisation of ‘values’ in the filter design study was very broad, to include not only ethical, moral, religious and other types of values ‘proper’, but also beliefs, preferences, experiences, choices, satisfaction, quality of life, etc. All of these have some mental representation and the capacity to affect decisions and behaviours (as demanded by the test synthesis question). The filter has been found to be highly effective in retrieving material from all of the above research fields. It was developed using principles of objective search filter design, in this case word frequency analysis, and had demonstrated very high levels of sensitivity, specificity and precision (*op.cit.*). I developed a pragmatic Mind-Body search strategy by selecting relevant Medical Subject Headings. Its aim was to identify research on supposed direct effects of the mental on the physical, unmediated by behaviours. (Further details of the search strategies are provided in the Appendix to Chapter 4.).

The impact of change of behaviours and mental contents on the incidence, trajectory, experience, recovery or mortality from cancer, and the effectiveness of services and interventions to help achieve those changes, was to be relativised to the impact of biomedical factors and biomedically-based services

and interventions. As my intention was nonetheless to focus on the former type of factors, no separate search strategy was developed for biomedical studies. I used false positives from the other search strategies. They are always in abundance.

Cancer studies were targeted through the Medical Subject Heading “Neoplasms”.

Finally, retrieval was limited to the three years preceding the running of the search – between June 2007 and June 2010. No language limits were placed.

This approach yielded 18,456 publications. From these, I sampled one hundred publications using a random numbers generator ([www.random.org](http://www.random.org)). I intended to extract from/ code all one hundred papers but the process produced much more data than I had envisaged. Certain patterns showed much earlier than expected. For certain other patterns to emerge, 100 studies would have still been an inadequately small sample. Relevant saturation was thus reached very early (at the seventeenth study). The approach of moving from the pool of 100 randomly sampled studies to the final purposive sample of analysed studies is discussed in Chapter 5.

### ***3. Features of the performance, observation and exploration of the extraction-coding***

#### **3.1. Standard and non-standard features of the performance of the extraction-coding**

Standard extraction-coding of information from the sampled papers included the identification and processing of findings, concepts and methodological features of the study of interest. Ostensibly, these findings, concepts and methodological features were to serve the purposes of the test research synthesis. My primary interest was, however, in the *process* of obtaining them. I also analysed some of the standard material relative to the thesis questions as opposed to the test synthesis question. Details of the standard extraction-coding can be found in Table T 4.2.

I envisaged non-standard practices of extraction-coding as tools for looking closer into the nature and drivers of transformations. Their application would also be a test of opportunities for methodological improvements in health research synthesis work. Non-standard extraction-coding was non-standard by virtue of adhering to the following five 'injunctions':

- *code all text* (also referred to as 'the injunction for comprehensiveness');
- *multiple code as a default strategy* (also referred to as 'extensive multiple coding');
- *first code as close as possible to the original, with no or minimal transformations, then experiment* ('close to the source first, experiments afterwards'); *combine critical appraisal with critique* ('double-edged critical analysis');
- *be extremely transparent* ('extreme transparency').

What makes these injunctions unusual is the degree to which and the actual processes by means of which the virtues underlying them are realised (e.g. the virtue of giving a complete account of one's steps so as to enable criticism and replication). If we look at the underlying virtues in the abstract, we can just as well say that those are the most standard research injunctions imaginable. Each of the injunctions is discussed below.

I conceptualised the injunction for **comprehensiveness** as a tool to help explore the potential contribution of any study element to transformation decisions in research synthesis. At least in theory, any feature of a study and its presentation may affect the particularity of a finding, its representation and thus its similarities and differences from other findings. I wanted to see the effects of extracting 'more' data (most of which would be typically considered meta-data, meta-information about the core data) on transformation decisions. My expectations of the effects were contradictory. On the one hand, more meta-information could furnish further lines of similarity between findings and markedly increase opportunities for transformations. More synthesis opportunities would thus arise and more varied knowledge and/or greater uncertainty from the multiple re-uses. On the other hand, further lines of difference between findings could become manifest and opportunities for

transformations markedly constrained. Much fewer syntheses would then be sanctioned as possible or rigorous. Would the similarities or differences weigh more? What will proposals for methodological improvements look like, since in both cases something of the certainty of synthesis knowledge would be lost?

I introduced the injunction of **extensive multiple coding** so as to bring to life the process of transformation – to make it happen again and again with greater or lesser variation, and so as to provide data about the openness of the same source material to varied, equally justifiable, transformations. As in the case of comprehensiveness, I had contradictory expectations of the effects of extensive multiple coding and the likelihood of it contributing to methodological improvement in research synthesis. On the one hand, it could substantially increase rigour (by allowing reasoned selection from a range of options and the more reflective generation of transformations). It could also open up the richness of a study's findings and increase their value by enabling more opportunities for re-use. On the other hand, extensive multiple coding could end up clashing with needs for actually achieving a coherent synthesis by creating too many analysis opportunities.

The injunction for extensive multiple coding went hand in hand with a supporting injunction to code as close as possible to the original first, with no or minimal transformations, and then experiment with larger transformations – **close to the source first, experiments afterwards**. It was not the chronology that mattered – I could start with a 'big transformation' if such was triggered before any other – but the fact that each unit of information *had to* receive a minimalist coding involving no or a minor transformation. The aim of the injunction was to help track transformations occurring at a variety of levels and compare them to a baseline of unproblematic (or at least as unproblematic as possible) transformations. Coding as close as possible to the original was mandatory: apart from providing a comparison background of unproblematic transformations, it ensured comprehensiveness of the data extraction. I sought more interpretive transformations consistently, but did not force them if they were not forthcoming. In setting up this injunction, I hypothesised that synthesis studies with a greater number of minimal transformations are likely to be more

trustworthy than synthesis studies with a smaller number of minimal transformations. I later revised this expectation.

The injunction for **double-edged critical analysis** – to combine critical appraisal with critique – used a distinction drawn by Dixon-Woods et al. (2006). Briefly, critical appraisal addresses methodological rigour and tends to use critical appraisal checklists. This is the only type of criticism that mainstream synthesis methods incorporate. Critique involves challenging “the ways in which the literature constructs its problematics”; the epistemological and normative assumptions it draws on; the traditions that have guided particular research fields; the particular forms of discourse available; the factors influencing the proposed solutions (Dixon-Woods et al., 2006: bmc 2, 6, 9). A critique may start as a challenge to a particular study, but there is always something more generic about it that touches on the broader research field. A critical appraisal is of a particular study only. I use ‘double-edged critical analysis’ to incorporate both activities. I took both of them to be drivers of transformations whose workings needed to be observed closely. Processes of critique help determine the meaning of studies and findings and similarities and differences between them and thus prepare or deny grounds for transformations. Critical appraisal gives reasons for adjusting the credibility and strength of claims and thus, too, prepares the grounds for transformations. Again, my expectations were unsettled if a much higher degree of critical appraisal and critique than currently employed had a chance of becoming a routine feature of research synthesis studies. Theoretically, this was appealing – music to the ears of anyone concerned with research rigour. But extensive critical analysis could also end up identifying too many and too large dissimilarities across studies, too many and too large quality failures and, eventually, hamper transformations.

Relative to the main aims of the thesis, **extreme transparency** was there to sensitise me to yet more features of the processes of transformation and their fleeting enablers and controllers (e.g. associations coming from background knowledge, recourse to material from the text other than that in the immediate context, subtle triggers or warnings against certain transformations). I expected

that in some cases increased transparency will be heuristically productive and lead to more transformations – for instance, by making it more likely to catch easy-to-lose ideas and following their leads to transformations. In other cases, I expected it to be corrective and limit transformations – for instance, by helping to detect biased arguments in intended transformations.

Extreme transparency was the injunction with the most ‘special status’ of all. One of the reasons was that apart from working independently, it had to be turned onto the workings of all other injunctions.<sup>88</sup> The other reason was that I saw it, unlike any other injunction, as an unquestionable component of my proposals-to-be for methodological improvements in health research synthesis. A core background assumption (bias?) of mine was that transparency in current research synthesis studies was not enough, or not of a good enough kind, and more of a better transparency must be both feasible and highly desirable.

I discussed a range of views on transparency in the literature review chapter (3.2.4) in the context of defining characteristics-cum-virtues of research synthesis. The concept remained, however, undefined. By way of preliminary definition, I took transparency to be

*the (realisation of the) goal of representing the details of the setup and workings of a research process, whether technology-driven or primarily mental, which involves the enlisting of all methods, approaches, tools, etc. used and describing as much as possible of 1) what is not typically represented as a method, approach, tool, etc. but has been employed in a particular study to drive the process forward while other plausible options were available, and 2) what is not represented in standard descriptions of a method, approach, tool application, etc. but, as above, has been employed while other plausible options were available.*

Unsystematic but long running observations of mine preceding this work (largely influenced by cognitive and social psychology thinking) had led me to conclude that relative to such a conceptualisation, certain aspects of transparency are

---

<sup>88</sup> To an extent, the critical analysis injunction worked similarly, but I took it that there is a ‘final’ level where transparency only remains, not followed by a critical analysis of how transparent something is. But perhaps such a critical analysis happens, gives an ‘all clear’ and remains unnoticed?

consistently neglected in research synthesis<sup>89</sup> (no detailed representation is given of methods, approaches, etc. where numerous, or at least several, other options are available), while some of the typical transparency descriptions are superfluous (for instance because our cognitive apparatus cannot function otherwise). This is a strong and controversial claim. It needs to be adequately substantiated. I will not be able to do it here, as its defence requires a stand-alone study, if not a research programme. While it is not a claim which is incorporated in any key argument of the thesis, which would make its adequate justification crucial, it has had an effect on many of my decisions and emphases – from constructing a methodology of certain features (including the injunctions), through the choice of issues to report on, to the presentation style.

My valorisation of transparency was not without its boundaries. I expected that, most likely, I will hit them much sooner than I hoped. At the very least, when we (re-)engage in serious thinking about thinking, our own on a very particular occasion or thinking in general, we (re-)discover what an opaque process it is. At the very least, transparency in research synthesis would inevitably be hitting the boundary of an inability to reconstruct our own mental processes. One of my aims in going for an extreme transparency injunction was precisely to explore the boundaries of the extremeness and formulate a realistic methodological recommendation.

For example, I saw the exploration of the boundaries of transparency in transformations (and there could be transparency of data extraction, transparency of critical analysis, transparency about transparency, etc.) as concerned with the following questions: To what extent is it possible to specify *which* discrete and (relatively) simple elements within a finding gave grounds for a particular transformation of this finding? To what extent is it possible to specify *how* these elements became part of a more or less complex derivation of a transformed finding which preserved the truth value of the original finding either categorically or probabilistically? And, finally, *what* are the known and unknown steps in this derivation?

My views on the boundaries of transparency were amongst those that changed the most as a result of the case study work. As a background assumption,

---

<sup>89</sup> And research more broadly.  
167

however, transparency was a major determinant of methodological and presentational choices.

Going back to all five injunctions, the main prompts for choosing them were as follows. Comprehensiveness, transparency and some form of critical analysis are all virtues of health research synthesis. As commented in Chapter 1, Section 6, these are, however, applied in very particular ways. A question arises whether further ways can be added. This was an attempt to do so. Multiple coding was a practical translation of theoretical ideas of non-essentialism and pluralism in the philosophy of science, mainly as in Dupré (1993; 2001). It was a particular attempt of having current philosophical ideas about science feed back into actual research practice. The injunction to code close to the source first was another expression of a concern with transparency and rigour and an attempt to realise those virtues in new ways.

The injunctions were operationalised as parameters in the Analysis Framework (such as critical analysis prompts or prompts for the generation of alternatives/lateral thinking). More consequentially, they were mental notes that structured my internal process of extraction-coding which, in turn, generated outcomes that materialised them (e.g. a large number of transparency/critical analysis memos, a large number of multiple codes, transformations at a variety of levels, etc.).

The extraction-coding was performed initially in QSR NVivo 8 for more qualitative data (QSR International, 2013) and Excel for more quantitative and categorical data. It was later moved to EPPI-Reviewer 4, specialised research synthesis software (EPPI-Centre, 2013). These are all standard health research (synthesis) tools combined in a non-standard, and in certain respects sub-standard, fashion. Data collection and analysis software is discussed in greater detail in Section 3.3.

### **3.2. Analysis Framework underpinning the extraction-coding**

The original version of the Analysis Framework underpinning the extraction-coding from recent publications on cancer comprised 205 parameters. These were extended further during extraction-coding.



The Analysis Framework combined parameters generated on the basis of the literature review and my previous research experience, and parameters operationalising more freely questions and hypotheses of this thesis when little directly relevant literature or experience were available. Five main versions of the Analysis Framework were developed, in Excel, without direct testing on studies. Testing on an actual publication made it immediately clear that I needed to change platform (rather than Excel only, a combination of NVivo and Excel). Further revisions were introduced in the process of transferring Excel categories into NVivo codes (prompted primarily by the fact that NVivo allows hierarchical ordering which Excel does not and by the capacity of any reorganisation to reveal new relationships and omissions). A second 'initial' testing on two studies was then performed. As a result of it, categories were further revised and some were 'quarantined' as currently ineffective (practically none of the quarantined categories was used later – 40 days of isolation turned into eternal damnation). The Analysis Framework was thus reasonably stable from the third analysed study onwards. Smaller revisions were introduced throughout the analysis, with their number substantially diminishing after the tenth study.

There were five broad types of parameters included in the Analysis Framework.

The first type of parameters served *to bring to life the processes of extraction-coding and transforming* core contents that is conventionally extracted for the purposes of research synthesis (such as findings, information on research aims, methods, settings, etc.).<sup>90</sup> This was standard research synthesis work with a non-standard aim – not to lead to an answer to the synthesis question but to have it happening so as to observe the extent to which processes of extraction-coding are direct or transformative, how transformations happen, and how they can be made more rigorous.

---

<sup>90</sup> As I found no sufficiently abstract analysis in the literature, I took that most synthesis researchers would agree that *prima facie* the core information they extract and which allows them to make the claims they are making in a synthesis is on findings (possibly including interpretations), study design and method, study aims and objectives, populations, settings and (some) concepts. I used this as a starting point to specify an understanding of 'conventional core contents'. 'Conventional' was added as for some 'non-conventional' synthesis methods information on theories and other matrices can also be 'core'.

The second type of parameters aimed *to compress the varied conventional contents* extracted for the test synthesis into *more formal types and features*. I expected this to illuminate lines of consequential similarity/ difference between various units of information.

The third type of parameters served to elicit a broad range of *matrices of the source studies* (factors that have shaped findings other than how the world is, see end of Chapter 3 for a brief discussion) and explore their effects on the combinability of findings.

Parameters prompting me to identify *matrices of the test synthesis study and the broader thesis work* (i.e. factors that were shaping my own findings) were the next type of parameter of interest.

Finally, 'paired parameters' were included that aimed to compare output from the test synthesis with input from the primary studies. Conclusions could then be drawn about the transformations that led from one to the other.

Table T 4.1 below presents the five types of parameters and the questions for which they enable observations or data collection. The question formulations modify the initial specification of the broad thesis questions as given in Chapter 1, Section 5.4.

**Table T 4.1: Types of parameters in the Analysis Framework**

Type of parameters	Related questions
<p><b>1. Parameters which prompt to identify, extract, describe and analyse conventional core contents for the synthesis study (on behaviour and 'mind' change in cancer)</b></p>	<p>How direct/ transformative is the extraction-coding of the material conventionally extracted-coded in synthesis studies? What is involved in the process? How reliable and transparent is it? Can it be made more so?</p> <p>How does the conventional core material compare to what is extracted-coded through the range of parameters used in this study? How useful is this other material for informing processes of transformation of the conventional core material?</p>
<p><b>2. Parameters that serve to describe formally, theoretically the conventional core contents of a synthesis</b></p>	<p>How can we describe the core contents conventionally extracted-coded in research synthesis studies with a view to similarities/ differences between different types of it? E.g. what types of findings, along a range of dimensions, can be identified?</p> <p>How do different types of findings and interpretations differ in terms of their distance from (raw) data? How does the extraction-coding and the transformations inherent in it further increase, or not, this distance? What are the implications of this for our trust in different types of synthesis studies?</p> <p>(see also Parameters Type 3) How are matrices entangled with the core material? For instance, to what extent can a finding be separated from the theory, method and worldview assumptions underpinning its generation?</p>

<p><b>3. Parameters that prompt to identify, extract information about, describe and analyse the matrices of the source studies</b></p>	<p>What are the matrices of which one can identify 'sufficient' information in a health research report?</p> <p>How direct/transformational is the extraction of information about matrices from source studies? Is much interpretation, critical analysis and external knowledge required?</p> <p>Can one develop specific enough hypotheses as to how matrices have shaped a finding and how it could have been different?</p> <p>Is extracting more information on matrices likely to enhance the confidence and rigour with which we bring together core material? Or, on the contrary, will it demonstrate how few pieces of information are unproblematically combinable?</p> <p>Even if desirable, would it be feasible to expand existing synthesis methods in the direction of more extraction-coding of information on matrices?</p>
<p><b>4. Parameters that prompt to identify, describe and analyse matrices and processes of the test synthesis study and of the broader thesis work as activated and unfolding during the extraction-coding</b></p>	<p>What further than the conventional background and method information accompanying a research synthesis study may be needed so that its own matrices (which too will affect transformations) are more clearly and accurately represented?</p> <p>How feasible are such increased self-reflexivity and transparency? How effective are they in tracing and guiding transformations? Could they become a basis for methodological improvements?</p>
<p><b>5. 'Paired parameters' that aim to explore input relative to output (units of information as appearing in the source studies relative to their transformed version in the synthesis study)</b></p>	<p>What can we learn about processes of transformation by comparing various types of input to and output of a synthesis? To what extent is the output predictable from the input?</p> <p>Can we use knowledge of the predictability or otherwise of transformations to facilitate and improve research synthesis? E.g. can we improve indexing and linking practices in bibliographic databases?</p>

Table T 4.2 below illustrates extensively specific parameters from the Analysis Framework coming under the first three of the broad types presented above. Specific parameters of type 4 (concerning the matrices and processes of the test synthesis study and the broader thesis work) and of type 5 (paired input-output parameters) are not included as they were used to a very limited extent. Pre-formulated prompts for identifying subliminal matrices of the test synthesis turned out not to be particularly effective (e.g. prompts for ascertaining my levels of familiarity with topic/method or prompts aimed to capture my overall attitude and potential biases towards certain research questions, types of studies, methods, etc.). Unprompted memos, annotations and entries in a detailed log were used much more extensively. These served as raw material for further analysis, including further coding. Paired input-output parameters (such as on the contents, form, centrality, quantitative and modal qualifiers of findings, concepts, entities, indexing terms, etc.) were used rarely. The depth of data extraction has left synthesis concepts and findings underdeveloped and the intended analysis of input-output relationship was not performed. This is discussed further in the Findings and Discussion chapters.

In the NVivo and EPPI-Reviewer environments where the actual data collection and analysis were performed, the specific parameters illustrated below were organised differently – around structural elements of a research article rather than around the five broad parameters. It needs to be noted that specific parameters were seen as naturally belonging to more than one category. That is, not only the source material was multiply coded. Elements of the Analysis Framework also had multiple memberships. "See also" or "shared with" notes were added (for instance, findings may reveal new aspects of a core concept). A disadvantage of such an approach was that shared codes had to be updated separately – changes in one of them were not reflected in automatic changes in the other(s). Such synchronisation was often not achieved. Current health research software is not well suited to extensive multiple coding. Software and its use here are addressed matter-of-factly in the section immediately after the table and, more critically, in the Discussion Chapter.

**Table T 4.2: Examples of specific parameters from the Analysis Framework**

*Note:* This is a long table including varied information (while being only illustrative of a much more extensive Analysis Framework). A consideration that may help in going through it is that parameters from the Analysis Framework were both instructions and ‘boxes’ for storing information. For instance, ‘Secondary aims and objectives’ is both an instruction *Identify secondary aims and objectives/ Derive on the basis of indirect claims* and a sign that in the actual software environment relevant information was stored under this heading.

Type of parameter	Examples of specific parameters
<p><b>1. Parameters that prompt to identify, extract, describe and analyse conventional core contents for the synthesis study</b></p>	<p><b>1.1. Parameters that broadly map the domain of findings</b></p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> <li>• Explicit key aims and outcomes of the study</li> <li>• Secondary aims and outcomes</li> <li>• Research design, in broad terms</li> <li>• Brief analytical description of study – on the basis of critical comparison of study aims and actual outcome, what would you say this study is and does?</li> </ul> <p><b>1.2. Parameters concerning the primary knowledge carriers in an empirical report</b></p> <p><b>1.2.1. Empirical findings (data, evidence)</b></p> <p><b>1.2.2. Concepts</b></p> <p><b>1.2.3. Interpretations of findings</b></p> <p><i>Examples of parameters concerning concepts:</i></p> <ul style="list-style-type: none"> <li>• Explicit definition, description or brief operationalisation</li> <li>• Broader context that gives the concept further meanings</li> </ul>

**1. Parameters that prompt to identify, extract, describe and analyse conventional core contents for the synthesis study (cont.)**

- Types
- Types of data collected about it
- Other concepts/variables relative to which it is explored
- Further characteristics from parts of the paper other than main locus where the concept defined
- Author's meta-comments on concepts, e.g. in strengths and limitations

**1.3. Parameters concerning the entities, phenomena and contexts that are sampled and/or to which the study findings and interpretations pertain**

In health research studies on humans, those entities, phenomena and contexts are typically:

- Diseases, health conditions and their trajectory, as objective entities and processes; the phenomenology of their being lived, experienced, managed and encountered by people (relevant parameters mostly under features of study sample and findings)
- Interventions (relevant parameters mostly under methods and findings)
- Health and well-being outcomes (relevant parameters mostly under concepts and findings)
- Populations, individuals and groups
- Settings/contexts – to include both the narrow study context and the broader context within which the populations, individuals, groups and interventions studied are situated

*Examples of parameters concerning study context:*

- Broad context – country, region, type of healthcare system, culture and values – partly pre-formulated, partly emerging features; from main locus where setting discussed and as dispersed throughout the text
- Narrower setting, e.g. hospital, GP practice, community centre – hardware features (geographical location, size, number of patients served, etc.) and software features (interpersonal dynamics, policies, etc.) – partly pre-formulated, partly emerging; from main locus and as dispersed throughout text
- Features of context explored in relation to which variables of interest? Correlate with which of these? Which parameters are reported but their relationships with the findings are not explored or reported?

<p><b>1. Parameters that prompt to identify, extract, describe and analyse conventional core contents for the synthesis study (cont.)</b></p>	<ul style="list-style-type: none"> <li>• Based on the above information, how easy/difficult is it to make a claim about the similarity/ difference of context and the relevance of certain features to judgements of combinability/compatibility?</li> <li>• Authors' claims concerning strengths and limitations of carrying out the study in a particular context and generalisability of findings; attention to the rhetoric of those claims</li> <li>• Prompts and spontaneous comments concerning context and its representation – e.g. further information required, potential relationship with study variables, features which may have had a stronger explanatory potential in comparison to the reported ones, etc.</li> </ul>
<p><b>2. Parameters that serve to describe formally, theoretically the conventional core contents for a synthesis</b></p>	<p>The parameters of this group aimed to prompt the 1) identification (from the report) or generation (on the basis of information from the report) of <i>types</i> of entities and scientific artefacts comprising the conventional core contents of a synthesis study, 2) the identification of more or less standard <i>features</i> to describe entities and scientific artefacts comprising the conventional core contents, and 3) the generation of further more <i>abstract re-descriptions</i> of such types and features.</p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> <li>• Types of 'work' (study, paper)</li> <li>• More abstract re-descriptions of aims</li> <li>• Types of data</li> <li>• Types of findings</li> <li>• Features of typical entities, phenomena and contexts – left to emerge from the source texts <ul style="list-style-type: none"> <li>- Of disease entities (in this case cancer) – cancer type, stage, time since diagnosis, etc.</li> <li>- Of interventions – stage, type, mode of delivery (e.g. of a drug), elements of services, lifestyle interventions, etc.</li> <li>- Of patient groups – e.g. diagnosis, comorbidities, age, sex, ethnicity, social class, educational status, clinical variables, behavioural, phenomenological, lifestyle variables, etc.</li> </ul> </li> </ul>



**3. Parameters that prompt to identify, extract information about, describe and analyse the matrices of the source studies**

**3.1. Matrices Type 1: coming from the means of obtaining scientific knowledge and from the visions of what knowledge is to be obtained**

**3.1.1. Research aims and outputs (hypotheses, concerns, questions, etc.)**

**3.1.2. Methodologico-theoretico-conceptual-classificatory matrices**

**3.1.3. Ontological, epistemological and axiological foundations – matrices incorporating fundamental beliefs about the world, knowledge and humans**

*Examples of parameters concerning ontological, epistemological and axiological foundations:*

- Most basic assumptions about the world, knowledge, humans, the aims of science and research, a researcher's role, etc. – direct claims as well as inferring from 'traces', such as interpretation of supposedly conventional signs (e.g. a concern with patient voices as associated with an activist agenda)
- Overall orientation and concerns – from direct claims and inferred from 'traces' (e.g. theoretical references in text, journal title, contextualisation within other research, etc.)
- Broad, far-reaching societal goal to which the study is trying to contribute
- Prompts and spontaneous comments concerning alternatives to what posited as fundamental/self-evident, the reliability of deducing positions from the traces, etc.

**3.1.4. Narrower grounding – clearly stated background beliefs and auxiliary assumptions**

*Examples of parameters:*

- Elements from other theories (e.g. particular concepts, laws/principles, mechanisms, metaphors, etc. from these)
- Elements from previous research (e.g. particular findings, claims about important gaps)
- Practice-based knowledge
- 'Common sense' beliefs
- Local and practical needs, affordances and constraints
- Current norms and legislation
- Ethical concerns

**3. Parameters that prompt to identify, extract information about, describe and analyse the matrices of the source studies (cont.)**

**3.1.5. 'Quieter' supporting premises incorporated in the study fabric – vaguer than the above, implicit, potentially hidden**

*Examples of parameters:*

- Any concerns about the relationship between explicit study foundations and grounding on the one hand and its aim and methods on the other? Between aims and methods?
- Any concerns about the move between findings and interpretations, interpretations and conclusions?
- How are contradictory/conflicting and converging findings addressed? On what grounds are the findings of 'own' study preferred or interpreted with caution?
- Prompts and spontaneous comments concerning, for instance, possible alternatives to the supporting premises thus identified, the effectiveness of deducing missing premises, the possibility for 'saving' arguments of imperfect logic through minor adjustments, etc.

**3.2. Matrices Type 2: coming from the knower – the person-people-community who obtain scientific knowledge**

*Examples of parameters:*

- Statements/extended descriptions of a researcher's history and relationship to the question researched
- Sideline information – e.g. on background and academic credentials
- Statements/extended descriptions of a researcher's role and position in the study – e.g. interaction with and response to study participants and material collected
- Statements/extended descriptions of practices of guarding against bias – e.g. double coding, participant validation, discussions with colleagues, etc.
- Prompts and spontaneous comments – e.g. concerning generic biases of human cognition that seem to have escaped the control of a method or (self-)reflexivity of a researcher; the assumptions taken for granted in a research community; the difficulty/impossibility of separating conventions of representation and actual behaviours, etc.

<p><b>3. Parameters that prompt to identify, extract information about, describe and analyse the matrices of the source studies (cont.)</b></p>	<p><b>3.3. Matrices Type 3: coming from epistemically relevant locations and orientations</b></p> <p><b>3.3.1. Study setting</b> (the context of the researched objects/ participants, with or without the researchers)</p> <p><b>3.3.2. Researchers' context</b></p> <p><i>Examples of parameters:</i></p> <ul style="list-style-type: none"> <li>• Context of the research programme/ project – in terms of stakeholders (funders, users, advisors) and ethos of a place – most likely from sideline information (e.g. affiliations, acknowledgements, grant-giving body)</li> <li>• Perceived attitude towards this type of research in the broader community – amongst health professionals, basic researchers, patients and the wider public, policy makers</li> <li>• Spontaneous comments about current influences on research, if UK study</li> </ul> <p><b>3.3.3. Uses, users and contexts of use</b></p> <p><b>3.3.4. Broader social context?</b></p> <p><b>3.4. Matrices Type 4 – as coming from the nature of (scientific) representation</b></p> <p><b>3.4.1. Parameters helping to identify the 'work' needed to generate a 'most complete and accurate' representation of information from the source studies</b></p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> <li>• Compare multiple representations of the same objects within the same study (e.g. aims, concepts, findings of the study)</li> <li>• Compare a variety of representations of standard processes and practices from across studies – e.g. various reporting of sampling in trials or of analysis of interviews; compare against what you know is going on</li> </ul> <p><b>3.4.2. Representations of the practical and mental processing of research material where full transparency/ procedural objectivity is problematic</b></p> <p><i>Examples:</i></p> <ul style="list-style-type: none"> <li>• Claims about the mental processes involved, their complexity and strongest determinants – e.g. how interpretative/ transparent, how empathic/objective, key considerations taken into account</li> </ul>
---	---

**3. Parameters that prompt to identify, extract information about, describe and analyse the matrices of the source studies (cont.)**

- Claims about the following of good practices and fully rational process – what happens at the intersection between the prescriptions/ambiguities of a method and their practical realisation/resolution
- Consider difficulties of separating stages, factors, etc. and presenting a linear process – e.g. use overlap of coding between phases of a research study as an indicator
- Parameters taking a note of rhetorical devices (generally under parameters concerning core contents, e.g. under Method – author’s comments on strengths and limitations)
- Parameters that explore the extent to which standard descriptors of entities, phenomena and contexts (e.g. socio-demographic information) are helpful in making decisions about transformations.

### 3.3. Software used in the extraction-coding

As mentioned above, the extraction-coding was performed first in QSR NVivo (Version 8), one of the premier software packages for analysing qualitative and mixed methods research, and in Microsoft Excel for more quantitative and categorical data. It was then transferred for all data types to EPPI-Reviewer 4, the literature review software of the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) at the Institute of Education, University of London – one of the key UK institutions engaged in the production of research synthesis studies, with a particular focus on syntheses bringing together a large variety of types of studies. As described on the software webpages, EPPI-Reviewer is “for all types of literature review, including systematic reviews, meta-analyses, 'narrative' reviews and meta-ethnographies” (EPPI-Centre, 2013).

Initially, I intended to use an entirely tabular format to extract data for the purposes of the test research synthesis and collect data for the broader thesis work. The Analysis Framework was thus first generated as an Analysis Table in Excel. Traditionally, synthesis studies have been extracting data in tables in Word or tables and boxes in software packages. Extracting data into a table was seen as the most appropriate format to explore transformations.

It quickly became clear, however, that this format is too constraining for the amount of data needed in view of the questions asked. I switched to NVivo as the primary technological environment. It was the tool of choice when detail and context were needed. It also enabled effective hierarchical representation of information. Excel was used as supporting software for numerical or categorical data. Such data can be coded but cannot be represented effectively in NVivo.

At the point of starting my work, there was no obvious choice of specialised software for research syntheses using highly heterogeneous material. Such work seemed to be done primarily in word processing documents or split between different packages depending on the particular type of material analysed. The most popular research synthesis software, Cochrane’s RevMan (The Cochrane Information Management System, 2013), met the needs of mainstream quantitative systematic reviews but was not adapted to more conceptual and theoretical work. Its use was going to constrain the process

similarly to the tabular format considered initially. I had experience of using EPPI-Reviewer but considered the combined functionalities of NVivo and Excel a better choice (e.g. better memoing and linking functions, easier entry of numerical and categorical information). Other options considered were the SUMARI tool (System for the Unified Management, Assessment and Review of Information), the Joanna Briggs Institute's "premier software for the systematic review of literature" (The Joanna Briggs Institute, 2013) and MAXQDA (2013). All specialised research synthesis software however, apart from RevMan, is paid for and not available through University of Exeter licenses for student use. As, at least relative to the synthesis software I was familiar with, the combination of NVivo and Excel served me better, I decided to stay with it.

Later in the study I moved from NVivo plus Excel to EPPI-Reviewer 4. The main reason was that after 12 heavily coded studies the use of NVivo was becoming unwieldy. A 'lighter' software was needed. As EPPI-Reviewer uses cloud technology, meaning that the processing is done on servers external to one's own computer, the size of files was not a barrier to the efficient processing. Transferring into EPPI-Reviewer reduced the effectiveness of some data collection processes. It also created difficulties of combining data and limited opportunities for presenting same-format data in the analysis. But it had substantial advantages in terms of ease of coding when large amounts of data had accumulated.

The process of performing the extraction-coding in these three environments was also a process of identifying software features that constrain good research synthesis and identifying features of 'the perfect research synthesis software'. Some of these are considered in the Discussion Chapter.

The choice of software has affected the scope of findings about transformations in this study. Had I used a fully tabular format for data extraction, which allows for no direct access to the original data and reflects standard practice until very recently, I will have claimed that greater than the here observed degree of transformations was under way and that their reliability was lower. In software like NVivo or EPPI-Reviewer, source studies can be accessed within the software. This direct access to the original data reduces the degree of transformation needed at this early stage of a synthesis and is likely to improve

the rigour of all transformations. Arguably, when a researcher wishes to transform the already transformed, it will be much more likely that they make a few extra clicks on the same screen rather than go back to the original paper in another programme or as a physical object. Changes have been rapid. A technological challenge to the rigour of transformations has been overcome. Claims about transformations made here are thus in line with what seems to be current practice. They are likely, however, to underestimate the extent and overestimate the rigour of transformations in many (the majority of?) synthesis studies that are continuing to exert effects on services and research.

#### ***4. Further approaches and tools used***

Some of the data produced by the extraction-coding could be used, with minimal further analysis, to answer questions posed by the thesis. Other datasets, however, required extensive further processing through supplementary tools and approaches before they could be related clearly to the thesis questions. This section outlines the main supplementary tools and approaches.

##### **4.1. Approaches to exploring types of, similarities and differences between, and grounds for (non-)combinability of findings**

The largest set of tools for further analysis of data obtained through the extraction-coding addressed similarities and differences between findings. These were investigated along the continua of 'formal – thematic', 'standardised – non-standardised' and 'paradigmatic – non-paradigmatic'. That is, similarities and differences were sought between findings in terms of formal characteristics, such as composition and structure, and in terms of thematic coverage. Next, levels of similarity and difference were explored between supposedly highly standardised parameters (e.g. socio-demographic information) from different studies, and between 'normal', non-standardised findings from the same and different studies. Finally, similarities and differences between findings were related to the quantitative or qualitative 'paradigm' in which they were seen as embedded or were investigated independently of such framing. Of the tools

used, the Vocabulary of Elements of Findings was the most innovative and extensively articulated.

#### 4.1.1. *The Vocabulary of Elements of Findings*

I developed the Vocabulary of Elements of Findings as a tool for representing and exploring the composition and structure of findings in health research reports. Its intended uses were two. First, it was to serve as a tool for exploring compositional and structural similarities and differences between findings – put more precisely, similarities and differences in the composition and structure of propositions represented through standard sentences or as encoded in tables, graphs and images but convertible into standard sentences. The ‘composition’ of a finding was seen as the set of elements within a finding, irrespective of their relationship to one another. ‘Structure’ was used to add this further concern with relationships and a potential hierarchy.

Second, I intended to use the Vocabulary to explore opportunities for and limitations to increasing the transparency of transformations in synthesis studies. I had operationalised the exploration of the boundaries of transparency in transformations (see 3.1) as concerned with the following questions: *Which* discrete and (relatively) simple elements within a finding gave grounds for a particular transformation of this finding? *How* did those elements become part of a more or less complex derivation of a transformed finding which preserved the truth value of the original finding either categorically or probabilistically? *What* are the known and unknown steps in this derivation? The Vocabulary offered a possibility to split findings reliably into discrete and relatively simple elements. It was thus a step towards exploring the boundaries of transparency in transformations. The better the Vocabulary performed across different and difficult cases, the higher the likelihood that an increased transparency of transformations was possible.<sup>91</sup> This duplicity of aims contributed to an occasional redundancy in the coding, arising in the first place by the interdependence of several of the key elements of the Vocabulary.

---

<sup>91</sup> The opposite assertion was much weaker. If the Vocabulary did not perform well across different and difficult findings, this was first going to be seen as a negative reflection on the qualities of the Vocabulary rather than as evidence that increased transparency of transformations was not possible. Transparency was thus part of a largely protected core of assumptions.



The Vocabulary was developed from an initial set of 20 findings – single sentences or clauses representing *prima facie* different finding types. These were sampled from the pool of findings generated through the extraction-coding of the first 10 reports in the case study. The approach involved giving abstract labels to semantic units – individual words, or phrases comprising a main word plus qualifiers. The process was inductive and iterative and was guided by the questions *What is there in research? What does research capture from the world? What is there in the world that is represented in research?* I expected to identify and label elements like theory, method, fundamental worldview assumptions, theoretical entity, agent, object from the world, capacities, etc. As will become evident, those expectations were quickly overturned. Initially, I labelled ‘surface elements’ (elements couched in words or numbers) but also attempted to elicit ‘silent elements’ (such as missing indicators of an agent or knowledge process – ‘the researcher’, ‘analysed’, ‘reported’). Gradually, I decided to focus on surface elements and only formalise silent elements opportunistically. This was primarily because the omission of some types of elements (e.g. indicators of the researcher as an active agent) is a widely shared feature of findings and would not be discriminating; because of substantial limitations to the reliability of recovering silent elements; and because of the fact that, if silent elements are elicited ‘properly’, one may well begin to see “the world in a grain of sand” – the whole study in a particular finding.

To get a more specific idea of how the labelling proceeded, consider the following simple finding from the study of Whitaker, Brewin and Watson, 2008, Study 2 in the analysis (see Table T 5.1 in Chapter 5 for a schematic representation of the studies analysed):

***Thirty patients (23%) reported an intrusive cognition.***

(“Intrusive cognitions”, or “intrusions”, are largely negative memories, images and thoughts that tend to come abruptly, to be difficult to dispel and to recur repeatedly.)

The units to be formally described were:

- *thirty*

- *patients*
- *23%*
- *reported*
- *an intrusive cognition*

*Thirty* and *23%* were labelled ‘quantifiers’. *Patients* was labelled a ‘carrier’. *Reported* was labelled a ‘knowledge something’ or, alternatively, an ‘expression’. *Intrusive cognition* was labelled an ‘obsessive something’. Nomenclature is clarified further below.

The process went through a number of iterations. Formal descriptors (labels, types of elements) were being redefined, added or discarded until the rules of token-type assignment had become unambiguous and easy to follow, or it became clear that the tokens of certain types could be included, without exception, under certain other types. 9 key and 16 sub-elements were formulated in the Vocabulary development phase. The key elements were named ‘obsessive something’, ‘carrier’, ‘expression’, ‘relationship’, ‘knowledge something’, ‘privileged something’, ‘subsidiary something’, ‘quantifier’ and ‘qualifier’. The application of the Vocabulary to over 300 further findings suggested one further key parameter – ‘indexical’. The number of sub-elements, however, was extended considerably and can continue to be extended almost indefinitely. 4.1.1.1 presents briefly the ten key elements of the Vocabulary of Elements of Findings. In 4.1.1.2, I outline the rules for applying the Vocabulary – performing the coding. 4.1.1.3 offers a conceptual discussion of several high level features of the Vocabulary.

#### *4.1.1.1. Key elements of findings as per the Vocabulary of Elements of Findings*

The **obsessive something** was taken to be the key ‘thing from the world’ or the key research variable the study reports about and relative to which all other things and/or variables are positioned (and with which researchers involved in the study do appear obsessed – they think and talk too much about it, relate everything to it, see it as explaining three-quarters of the world, as touching on most of the age-old questions of humanity and opening a vast array of

extremely interesting new ones). Some examples of obsessive somethings from the case study papers are intrusive cognitions (Study 2, Whitaker, Brewin and Watson, 2008), cancer pain (Study 15, Al-Rowaili et al, 2009), oxygen therapy (Study 14, Jaturapatporn et al, 2010), post-traumatic growth (Study 7, Thombre, Sherman and Simonton, 2010), work changes (Study 5, Mols et al., 2009), etc. Overall, there is a clear single obsessive something in most studies, but there are exceptions. For instance, in Study 10 (Rapp et al., 2008) the relationship between cancer risk and weight change was explored. The variables appeared equally dominant. In such cases I generally coded for obsessive something 1 and 2 rather than posit one of the parameters as more central.

The **carrier** was taken to be (a representation of) the entity, substance or collection of these which somehow 'has' or 'gives' the obsessive something. It can, for instance, contain or capture it – such as the receptors to which insulin-like growth factor 1 (IGF-1) bind and which can be found on normal colonic mucosal cells and colon cancer cells (Study 3, Thomas and Davis, 2007). It can emit it – such as patients fitted with a breast wire who give off radiation (Study 8, Meades et al., 2010). It can experience it – such as individuals or groups who experience intrusive cognitions (Study 2, Whitaker, Brewin and Watson, 2008). Learning about the obsessive something is also a way of learning about the carrier but the carrier is explored in other ways, too, so that this knowledge can be related to knowledge about the obsessive something. Nevertheless, the carrier is, for the purposes of a particular study, more stable, fixed, known than the obsessive something.

**Expression** was used to designate (a representation of) a single property, behaviour, state, pattern of occurrence, etc. of an object of study which it demonstrates in the process of being studied through observation, query, experimentation or interpretation. Syntactically, expressions are typically of the obsessive something or carrier, but they can also be of any of the other 'somethings' specified below ('privileged', 'subsidiary' or 'knowledge something'). 'Expression' is the concept corresponding to the most frequent element of findings. An expression is the proper finding within the finding, the evidence within the evidence – (representations of) the very properties, behaviours, states, etc. which were sought in the study or which the study

delivered unexpectedly rather than the objects they are properties, behaviours, etc. *of* or the processes and tools through which they were revealed.

**Relationship** denoted, for instance, an outcome of a comparison (smaller, higher, also, was no longer, different), (a representation of) a causal or explanatory connection (because, based on, has affected, threatened, led), statistical covariation (significantly associated with), etc. Relationships can be thought of as a particular form of expression where properties, behaviours, patterns of occurrence, etc. of the obsessive something are explored in relation to variables, properties, behaviours, etc. external to it.

**Knowledge something** was used to label (a representation of) a knowledge process (e.g. explore, analyse, report), research tool (scale, QIAGEN DNA Isolation kit, narrative analysis), fundamental research entities (e.g. data, results), background literature (e.g. references), knowers (the researcher, we, our) and similar drivers, preconditions, deployments or outputs of knowledge generation processes.

**Privileged something** was used for ‘things from the world’/ variables whose associative and causal relationship with the obsessive something were given prominence in a study – such as standardised socio-demographic and disease-related characteristics (age, gender, education, ethnicity, stage of disease, etc.) and non-standardised variables incorporated in a study-specific model.

**Subsidiary something** labelled ‘less important’ things/ variables whose associative and/or causal relationships with the obsessive something were explored or speculated about. These were given limited attention in a study report and often left quite vague.

‘Qualifier’, ‘quantifier’ and ‘indexical’ are probably self-explanatory. **Qualifier** was used primarily to denote evaluations, attitudes and, linguistically, adverbial phrases of manner (e.g. significantly, moderately, relatively high, as expected, interestingly, only).

**Quantifier** stood for sizes, quantities, frequencies, etc. and, symbolically, for precise numbers and loose expressions such as ‘some’, ‘many’, etc. Quantifiers tended to either accompany carriers (as in *Six out of eight patients used oxygen all the time* – Study 14, Jaturapatporn et al, 2010), or further specify an

expression (as in *Examples of such symptoms included ... feeling irritable (0.35), diarrhea (0.39), and having sweats (0.27) at T2* – Study 12, Molassiotis, Wengström and Kearney, 2010).

**Indexical** stood for (the reference behind) pointing words and phrases such as ‘this’, ‘that’, ‘the latter’.

The Vocabulary did not contain designations for negations. It did not have a grammar connecting its elements either.

#### 4.1.1.2. *Rules of coding*

The main rule for applying the Vocabulary was that no word should remain unacknowledged. Everything in a finding, whether a natural language sentence or proposition represented elliptically in a table, was to be coded. Overall, I coded elements following the order in the natural language sentence or tabular representation. As explained in the beginning, I only coded silent elements opportunistically, after I assessed initial attempts to elicit them more consistently to be of limited value.

In many cases alternative formalisations were possible. For instance, a knowledge something like ‘reported’, let’s say as in *reported an intrusive memory of girlfriend leaving him over 50 years ago* (Study 2, Whitaker, Brewin and Watson, 2008), can also be seen as an expression of a carrier. The flexibility of the Vocabulary arose from the interdependence and orientation-towards-the-other of several of the key elements of the Vocabulary (discussed in 4.1.1.3). Coding alternatives were recorded opportunistically, with the exception of cases where a knowledge something was present in one of the options (knowledge somethings were involved in a hypothesis about differences between qualitative and quantitative findings). This openness to alternatives does not mean arbitrariness. Elements were well specified. The surface structure of a finding – e.g. the syntax of a sentence or layout of a table – privileges one coding over another in the majority of cases. Much more rarely are surface structures ambiguous and prompt multiple formalisations. Also,

reformulations cannot be too numerous, as only some functional equivalences are there or changes of perspective possible (e.g. a quantifier cannot become a carrier).

A simple example of coding is given below. More complex ones are given in the conceptual discussion in 4.1.1.3. In each of the coding tables presenting the examples, the left-hand column contains the units into which the finding was broken up. The right-hand column contains the element of the Vocabulary ascribed to a particular unit. In some cases, silent units derived from other parts of the study report were added to the left-hand column to clarify a reference.

This first example of a coded finding is, apart from simple, also typical in a number of ways. It has a fairly typical knowledge something and expressions although a somewhat unusual obsessive something. There seems to be no ‘classic’ finding to give as a first example. The coding of each finding had something which distinguished it from the great majority of findings.

*Six symptom clusters were identified at baseline* (Study 12, Molassiotis, Wengström and Kearney, 2010: 850).

Symptom clusters	Obsessive something, composite
Were identified at baseline	Knowledge something
Were identified	Expression, covering [presence, existence, occurrence]
Six [were identified]	Expression 1
[Were Identified] at baseline	Expression 2

“Symptom clusters” was the obsessive something in the study of Molassiotis, Wengström and Kearney, 2010 addressing symptom cluster patterns in the first year after cancer diagnosis. This is a case of a ‘composite’ obsessive something – the set of symptom clusters included several somewhat dynamic individual clusters, such as gastrointestinal, respiratory and body image

clusters. 'Obsessive something, composite' is one of the subtypes of the obsessive something.

"Were identified at baseline" is a knowledge something – a reference to some 'work', tool, artefact, agent, etc. that had a causal role in bringing about the finding reported. An alternative formulation of the same finding which ignores the knowledge generation process and drivers may be, for instance, '*Six symptom clusters were present at baseline*'.

Three expressions follow in the coding table – "identified", "six [identified]" and "[identified at] baseline".<sup>92</sup> The contents of these three expressions partly overlaps with the contents of the knowledge something. As indicated above, this was due to the interdependence and orientation-towards-the-other of several of the key elements of the Vocabulary, allowing for the same claim to be formulated from the perspective of at least two concepts, and due to the duplicity of aims for the Vocabulary, preventing the prioritisation of a single coding pattern in ambiguous cases.

'Were identified' was coded as 'a covering expression' concerning 'presence, existence, occurrence'. 'Covering expression' was a frequently occurring subtype of expression, and 'presence, existence, occurrence' was a label for a frequently occurring subtype within this subtype. As much as the notorious predicate of existence immediately recedes into the background of more specific findings, it is a finding in itself that there are symptoms which cluster in the first year after cancer diagnosis as opposed to having a diversity of individual unrelated symptoms. The ways in which this presence of clusters is specified here is through its temporal fixation ('at baseline', namely towards the end of the first month after cancer diagnosis – Expression 2) and through the number of forms it takes ('six' symptom clusters – Expression 1).

---

<sup>92</sup> The parts of the expressions in the square brackets are in this case 'old information'. I used square brackets containing various types of additional information often in the coding, as the examples in this section will illustrate. For the moment, I have standardised only partially the types of information added and the particular labels.

#### *4.1.1.3. Some conceptual issues associated with the key elements*

The Vocabulary of Elements of Findings emerged as much more minimal at its highest level and as much more uniform across finding types (in particular quantitative and qualitative findings) than I had expected. One of the reasons was that I did not want to make too early judgements, choices and distinctions associated with certain big philosophical debates and challenges. Thus, the naming and conceptualisation of the key elements are characterised by extreme openness to alternative views. Commitments to more particular philosophical or methodological claims are captured at the level of sub-elements. This section outlines how the abstaining from assumptions worked at the level of key elements and what directions the formulation of sub-elements took.

*At the level of key elements, the Vocabulary remains open to various positions on the relationships between things in the world – things in research – phenomenal experiences – representations through language and other symbolic systems, which can give quite varied answers to the question of ‘what is there in a finding?’.*

I aimed to name and conceptualise key elements of the Vocabulary in such a way so that they can accommodate both a world-focused perspective concerned with ‘things in the world’ and a research-focused perspective concerned with variables and symbolic representation. At the very start of generating labels for units of findings, it became clear that taking only one of these perspectives will misrepresent ‘what there is in a finding’. This duplicity of perspectives is particularly visible in the choice of the awkward word ‘something’ (rather than ‘entity’ or ‘research variable’, for instance). A ‘something’ made room for the divide in philosophical thinking about the reality of theoretical entities. Its inclusiveness allowed not to commit to a claim whether a thing spoken about in a research report really existed in the world or was a theoretical construct. In addition, it could refer quite naturally both to the thing or research variable studied and its representation in a research paper, whether through words, formulas, graphs, images, etc.



In some cases the vagueness of ‘something’ was just what was needed: no easy claim could be made as to *what* was represented in a finding (e.g. a thing from the world or a theoretical entity). In other cases, there was a much greater certainty about the *what*. A sub-element was then created. Consider the phrase “adjustment to cancer diagnosis”. When a research paper claims that health professionals should monitor patients’ adjustment to cancer diagnosis so that counselling referrals can be made when needed, this is a statement about a thing from the world (be it from the world of lived experience). When the abstract of a research paper highlights as a key finding that *No difference was found in adjustment to cancer diagnosis in patients who are practising Christians as opposed to practising Buddhists*, the thing from the world and the research construct cannot be neatly separated. It is a matter of perspective and/or theoretical commitment to express a preference for one or the other description. When a finding is that *The average Mini-MAC adjustment to cancer score of the intervention group was lower than that of the control group*, where Mini-MAC is a scale for assessing patient adjustment to cancer diagnosis,<sup>93</sup> then the obsessive something is represented from a strongly research perspective. In the first case, ‘obsessive something, world perspective’ was used. In the third case, ‘obsessive something, knowledge perspective’ was used. The majority of cases were of the second type. These were labelled with the generic ‘obsessive something’.

A further set of philosophical debates whose complexity was both acknowledged and avoided through the conceptualisation of key elements addressed questions of whether, when we are talking about evidence, we are talking about propositions, phenomenal experiences, or objects and events. In the brief presentation of the key elements in 4.1.1.1, (*representation*) – a concept to include or omit depending on context and stance – was used frequently to reflect some of the alternatives in the above set of debates. Ideally, further distinctions would have been drawn, primarily between representations of the world by means of research entities, such as variables, data, models,

---

<sup>93</sup> Mini-Mental Adjustment to Cancer Scale – Watson, M., Law, M., Dos Santos, M., Greer, S., Baruch, J. and Bliss, J., 1994. The Mini-MAC: further development of the mental adjustment to cancer scale. *Journal of Psychosocial Oncology*, 12, pp. 33–46.

etc., and representations (of the world and of research entities) by means of a natural language or another symbolic system. However, as in this work findings were analysed only in terms of their composition and structure as represented symbolically, mostly as standard sentences and tabulated information in study reports, these otherwise important distinctions were not brought into a foreground.

*The Elements in the Vocabulary are closely interdependent and 'oriented-towards-the-other', i.e. definable through other elements in the Vocabulary. Coding alternatives arise, as the same proposition can be represented from the perspective of different key elements.*

The above feature of the Vocabulary is most clearly illustrated by attending to the 'owner', the 'locus', the 'originator' of an expression – whose expression is an expression? The brief answer is – of many. I aimed to formulate a concept which embodies, rather than glosses over, a central difficulty of adjudication in science and research. I will call it *the 'whose effect' difficulty*. Using concepts from the Vocabulary, it may be described as follows. The aim of a research study is to identify some stable or traceably dynamic property of an obsessive something (to use one of the examples above, the composition of symptom clusters in the first year following a cancer diagnosis). But it is practically never definitive enough if a finding tells us more about 1) the obsessive something; 2) the particular milieu from which the obsessive something was taken or in which it was explored, i.e. the carrier (in this case a particular sample of patients); 3) the approach used to study the obsessive something (the set of all relevant knowledge somethings); or 4) about factors that could not be articulated and/or controlled in the study. An expression is thus an attention-grabbing intersection typically between an obsessive something (or more rarely, one of the other somethings), a carrier, a set of knowledge somethings and study-relevant factors that could not be articulated and/or controlled. An expression is a finding about all these elements and a thorough data extraction requires it to be attributed to all (or at least the first three) of them.<sup>94</sup>

---

<sup>94</sup> In discussing expressions and when I am not focusing first and foremost on their being an intersection of effects, I attribute an expression to the obsessive something.

For instance, in the example

*Some people saw themselves as different to who they once were, suggesting that a disruption to their sense of self had occurred ...*

(Hubbard, Kidd and Kearney, 2010: 137).

'saw themselves as different' is syntactically an expression (in this case perception) of the carrier (people with colorectal cancer in the first year following diagnosis). It is also an expression of the obsessive something (biographical disruption, not mentioned in this sentence): namely, biographical disruption takes the form of a discontinuity of identity. It is also an expression of the privileged something (sense of self): the sense of self undergoes a disruption in the context of a recent colorectal cancer diagnosis. Finally, it expresses those knowledge somethings which ground the study in biographically informed approaches to understanding chronic illness.

The concept of 'expression' thus bundles together the study object and context and avoids committing to a principled position on the 'whose effect' difficulty. It treats the resolution of that difficulty as pertaining to a particular case rather than as a matter of principle.

*The key elements of the Vocabulary do not discriminate between elements of lower or higher level of interpretiveness, lower or higher level of theory-ladenness or greater or lesser proximity to raw data.*

Again, the concept of expression is the clearest example of the above feature of the Vocabulary. A representation of the raw data concerning an obsessive something, carrier, etc. is treated as an expression of that obsessive something, carrier, etc. just as much as an interpretation that is several levels removed from the raw data. I wanted to avoid distinctions in this direction primarily to avoid certain claims of reliability associated with them – 'if it is conceptually and formally closer to the raw data, it is more reliable'.<sup>95</sup> Perhaps yes, perhaps no.

---

<sup>95</sup> Even though at the start of the thesis I endorsed them.

The coding for the next example (the full version of the finding above) shows four layers of interdependent expressions, at a different level of interpretiveness/ theory-ladenness/ distance from the raw data.

*Some people saw themselves as different to who they once were, suggesting that a disruption to their sense of self had occurred:*

*One of the questions is that the thing is, do you look in the mirror and think that you look different or something? Well I do feel some days I'm the same person as I was. In some respects I'm stronger; in other respects I'm weaker. I think I'm stronger mentally and physically weaker but I mean you look in the mirror and sometimes and you think it isn't me sort of thing but it's, can be strange. But the other day I was getting a contact lens thing and the woman says, 'Go and put your contact lenses for me son'. I was at this big mirror right in front of me and it was like looking at a stranger, it was incredible because you're right up the mirror like that. (P28, I3, employed, male)*  
(Hubbard, Kidd and Kearney, 2010: 137).

Some	quantifier
People	Carrier-cum-knower, composite
Saw themselves as	Knowledge something [weak]
[were] different	Expression, covering [comparison]
Saw themselves as different to who they once were	Expression 1, layer 1
Suggesting that	Knowledge something
Sense of self	Privileged something
A disruption had occurred	Expression 1, layer 2
One of the questions is that the thing is, do you look in the mirror and ...	Expression 1, grounding (raw data)
(P28, I3, employed, male)	Carrier-cum-knower

On the basis of the ‘grounding expression’ – the raw data representation, in this case an interview quote – the authors formulate two more general interdependent expressions. Expression 1, layer 1 is an expression of a subsample of the carrier (“some people”) – they perceive themselves as being different to who they once were. Expression 1, layer 2 is an interpretation of Expression 1, layer 1 – the sense of self (a privileged something) has become disrupted. The covering expression, which derives directly from the raw data, is of comparison ([were] different).

Expressions can thus be at a different proximity to the raw data: be (representation of) the raw data itself; a more general claim based on the raw data that remains within the framework of concepts appearing in the raw data as in Expression 1, layer 1; a more general claim based on the raw data which, however, shifts the reference framework within which the raw data were stated as in Expression 1, layer 2. Similarly to the case of expressions, relationships were coded at different levels, with very general, somewhat ‘metaphysical’ concepts (e.g. association, causality) at the one end and raw data at the other:

*Results indicated that caregiver PTGI [Post-traumatic Growth Index] scores were significantly associated with the patient’s recurrence status ( $p < .05$ ) and marginally associated with time since initial diagnosis ( $p = .08$ ) (Study 7, Thombre, Sherman and Simonton, 2010: 180).*

Results indicated	Knowledge something
Caregiver	Carrier
PTGI scores	Expression 1
Patient’s recurrence status	Privileged something 1
Time since initial diagnosis	Privileged something 2
Associated	Relationship, covering [association]
Associated [PTGI scores; patient’s recurrence status]	Relationship 1, knowledge perspective [expression 1; privileged something 1]

Significantly associated [PTGI scores; patient's recurrence status]	Relationship 1, knowledge perspective, layer 1 – with qualifier
Significantly associated ( $p < .05$ ) [PTGI scores; patient's recurrence status]	Relationship 1, knowledge perspective, layer 2 – with quantifier
Associated [PTGI scores; time since initial diagnosis]	Relationship 2, knowledge perspective [expression 1; privileged something 2]
Marginally associated [PTGI scores; time since initial diagnosis]	Relationship 2, knowledge perspective, layer 1 – with qualifier
Marginally associated ( $p = .08$ ) [PTGI scores; time since initial diagnosis]	Relationship 2, knowledge perspective, layer 2 – with quantifier

Going back to the last but one example, it also demonstrates that interdependent expressions arising from the same raw data need not come into a clear hierarchy where a more abstract expression is more closely related to another more abstract expression than to the raw data. For instance, the covering expression ([were] different) is directly connected, through an identity of concept, with the level of raw data (*One of the questions is that the thing is, do you look in the mirror and think that you look different or something? ...*) and Expression 1, layer 1 (*saw themselves as different to who they once were*) but not with Expression 1, layer 2 (*disruption to their sense of self had occurred*), regardless of the fact that the latter is based on Expression 1, layer 1.

The relationships between levels and layers<sup>96</sup> of expressions can be quite varied: deductive, inductive, abductive, analogical, metaphorical, tautological, etc. This means that an expression building on another expression can be a valid representation of the state of affairs even if the expression underneath it is not. Luckily or worryingly, if our reference expression was derived from an

---

<sup>96</sup> I use 'layers' for levels within levels.

invalid expression through an inference type that does not preserve truth, it can end up being a valid representation of the state of affairs. Much further work is needed to specify the ways in which expressions at different layers and levels relate to one another as well as how this variety of levels of interpretiveness, theory-ladenness and distance from raw data is realised in the context of the other key elements of the Vocabulary.

*The key elements of the Vocabulary aimed to circumvent metaphysical commitments associated with stability and dynamics.*

The key elements of the Vocabulary also aimed to avoid alignments with an ontology of stability or dynamics. For instance, in philosophical discussions of empiricism the term “occurrent properties” seems to be used similarly to the term ‘expressions’ as defined here. Another word was preferred so as to open more space for what things ‘do’ as opposed to what features they demonstrate. ‘Expression’ was seen as avoiding the narrowness of connotations of ‘property’, ‘feature’ and ‘characteristic’ (connotations of essential, static) and of ‘behaviour’ (dynamic, intentional), arising even when the definitions of those terms explicitly disavow such boundaries of meaning.

In the case of the ‘something’ concepts, I constrained the concept nominally more than I wished. Some of the somethings are processes rather than things – for instance, post-traumatic growth and weight change in the case of the obsessive something or research processes such as ‘analyse’ and ‘report’ in the case of the knowledge something. If I was not running the risk of creating an overly bizarre vocabulary, I would have used ‘something-someflowing’ to emphasise this fact. I preferred not to write like Heidegger on this occasion.

*Finally, the Vocabulary encompassed a broad range of sub-elements and is open to the formulation of many more sub-elements.*

While I kept the highest level of concepts in the Vocabulary minimal, I was generating a broad range of concepts at lower levels and attempting to specify their relationship to philosophical and methodological discussions about how

science and research work. In some cases, a degree of standardisation of subtypes was achieved.

For instance, 'knowledge perspective' was used consistently to accompany key elements where these were strongly represented in research terms or as things from the world of research:

- At the beginning of 4.1.1.3, I gave the example of a strongly knowledge perspective towards the obsessive something – *the average Mini-MAC adjustment to cancer score* as an 'obsessive something, knowledge perspective' in contrast to the ambivalent obsessive something of *adjustment to cancer diagnosis*.
- In the context of relationships, 'relationship, knowledge perspective' was recorded when relationships between scientific artefacts were reported which themselves took the form of scientific artefacts *par excellence* – e.g. comparison of the means of two groups using a t-test, with the relationship reported as a t-test statistic and a p-value for its significance.
- The knowledge element was also added to the 'carrier' label, though in a somewhat different form – as 'carrier-cum-knower'. This was used when the carrier was a person and, apart from 'having' the obsessive something, also reported on it. For instance, a study participant experiences anxiety but is also the one to report the anxiety. In this way the study participant is also a 'tool' for obtaining knowledge, a knowledge something.

Three further types of the obsessive something were frequently applied, in addition to 'obsessive something, knowledge perspective':

- 'Obsessive something, aspect' was used when the obsessive something was discussed through its components or the aspects it revealed in particular situations. For example, in Study 9 of the case study papers (Andrykowski, Donovan and Jacobsen, 2009), the obsessive something was "shift in fatigue ratings", with two main aspects being studied – "shift



in most fatigue” and “shift in average fatigue”. When each of these was mentioned separately in a finding, ‘obsessive something, aspect’ was coded for.

- Another stable type of an obsessive something was ‘obsessive something, variant name’. It was ascribed when a synonymous word or phrase was used in place of the study’s own preferred nomenclature (e.g. “intrusions” rather than “intrusive cognitions”).
- ‘Obsessive something, composite’ was the third of these types (see the example in 4.1.1.2 on symptom clusters).

A stable sub-element but difficult to code consistently was ‘relationship, through juxtaposition’. Not infrequently in research papers, relationships are left to the reader to construct from juxtaposed parameters. For example, from *Analgesics were given orally in 70.5% of the patients, 26.5% were prescribed fentanyl patches, and 3% intravenous morphine* (Study 15, Al-Rowaili et al., 2009: 39) conclusions can be drawn that in the study setting oral analgesics were used more often than any other form of pain medication, that fentanyl patches were used more often than intravenous morphine, etc. In such cases ‘relationship, through juxtaposition’ was coded for. As it was a silent element, it was not coded consistently at this stage.

At the other end of the spectrum, subtypes of knowledge somethings are clearly many and still in the process of becoming consolidated. I have been testing out a variety of labels, aimed to foreground a typology of knowledge somethings in health research findings, but do not consider them sufficiently rich and informative.

The two examples below illustrate some such test labels. The examples are unusually rich in knowledge somethings: there are three in each, and an additional ‘knowledge perspective’ in the second example, while most findings have one or no knowledge something. After each of the coding tables, I give a

version of the same finding without or with a toned down knowledge something (complete avoidance is not always possible).

*“However, the interviews do not suggest that this experience led to loss of self as defined by Charmaz (1983)” (Hubbard, Kidd and Kearney, 2010: 140).*

However	Knowledge something [expectations contradicted] [weak]
The interviews do not suggest that	Knowledge something [method as agent]
This experience	Subsidiary something
Led	Expression, covering [being a cause/contributing factor]
Led to loss of self	Expression 1
As defined by Charmaz (1983)	Knowledge something [background theory]

Without a knowledge something this finding could read: *‘This experience [of being diagnosed and living with cancer] did not lead to loss of self’.*

*“Twelve patients showed comparable results between measurements and estimated doses” (Study 8, Meades et al., 2010: 529).*

Twelve	Quantifier
Patients	Carrier
Showed results	Knowledge something [data, findings]
Measurements	Knowledge something [data, findings]

Estimated doses	Knowledge something [theory]
Measurements [of radiation obtained]	Expression 1
Comparable results	Relationship, covering [similarity]
Estimated doses	Privileged something
Comparable results between measurements and estimated doses	Relationship 1, knowledge perspective [expression 1; privileged something]

A reformulation of this finding with a toned down knowledge something could be: *'Twelve patients received doses of radiation close to the ones expected'*.

With the exception of subtypes labelled for a 'knowledge perspective' and 'obsessive something, aspect', sub-elements of the Vocabulary were not used for the analysis presented in Chapter 5. I considered the method insufficiently reliable at the level of most subtypes.

The above discussion has certainly demonstrated the need for much further work on developing the Vocabulary of Elements of Findings. I hope it demonstrated its potential, too.

#### 4.1.2. 'Simple' approaches

Three further main approaches were applied to study types of, similarities and differences between, and grounds for the (non-)combinability of findings: exploration of types and labels of findings for 'lines of tension', exploration of the degree of thematic differences across findings, and exploration of the contents of highly standardised parameters, namely socio-demographic and cancer-related parameters. These approaches were technologically simple (though not quick to execute). They involved basic techniques of classification and, in the latter two cases, calculations of proportions.

The first approach probed into one of the most typical strategies for asserting how challenging it is to combine findings in health research syntheses on broad

topics. This strategy involves the presentation of long lists of varied labels for types of studies, evidence, data, etc. that tie the mind in knots and prime effectively subsequent statements of serious challenges to combinability. Here is a typical example - an excerpt from Pope, Mays and Popay (2007: 3) which continues well beyond the quote here:

*Every year a vast number of health-related research studies are carried out. There are randomised controlled trials of new treatments, surveys of patient experiences, evaluations of interventions designed to improve health and/or reduce health inequalities, analyses of routinely collected hospital episode statistics, economic evaluations of the costs and effectiveness of health care, case studies of new practice, research on the experience of service users, patients and health care professionals and so on. This research is conducted by researchers from a diverse range of disciplinary and professional backgrounds, including statistics, epidemiology, psychology, sociology, anthropology, political and economic science, geography, public health and the medical, nursing and therapy professions, and it draws on a variety of different theoretical and methodological approaches.*

The numerous classes of types of findings, data and studies that were generated in the process of extraction-coding were further grouped together. Attempts were made to identify and create 'groups of tension' – i.e. groups in which the different types of findings fell at very different points on a shared continuum, to the extent to which this would thwart the bringing together of such findings.

The second simple approach was of exploring thematic differences between findings. The application of the injunction to code first as close as possible to the original, with no or minimal transformations, allowed for minor differences between units of information to be preserved. The function of NVivo which registers number of "references" (units of information) under a code permitted to identify the number of codes which had one, two or more units of information within them. The larger the proportion of codes with only a single unit of information, the larger the degree of thematic differences across studies.

Finally, the third simple approach explored similarities and differences across studies in terms of supposedly highly standardised parameters – basic cancer-related information (such as cancer type, stage, treatment type, etc.) and socio-demographic parameters (such as age, gender, socio-economic status, etc.). Similarities and differences in the collection, contents and representation of such information were explored.

#### **4.2. Relating findings about types of findings to findings about the presence and effects of matrices**

All the approaches described above – the extraction-coding through the Analysis Framework, the development and application of the Vocabulary of Elements of Findings, the simple approaches – aimed to identify various types of similarities and differences between findings and various types of matrices (factors that shape findings and thus determine similarities and differences between them). Within the conceptualisation of this study, these are two sides of the same coin. Similarities and differences of findings will be partly traceable to similarities and differences in matrices, and similarities and differences of matrices will be partly traceable to similarities and differences in findings. As the approaches took different routes to the same questions, it was going to be informative to relate their findings.

Types of findings, as identified through applying the Vocabulary of Elements of Findings, were to be mapped onto types of findings as generated 1) by transforming and grouping findings during the extraction-coding for the purposes of the test research synthesis and 2) by applying Parameters 2 in the Analysis Framework, which aimed to enable the development of formal, theoretical descriptions for the core contents of synthesis studies, including findings, data and evidence (see Table T3.2.). In turn, findings about matrices as obtained by applying Parameters 3 in the Analysis Framework were to be compared to findings about matrices as obtained through applying the Vocabulary of Elements of Findings. This intention was superseded after the Vocabulary was fully developed. Initially, expectations were that it would contain a fair number of elements for traces of matrices. As its description in 4.1.1.1 shows, this was not the case.

Finally, as the quantitative-qualitative paradigms debate was the only meta-scientific debate used to illuminate the combinability of highly heterogeneous evidence, findings about the composition and structure of findings, as identified through the Vocabulary of Elements of Findings, were to be related to their quantitative or qualitative status. Findings were to 'inherit' the status of the studies they came from. Unlike the typical practice of ascribing a single label – quantitative, qualitative or mixed methods study, studies were to be given a primarily quantitative/ primarily qualitative label on four features: 1) study topic, 2) method of data collection and type of source material, 3) method of data analysis, and 4) presentation style.

Those initial intentions were modified substantially as findings from the core method (the application of the Analysis Framework) and the contents of the main supporting tool (the Vocabulary of Elements of Findings) strongly defied expectations. The world of research had other things to show while casually disconfirming hypothesis after hypothesis. In fact, only the relating of quantitative and qualitative status of findings to their composition and structure was performed. The particulars of the approach are given along with the presentation of its findings in the Findings Chapter.

#### **4.3. Further generic methods used to complement the Analysis Framework and the Vocabulary of Elements of Findings**

A number of generic methods were used opportunistically to obtain further data concerning questions whose exploration was embedded in the Analysis Framework and the Vocabulary of Elements of Findings. This was either to counteract deficiencies of methods which became visible in their deployment (a highly probable outcome in the case of novel methods) or to use affordances which were not foreseen. For instance, it became clear that the approaches for exploring the multiplicity of findings incorporated in the Analysis Framework require much further operationalisation. A simpler method of capturing this multiplicity was devised. Or the process of formalising findings using the Vocabulary showed a further, and very telling, way of exploring missing

information in studies, in addition to the ones embedded in the Analysis Framework. Brief comments on those methods are given in introducing relevant findings. The use of such additional methods was kept to a minimum. They always related to questions that were already asked and operationalised, with the operationalisations found to be suboptimal or ineffective as the study progressed.

## Chapter 5: Case study findings

### ***1. Contents and structure of the chapter***

This chapter presents the empirical contents of the thesis as well as modifications of positions set out in earlier chapters. Both the data and re-conceptualisations concern various aspects of the question of how findings are transformed for the purposes of health research synthesis studies. As transformations are seen as a way of smoothing out differences between findings, provided relevant similarities between these findings are in place, this is also a chapter on similarities and differences between findings.

Section 2 describes the sample of papers analysed for obtaining the empirical data. Section 3 describes the overall deployment and outcomes of the analysis. Section 4 takes a step back to revise the initial conceptualisation of 'transformations' in view of the findings that emerged in the analysis. The subsequent sections focus on individual elements of the re-conceptualisation of 'transformations', generally by way of describing these elements, quantifying them, and identifying subtypes within them. Section 5 presents four different perspectives towards the question of how similar/ different are the findings with which research synthesis needs to deal. Section 6 makes tangible the multiplicity of findings. Section 7 addresses that which is missing from findings. Section 8 discusses the effectiveness of the tools for enabling and controlling transformations used in this study.

This chapter contains more methodological information than is typical of 'results' sections in empirical studies. A number of the techniques used to obtain the results could be fully specified only after an advanced stage of the analysis has been reached (completion of its most extensive component – the extraction-coding in NVivo, Excel and EPPI-Reviewer). The description of these techniques in the Methods Chapter would have made for 'too many results' there. Another reason for the untypical proportion of methodological information is the experimental character of the methods. The discrepancies between pre-specified study design and its actualisation in practice, which are normally discussed in results sections, were many. To avoid breaking down the



presentation too much, however, methods not described previously are presented briefly before their respective findings, with most of the detail provided in the Appendix to Chapter 5.

## ***2. The sample of papers***

### **2.1. Outcomes of the sampling process and key features of papers**

As described in the Methods chapter (Section 2), 100 papers were first sampled randomly from within the search strategy retrieval (18, 456). From this pool of 100, papers were sampled purposefully until saturation. The first and thus anchoring paper was “How Should We Design Supportive Cancer Care? The Patient’s Perspective” (Casarett et al., 2008). Of all sampled titles, I judged this title to correspond most closely to the test synthesis question. The process of selection involved varying the features of papers instantiating parameters such as topic, method, stakeholders under study, sample size, etc., while also taking tendencies within the broader sample into account (e.g. more papers were sampled from more frequent topic types). I considered saturation to have been reached after the comprehensive extraction-coding of 17 papers. Saturation was constructed in terms of formal features of papers (such as what the elements of a research paper on the health of humans are and how they are linked together), in terms of approaches to their handling under the Analysis Framework, and in terms of patterns of data relevant to the questions of the thesis. Saturation did not relate to any thematic threshold concerning the test synthesis question.

13 of the papers were primary research studies, 2 were reviews (one systematic and one non-systematic), 1 involved secondary data analysis, and 1 was a ‘research directions’ paper proposing a vision for the future of a field. 4 papers addressed issues such as experiences, perceptions, preferences and quality of life of cancer patients and informal caregivers; 4 addressed quality and safety of care, organisation of services and provider skills; 3 explored illness in a spiritual, existential and intensely emotional context; 2 were basic biomedical research papers; 2 were papers on predictors (clinical and socio-demographic) of cancer-related outcomes, and 2 were methodologico-theoretical papers. All papers touched on patient-relevant outcomes. 4 studies were set in the UK, 2 in

the US, 1 in Australia, Austria, Canada, India, Israel, the Netherlands, Saudi Arabia and Sweden. The reviews and the research agenda article included papers from various countries. One journal was found to have been sampled 3 times after which the variety of journals was explicitly considered, too. One author was noticed to appear twice, in both cases at third place (still a lottery winning chance for a sample that started with over 18,000 papers!). The methodologies of the sampled papers are discussed in the next section. Table T 5.1 below presents a snapshot of their main features.

## 2.2. Quantitative/ qualitative status of papers

As per intentions set out in the Methods Chapter (4.2), studies had to be given a primarily quantitative/ primarily qualitative label on four features – 1) study topic, 2) method of data collection and type of source material, 3) method of data analysis, and 4) presentation style. In the end, only a ‘study topic’ and an ‘integrated method’ feature were used. This was because in this particular sample there was only one study where a qualitative method of data analysis went with a more quantitative style in presentation, and only three studies (the two reviews and the research agenda study) where the relationships between method of data collection, type of source material, and method of data analysis were relatively more complex. These particularities were handled individually. Table T4.1 presents three partly overlapping quantitative/ qualitative status ascriptions (see shaded columns). Alternatives were generated so as to enable several types of testing of hypotheses concerning differences between the findings of quantitative and qualitative studies. Dramatic differences in structure had to be observed across classifications to be considered reliable.

The ‘preferred’ classification, seen to reflect best the internal variety of studies, comprised 4 ‘hybrid’ studies, 9 quantitative studies and 4 qualitative ones. Hybrid status was assigned if there was a supposed internal tension (e.g. as in Study 7, where religious coping and post-traumatic growth, seemingly typical qualitative research topics, were explored through scales).<sup>97</sup>

---

<sup>97</sup> Hybrid status was *not* assigned when the method was strongly associated with one of the ‘paradigms’ and the topic was ‘common’ – one which can be, or has traditionally been, approached by both quantitative and qualitative studies. Methodologico-theoretical issues or the  
210

In the first of the alternative classifications, hybrid studies were split into primarily quantitative or primarily qualitative through prioritising their method. 12 studies were labelled as quantitative and 5 as qualitative.

In the second of the alternative classifications, hybrid studies were split into primarily quantitative or primarily qualitative through prioritising their topic. 9 studies were labelled as quantitative and 8 as qualitative.

**Table T 5.1: Key features of sampled papers**

**NB: The numbers assigned to papers in Column 1 are used from here onwards as their main identifier (e.g. quotes are attributed to Study 1, 7, 12, etc.)**

The order of parameters (apart from those on quantitative-qualitative status, which draw on the first two parameters) follows the priority given to them in achieving variety through the purposeful sampling. I.e. first, studies had to differ by topic and method, then by stakeholders studied, then by country, etc.

Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
1. Casarett D, Fishman J, O'Dwyer P J et al. <b>How Should We Design Supportive Cancer Care? The Patient's Perspective</b>	Experiences, perceptions, preferences, quality of life	Quantitative, questionnaire-based	Hybrid	Quan	Qual	Patients receiving treatment for cancer	US	300	2008	<i>Journal of Clinical Oncology</i>	Ok
2. Whitaker K L, Brewin C R and Watson M. <b>Intrusive cognitions and anxiety in cancer patients</b>	Experiences, perceptions, preferences, quality of life	Mixed – quantitative, questionnaire-based, and qualitative  Quantitative dominates	Hybrid	Quan	Qual	Prostate cancer patients	UK	130	2008	<i>Journal of Psychosomatic Research</i>	Ok

Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
3. Thomas R and Davies N. <b>Lifestyle During and After Cancer Treatment</b>	Basic biomedical and clinical research relevant to lifestyle	Review of quantitative (basic biomedical and clinical) research; discursive (non-systematic) approach to review	Quan	Quan	Quan	Papers – concerning cancer patients but also food, chemical elements and compounds, and basic processes	UK authors, origin of reviewed papers varies	Number of studies not reported	2007	<i>Clinical Oncology</i>	Some (at times serious) difficulty with the basic research
4. Paul C L, Clinton-McHarg T, Sanson-Fisher R W et al. <b>Are we there yet? The state of the evidence base for guidelines on breaking bad news to cancer patients</b>	Quality and safety of care, organisation of services, provider skills	Systematic review of state of the research (mainly in quantitative terms) rather than findings (which appear mixed quan and qual)	Quan	Quan	Quan	Papers – on patients and providers	Australian authors; English-language articles	245 articles	2009	<i>European Journal of Cancer</i>	Ok

Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
5. Mols F, Thong M S Y, Vreugdenhil G et al. <b>Long-term cancer survivors experience work changes after diagnosis: results of a population-based study</b>	Predictors – clinical and socio-demographic	Quantitative, clinical data-based, secondary data analysis	Quan	Quan	Quan	Cancer survivors	Netherlands	403	2009	<i>Psycho-Oncology</i>	Ok
6. Bennett B, Friedlander M and Lloyd A. <b>The Experience of Cancer-Related Fatigue and Chronic Fatigue Syndrome: A Qualitative and Comparative Study</b>	Experiences, perceptions, preferences, quality of life	Qualitative, interview-based	Qual	Qual	Qual	Women with cancer and chronic fatigue symptom (comparator without cancer)	Australia	12	2007	<i>Journal of Pain and Symptom Management</i>	Maybe the easiest relative to my knowledge
7. Thombre A, Sherman A C and Simonton S. <b>Religious Coping and Posttraumatic Growth Among Family Caregivers of Cancer Patients in India</b>	Illness in a spiritual, existential and intensely emotional context	Quantitative, questionnaire-based	Hybrid	Quan	Qual	Family caregivers	India – setting US team	58	2010	<i>Journal of Psycho-social Oncology</i>	Ok

Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
8. Meades R T, Svensson W E, Frank J W et al. <b>Carcinoma of the breast wire localisation post nuclear medicine sentinel lymph node imaging. Are radiologists receiving a significant dose?</b>	Quality and safety of care, organisation of services, provider skills	Quantitative, applied physics research	Quan	Quan	Quan	Providers – radiologists	UK	2 radiologists, 12 patients	2010	<i>European Radiology</i>	Had difficulty
9. Andrykowski M A, Donovan K A and Jacobsen P B. <b>Magnitude and Correlates of Response Shift in Fatigue Ratings in Women Undergoing Adjuvant Therapy for Breast Cancer</b>	Methodological and theoretical	Quantitative, questionnaire-based	Quan	Quan	Quan	Patients (women), focus on method	US	102	2009	<i>Journal of Pain and Symptom Management</i>	Ok, though some difficulty with details of statistical analysis
10. Rapp K, Klenk J, Ulmer H et al. <b>Weight change and cancer risk in a cohort of more than 65 000 adults in Austria</b>	Predictors – clinical and socio-demographic	Quantitative, clinical data-based	Quan	Quan	Quan	Patients	Austria – setting, German, Austrian and US team	Over 65,000	2008	<i>Annals of Oncology</i>	Ok

Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
11. Dwivedi R C, Nutting C M, Rhys-Evans P et al. <b>Future Perspectives for Health related Quality of life (HRQOL) studies in Head and Neck Cancer</b>	Experiences, perceptions, preferences, quality of life	Research agenda setting based on a review, discursive approach	Qual	Qual	Qual	Research tendencies and papers	Italian journal UK team	Number of reviewed papers not reported	2009	<i>Giornale Italiano di Medicina del Lavoro ed Ergonomia</i>	Ok
12. Molassiotis A, Wengström Y and Kearney N. <b>Symptom Cluster Patterns During the First Year After Diagnosis with Cancer</b>	Methodological and theoretical	Quantitative, questionnaire-based	Quan	Quan	Quan	Patients, but mostly theoretical focus	UK	143	2010	<i>Journal of Pain and Symptom Management</i>	Ok, though some difficulty with details of statistical analysis
13. Goldberg Y, Porat R M, Kedar I et al. <b>An Ashkenazi founder mutation in the MSH6 gene leading to HNPCC</b>	Biological substrate	Quantitative, basic biomedical research	Quan	Quan	Quan	Genes	Israel	4 families	2010	<i>Familial Cancer</i>	Maybe the most difficult of all



Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
14. Jaturapatporn D, Moran E, Obwanga C et al. <b>Patients' experience of oxygen therapy and dyspnea: a qualitative study in home palliative care</b>	Experiences, perceptions, preferences, quality of life	Qualitative, interview-based  Low on richness of description and evocative power	Hybrid	Qual	Qual	Patients	Canada – setting  Thailand – university of first author	8	2010	<i>Supportive Care in Cancer</i>	One of the easy papers
15. Al-Rowaili A , Al-aqeel S A, Al-Naim LS et al. <b>Appropriateness of Cancer Pain Management in Saudi teaching Hospital</b>	Quality and safety of care, organisation of services, provider skills	Quantitative, clinical data-based	Quan	Quan	Quan	Patients and providers	Saudi Arabia	160	2009	<i>The Gulf journal of oncology</i>	Ok
16. Hubbard G, Kidd L and Kearney N. <b>Disrupted lives and threats to identity: The experiences of people with colorectal cancer within the first year following diagnosis</b>	Illness in a spiritual, existential and intensely emotional context	Qualitative, interview-based	Qual	Qual	Qual	Patients	UK (Scotland)	18	2010	<i>Health</i>	Ok

Paper authors and title	Broad thematic category	Broad methodological category	Quan, qual or hybrid	Quan or qual, method priority	Quan or qual, topic priority	Main stakeholders under study	Country of study	Sample size	Year	Source	How easy to process?
17. Hök J, Wachtler C, Falkenberg T et al. <b>Using narrative analysis to understand the combined use of complementary therapies and bio-medically oriented health care</b>	Illness in a spiritual, existential and intensely emotional context	Mixed – research with arts and literature	Qual	Qual	Qual	Caregiver	Sweden	1	2007	<i>Social Science &amp; Medicine</i>	Ok

### **3. The extraction-coding process and outcomes**

#### **3.1. Basic features of the extraction-coding process and outcomes**

As discussed in the Methods Chapter (Section 3.3), the extraction-coding was first performed in NVivo 8 in combination with Excel (12 studies) and then in the synthesis research specific EPPI-Reviewer 4 (5 studies). Five untypical injunctions had to be taken into account: for comprehensiveness, extensive multiple coding, extreme transparency, double-edged critical analysis, and for first coding as close as possible to the source material and experimenting with a variety of more distant transformations afterwards (Chapter 4, 3.1).

It took between one and four days to process a paper, and typically two – two and a half. Coding was mostly done at the level of sentences and clauses, occasionally at the level of words and larger units of information. There was no gradual increase in the intensity of coding (number of codes per paper) from the beginning towards the end of the case study, judging by NVivo data.<sup>98</sup> The extraction of numerical data was done for each individual data point for studies 1 to 12 (the NVivo-Excel sample). After this, only cancer-related and socio-demographic numerical data were extracted in detail in keeping with the streamlined analysis intentions.

The text below describes further the processes and outcomes of the extraction-coding, separately for each of the three software environments. The largest amount of meta-data (number of codes and units of information) is provided for NVivo, which appears to have superior functionalities in that respect.

##### *3.1.1. Extraction-coding in NVivo*

1514 codes grouped under 25 main codes were generated in NVivo. These included codes of the Analysis Framework, set up on the basis of the literature review and expectations from practical experience (approximately 205 codes), as well as thematic and formal elements prompted by the analysis of the 12 papers.

---

<sup>98</sup> The statistic on number of codes per paper was immediately available in NVivo. If it can be obtained automatically in EPPI-Reviewer, I have not identified how as of yet.

The largest numbers of sub-codes were generated under Findings (260), Methodological (249) and Conclusions and Suggestions (167). The smallest numbers of sub-codes were under Classification and Coding issues (none), Vocabulary (1), Unclear points (2) and Ethical Issues (2).

Of the 1514 codes, a third (505) were empty. Some of these were ineffective pre-formulated codes from the Analysis Framework. Others were more abstract 'tags' (in the sense described in Chapter 3, Section 3). Still others were high level codes where the coded contents appeared only at lower levels (e.g. *Variables* was such an empty high level code with the contents coming under subtypes of variables). In a minority of cases (I identified 15, approximately 1% of the codes), empty codes represented omissions of coding – a category was created on the basis of some contents but the latter was not coded thereafter. Put flatly, these were errors on my part. They are correctible and they are not something that is reported in research publications. I report them as a background consideration. Individual errors (which of course are many more than these since it is a single point in the work) are tightly controlled for in most synthesis studies, through double or triple extraction for instance. This chapter discusses many other factors that potentially limit the reliability of synthesis studies and for which no control is instituted.

The ineffective pre-formulated codes were primarily ones found to require too multidimensional and intense critical analysis before they can be filled in (such as what alternative explanations for findings can be offered) and a much more advanced stage of the test synthesis study than could be achieved (such as codes aimed at exploring transformations by comparing preliminary synthesis output with source study input). As both these types of codes relate to explorations of the process of transformations, some of the intended analysis of transformations could not be performed.

1009 codes were thus effectively used across papers. There were 2294 units of information under them.

(By way of crude comparison, a primary qualitative study where I did the coding (Munday et al., 2009), used NVivo and performed more multiple coding than I had done in previous work – i.e. at least the researcher and software are the same – used 37 sources, and had 73 codes and 758 units of information. That

is, the injunctions have contributed, along with other relevant parameters in terms of which the two studies differ, to almost 14 times more codes in the current study (73 vs. 1009) for approximately half of the sources (37 vs. 17). Approximately 30 times more codes could then be expected for the same number of sources. In terms of units, the injunctions have contributed to slightly over 3 times more units of information in the current study (2294 vs. 758) for approximately half of the number of sources. Approximately 6 times more units could then be expected for the same number of sources. Note again that this is a very crude comparison, as the comparator study was not a synthesis study.)

761 codes (50.3%) had units of information from a single paper. As discussed above, 505 codes (33.4%) were empty. 50 codes (3.3%) had units of information from 4 or more papers, with a maximum of 11.<sup>99</sup> Some of these 'richer' codes were highly inclusive categories addressing aims of the thesis rather than the test synthesis (e.g. Keywords for Paper, to help explore how themes highlighted by authors or indexers lose or gain importance in the process of moving from the primary into the synthesis study). Others of the richer codes were for issues of limited interest to both the thesis and test synthesis but frequently reported in health research papers (e.g. research ethics). The richest codes of interest to the test synthesis were on sampling and on temporal and dynamic features of phenomena (such as frequency, duration, intensity, etc.).

In terms of units of information within codes (as opposed to number of papers from which these were taken), 511 codes (33.8%) contained a single unit of information. 505 (33.4%) were empty. 190 (12.4%) of the codes had 4 or more units of information, with a maximum of 53. The codes which received the largest number of units of information were broad codes on nomenclature and descriptions of methods.

Further detail on numbers of papers and units within codes can be found in Appendix to Chapter 4.

---

<sup>99</sup> These numbers treat each code as unrelated – i.e. with no codes below and above. Otherwise each of the main codes and a large number of other codes would have had many more units of information from many more papers.

NVivo was the environment where the largest number of separate tagging documents, with various critical, reflective, heuristic and pragmatic notes in them, were produced – a total of 19.

### *3.1.2. Extraction-coding in Excel*

The Excel extraction-coding was used to extract numerical or categorical data as a complement to the NVivo extraction-coding. Thematically, there were two main types of such data extracted. One concerned parameters on the basis of which similarity between studies, populations, health conditions, settings, interventions, outcomes and methods are determined and groups formed. The other were data representing numerical findings about the phenomena of interest to a study, often in relationship to the aforementioned properties (e.g. hazard rate ratios for developing cancer relative to degree of weight change from baseline, Study 10).

### *3.1.3. Extraction-coding in EPPI-Reviewer*

The extraction-coding in EPPI-Reviewer combined features of the NVivo and Excel processes, as the software is well adapted to handling numerical information, unlike NVivo. The framework of codes from NVivo and the categories from Excel were not transferred into EPPI-Reviewer. This was to allow for new conceptualisations to emerge. The novelty thus achieved was primarily of lower level conceptualisations and not for the overall framework.

Differences in the software environments notwithstanding, findings from the three platforms pointed in the same directions. They are discussed together in the remainder of this chapter. But not quite as expected.

## **4. A conceptual transformation of transformations**

The initial conceptualisation of ‘transformations’ in this work was framed strongly in terms of similarities and differences. Transformations were thought of as driven by differences in findings, provided some right type of similarity was

there, too, and having as an outcome a right type of reduction of those differences. This understanding was then intertwined with the concept of matrices – those frameworks, causes, factors which drive the generation of and co-constitute findings, contribute to their similarities and differences, and enable and constrain their combinability. The Analysis Framework was organised, to a large extent, around prompts aiming to detect key matrices in a study report. The process of accumulating and analysing data suggested, however, that a conceptualisation in terms of matrices can contribute little to a descriptive account of transformations. A primary reason was the rather limited degree to which presumed key matrices, with the exception of methods, were detected in source studies.

I then built on the emerging empirical findings and on ideas from the introductory description of transformations to formulate a revised understanding of the concept. It focused on the capacity of a finding for meaningful transformation. At the very least, this capacity was taken to depend on:

- the contents of the set of findings **S** whose combinability or compatibility with **f** ('our finding') is in question;
- the extent of differences and similarities (thematic, formal and factual) between **f** and the other findings in **S**;<sup>100</sup>
- the presence of characteristics of a finding that come into a 'logical contradiction' with specifications made in the synthesis question;
- the multiplicity of **f** – the extent to which **f** can be broken down into smaller findings (called here 'inherent multiplicity') and the extent to which it can be interpreted in different ways by different users and relative to different aims ('relational multiplicity');
- the extent of missing and available pieces of information in **f**;
- the breadth of the synthesis question and its distance from **f**;
- the effectiveness of the tools used to enable transformations and of those used to constrain them;

---

<sup>100</sup> Initially, I considered excluding factual differences. Transforming them, provided sameness of thematic and formal characteristics of a finding, appeared to mean changing the data. However, one can envisage cases where extreme findings are adjusted in accordance with some weighting principle that takes into account size of sample, reliability of method, etc.

- the familiarity of the researcher with the research area of **S** and that of the synthesis question;
- a still-to-be-specified set of environmental influences on cognitive processes and idiosyncratic cognitive characteristics of the researcher – such as order of presentation of findings, recent salient information available to the researcher, his or her capacity for lateral thinking, chance juxtaposition of information, etc.

The primary reason for offering a re-conceptualisation of transformations at this stage is to explain the interconnectedness between the seemingly disparate empirical findings that follow. As mentioned above, this re-conceptualisation was not only derived from the data but through independent re-thinking of more abstract claims, too. Many holes in the empirical material remain as a result. No sections on transformations in relation to the synthesis question, researcher's expertise, logical contradictions, sets of findings, or environmental and idiosyncratic personal factors affecting cognitive processes are included. Although some relevant data were collected on the researcher's expertise (and that particular subject is always available for more data collection), these were, ultimately, considered insufficient, unreliable or requiring the development of further analysis methods. To the remaining elements of the re-conceptualisation, a separate section is dedicated. Section 5 is on differences and similarities of findings; Section 6 is on the multiplicity of findings; Section 7 is on 'omissions, silences and half-silences'. Section 8 offers observations on the tools for enabling and controlling transformations used in this study.

### ***5. Differences and similarities between findings***

Four types of findings on differences and similarities between findings are presented: findings on compositional and structural similarities and differences between quantitative and qualitative findings explored through the Vocabulary of Elements of Findings (5.2); a small set of other formal similarities and differences obtained through an analysis of typologies of findings and descriptors of data, evidence, studies and findings (5.3); quantification of the scope of thematic similarities and differences between findings (5.4); and findings on similarities and differences of supposedly highly standardised



parameters of health research studies on cancer (socio-demographic information and cancer-related variables). These represent different ways of exploring where the most relevant similarities and differences between findings lie, as far as transformations aimed to make them more combinable are concerned, and what their extent is.

## 5.1. Starting points

### 5.1.1. A (conventional) finding

In the Postscript to Literature Review – Prescript to Methods, the concept of a finding was chosen to represent the main carrier of empirical knowledge in a health research paper but was left vague. The processing of study findings in the analysis suggested the following refinements.

For the purposes of this work, *a (conventional) research finding is taken to be any proposition, data point, material object or a closely connected set of these that IS and/ or represents primarily (a framing of) some empirical contents obtained through the methods of a specifiable research study.*<sup>101</sup>

---

<sup>101</sup> Here is a clarification of some of the choices of words in this definition:

- *Primarily*, in ‘primarily empirical contents’ takes into account what is now a received idea that empirical knowledge cannot be entirely theory-free as well as of boundary cases where a description of a method, theory, background assumption, etc. appear in a proposition together with empirical knowledge.
- *IS and/or represents* tries to accommodate the fact that we speak of findings as: 1) things (whether material objects, data points or propositions) procured by researchers in a laboratory, a field, etc.; 2) representations of these things for further manipulation and the outcomes of such further manipulation; and 3) representations of 1) and 2) aimed at being shared with others. The *and/or* tries to accommodate views which distinguish between some finding ‘in and of itself’ and its representation and ones which see a representation as integral to a finding, with no finding being there before it is represented.
- *A framing of* has been added to acknowledge positions that some (understood as the ‘same’) empirical contents can be there or be represented in various ways – e.g. through the highlighting of different features, the application of a different theory, conceptualisation, analysis, the use of different words and sentence structure, etc. ‘IS and/or represents’, as described above, implies but is not explicit about such a possibility.

More narrowly, the findings which the papers analysed in this study provided stand anywhere on a continuum

going from

representations of raw data, typically quotes (e.g. *At work you have to force yourself to stay awake of course*, Study 6);

through

findings that could (appear to) be purely observational claims, e.g. *Analgesics were given orally in 70.5% of the patients, 26.5% were prescribed fentanyl patches, and 3% intravenous morphine* (Study 15)

to

interpretations that show a high degree of recourse to external sources and transformations of the raw data, but nevertheless point to some empirical contents from the current study, be it framed entirely within the language of those external sources, e.g. *Using Bury's (1991) definition of biographical disruption and Charmaz's (1983) concept of 'loss of self', several of the people in this study appeared to experience cancer as biographical disruption because it represented a threat to their identity* (Study 16).

The latter are findings whose evidential status may be questioned from certain 'positivist' positions, but are considered typical qualitative research findings and are treated as such here.

### 5.1.2. A basic finding

A 'basic finding' is understood within the framework set by the Vocabulary of Elements of Findings, as containing a single 'expression' (that is, representation of a single feature, state, behaviour, etc.) or a single 'relationship'. This largely syntactic view of a basic finding does not discriminate in terms of the complexity

- 
- *Obtained through the methods of a specifiable research study* guards against the possibility that any empirical fact asserted in a research study, including those vaguely attributed to 'research' or those whose relationship to the methods of a particular study is unclear, are treated as a 'research finding'.

of elements (e.g. a finding can be about a gene but also about a hospital) and interpretiveness (e.g. a finding can be 'purely observational' or heavily laden with sociological theory).

### *5.1.3. Boundaries of findings*

Two approaches to the setting of boundaries of findings were resorted to in the exploration of similarities and differences between types of finding.

One of them was to pick/construct units of meaning that gravitated around a normal sentence. Often these were coterminous with a normal sentence but not necessarily. Several sentences could be extracted together as a single 'composite' finding. Elements of a single sentence could be extracted individually if perceived as separable and important. This was the approach used in the extraction-coding in NVivo, Excel and EPPI-Reviewer.

The other approach was to set narrow and more rigid boundaries for findings. For connected text, a finding was an individual sentence or a verbless heading. For tables, a finding included the data for a feature or entity with a limited number (0-5) of subtypes beneath (e.g. education with its five levels). This was the approach used in the formalisation of findings through the Vocabulary of Elements of Findings.

## **5.2. Similarities and differences between quantitative and qualitative findings**

In response to by far the most prominent issue in the debate on combinability of studies in health research synthesis, the search for similarities and differences between quantitative and qualitative findings gained priority. It was carried out through the Vocabulary of Elements of Findings and was thus a search for compositional and structural differences between quantitative and qualitative findings. The development of the Vocabulary had already re-scaled expectations about the likely extent of such differences. Its 10 key elements ('obsessive something', 'carrier', 'expression', 'relationship', 'knowledge something', 'privileged something', 'subsidiary something', 'quantifier', 'qualifier')

and 'indexical') were generic. They appeared indispensable in any type of research even if not in any one finding. They did not appear to cluster in stable patterns. Such impressions had to be, however, formally verified. As the refinement of the approach could only happen at an advanced stage of the analysis, it is briefly described here rather than as part of the initial methodological set-up. Details are provided in the Appendix to Chapter 5.

Findings were sampled from each group into which findings were classified during the extraction-coding, with the exception of some of the lowest level groups in EPPI-Reviewer. The initial sample consisted of 327 items. 26 of these were excluded as a result of applying the conceptualisation of 'conventional finding' given in 5.1.1. The final sample contained 301 findings.

The data – the formalisation for each finding – were transferred into IBM SPSS Statistics, Version 19 (IBM, 2013). New variables for each of the key elements of findings were computed, to reflect the number of times a particular element appeared in a finding (e.g. Finding 1: obsessive something – 1, carrier – 1, expression – 2). Each finding was also ascribed a label for quantitative, qualitative or hybrid status reflecting the classifications of the sampled studies described in 2.2. As per the 'preferred grouping', there were 128 quantitative, 90 qualitative and 83 hybrid findings in the sample. As per the alternative grouping where method was prioritised, there were 182 quantitative and 119 qualitative findings. As per the alternative grouping where topic was prioritised, the numbers were of 150 quantitative and 151 qualitative findings.

Five well established claims about distinguishing features of quantitative and qualitative studies were operationalised into hypotheses about the relative prevalence of certain elements. Frequencies for the Vocabulary elements participating in these hypotheses were obtained, for each of the groups of the above categorisations. For the remaining elements frequencies were still obtained to explore the possibility for unexpected differences of composition.

The first claim tested was that quantitative studies are markedly more 'positivist' and/or 'naively realist' than qualitative studies, which are, purportedly, more sophisticatedly realist (e.g. critically realist) or constructivist. This was operationalised as the task to compare, between groups, the number of findings

containing at least one knowledge something (which I had conceptualised as indicator of the mediated or uncertain character of knowledge) and the number of findings formulated as if of “a world out there” (a common, somewhat pejorative, description of the representational conventions of quantitative research). Let me remind some of the features of knowledge somethings discussed in Chapter 4 (Section 4.1.1). For instance, “Noise related to the equipment was another minor disadvantage of oxygen usage” (actual formulation of a finding from Study 15) spoke directly about ‘the world’. ‘Noise related to the equipment was *reported to be* another minor disadvantage of oxygen usage’, or ‘*Analysis suggested* that noise related to the equipment was another minor disadvantage of oxygen usage’ would be examples of formulations that contain a knowledge something. In contrast, “The monthly personal monitoring dosimetry readings were lower than the minimum detectable radiation dose for both the chest-badge and finger-ring dosimeters (10  $\mu$ Sv and 300  $\mu$ Sv, respectively)” (actual formulation of finding from Study 8) is a finding expressing by means of several knowledge somethings what can also be expressed through none, as in ‘The radiation received by radiologists performing breast wire localisation was very low’. I hypothesised that most qualitative studies would contain a knowledge something, while most or all quantitative studies would have none.

As per the preferred grouping of studies (quantitative, qualitative and hybrid), 59.4% of the quantitative findings, 47.8% of the qualitative findings and 51.8% of the hybrid findings referred to some knowledge something. The numbers for alternative classification 1 (method prioritised) were 55.5% for quantitative and 51.3% for qualitative findings, and for alternative classification 2 (topic prioritised) were 57.3% for quantitative and 50.3% for qualitative findings. Thus, regardless of the manner of classifying studies, quantitative studies were, contrary to expectations, somewhat more likely to make references to scientific artefacts and processes that mediate knowledge about the world. The differences were, however, minor.

The second claim tested concerned the greater (acknowledged) role of the researcher in qualitative studies. It was operationalised as the hypotheses that

qualitative findings will be much more likely to contain a reference to a 'knower' engaged in a particular study and that those references will be much more extended in qualitative studies. This was not the case. A knower was present in only 7 findings, 5 or 6 of which were classified as quantitative as per all three classifications. Knowers were represented very simply – as 'we' or 'our' (we reported, our sample). As much as quantitative and qualitative studies may differ substantially in the extent to which they recognise the effect of the researcher on findings, describe relevant features and behaviours of the researcher, etc. such differences were not identifiable at the level of findings.

As per the third hypothesis, as it is often claimed that qualitative studies are much more focused on context and interconnections, qualitative findings were expected to contain many more privileged somethings, subsidiary somethings and relationships than quantitative findings. Privileged somethings were coded for in 32.8% of quantitative findings, in 43.3% of qualitative findings and 42.2% of hybrid findings (the differences of these numbers with the numbers from the alternative groupings ranged between 0.7% and 2.9%). Relationships (at least one) were coded for in 34.0% of quantitative, 44.4% of qualitative and 44.6% of hybrid findings (the differences of these numbers with the numbers from the alternative classifications ranged between 0% and 2.7%). The number of subsidiary somethings across the sample was very low (a total of 12 for all 301 findings). It was equal in quantitative and qualitative findings as per the preferred grouping and alternative classification 2 (5-5 and 6-6). Thus, the general tendency predicted by the hypothesis was observed (in the case of elements of higher frequency), but the marked predominance was not.

The fourth hypothesis tested the claim that the concepts of qualitative studies tend to be much more complex and vaguer. I translated this claim into a hypothesis that qualitative findings will have a larger number of codes for 'obsessive something, aspect' – the code reflecting that an obsessive something is studied through a particular component or perspective towards it. The numbers were slightly larger in qualitative findings as per the preferred classification and alternative classification 2, topic prioritised (in the former

case, 23.4% of quantitative findings, 28.9% of qualitative findings and 16.9% of hybrid findings contained an obsessive something, aspect; in the latter – 21.3% of quantitative and 25.2% of qualitative findings). The difference was very small as per alternative classification 1, method prioritised (22.5% of quantitative and 24.4% of qualitative findings contained this element of the Vocabulary).

Finally, I hypothesised that codes for ‘qualifiers’ will be much more frequent in qualitative studies, as a most likely pointer to evaluative (in a moral or emotional sense) elements in findings. The results disconfirmed this expectation, but the difference was unstable. In the case of the preferred and alternative classification 1 (method prioritised), qualifiers were more in quantitative findings (15.6% of quantitative findings, 7.8% of qualitative findings and 19.3 of hybrid findings contained a qualifier in the former case and 17.0% of quantitative and 10.1% of qualitative findings contained a qualifier in the latter case). In the case of alternative classification 2, the difference was almost lost (with 15.4% of quantitative and 13.2% of qualitative findings having a qualifier).

Table T 5.2 shows the relative presence in quantitative and qualitative findings of the remaining elements.

**Table T 5.2: Presence of the ‘other’ elements of the Vocabulary of Elements of Findings<sup>102</sup>**

	<b>Obsessive something</b> Group – No (%)	<b>Carrier</b> Group – No (%)	<b>Expression</b> Group – No (%)	<b>Quantifier</b> Group – No (%)
<b>Preferred classification</b> 128 quantitative 90 qualitative 83 hybrid	Qn – 60 (46.9%) Ql – 43 (47.8%) H – 47 (56.6%)	Qn – 55 (53.0%) Ql – 51 (56.7%) H – 47 (56.6%)	<b>Qn – 79 (61.7%)</b> <b>Ql – 73 (81.1%)</b> H – 69 (83.1%)	<b>Qn – 44 (34.4%)</b> <b>Ql – 16 (17.8%)</b> H – 28 (33.7%)
<b>Alternative classification 1 – method prioritised</b> 182 quantitative 119 qualitative	Qn – 90 (49.5%) Ql – 60 (50.4%)	<b>Qn – 81 (44.5%)</b> <b>Ql – 72 (60.5%)</b>	<b>Qn – 122 (67.0%)</b> <b>Ql – 99 (83.2%)</b>	<b>Qn – 61 (33.5%)</b> <b>Ql – 27 (22.7%)</b>
<b>Alternative classification 2 – topic prioritised</b> 150 quantitative 151 qualitative	Qn – 77 (50.7%) Ql – 74 (49.0%)	<b>Qn – 64 (42.7%)</b> <b>Ql – 89 (58.9%)</b>	<b>Qn – 98 (65.3%)</b> <b>Ql – 123 (81.5%)</b>	<b>Qn – 52 (34.7%)</b> <b>Ql – 36 (23.8%)</b>

<sup>102</sup> Group – Qn (quantitative), Ql (qualitative), H (hybrid)

No – number of findings of a particular type (Qn, Ql, H) in which a particular element found

% – percentage of the above number relative to the total number of findings of this type

**Comments to the Table:** The difference in quantifiers (in the direction of quantitative findings having many more of these) was not as impressive as expected partly because non-precise or common quantifiers (some, many, one) are reasonably well represented in qualitative research, but more importantly because many quantifiers were coded as part of a particular type of ‘expression’ which was then merged with expressions rather than quantifiers.

One possible explanation for there being a larger percentage of expressions in the pool of qualitative findings is the fact that knowledge somethings were more in quantitative studies. As discussed in 4.1.1.2, there is a stable pattern of coding where coding for a knowledge something is equivalent to coding for one or more expressions. If a knowledge something was present, I tended to code first for a knowledge something rather than an expression as it participated in specific hypotheses. In the quantitative analysis, I used only one coding string per finding rather than all alternatives I had generated. It tended to be the first string, as typically this was closest to the surface syntax of a sentence.



Overall, compositional differences between quantitative and qualitative findings were found to be minor. This was unsurprising – the process of coding with the Vocabulary of Elements had indicated quite clearly that, through the lens of this method, quantitative and qualitative findings are first and foremost findings. The hypotheses tested above were knowingly weak. I nevertheless needed to perform the quantitative verification so that I was not swayed by impressions. ‘Knowingly weak’ can only be said *post factum*.

### 5.3. Other formal similarities and differences between findings

A small set of formal similarities and differences was obtained by probing into the other most typical way of arguing for challenges to the combinability and compatibility of findings – the presentation of long lists of varied labels for types of studies, evidence and data (see Methods Chapter, 4.1.2).

260 types of findings (some of them single-member ones) were specified during the extraction-coding. Over 50 descriptors for data, evidence, findings and studies were identified in the analysed texts or generated by applying further labels to these (e.g. “epidemiological data”, “cohort study data”, ‘emotive data’ ‘associative, correlative, causal data’, ‘data of greater or lesser public interest’, etc.). These were used to create ‘groups of tension’ where types had a dimension in common but fell at very different points on its continuum.

Five types of formal differences between findings were elicited. The first – *dramatic difference of magnitude* – is well recognised in the ‘heterogeneity of findings’ debate in the context of meta-analysis (see Chapter 1, 2.2). The remaining types of differences also appear to be sufficiently acknowledged though not in a structured way. I will describe them only briefly, without this being a comment on their importance. First, we have a difference of *reliability* between findings on the same issue. For instance, self-reported clinical data are less reliable than clinical data from medical records which are, in turn, arguably less reliable than clinical data collected specifically for the purposes of a particular study. The paradigmatic case of difference of reliability in the context of research synthesis is that of findings from RCTs and findings from observational studies. Then, we have a *difference of degree of detail*, such as

from ‘thicker’ and ‘thinner’ descriptions of emotional experiences of illness. Then, we have a *difference of scope*, such as in the case of longitudinal and cross-sectional findings, or descriptive and associative findings on the same phenomenon. Finally, we have a *difference of level*, as with findings directly about a phenomenon and with findings about the method used to obtain findings about that phenomenon.

As informative as these types of differences may be, the approach of accumulating strikingly different labels for types of findings still bears the burden of showing some substance.

#### **5.4. Scope of the thematic differences between findings**

Perhaps that substance is, after all, in the thematic differences between findings. The proxy measure used for these was simple – the frequency with which more than one unit of information was coded under the same NVivo code (already reported in 3.1.1, in describing the extraction-coding in NVivo). By virtue of being in direct relationship to the injunction to code first as close as possible to the original, with no or minimal transformations, and only then experiment, this frequency is an indication of the degree of thematic variability within and across the sampled studies. To remind, 511 codes (33.8%) contained a single unit of information; 505 (33.4%) were empty; 308 codes (20.4%) had 2 or 3 units and 190 (12.4%) had 4 or more units of information.<sup>103</sup> If we take the used rather than all codes as 100% (i.e. exclude the empty ones), 50.7% of the codes had a single unit of information, 30.5% had 2 or 3 units, and 18.8% had 4 or more units of information. This suggests a substantial variability of themes within and across studies.

The hierarchical structure of the coding, where any inclusion in a higher level is an indication of a shared meaning (and there were only 25 main codes, out of 1514), is a reminder that this individuality of meaning is relative. Also, those studies were intentionally sampled as highly heterogeneous. But they are nevertheless studies obtained for the purposes of the same synthesis question. They were retrieved using restrictive search strategies. Each study had at least

---

<sup>103</sup> The information on 2 and 3 units was available only in the Appendix. The percentage was calculable from the other information though.

one counterpart of the same broad theme, with two groups of four thematically similar studies and one group of three. There is a thought to be kept in mind that findings on the same or similar health research topic may be, in their untransformed or minimally transformed state, more thematically individual than their being 'on the same or similar topic' would suggest.

## **5.5. Similarities and differences in supposedly highly standardised descriptors**

Maybe the above thematic variability within and across studies is expected. After all, a defining feature of good research is its originality. Another defining feature of research, however, is its high degree of standardisation. This section presents findings about similarities and differences in the collection and representation of two types of supposedly highly standardised information in health research papers on cancer: 1) basic cancer-related information (e.g. on cancer type, stage, treatment, etc.) and 2) socio-demographic data on participants. These contribute various elements, be it background ones, to any finding in a health research paper on cancer in humans.

### *5.5.1. Similarities and differences in the collection and representation of basic cancer-related information*

Differences in the reporting (and, potentially, collection) of data on 6 basic cancer-related variables were explored. These variables were cancer type (e.g. breast, lung, stomach), (potential) cancer phase (e.g. predisposition & risk, primary prevention, occurrence), cancer stage (e.g. early, locally advanced), time since diagnosis, treatment phase (e.g. active surveillance, on treatment, post treatment) and treatment type (e.g. surgery, chemotherapy, hormones). Data from the Excel extraction-coding worksheets (Studies 1-12) and EPPI-Reviewer (Studies 13-17) were used.

*Availability of information across those standard parameters.* The reporting of data on cancer-related variables was most consistent in the case of cancer type. Of the 17 studies analysed, 15 were explicit of the cancer(s) they addressed. The two exceptions were Study 4, a review on breaking bad news to

cancer patients where an ‘all cancers’ can be safely presumed, and Study 17, a narrative analysis aiming to understand the combined use of complementary and biomedical care by a patient and her spouse. Cancer stage was recorded for 9 studies, treatment type for 6,<sup>104</sup> time since diagnosis for 5 studies, cancer phase for 4, and treatment stage for 2 studies. That is, data on basic cancer-related parameters different to cancer type were reported in between 16.7% and 52.9% of the analysed studies. No study reported all of them. This does not necessarily mean that information on those parameters could not be deduced, at least approximately. For instance, if a cancer has been reported as “advanced” in ‘cancer stage’, some of the categories from ‘(potential) cancer phase’ become irrelevant, such as primary prevention. Also, information may have been omitted if it was reported in a non-standard place or in a non-standard way.<sup>105</sup> Such disclaimers notwithstanding, substantial variability in the reporting of seemingly basic cancer-related information was observed in the sampled studies.

*Identity of study-level categories.* As standard as they may seem, the above parameters and their subdivision into categories were formulated for the purposes of the test synthesis study. This was necessitated by differences in designations across studies – small in the case of the broad parameters, e.g. time since diagnosis, substantial in the case of specific categories, e.g. designations for cancer stages. In the case of cancer type and treatment type, the sample did not show a notable variability in labels for specific categories – e.g. breast cancer and chemotherapy were breast cancer and chemotherapy across studies, at least in this sample. In the case of treatment type, however, combinations of treatments were sometimes reported together (e.g. chemotherapy and radiotherapy), while at other times the percentages accumulated to the respective individual categories. In the case of cancer phase, where 4 synthesis categories with data from more than one study were formed, all source study categories were different (e.g. the synthesis level category of ‘progression phase’, included data on “progressive early prostate

---

<sup>104</sup> If only one aspect of the treatment was reported which coincided with the focus of the study (e.g. the use of oxygen therapy in Study 14), this was not considered sufficient to treat ‘type of treatment’ as reported.

<sup>105</sup> A further level of testing requires re-reading all studies with a focus on cancer-related and socio-demographic information only.

cancer” from a study reviewed in Study 3 and “2 to 3 weeks after diagnosis, followed up for 1 year” from Study 12). In the case of cancer stage, where again 4 categories with data from more than one study were formed, 1 category showed a partial overlap of designations (2 out of 5 studies referred to ‘their’ cancers as “early”). More consequentially, the criteria by which cases were ascribed to categories (such as what criteria were used to diagnose breast cancer or what counts as early stage cancer in this study) or the details of a category content (such as what exactly chemotherapy in this case involved) were not reported. Study 10 appeared to be the only one noting the classification system according to which a type of cancer was determined. This is 1 study on 1 parameter out of 17 studies and 6 parameters.

#### *5.5.2. Similarities and differences in the collection and representation of socio-demographic information*

In the case of socio-demographic information, differences in the reporting of (and, potentially, collection of data on) 7 parameters were explored: sex/gender, age, ethnicity/ race/ nationality, marital status, educational status, socio-economic status, and employment. Data from the 14 primary studies within the sample were analysed. Reviews were excluded as I did not expect them to report extensive detail from the (many) studies they covered. The bar of completeness of information set for primary studies would have been too high for them.

*Availability of information across those standard parameters.* Sex/gender was reported across all 14 primary studies. Age was reported in 13 of them. Marital status appeared in 6; education in 5; breakdowns of ethnicity/nationality/race were available in 4 studies; socio-economic status and employment were reported on in 3 studies. Within categories, of the 13 primary studies which reported age, 8 reported a mean value, 7 a standard deviation, 5 reported range, 4 reported age bands, and 1 reported year of birth of individual participants. Again, this does not mean that information on those parameters could not be deduced at some level of precision.

*Identity of study-level categories.* Study-level categories were (expectedly) identical for sex/gender and for mean, standard deviation and range of age (a

man is a man and a standard deviation is a standard deviation quite reliably). The 4 studies which used age bands to report age cut them off differently, with 1 overlap (2 studies had a category of “> 85”). Similarly, the breaking into categories was different for each of the 5 studies which reported marital status. In the 5 categories for education, out of 18 study-level categories (e.g. “did not graduate from high school”; education level “low”, etc.), 2 were identical (the designation “high school”). The most consistent study-level category of ethnicity across studies was “other”. In the 3 studies that reported on socio-economic status, 1 operationalisation was of household finances at the end of the month (with categories such as “just enough to make ends meet”), 1 was of caregiver income in Indian rupees, and 1 of annual household income in US dollars. Details of category contents for non-obvious or location-specific categories and criteria by which boundary cases were ascribed were not reported.

In summary, findings differ formally and thematically in an aspect where they are supposed to differ primarily factually.

## ***6. The multiplicity of findings – illustrations of multiple meanings and uses***

Similarities and differences between findings are obviously a highly problematic issue. Not much of an original claim in the abstract – we could have just as well stopped with Plato’s “Parmenides” – but hopefully the specific illustrations were telling of the nature of research synthesis and scientific research more broadly. Similarities and differences between findings are a highly problematic issue not least because of a finding’s multiplicity. This section will aim to make the latter tangible. Theoretically, the multiplicity of findings is known all too well to anyone who has thought in some depth about language and communication even if not research synthesis. But it seems to slip out of view. Research synthesis studies are in the paradoxical position of being premised on and making use of the multiplicities of findings (if a finding could not have somewhat different meanings to the ones initially intended, most synthesis studies would not be possible), but never facing the implications of this multiplicity beyond the one that has been of use to them. If research synthesis studies were taking the multiplicity of findings seriously, a reversal of the process would have been

discussed much more frequently – starting from primary studies and proactively suggesting what information they can contribute to a range of questions that concern us, rather than starting from a synthesis question and filtering information relative to it only.

In bringing the multiplicity of findings to the limelight, I also make the point of specifying it in an ‘extreme but (largely) realistic way’. By ‘extreme’ I mean that I break up a finding or generate interpretations well beyond what a normal synthesis study would do. In the case when I break up a finding, I do so up to basic findings, where ‘basic finding’ is understood within the framework set by the Vocabulary of Elements of Findings as containing a single expression or a single relationship (see 5.1.2. on starting points). By ‘largely realistic’ I mean that if a sufficient number of similar studies and the technology to enable the data extraction and storage were available, an average synthesis researcher would proceed in very similar ways to the ones outlined (or so I believe).

Three aspects of the multiplicity of a finding were specified:

- **the multiplicity of boundaries, contents and meaning of ‘the same’ finding** – understood as the fact that we can break a research finding meaningfully at a number of different points, and/or combine it meaningfully up to many different points. Formulated differently – the fact that a posited research finding (e.g. coterminous with a normal sentence) can give rise to two or more self-contained units of the right type of relationship to the source finding, where a ‘self-contained unit’ is understood as one that can be expressed in a proposition and ‘the right type of relationship’ is understood as one where the movement from the source finding to the self-contained unit(s) can be justified within the relevant research context.
- **the multiplicity of meanings and uses of ‘the same’ finding** – a multiplicity that is closely related to and partly overlapping with the above multiplicities but also emphasising the fact that some of the meanings and uses of findings cannot be seen as ‘inherently contained’<sup>106</sup> within

---

<sup>106</sup> If we take extreme non-essentialist positions, no meaning or use of a finding is simply inherent within them. However, even from such positions one can accept that in certain contexts the majority of those perceiving a research finding are likely to say that it is a ‘finding about Xi’,

them but come from their relationships with other carriers of knowledge and with user's intentions.

- **the multiplicity of representations of 'the same' finding.**

Multiplicity is thus sometimes seen as primarily a function of the finding (inherent multiplicity) and sometimes as a function of the study, its intended secondary uses and its users (relational multiplicity).

Although I had incorporated the intention to quantify the multiplicity of findings in the study methodology, I could not realise it because of an unaccounted for limitation of the software. Multiply coded findings could not be retrieved unless the software (both NVivo and EPPI-Reviewer) was queried finding by finding and the coding (multiple or singular) for each finding checked. This prevented the quantification, as means, of the multiplicity of boundaries, contents, meanings, uses and representations of 'the same' finding. Instead, its tangible reality is illustrated through two examples and a lower level quantification. In the first example, multiplicity is seen primarily as a user-independent fact of a complex finding being composed of a large number of clearly separable independent elements. In the second, multiplicity is seen primarily as user-dependent affordances for multiple interpretations offered by the same data.<sup>107</sup>

### **6.1. Multiplicity of contents, meanings and uses in a negative finding of 'no difference'**

Take the following example of a multi-component finding:

*Patients who preferred alternative services were no different than others with respect to demographic variables (eg, age, sex, income), functional status (ECOG score, activities of daily living), social support (Medical Outcomes Survey score), quality of life (Functional Assessment of*

---

where  $X_1 = X_2 = \dots = X_i$ , and 'it is a finding that can be used for  $Y_i$ ', where  $Y_1 = Y_2 = \dots = Y_i$ . I treat those meanings as 'inherently contained'.

<sup>107</sup> I do not illustrate the multiplicity of representations of findings as, on subsequent analysis, it was seen as more relevant to processes of selection in research synthesis rather than processes of transformation.



*Cancer Therapy-General G score), or physical or psychological symptom burden (physical and psychological subscales of the GDI, respectively) (Study 1).*

The details of an extreme but (largely) realistic estimate of the number of meanings and uses of this finding can be found in Box B 5.1. This estimate comes up with 12 basic findings from the above sentence, 120 versions of the basic finding (each of the basic findings extended with one of 10 key features of participants and context gleaned from the broader text) and 5 pointers (mentions of the scales used) to methodological findings in the broader text. The first line synthesis studies that can be envisaged are 14 'same topic' studies (with the number so large because the alternative services explored were 6, and because alternative services can be employed not only in palliative care, as in the example study, but in healthcare more generally), 3 'transformed topic' studies where the distance from the original topics is greater and the topics reflect the researcher's (my) interests (e.g. attitudes to practical services with limited human interaction), and 1 broad methodological synthesis study. In total, this makes 2045 units of information (17x120 units of information to be used in thematic syntheses and 5 units of information to be used in a methodological synthesis). The unique units of information are, however, 34 (12 expressions, 10 features of sample and context, 5 methodological pointers and 7 service types).

## Box B 5.1: Multiplicity of contents, meanings and uses in a finding of ‘no difference’

**The finding:** *Patients who preferred alternative services were no different than others with respect to demographic variables (eg, age, sex, income), functional status (ECOG score, activities of daily living), social support (Medical Outcomes Survey score), quality of life (Functional Assessment of Cancer Therapy-General G score), or physical or psychological symptom burden (physical and psychological subscales of the GDI, respectively) (Study 1).*

### **Steps in the estimate of its multiplicity:**

1. Following the principle that a basic finding contains a single expression or a single relationship, we can count nine expressions that are explicit in the sentence and three further expressions to which it alludes (shown in square brackets),<sup>1</sup> i.e. there are **12 extractable basic findings in the ‘bigger finding’**:
  - Age of the group of interest is no different than the criterion
  - Sex of the group of interest is no different than the criterion
  - Income is no different than the criterion
  - [Race/ethnicity] is no different than the criterion
  - [Marital status] is no different than the criterion
  - [Education] is no different than the criterion
  - Functional status as ECOG score is no different than the criterion
  - Functional status as Activities of daily living is no different than the criterion
  - Social support as Medical Outcomes Survey score is no different than the criterion
  - Quality of life as Functional Assessment of Cancer Therapy-General G score is no different than the criterion
  - Physical symptom burden as Physical subscale of GDI is no different than the criterion
  - Psychological symptom burden as Psychological subscale of GDI is no different than the criterion.
2. These are, however, expressions in a particular group, in a particular context. Thus each of the above basic findings has to be ‘tagged’ with each of the key features of participants and context reported in the study.

To keep to the ‘realistic’ in the ‘extreme but (largely) realistic data extraction’, where one does not take every feature of a group and context as relevant, I take as key 7 of the features reported in the table on patient characteristics (all demographic characteristics and the type of cancer) and three of the characteristics of the context described in the

text (country – US, particular setting – oncology clinics, health services programme – Medicare). That is, there are **10 key ‘tags’** to be attached to each of the basic findings. This makes a total of **10x12 = 120 versions of the basic findings**.

3. At the level of knowledge somethings, there are five scales/subscales mentioned in this sentence – ECOG, Medical Outcomes Survey, Functional Assessment of Cancer Therapy-General, Physical subscale of GDI and Psychological symptom burden of GDI. Although no specific information is available from the sentence, there are **5 pointers to methodological findings** in this study.

4. There are **14 ‘close-at-hand’, ‘same topic’ potential research synthesis studies**, if we take the broad topic of this study to be ‘attitudes to alternative services’.

These include two ‘same topic’ studies at the generic level – ‘alternative services’ – on alternative services in palliative care and alternative services in healthcare in general.

The remaining are 2x6 ‘same topic’ studies at the particular level. The alternative services explored in this study are 1) vouchers for practical assistance at home, 2) transportation, 3) peer support, 4) meal delivery, 5) case management and 6) family care. The doubling is because, again, the attitudes to those services can be explored within the context of palliative care and healthcare in general.

Again following the ‘realistic’ in the ‘extreme but (largely) realistic data extraction’, where a researcher’s interests make a difference, I will add **three ‘transformed topic’ potential synthesis studies**, to cover all healthcare fields – on attitudes to practical services with limited human interaction (the examples from this study will be transportation and meal delivery), human interaction-intense services (the examples from this study will be peer support and family care) and ‘formalising the informal services’ (vouchers for practical assistance at home and, possibly, case management). I will also add a topic for a **broad methodological synthesis study** – standardised measures of functioning and experiences used in palliative care research.

Thus, in an extreme but largely realistic scenario 18 synthesis studies were envisaged, i.e. 18 different uses of the information from the ‘starting point finding’ above.

**For the potential thematic syntheses, the example finding will provide 17x120 = units of information** (the estimate of 120 is as per 2. above).

**For the potential methodological synthesis, the example finding will provide 5 units of information** (as per 3. above).

## 6.2. Multiplicity of meanings and uses – multiple affordances for interpretation offered by the same data

The other well known but sometimes vaguely appreciated or sent into infinity multiplicity of findings comes from their potential for multiple theoretical and conceptual re-framings.

Let us take the following extended excerpt (to include both raw data – interview quotes, and interpretations) from Study 16:

*The experience of threats to identity manifest as ruptures in the taken-for-granted assumptions that comprise an individual's social world. This includes bringing to the fore pain, suffering and death which in turn are experienced as threats to identity. An example is when the onset of illness brings to the fore one's own mortality and acute awareness of the ageing process. This is evident when individuals perceive that their illness has affected their appearance and self-image:*

*"I hate myself at the moment. I really hate going out because I look ... oh my skin at the moment ... my face, I'm coming out in spots and I don't like the way I look. I look like a little old lady and I don't feel like one usually ... I've never classed myself as old ... I look at that photo [before cancer diagnosis] and think now that's me ... this [talking about the present] is not me." (P6, I2, unemployed, female)*

*Looking older threatened P6's previous identity and made her think that she was no longer the same person as she was prior to her diagnosis. She referred to a photograph of her former self which served to re-enforce her perception that she was no longer the same person. Her negative perception of her self-image was re-enforced by her GP who had written on a form that she had read that she was a 'frail, underweight female who now looks older than her years'. With reference to Bury's (1991) concept of biographical disruption, cancer took on symbolic significance for P6 because it affected how she saw herself and how others saw her. Further, her experience demonstrates the ways in which others, in this instance, health professionals, can influence the ways in which an individual perceives their identity.*

Taking into account the context of the whole study, I could reconstruct the following main functions and messages of the excerpt above as reflecting the authors' intentions:

- to illustrate and assert the effect of illness on appearance, self-image and awareness of aging – with these effects being such that they constitute an identity threat;
- to assert the potential influence of health professionals on perceptions of identity;
- to assert that the effects of cancer on the interviewee's life are to be taken to mean that cancer took on symbolic significance for her and that it acts as biographical disruption.

Next are some possible re-uses of the same excerpt for synthesis studies – the outcomes of an extreme but (largely) realistic generation of associations:

- the interview quote shows evidence of 'verbal self-harming' (*"I hate myself at the moment", "I don't like the way I look"*); possibly the silences are meaningful in this direction, too - *"I really hate going out because I look ... oh my skin at the moment ... my face"*);
- the description of the consequences of reading the GP notes prompts an interest in *non-biomedical information in patients' records and issues of access to them*;
- that same description may be used as data for a study on *health professionals' words that hurt – for better or worse*;
- more directly, the interview quote provides evidence of the effects of colorectal cancer and respective treatments on appearance, and the centrality of those effects for patients;
- the concern with appearance of this interviewee, coupled with recurring comments in the study of the effect of cancer on social life, raises the

question if the effect of cancer on social life may be mediated in a significant way by appearance concerns;

- elements of the first quote along with the second quote and its framing may be used in a study of illness and crossing the young-old boundary (*“I look like a little old lady and I don’t feel like one usually ... I’ve never classed myself as old ...”*; *Her negative perception of her self-image was re-enforced by her GP who had written on a form that she had read that she was a ‘frail, underweight female who now looks older than her years’*);
- This excerpt provides data for an exploration of photos as an identity anchor point.

I will not be quantifying or evaluating the multiplicity of this research finding. My point has been to illustrate another aspect of the multiplicity of findings and provide some boundaries to it as a reference point – much broader than the ones currently applied in research synthesis but not unthinkable. If I were the researcher extracting data from this excerpt, this would be the point before which I would feel there is more that is interesting in this finding and beyond which I would consider I am “data dredging”. I will come back to the issue of multiplicity in Section 8, in discussing the application of the injunction for multiple coding.

## **7. Omissions, silences and half-silences**

The relationship between missing and available information was another determinant I posited of the capacity of a finding for meaningful transformations. I use ‘missing’ or ‘omissions, silences and half-silences’ to stand for anything that is ‘not (sufficiently) there’, in a piece of finding or in the broader study as relevant to this finding – missing altogether or intentionally or unintentionally left vague. Omissions, silences and half-silences are sometimes only a function of the study, in the sense that most of its competent and careful readers will consider certain additional information as highly relevant. Sometimes they are a

function of the study and its intended secondary uses: that is, most of the competent researchers working to some synthesis specifications will require certain additional information, even if it was not an omission in the terms of the primary study. Finally, they can be a function of any of the above and the sensitivities of a competent user: different types of users, even if competent, will be responsive to different omissions, silences and half-silences. If the original study failed to provide a needed element or provided an imperfectly specified one, a transformation could help in 'creating' the missing, or a substitute for it, from other information available in the research report. The 'missing' was conceptualised as a negative determinant of transformations, too – it would block the possibility for certain transformation as the acceptable directions would be insufficiently determined.

A range of codes, tagging documents and tags were used to explore what was not there. Just as with people, there are so many ways for a piece of information to be missing and to be missed. The section on supposedly highly standardised types of information has shown some of these. Further to these, there are often discrepancies between intentions and outcomes, e.g. a study sets out to explore certain associations but does not report on them. There are "deletions" of the context of production of a finding. There are missing elements that are automatically supplied by the reader and as such are not really missing in an extended understanding of what makes a text (such as objective descriptions of gruesome experiences prompting conclusions about the importance of the research). There are silences behind associations with standard parameters which health researchers and readers appear to have accepted comfortably. For instance, associations between outcomes and time elapsed are frequently sought with no speculation offered of what may have filled this time to produce change or how something 'dissipated' in the course of time so as to make a difference to outcomes.<sup>108</sup> There are half-silent findings on relationships – the numbers are laid out and the reader has to make and interpret the comparison.

I will report in greater detail findings concerning omissions, silences and half-silences from an analysis supporting the application of the Vocabulary of Elements of Findings. As the formalisation of findings required a 'perfect'

---

<sup>108</sup> The same applies often to sex, age, etc.

understanding of their meaning, I took note of my uncertainties of that meaning, perceptions of missing information, and 'wishes' for a further explanation. The process was not strictly methodologised – I would record an issue if it arose spontaneously. The results provide only a rough orientation of the frequency of missing information at the level of findings. Maybe more importantly, they demonstrate how sensitivity to missing information is gradually lost with the advancement of a study, if left to arise spontaneously.

Four main types of omissions, silences and half-silences were distinguished by grouping cases encountered in the process of applying the Vocabulary of Elements of Findings. A fifth group was also formed, but I attributed my sense of missing information to unfamiliarity with the topic. Findings from this group are treated as irrelevant here as I was not fulfilling the requirement for being a competent user outlined at the start of this section.

The first of the four groups was of lack of clarity of the linguistic expression when this had implications for understanding the research. ("Economy of language" omissions or lack of clarity which appeared simply as clumsy sentences were not considered relevant.)

For instance, the following sentence from Study 4:

*Much of the intervention effort was directed towards improving provider skills rather than patient outcomes (9.8% of studies)*

leaves a hesitation if the 9.8% refers to the percentage of studies on improving provider skills or those on patient outcomes.

The second type of omissions, silences and half-silences covered cases of unreported research knowledge whose presence may be expected in such a context and which may be missing for three main reasons: 1) it is unremarkable and/or has been considered dispensable, in a balancing act between brevity, clarity and comprehensiveness; 2) relevant data have not been collected; and



3) it undermines something of the (strength of) claims or procedures of the study. In the first case, the omission is justified, or at least understandable. In the latter two cases, credibility concerns arise.

For instance, the following sentence from Study 12 (or any of the sentences in the text)

*Sample size in subsequent assessments was 125 at T2, 123 at T3, and 113 at T4 (attrition 20.1%, total number of assessments = 504)*

does not report explored or supposed reasons for attrition in a situation where it is not minimal but covers over 1/5 of the original sample.

Third, we have cases where the expected reporting is complete but something potentially unusual or 'disturbing' about it, which may be commented on, does not receive any attention.

For instance, the fact that the confidence intervals in two of the cases below include 1 weakens the claims made but this is not commented upon:<sup>109</sup>

*In men, there was an inverse association between high weight gain (versus stable weight) and all cancers combined (HR 0.76; 95% CI 0.57–1.02) (Table 2) which appeared to be driven by a strong inverse association between high weight gain and prostate cancer (HR for high weight gain and prostate cancer 0.43; 95% CI 0.24–0.76; HR for high weight gain and all cancers except prostate 1.00; 95% CI 0.71–1.40) (Study 10).*

A fourth category was added – 'silencing through sentence structure and formulation' – to include cases where the phrasing of the finding is such that certain elements of it receive a much greater importance than others or that the finding as a whole begins to appear formulated in a biased way.

---

<sup>109</sup> If the confidence interval for hazard ratio (HR) includes 1, the result is not statistically significant.

For instance, the exception below receives much less attention than the general tendency:

*They also reported higher HRQL scores compared with those with reduced work hours or who have stopped working (all p-values < 0.0001; except spiritual well-being  $p = 0.61$ ) (Study 5).*

Or the phrasing of

*However, not all patients preferred alternative services (Study 1)*

sounds somewhat strong for a case when the 'not all' covers 42% of participants (20% not having a clear preference and 22% having the opposite preference).<sup>110</sup>

A total of 45 omissions, silences and half-silences were recorded (15; 18; 10; 2 for each of the four groups). With a single exception, these were in the first two-thirds of the findings (up to 199 from 301). In view of the fact that the drive for registering an omission or silence was my spontaneously felt need for clarification/ more information, such a distribution suggests that one gradually loses sensitivity to missing information and the motivation to seek for more. In the set of findings where I actively recorded omissions, I felt that 22.1% of the findings were missing something. How important this is for a particular synthesis study is unclear.

## **8. Tools for enabling and controlling transformations**

The last factor affecting the capacity of a finding for meaningful transformations I will address is the effectiveness of the tools used to enable or control transformations. In this study, these were the injunctions underpinning the extraction-coding: for comprehensiveness, extensive multiple coding, extreme transparency, double-edged critical analysis, and for coding as close as possible to the original first and experimenting afterwards. I will start by

---

<sup>110</sup> There is a sense in which the 3<sup>rd</sup> and the 4<sup>th</sup> categories appear very similar. The distinction I am making is that one would not expect a reformulation of the sentence in 3 to give more voice to the silent but a separate explanation, while in the case of 4 a more balanced representation requires a reformulated sentence and not a comment to that sentence. Category 4 was part of a 'rhetorical category' which went beyond omissions and silences.

presenting overall impressions concerning the effectiveness of the injunctions. These are, indeed, impressions based on limited operationalisations. The expansion of this analysis is a matter of much further work. I will then provide further observations on the extensive multiple coding and, jointly, on transparency and double-edged critical analysis. Brief information on the application of the comprehensiveness injunction was already given in 3.1. The limiting effects of staying as close as possible to the original text can be seen in the discussion of thematic differences between findings in 5.4.

### **8.1. Expectations and actualities**

At the outset of the study I had conflicting expectations whether the injunctions for comprehensiveness, double-edged critical analysis and transparency would enable or constrain transformations. For instance, the comprehensiveness and double-edged critical analysis appeared likely, on the one hand, to 'open up' a finding – by providing alternative perspectives towards it (the critique part of the double-edged critical analysis) and by contextualising it better and thus showing its richness (the comprehensiveness injunction). In both cases a finding's capacity for transformation would increase. On the other hand, it was also plausible that the application of the injunctions for comprehensiveness and double-edged critical analysis would demonstrate the fixedness of a finding within a certain context. This would decrease a finding's capacity for transformations. In turn, extreme transparency was a way to capture fleeting associations which, once pinned down and developed, might trigger further transformations. But extensive transparency could also highlight biased directions of interpretation and thus be a delimiter of intended transformations. It was an empirical question which of those two types of potential would be realised or prevail. The injunction for coding as close as possible to the original first and experimenting afterwards was seen as necessarily dual – both controlling and enabling transformations. As for the injunction for multiple coding, it could only enable transformations. It was, however, unclear to what extent. It is one thing to code multiply an isolated finding, as was the case in Section 6, another to do it consistently and relative to a particular synthesis question.

Transformations of findings in the test synthesis were certainly more than in a typical extraction-coding. As a whole, the injunctions then served as transformation-enabling rather than transformation-constraining tools. My perception was, however, that this was entirely due to the multiple coding injunction and the greater number of pieces of information rather than to any of the other injunctions – e.g. seeing more of the richness of a finding through a critique of it or through having more information on the matrices in which it was generated.<sup>111</sup> I did not experience the injunctions for transparency, double-edged critical analysis and comprehensiveness as having the opposite effect either – of constraining already intended transformations. As constructed and applied here (and this was in a minimal way), these three tools made no substantial difference to the capacity of a finding to transform.

Those issues are taken up again in the Discussion. I will now provide further detail of the application of the injunction for multiple coding, which was the only one clearly responsible for the increased number of transformations.

## **8.2. Observations on the application of extensive multiple coding**

The capacities of the software packages did not permit an estimate of the average number of times the same unit of information was coded. My informal impressions were that after the fourth recoding of the same piece of information I would begin to consider the returns very limited and move to the next piece of information, and that I would code a substantial number of pieces of information only once, regardless of the injunction (discussed below). Much of the multiple coding followed formal categories – e.g. ‘strength of findings’, ‘valence of finding’ (is it positive or negative relative to what is ‘good’ for patients), or semi-formal/ semi-thematic categories – e.g. ‘something common and unpleasant, not fully understood, not effectively managed, etc.’ Some of the multiple coding arose from the fact that the higher level categories did not have clear boundaries or tended to mix quite often (e.g. the broad thematic category of ‘findings about phenomena’ often crosses that of ‘findings about relationships’).

---

<sup>111</sup> The comprehensiveness injunction thus worked in an enabling manner in a superficial way – simply providing more information to be coded. In contrast, I expected that if it enables transformations, it will enable them by providing information that contextualises findings better and thus opening up a richness of associations for them.

The above two types of multiple coding were generally unproblematic provided appropriate operationalisations, but carrying them out could not be effectively sustained. First, they are repetitive and mechanistic and thus burdensome. More importantly, relevant formal and semi-formal/ semi-thematic categories grew to a substantial number and only a reverse process (of the researcher being queried how a certain piece of information stands relative to a category rather than proactively ascribing it) would have ensured completeness of coding.

Much clearly available multiple coding was not performed as it involved simple reversal or extrapolation of information (e.g. 'accepted to participate' is the opposite to 'declined to participate', or 'reasons given for non-participation' means that attempts at discussing reasons, which had a category of their own, have been made). Clearly available multiple coding was also not performed because the whole high level category could be consistently treated as providing information on another high level category (e.g. the findings, which go under Findings, are also findings about the variable and should thus go invariably under Variable, too).

The spontaneous lower level thematic multiple coding was more limited than I had expected. In some cases, I experienced findings as fully fixed. For instance, I would not spontaneously code the finding *Of the 34 intrusive cognitions, 6 lasted only seconds, 23 lasted minutes, and 5 lasted hours* into something different to 'frequency, level, duration, dynamics'. Many findings appeared to have quite a limited capacity for thematic transformations, maybe because the source study provided a very strong framing that was difficult to escape in the process of extraction-coding. Overall, I experienced a strong tendency to slip back into singular thematic coding. The framing of a paper leads you very effectively once left to lead you. Multiple coding requires an intensity of critical and creative thinking – a constant need to overcome the grip of the source study – which is hard to sustain.

Nevertheless, when the internal recognition of the need to code more than a handful of studies was temporarily suspended, the multiplicity illustrated in

Section 6. 2 above would easily come forth.<sup>112</sup> The grip of a study seemed to relax once the mental framing of having to extract all data from all sampled studies was relaxed. The affordances for multiple thematic coding thus appear more limited by the context of data extraction than by the study matrix. The multiplicity of findings and papers opens up through an injunction for multiple coding, but the process has to be set up and implemented quite differently, as I will argue in the Discussion. Otherwise we will always come out defeated – by limits to human attention and memory, a sliding slope of creativity and individual limits to the scope and directions of that creativity, boredom in the face of repetition, and a tendency to automate within a narrow range of actions.

### **8.3. Observations on the application of the injunctions for transparency and double-edged critical analysis**

In introducing the concept of ‘tagging’ in Chapter 3, I presented what were largely ‘later ideas’ – as developed from observations of how the injunctions for extreme transparency and double-edged critical analysis worked in practice. These two injunctions are highly intertwined and the processes that realise them are often inseparable. Critical analysis tends to be presented transparently as a series of steps, rather than only as an outcome in the form of a critical conclusion. If not, the critical conclusion has the sound of ‘I don’t like it’ and ‘I like it better that way’ (or ‘I don’t like you and your ways’). Transparency, in turn, can reveal inconsistencies and paths not taken, which gives rise to critical analysis. As described in reference to tagging, the application of the injunctions for double-edged critical analysis and extreme transparency took the following forms:

- critical appraisal comments addressing the methodological rigour of studies;
- ‘critiquing comments’ addressing its foundations;
- meta-critical comments on the critical appraisal and critiquing comments.

This type of tag showed that double-edged critical analysis did not have

---

<sup>112</sup> Even if one would still not code multiply the above finding, a range of uses can be envisaged (typical duration of an intrusive cognition, contents of shorter and longer intrusive cognitions, personality characteristics associated with these, etc.).

enough of an edge and, ideally, should be ‘triple-edged’ critical analysis;<sup>113</sup>

- ‘research process’ tags – notes which registered the development and origins of the ideas and conclusions of the study, both the test synthesis and the broader thesis. These could be notes on, for instance, preliminary observations, operationalisations, change of response to the material, pragmatic issues, e.g. behaviour of the software, etc.;
- ‘self-monitoring for rigour’ tags – observations on problematic processes and decisions and actions implicated in them. These were to serve as an ‘audit trail’ for problematic decisions but also aimed to help explore the source of difficulties and, as a result, possibly enable refinement of problematic processes;
- ‘self-monitoring for tendencies of thought’ tags – notes that aimed to capture broader interests, concerns and proclivities of mine which were triggered by the material analysed and which are likely to be shaping my interpretation and uses of it.

The processes and outcomes of applying the injunctions for double-edged critical analysis and extreme transparency remained in a broadly rationalist framework – for instance, overall, I was not recording ‘irrelevant thoughts’, such as on whether I was hungry, anxious, blissfully working or heart-broken. I would, however, record occasionally if a process has started feeling boring or burdensome. This was seen as relevant to the quality with which it was performed and its potential for becoming a normative requirement (no irrevocably boring processes were to be admitted as a methodological step).

My observations on the effectiveness of the injunctions for extreme transparency and double-edged critical analysis are again, as with multiple coding, impressionistic rather than structured. I coded the study log in NVivo and thus performed a formal analysis, too. However, a truly fruitful analysis of these observations needs further operationalisation and theoretical backing from debates on critical thinking, introspection and self-reporting. I offer these observations in the spirit of thinking that some evidence is better than no evidence. Also, although generally unsurprising, they negated strong

---

<sup>113</sup> Critical appraisal and critique were taken to comprise ‘double-edged critical analysis’. Self-reflexivity on these could be seen as extending it into a ‘triple-edged critical analysis’.

expectations with which the study started – for instance, that the transparency of current synthesis studies can be markedly improved within the existing overall model of research synthesis. I had high expectations of identifying a point of greater rigour and heuristic potential somewhere between current practices and the extreme transparency I was testing out.

As far as the effectiveness of applying the injunction for double-edged critical analysis is concerned, it became evident that the critiquing element of it needs to be supported by a much more tightly specified analytical structure than general ideas from the meta-scientific literature, by much more information external to the study report and, ideally, by expert support.

As far as the effectiveness of applying the injunction for extreme transparency is concerned, the following main impressions crystallised. As expected, many of the transformations were swift, automatic and subliminal and their drivers were hard to trace. Trying to capture transformations in progress and to elicit their causes was sometimes effective but not to the extent that would allow me to claim that current practices are failing to use a readily available potential. I felt that all too often transformations would happen while I was preparing to register them rather than when I was ready to do so. As if I would blink just when my eyes had to be wide open. Similarly, eliciting drivers of transformations seemed to hit walls of rationalisation and folk theories of thinking more often than I had hoped. Next, I found much transparency to be painful, and not only because it was tedious. I was tempted to rationalise suboptimal decisions more frequently than I had expected. Somewhere in a vague experience of disquietude I could sense that I was withholding relevant information but it would take effort to see the shape of what I was withholding. An expectation of transparency forces one to write down the shortcuts one has used, to do the whole work properly, or play a game of transparency. The first two options are costly. It is understandable why a substantial part of our transparency is a game of transparency, even without us being fully conscious of it. And it is too much to expect that science and research will move to a completely different standard of honesty in reporting. It will be ugly, as the social studies of science have seen well. Finally, a requirement for extreme transparency questions the boundary between relevant and irrelevant factors in research. Maybe it is, after all, relevant if one is hungry, anxious, blissfully working or heart-broken. But I doubt we would want



research reports to tell us all this although we may be interested in psychology of science research on such issues.

With the benefit of hindsight, these findings about the limits of transparency could not have been more self-evident. And yet they are not part of the discourse on transparency in research synthesis or the critical analysis of our transparency practices. In the Discussion section, I expand the analysis of where our current synthesis methods stand in terms of transparency and try to move beyond the self-evident – and pessimistic – conclusions following from the findings presented here.

## Chapter 6: Discussion

### **1. Structure of the chapter**

This is the final chapter of the thesis. Section 2 summarises the intent, approach and main findings of the work. Section 3 highlights its contributions to knowledge. In Section 4, these contributions are located within the debates on evidence in the philosophy of evidence-based medicine, Big Data, and quantitative and qualitative research. Section 5 re-examines the key concepts of the work – evidence and transformations. Section 6 lays out the study limitations. Finally, Section 7 channels findings about challenges to the process of rigorous and heuristically promising transformations into a radical vision for the future of research synthesis. I claim that if there is a normative ideal of research synthesis we need to aspire to and compare our current efforts with, it is very similar to that vision. I am, however, pessimistic of the likelihood of its approximation in practice.

### **2. Summary of the intent, approach and findings of the thesis**

#### **2.1. Intent and approach**

This work aimed to probe the extent to which we can place trust in and improve on one of the major knowledge technologies of health research – a family of disparate and often non-communicating methods and study designs here referred to as methods of research synthesis. This technology was taken as demanding careful and urgent philosophical attention as its outcomes are increasingly incorporated in clinical practice in a culture of evidence-based care and, at least theoretically, it produces the only fully legitimate type of research knowledge underpinning health policy decisions. When evidence is used in making evidence-based decisions and recommendations, at least in a health context it is assumed to be synthesised evidence whenever there is sufficient evidence to synthesise. I saw current philosophical debates as placing disproportionate emphasis on criticising preferred types of evidence (RCT evidence) and arguing for the value of ‘other’ types of knowledge (e.g. observational evidence, mechanistic evidence, causal claims, clinical wisdom),

while being oblivious to the variety of approaches to and complex procedures of evidence synthesis.

I claimed that 'transformations of evidence' are the element of research synthesis methods which has to bear the heaviest burden of proof of reliability, if the trustworthiness of the outcomes of such methods was to be established solidly. The very existence of this element of health research synthesis methods has not been clearly acknowledged in the literature although, subliminally, the formal methodological debate and informal discussions in the community constantly refer to various instantiations of it. An argument by way of illustrations was made that such processes of transformation are 'very real'. Transformations for the purposes of research synthesis were conceptualised minimally as driven by differences in evidence, provided some right type of similarity was there, too, and having as an outcome a right type of reduction of those differences or, from the opposite perspective, an increased combinability or compatibility of pieces of evidence.

The aim of the thesis thus became to describe how transformations of evidence happen and what drives them, and to do so through close attention to the first major stage of a research synthesis process where transformations are performed – that of 'data extraction'.

The approach I took combined two main features.

First, ideas from the 'meta-scientific' literature, mainly about factors that make evidence what it is and potentially fix it to a particular context, were brought to bear on observations of the process of data extraction. 'Meta-scientific' was understood to include the literature from fields external to science that explore science, such as the philosophy and social studies of science (meta-scientific literature in the narrow sense), and the critical methodological and theoretical literature internal to health research (referred to as 'meta-methodological' literature if a distinction was needed). I found most of the ideas from the meta-scientific literature to be more abstract than I had expected, to have been developed primarily outside health research contexts, and/or to have been operationalised in ways that were not relevant to this work. I added to those ideas my own experiential knowledge of health research and turned them into

prompts and parameters for analysis of the contents of research papers and the process of extraction of information from such papers.

Secondly, the approach aimed to apply the self-professed virtues of mainstream research synthesis (such as transparency, exhaustiveness, critical analysis, rigour) in ways and to aspects of the process of data extraction that have not, so far, been covered by those virtues. Exploring ways of enhancing transparency was a primary concern.

Put differently, the attempt to give a descriptive account of the processes and drivers of transformation of evidence in health research synthesis studies was, on the one hand, an experiment of whether this can be framed by bringing together and operationalising disparate ideas from the meta-scientific fields. On the other hand, it was a test of how faithfully the virtues of research synthesis are applied in current synthesis work.

A less explicit red thread running through this work may be worth mentioning. It was of attention to the mundane and trivial, to what looks too minor to matter. The thesis aimed to draw attention to the automatic, highly repetitive and/or low level decisions taken in research. These have no glamour, no drama and thus tend to evade critical analysis while at the same time having, potentially, a profound influence on research outcomes.

## **2.2. Summary of findings**

The novelty of this thesis was in its methods and not only in its data and claims. Three main types of findings will thus be summarised: methodology-relevant findings concerning the feasibility and effectiveness of the overall approach and specific steps within it; substantive findings about questions of transformation and types and features of evidence; and findings which are simultaneously about method and topic.

### *2.2.1. Findings on the feasibility and effectiveness of methods*

At its most general, the methodological work in this study was about developing and applying a reliable framework for the analysis of transformations and types

of evidence, well grounded in the meta-scientific literature. I found the realisation of this intent problematic. There was less complementarity, cross-referencing and/or mirroring between the meta-scientific and meta-methodological literature and between niches within each of these literatures than I had envisaged. It was an unstable assumption that the studies of science and the critical discussion in health research come together into a coherent enough meta-scientific field on the grounds of their meta-position to actual scientific practice. Debates are disconnected. They are also prone to excessive simplification of arguments upon which they build castles or whose supposed castles they seek to demolish (see 3.1. below). Concepts are used quite differently between the meta-scientific and meta-methodological debates (e.g. relevance, transparency, quality) and key concepts are not shared (e.g. propositions, observations, properties, coding, homogeneity/ heterogeneity). Philosophical debates that are *prima facie* relevant to the field of research synthesis and evidence-based medicine – e.g. on unity and disunity of science, interdisciplinary work, total and optimal evidence, quality in science, classification, pluralism, etc. – have neither been noticed by it, nor have noticed it. This much more limited than expected level of complementarity, cross-referencing and/or mirroring of debates was observed even between fields that are much closer and share many more researchers than philosophy and health research synthesis. For instance, issues of data and study integration, as discussed in the mixed methods literature, do not seem to have found a way into the research synthesis literature.

Secondly, I found certain meta-scientific debates I had posited as central to the thesis little articulated and associated with limited community agreement. For instance, I expected that a more structured discussion of types of data, findings, evidence, etc. would be readily identifiable. It was not. I thus formulated more contentious hypotheses concerning similarities and differences between evidence than befitting a 'reliable' and 'well grounded' framework of analysis. Similarly, I expected that there would be a number of well known models of the structure and composition of pieces of evidence in the philosophical literature. Again, this was not the case and led to the development of the somewhat unconventional Vocabulary of Elements of Findings.

Third, the most direct operationalisations of certain abstract claims of long history and substantial intuitive plausibility fell apart very easily (e.g. of the differences between quantitative and qualitative findings). Much more contrived ways had to be sought to preserve the thrust of such claims rather than dismiss them as unsubstantiated. As a result of such features of the meta-scientific debate, the reliability and good grounding in the meta-scientific literature of the analytic approach is debatable (see also Section 6 on limitations).

### *2.2.2. Main substantive findings from the perspective of transformations*

Substantively, one of the key sets of findings from this work concerned the frequency with which transformations are demanded in research synthesis studies.

An approach of estimating the number of units of information coded in the lowest level NVivo codes showed that, in their original or minimally transformed form, health research findings on broadly the same topic are highly individual thematically – 50.7% of the used codes had a single unit of information in them, 30.5% of codes had two or three units of information, and 18.8% had four or more units of information. This means that at least in the case of largely heterogeneous studies by far the greater majority of findings – over four fifths of those obtained here – needs some level of transformation so that sufficiently high level, aggregated findings can be obtained. This is a finding about a sample of largely heterogeneous studies. It cannot be generalised to collections of homogeneous studies. Nonetheless it has important implications. It is precisely such pools of studies that are used to answer ‘big and complex questions’ that matter the most to individuals and society.

At what may be thought of as the opposite end of a continuum of expected variety – namely, supposedly highly standardised parameters such as cancer-related and socio-demographic information, an extensive variability across studies was found, too. While in some cases the measurement unit and subcategories were by and large fully predictable (such as age and gender), higher level categories often broke down into numerous lower level categories

with no or minimal overlap (as in the cases of cancer phase, education, ethnicity, etc.).

It may be argued that such findings are an artefact of the small sample of studies. Indeed, it seems inevitable that in a larger sample a crystallisation of standardised categories will be observed, though probably within a broader range of standards than we are inclined to assume.<sup>114</sup> Even if this is found to be the case though, the variability of particular instantiations of those standards across studies will leave studies being, ultimately, one of a kind. To start with one of the most consistently reported and most invariable in terms of differences in nomenclature categories observed here, cancer type, there are over 200 types of cancer (Cancer Research, 2011). A minimal number of sub-categories in all other standardised categories, of which only some were explored here, is enough to keep each study like no-other. This means that even the supposedly non-interpretive research syntheses, in which the source studies are as similar as possible, have to transform a broad range of low level parameters so that any aggregation of findings may be achieved. Some higher level of standardisation of categories may be beneficial. Appeals for greater standardisation of the reporting in health research are frequent and frustrated. But too strong a move in that direction will hide or suppress a state of affairs that deserves recognition – of hypotheses of associative and causal relationships that vary immensely across contexts and of a large degree of arbitrariness in setting cut-offs.

Another key set of substantive findings from this study concerned the reliability and validity of transformations. Transformations are performed on limited information as even extremely thorough data extraction ends up with much missing data. The study also illustrated that the affordances of pieces of evidence, in terms of multiplicity of contents, meanings and uses, are much more extensive than what is used. This is not only a matter of missed opportunity but of rigour, too – the preference for one transformation amounts to suppressing alternative transformations.

---

<sup>114</sup> There must be a limit to the difference-making ways in which you can be married or not, educated or not, even if it is not reached too easily ... Or maybe not.

Some of the findings underpinning such claims came again from the analysis of standardised parameters. Levels of missing information were much higher than expected. For example, cancer type was the most consistently reported cancer-related parameter and yet it too was not always mentioned (being unimportant to the concerns of a particular study). Data on the remaining cancer-related parameters considered (such as cancer phase, time since diagnosis, treatment phase, etc.) were reported in between 16.7% and 52.9% of analysed studies. Perhaps most importantly, the criteria by which cases were ascribed to categories or the details of a category content were practically not reported (with 1 study out of 17 noting the classification system used to determine cancer type, and with cancer type being 1 out of 6 parameters traced). Results were similar in the case of socio-demographic parameters. We have good reasons to believe that the completeness of reporting of RCTs, the material for traditional systematic reviews, will be higher – much effort has gone into developing reporting guidelines such as the CONSORT statement ([www.consort-statement.org](http://www.consort-statement.org)). It is, however, over-optimistic to expect that missing information will be a problem entirely dealt with in this type of studies and, from there, in their syntheses.

A range of further findings supported claims that transformations in research synthesis are performed on uncertain grounds more often than fitting a truly rigorous process. Information on matrices available in research papers was limited. Omissions, silences and half-silences at the level of individual propositions were found to be too many. Such were registered in over a fifth of the cases (44 out of approximately 200) where focused attention was given to a sense of *I wish that more was said!* What also happens, however, is that we forget to wish. For instance, when acute awareness receded into the background, omissions, silences and half-silences receded, too (only one was registered in a sample of 102 findings once internal expectations were adjusted). Sensitivity to what we would want to know in optimal research reporting conditions fades away quickly. It cannot be otherwise. The current main format of study reporting, research papers, cannot be expected to provide us with all this knowledge and we soon remember so.



As far as the multiplicity of contents, meanings and uses of pieces of evidence is concerned, there is too much, too varied information in at least some pieces of evidence. This information is not used proactively. Multiple coding is the exception rather than the rule in current research synthesis. Also, approaching a paper with an aim has no alternative in the direction of simply extracting information from papers for re-use. There are thus at least two ways in which current practices are suppressing or ignoring alternative transformations, with the implications this has for constraining the rigour of processes and underusing research knowledge. That said, the process of multiple coding also showed that some findings are best treated as unidimensional. The multiplicity of the complex, rich, clearly 'multiple' findings was not inexhaustible either. Such observations can be seen as having the potential to give grounding to criteria of reliability and validity of transformations. It did not appear as if we can go in any direction, indefinitely. A unidimensionality and finiteness, reassuring in terms of the possibility of trust in the research synthesis project, with all its transformations, could also be perceived.

The last set of key substantive findings from this work related to hypotheses about productive/ non-productive or impossible transformations. No evidence was found of fundamental, insurmountable obstacles in any of the directions considered. In contrast to certain well entrenched assumptions, health research findings appeared to be highly similar in terms of generic composition and structure. This is what the development and application of the Vocabulary of Elements of Findings suggested. Its elements were consistent across the different 'paradigms', disciplines, research fields, methods, functions, etc. from which findings for its development and testing were sampled (over 300). The tests for less noticeable differences of composition and structure between quantitative and qualitative findings were largely negative, too. Method effects of how the procedure for developing the Vocabulary was conceived and implemented are likely to have contributed to the observed degree of uniformity of findings. Also, the question of typological similarities/ differences between findings is inexhaustible. The approaches taken here and hypotheses tested are only a fraction of what can be tried out. Nevertheless, such empirical results

weaken claims about typologically-running incompatibilities of findings. There do not seem to be directions of transformation denied in principle.

### *2.2.3. Findings throwing light on both methods and substantive issues*

A set of findings that were both substantively and methodologically relevant concerned the applications of the injunctions for comprehensiveness, transparency, double-edged critical analysis and multiple coding. Substantively, I did not perceive the injunctions for comprehensiveness, transparency and double-edged critical analysis as having an effect of constraining or enabling transformations. The value of all injunctions was found to be ambiguous. They worked. More of the richness and complexity of studies came forth in comparison to any standard process of data extraction. But they were also found to be too unreliable and/or too burdensome, while providing no clear benefit in terms of rigorous facilitation or constraint on transformations. Metaphorically, the sleeping beauty of studies could not be fully awakened by injunctions, although eyelashes fluttered. The Prince should come with an appropriate infrastructure, too. From a methodological point of view, this was a further case where the study approach turned out to be less effective than anticipated. This is discussed again in the Limitations section.

## **3. Main contributions to knowledge**

### **3.1. Opening up questions and linking disconnected debates**

This work generated a range of specific findings concerning the frequency and drivers of transformations of evidence. But its main contribution is most likely elsewhere. It drew attention to and in certain respects constituted the field of research synthesis as an intriguing object for philosophical and broader meta-scientific investigation. It can, of course, be argued that this object is already receiving intense attention from the philosophy of science and other meta-scientific fields. After all, the nature and value of RCT evidence on the one hand and of alternative forms of evidence on the other can be seen as the two defining topics of the philosophy of evidence-based medicine. As much as it may seem like a minor shift in emphasis, from evidence (that tends to be

synthesised) to processes of evidence synthesis, the change in “puzzles” of interest is far reaching. I see this change as highly beneficial, not least because extensive meta-scientific research has already accumulated on ‘isolated’ medical evidence and that on the integration of medical evidence is close to non-existent. I take the view that if the philosophy of evidence-based medicine re-oriented itself towards issues of bringing evidence together, it would become much more relevant, influential, lively and engaging. As far as the philosophy of science more broadly and other meta-scientific fields are concerned, a clear structuring of a niche on bringing evidence together seems likely to open surprising conundrums and directions for investigation. This work tried to contribute to a re-thinking of the research priorities in the philosophy of evidence-based medicine and to provide some pointers in a broader meta-scientific subfield on information integration in an era of information overload.

The development of modern practices of health research synthesis is closely associated with the rise of availability of information and information processing technologies. It is still a relatively recent development (beginning in the late 80s – early 90s).<sup>115</sup> Constituted in the way proposed in this work, the field of research synthesis is non-existent not only as an object of philosophical attention but as a recognised research field, too. The similarities and connections between what have been suggested as its constitutive areas may appear obvious once drawn. Yet links between most of these areas are tenuous. Mutual awareness is vague, if at all there. Such a disconnection, although at times beneficial for methodological innovation, seems to be already functioning to the detriment of exchanges of good methodological practices, mutual theoretical enrichment and joint effort. If the proposal of this work for a broader construction of the field of research synthesis is taken as valuable, there may be immense opportunities for the meta-sciences to contribute to the self-reflective formation of a research field, the setting of its foundations and clarification of its terminology. If not, the meta-scientific work of the future has another interesting case study for investigating the principles by which science ‘lives’ – by observing how fields connected by parallel concerns, concepts and

---

<sup>115</sup> If we are looking for a specific date, we can choose between 1992, the year of the announcement of the evidence-based medicine movement, or 1988, the year in which Noblit and Hare published their booklet on meta-ethnography.

practices, yet disconnected in terms of actual exchange of knowledge, end up developing and gaining or losing prominence.

In spite of this state of pre-formation of the field as a whole, specific subfields, primarily the mainstream systematic review inclusive of meta-analysis, have come to have enormous power in clinical practice and health policy. If a decision about health policy is claimed to be evidence-based, and nowadays they are expected to be, the evidence which forms that base should be from a synthesis study or synthesis studies, provided more than one study is available. If this is not the case, the term is used without the principles of evidence-based medicine having been applied. The policy drive has also resulted in a growing impact of types of syntheses using non-RCT evidence. In this context, a high quality and highly relevant philosophical and broader meta-scientific debate will matter directly to the health and well-being of people and to processes in society.

In addition to the overarching appeal for philosophical attention to (health) research synthesis, this work proposed a range of specific ground-level questions arising from such attention. The topics of transformations, units of analysis/synthesis, and types of evidence were given priority, though with the caveat that an area cannot settle its priorities well before sufficient substantive work has accumulated. I also tried to bring to the fore a range of other issues that call for critical analysis – the debates on the characteristic features of research synthesis, the terminological variety, the numerous lower level decisions taken in research synthesis, etc. Such issues, apparently peripheral to the core interests of the work, fell into its visual field as they had some part to play in the foundations required to handle the core concerns. They were addressed to the extent of reaching some acceptably stable starting point and then closed off. An initial framing was nevertheless given for attending to a range of conceptually interesting and complex issues.

Similarly, links were drawn between a range of disconnected areas within the meta-scientific and health research synthesis fields. Attempts were made to identify, bring together or juxtapose debates that have a high level of overlap of claims and concerns but tend to be isolated (even within the same broad field), dramatically pitched against one another, or precarious in their own terms, and

thus difficult to analyse and present jointly. The aim was to draw a more comprehensive picture of research synthesis, but also to point to directions for enriching or challenging those debates through relevant counterpart debates. For instance, jointly considered and represented were the discourse of mainstream research synthesis and alternative methods; methodological handbooks, critical work and actual synthesis studies; theoretical literature in health research, philosophy of evidence-based medicine, general philosophy where philosophy of medicine was not enough, ideas from the social studies of science, elements of psychology of science debates; accounts based on avowed positivist, constructivist, realist, etc. positions.

It seemed that most frequently those debates were not, as charitable assumptions of inter-field exchange would have it, complementary, 'the same' but adapted to the needs of the particular field, or oppositional through a 'fair play'. Too often they misrepresented their source or critical target, primarily through crude simplification or lack of awareness of recent relevant developments (of a few decades). For instance, as pointed in the literature review, the 'paradigms debate' in the philosophy of science internal to health research and to applied social sciences research is far from a good up-to-date simplified version of relevant debates in the mainstream studies of science. Maybe we can ignore that it is based on a conceptualisation of 'paradigm' that is too far from that developed by Kuhn and yet consistently attributed to him. This is just one concept and one that is so variably interpreted anyway. But it is also the case that a fair number of elements of the paradigm accounts appearing in the meta-methodological debate and represented as 'what philosophy says' have been clearly discredited in the mainstream studies of science. There, they are no longer seen as descriptive of how science is done or have much more sophisticated formulations. Similarly, the representation of understandings of evidence in medicine and health research, as given in the philosophy of medicine, seems to be failing to represent. Even if we put to one side the argument made here, that thinking of evidence in evidence-based medicine should go, fundamentally, through considerations of research synthesis, the philosophy of evidence-based medicine fails to acknowledge the complexity of thinking about evidence in health research. This work has tried to point to some

such issues which require detailed critical analysis and a much improved inter-field exchange.

A related feature of this thesis is that it brought together entities and processes that tend to be discussed in isolation while having much in common. In some cases new concepts resulted. This aggregation of entities and processes could be seen in the very concept of 'transformations' – its instantiations included processes of reclassification; rephrasing and simplification; minor, uncontroversial interpretation; radical, controversial re-interpretation; re-use of the same information in a different area, etc. It can also be seen in the joint discussion of coding and extraction or the formulation of concepts like 'expression' or 'knowledge something' in the Vocabulary of Elements of Findings. This redrawing of lines of similarity and difference between entities and processes has variously helped to make more generic claims about the research synthesis process; challenge or refine definitions; borrow considerations and richness from 'other' debates; and question credibility claims associated with traditional study type distinctions (for instance, 'high on interpretiveness – low on credibility' is associated with qualitative syntheses; drawing attention to transformations, which are performed across synthesis study types, re-opens the question of credibility).

The forging of links between disconnected debates has hopefully drawn attention to features of the debates, too, rather than only to their objects. Practices of research synthesis are shared and established through particular types of discourses and rhetorics. These discourses and rhetorics then seem to get consolidated on their own terms, in search of a greater coherence and persuasive power, while obscuring some realities of practice. In turn, this has a stabilising and limiting effect on practice. For instance, the discourse/ rhetorics on transparency – the virtue of research synthesis whose extension was of primary interest to this work – is powerful in mainstream research synthesis and in current methodological texts, including qualitative ones. This is a persuasive and coherent discourse which subtends confidence in a range of methodological steps directed at achieving transparency. 'Research in action', however, shows that the quest for transparency meets serious obstacles as this

work has also illustrated. With the benefit of hindsight, every researcher knows this. But we know it mostly vaguely. And we know it mostly vaguely partly because the mainstream discourse/ rhetorics of transparency hides it, or, put more neutrally, lacks a good articulation of it. A counterpart argument can be developed with regard to the 'negative' discourse on transparency (on its limits) running in some alternative synthesis methods. The discourse is coherent, confident and persuasive. A closer look into practice and conflicting discourses breaks some of it. Its strength and disconnectedness from other discourses, however, prevents this closer look. Coherent conversation and practices concerning transparency reinforce one another in complex ways. As a result, practices of ensuring or avoiding transparency persist unchanged even when change is desirable and may be feasible.

This work was not in a position to address discourse features in detail or to make claims as to what is of the object, in this case research synthesis, and what is of language, research tradition and broader culture. But I hope it has implied that the exploration of research synthesis should also, and very importantly, be an exploration of the discourse and rhetorics of research synthesis. And it is not only the mainstream discourse and rhetorics that should be of interest, as is currently the case. Work that deconstructs the mainstream discourse and rhetorics is itself in a discourse and rhetorics. Just as this work is.

### **3.2. 'Empiricising' what tends to be discussed in the abstract; methodological innovation associated with this**

This thesis addressed philosophically and scientifically important questions (primarily about the drivers and extent of change of evidence during its aggregation, with a view to the trustworthiness of the outcomes) by drawing on philosophical and other meta-scientific ideas (primarily about the nature of carriers of empirical knowledge and the factors that shape and co-constitute them). In broad terms this combination of the scientific and philosophical/broader meta-scientific is fairly standard in the studies of science, not least because the distinction between the philosophical and scientific is problematic. The difference here was that the questions were addressed in the manner of a (health) sciences investigation. Data were collected through well

specified procedures. Data were collected until saturation or up to large samples (rather than what seems to be more typical in philosophical and other meta-scientific explorations, of exemplary cases described in depth). The principle of no commitment to positive results or specific outcomes was upheld (as opposed to strong commitment to a position put forward, which is the norm in philosophical and other meta-scientific work).<sup>116</sup> Standard approaches to quantitative and qualitative analysis were used. Reporting on method, process and findings also followed standard health research practices. Although it can be claimed that this is in line with practices in experimental philosophy, the issues addressed and methods used do not seem to have a precedent in experimental philosophy.<sup>117</sup>

The use of the 'health scientific' to address the philosophical and meta-scientific resulted in modifications of the scientific, too. Modified versions of health research methods, new tools borrowing from methods from other fields, and untypical material were used in addition to standard health research practices. For instance, the process of extraction-coding and the analysis of material collected through it included steps from framework, thematic and content analysis – well established analysis approaches in the health sciences. The method of coding of elements of findings was developed specifically for this study while borrowing from an external field – linguistic content analysis methods. Most untypical, however, were the realisation and effect of the five injunctions (comprehensiveness, transparency, double-edged critical analysis, multiple coding and minimal transformations). They led to the analysis process departing substantially from the standard. This was in spite of the fact that the basics of the extraction-coding process were unchanged and that all injunctions reflected both generic scientific and research synthesis-specific virtues. Yet the intensity with which they were used, the context in which they were applied, and/or their specific contents resulted in a different style of health research

---

<sup>116</sup> I use a roundabout expression as I am not sure that belief in this principle of no commitment translates too easily into no commitment. Of course it feels better when your approach works and when your findings fit your expectations or violate them in intriguing ways. It is interesting if scientists get over such 'failures' quicker than other scholars, provided messages that this is useful knowledge and we should have no bias for positive findings are strong in the field.

<sup>117</sup> The preferred object of experimentation there seem to be lay intuitions and assumptions, e.g. about consciousness, knowledge, moral responsibility, intentionality. Surveys seem to be the preferred method.



conduct. This was mostly in terms of allowing multiplicity of interpretations and meta-knowledge comments to proliferate beyond any communally accepted limit. As discussed in the Findings Chapter and the summary of findings above, this style was found to be largely ineffective within the current broad model of research synthesis. In Section 7 I will consider whether this innovation using tools as old as science may need different conditions rather than the scepticism it so naturally invites.

Throughout, I have aimed to position this work in the middle between philosophy and the meta-sciences on the one hand and health research on the other. I wanted its findings to feel interesting and credible to both philosophers and scientists. The risk was that they may have ended up being credible to nobody. As no similar 'empiricising' work seems to have been done in the philosophy of medicine or in meta-methodological research in the health sciences, the experimentation with such an approach and the testing of its credibility is in itself a contribution to knowledge in the area where philosophy attempts to be directly useful to science.

### **3.3. Conceptual innovation and refinement**

A range of conceptual innovations and refinements were proposed in this work. As discussed in 3.1, new concepts tended to be broad and to enable the perception of 'bigger pictures' and the formulation of more general claims. Examples include the very concept of 'transformation' of evidence; concepts from the Vocabulary of Elements of Findings, such as 'obsessive something', 'carrier', 'expression'; the concept of 'tagging' for various ways of generating and recording meta-knowledge in the research process, etc.

Refinements to existing concepts were also made. Most obviously, the concept of research synthesis was given a new interpretation. Its breadth was extended by including more research fields than typically subsumed under it. Specificity was preserved by emphasising some, but not exclusive, use of research reports; an output from which the input cannot be fully recovered; and a greater than a local baseline, but not necessarily high, articulation of a process of bringing knowledge together. The concept of 'transparency' was redefined in a way that aimed to draw attention to cases of superfluous transparency and

invisible non-transparency. 'Coding' and 'extraction' were given a closer look, similarities were highlighted and certain distinctions affirmed by the relevant discourses questioned. 'A finding', too, was refined in the direction of a greater breadth and compatibility with a range of epistemological positions.

A number of clearer distinctions between types were also made, articulating differences between commonly used meanings that tend to be entangled or appear in separate debates. This included, for instance, the articulation of different meanings attributed to 'quantitative' and 'qualitative' as per features of exemplar cases which are not, however, uniquely defining of the broad type; different types of multiplicity of a finding; the different ways for something to be missing, etc. As much as such distinctions are quite close to the surface, they are often not made while underpinning mismatched arguments.

In the case of less central concepts, their clarification may have burdened the presentation unnecessarily. If only most direct relevance to key questions of the thesis is taken as a criterion, it may have been better to suppress such conceptual clarifications in favour of readability. Yet in a field where there is substantial terminological variety and vagueness and some basic entities, processes and occurrences do not seem to have good names and descriptions, this was seen as contributing to broader aims of support for research in the field. As much as both the new concepts and refinements of old ones need extensive further critical analysis and enrichment, they have the potential of being useful tools in organising the discourse on health research synthesis.

### **3.4. Highlighting and challenging inconsistency where none have been noted**

This work challenged some unquestioned inconsistencies in health research synthesis and explored whether the reasons for this neglect are justifiable. All injunctions guiding the extraction-coding reflected a concern with an inconsistency. For instance, the study questioned the contradiction of research synthesis studies being premised on and making use of the multiplicities of studies and findings (if a finding could not have somewhat different meanings to the ones initially intended, most synthesis studies would not be possible), but never facing the implications of this multiplicity beyond the one that has been of

use to them. The thesis thus set out to explore how far a multiple coding of findings can go. Or it questioned the fact that a commitment to comprehensiveness, extremeness in literature searching as a way of achieving rigour is not paralleled, if not across the board then at least in some models of research synthesis, by a similar extremeness in data extraction. It thus set out to explore how far the comprehensive coding of studies can go.

I concluded that, overall, there are good reasons for current methods not to challenge those inconsistencies. Bringing to light the loci where a virtue is applicable and attempting to extend its command created trouble and gloom. One has seen both the inconsistency, with the effect this has on our trust in current health research synthesis methods, and the lack of promise of the obvious methodological revisions, with the effect this has on our trust in the progress of such methods. A sceptical argument seems fully justified. Yet maybe there is some way forward that also makes good on the above inconsistencies. I take this issue up in Section 7.

#### ***4. Contribution of this study to the 'nesting' debates***

For lack of focused philosophical and broader meta-scientific debates on (health) research synthesis, this work was linked with a different level of explicitness to three big debates: the debate in the philosophy of evidence-based medicine; the interdisciplinary debate on 'Big Data'; and the debate on the division between and the mixing of quantitative and qualitative research. Comments on the contribution of this work to such debates follow.

##### **4.1. The philosophy of evidence-based medicine debate**

As reviewed in detail in Chapter 2, the philosophical discussion on evidence-based medicine is most visibly a discussion of the nature of medical evidence. The thesis was partly an argument that by ignoring the methods and the whole culture of research synthesis, the philosophy of evidence-based medicine and the other meta-scientific fields are missing factors that are at least as defining of the nature of evidence used in clinical practice and health policy as the

centrality of the RCT is. A voluminous new set of questions relevant to the philosophy of evidence-based medicine was thus pointed at.

#### **4.2. The Big Data debate**

Similarly, the topic of research synthesis does not appear to have become part of the 'big data' debate and its concern with issues of hypothesis- versus data-driven research, data re-use and data sharing. At the same time it makes an interesting case study in all these directions. The data and methods used in research synthesis are generally 'small', but within the context of much data of potential relevance that are gradually filtered away – partly automatically through search strategies, partly mechanically through 'sifting' on the part of researchers. For good or bad reasons, the technological and methodological sophistication of big data research has not made its way into mainstream research synthesis. Methods of text mining, for instance, are only now beginning to be used (Thomas, McNaught and Ananiadou, 2011). Some further peculiarities of research synthesis as a (boundary) big data enterprise are that much of its data are extensively processed rather than (relatively) raw; that in some types of synthesis work the process is clearly hypothesis-driven while in others can be seen as data-driven; that although a process of data re-use and sharing is under way, the model is very traditional, with research papers being by far the most typical route used.

Because of some of the above characteristics, it is also possible to see research synthesis as an enterprise to be phased out or greatly modified by Big Data initiatives. For instance, at a recent public debate on Big Data, it was discussed how routinely collected data can begin to replace RCTs (Boyle, 2012). Go a step further and the traditional synthesis – meta-analysis of RCTs – will be replaced by the synthesis of routinely collected data. Personal biomedical data, as through personal genome sequencing, may render certain general conclusions for the effectiveness of interventions unnecessary. Some key issues associated with research synthesis, such as of combinability and transformations, can never be phased out. Many of the concerns raised here will remain enduringly valid. Yet the face of the enterprise may change thoroughly if it becomes more attuned to work on Big Data.

The (partly) opposite development is also possible. A greater attention to research synthesis may modify our concern with *Data* in Big Data. In a recent special issue of *Studies in History and Philosophy of Biological and Biomedical Sciences* on data-driven research, introduced as “the first concerted attempt to make sense of data-driven science within the fields of history and philosophy of science” (Leonelli, 2012b: 3),<sup>118</sup> O’Malley and Soyer (2012) attend to the distinction between data-driven and hypothesis-driven science and see the emphasis on data-driven science as only very generally informative of the dynamics of contemporary life sciences. Using molecular systems biology as a reference point, they suggest that the process of ‘integration’ is much more defining of today’s life sciences than the scale of data production. In addition, they suggest that attention to the distinctions between and interactions of data-driven and hypothesis-driven research will do comparatively little to illuminate integration and, more broadly, the dynamics of the contemporary molecular life sciences and scientific change. There is much more promise in addressing issues such as exploratory questioning, technological development, and the (transformative) transfer of already developed systems of knowledge production, such as tools, methods and explanations, into new research areas (*op. cit.*: 59, 64, 65).

How exactly research synthesis and Big Data relate, as scientific practices and discourses, thus appears a fruitful topic for philosophical and other meta-scientific debates. The two fields do not seem to have been linked in any sufficiently visible way so far.

### **4.3. The quantitative-qualitative divide and integration**

The debate on the differences and possibility for productive interaction between quantitative and qualitative research was the third of the major nesting debates for this work. Conceptually, it seems hard to enrich this so heavy debate – heavy on arguments asserting unbridgeable differences between quantitative and qualitative studies and the importance of preserving character; heavy on

---

<sup>118</sup> Callebaut (2012) suggests that in this respect scholars in the social studies of science, information studies and the history of science have done some serious work and are ahead of philosophers. His focus is on philosophers of biology who “have barely entered the scene” (*op. cit.*: 70).

arguments about relative advantages and disadvantages and amicable interaction; heavy on arguments about why this debate is unproductive, exasperating and needs to stop. To my knowledge, the particular claims made here have not been made elsewhere, but their implications (e.g. of reducing the energy and effort expended on this debate, of challenging polarising assertions, etc.) are all too familiar. I will thus try to summarise only briefly the particular claims made in this study.

The foregrounding of this study (through the literature review) challenged the quality of claims concerning quantitative-qualitative 'paradigms' made in the meta-methodological literature through comparisons with debates in the mainstream studies of science. If we continue to take broad epistemological paradigms as major determinants of similarities and differences between health research findings, the shortcomings of their descriptions will keep on affecting our theorising on the synthesisability of findings. Extensive professional philosophical critique of the philosophy of science internal to health research, open enough to see the particular needs of health research rather than only 'straighten' the debate, is much needed.

Maybe more importantly, however, findings from this study questioned, in unusual ways, if differences in broad epistemological paradigms are sufficiently relevant to the combinability of findings. This doubt is commonplace. But it tends to be voiced in the abstract, along the lines of commensurability of paradigms, or justified in the reverse – since we are carrying out mixed methods primary studies and syntheses of largely heterogeneous studies, then it must be the case that either paradigms are commensurable or they do not matter that much. Here, the challenge arose from the limited success of operationalising claims from the abstract paradigms debate. Using a range of operationalisations and a large enough sample of findings (over 300), this study was unable to identify with any clarity, at the level of findings, prominent features from the paradigm descriptions. The broad epistemological paradigms, as discussed in the meta-methodological literature, disappeared almost without a trace by the time a finding, as represented in a health research paper, was reached.

Of course, many criticisms of the operationalisations chosen in this study can be given. In view of the stability of the debate within the health and social

sciences, it is necessary to propose and test a broad range of alternative operationalisations. But if we give some credence to the lack of evidence for marked traces of the broad epistemological paradigms at the level of findings, we need to say something about the debate, too.

Three main options seem available. The position that is least damaging to the debate is somewhat contrived, but not inconceivable. We can argue that its key claims are solid but subtle and this work, just as any work that has discussed the combinability of health research findings in terms of paradigm differences, has missed the subtlety: the debate is primarily about differences of *studies and research*, not of *findings* from these. Indeed, in the case study, features of paradigms and traces of these were much more easily identifiable at the level of studies. This would show, however, that the paradigms debate is reticent as to whether distinctive features of studies, as determined by differences of broad epistemological paradigms, are gradually lost until a finding is represented and how this relates to the combinability of findings. Alternatively, we can argue that the broad epistemological paradigms are solid determinants of similarities and differences between findings but that their typical *accounts* are flawed. As a result, the 'wrong' features of paradigms had been operationalised in this work. The broad epistemological paradigms need to be reconceptualised and redescribed so that the true determinants of differences between quantitative and qualitative findings can be identified. Finally, it may be the case that other things matter more for how health research findings differ than the quantitative/qualitative matrix within which they were generated. Of all possibilities considered, this is most damaging to the quantitative-qualitative debate. For many scholars this will be no loss. But a contextual problem remains. If we give up a commitment to the explanatory model of quantitative-qualitative incommensurability, we are left with no well worked out alternative and an incontestable phenomenal experience of some health research findings being rather difficult to combine with some other health research findings. This debate was our only fully grown leg.

The empirical work also showed that the majority of health research studies are best described as some form of hybrid studies, rather than as quantitative and qualitative, even if they are not explicitly mixed methods studies. Such a claim

of grey zones between the quantitative and qualitative is not infrequent in health research. But hybridity is still thought of as the exception rather than the norm. This work offered some initial ideas of how quantitative/ qualitative/ hybrid status can be specified. Applying them to larger samples is one way of testing if a primary typological distinction in health research does not hide more than it reveals.

There are certainly very important epistemological questions discussed and positions articulated in what now appears as the quantitative/ qualitative and mixed methods debate in the health and social sciences. There is energy in the debate. It is also the most direct point of access of health researchers to thinking in the philosophy and social studies of science. At the same time there are well known problems with and challenges to it. The main challenges from this work, which do not seem to have been extended in the existing literature, came from highlighting the discontinuity of the quantitative/ qualitative debate from current thinking in the mainstream studies of science and from a failure to find evidence, sought through tight operationalisations, that differences between broad epistemological paradigms are traceable at the level of findings and in this sense are well specified or 'real enough'.

## ***5. A little more on the key concepts of the thesis***

### **5.1. The unit of research synthesis**

This thesis was framed as being about transformations of *evidence* – the predominant concept for a carrier of primarily empirical knowledge in the health research synthesis discourse, as shaped by the broader one of evidence-based medicine. 'Data' was also an early framing concept, suggested by the use of 'data extraction' for processes of taking out material from primary studies so as to include it in synthesis studies.

From the very beginning, however, the reference to evidence (and data) as the unit of analysis/synthesis in research synthesis was put into question. This was mostly in response to views that knowledge gained from research synthesis is purely empirical, theory-free, minimally interpretive. 'Evidence', as used in the



context of mainstream health research synthesis and health policy documents, has such connotations of absoluteness.<sup>119</sup> Writings on alternative synthesis methods challenge this image. They prefer to speak of findings, emphasising the roles of analysis, interpretation and theory in moving from primary data to 'knowledge'. But as this preference remains confined to their texts, it is only they that end up carrying the connotations of uncertainty and subjectivity. In addition, the insistence on a reference of 'findings', as associated with heavy analysis, theory and interpretation, ends up misbalanced in the other direction – research synthesis studies can and do use (almost) raw data, such as non-interpreted interview quotes or data on individual cases.

My concerns in illuminating the unit of analysis/synthesis in health research synthesis, in preparing to address its transformations, were thus two. The first was to bring forth that which is not purely of the object being studied, but comes from theories, auxiliary assumptions, researcher background, possibly cultural-historical contexts, etc. in research synthesis and to search for its place in judgements of combinability and processes of transformation of the primarily empirical. The second was to explore if source study material (i.e. the data from primary studies to be used for research synthesis) can be typologised reliably in terms of its distance from raw data – that is, in terms of all the transformations under way in primary studies, too.

As the literature review showed, in spite of the tendency in some methodological literature to oppose 'evidence' and 'data' as the 'raw' and 'purely empirical' to 'findings' as the 'processed' and partly theoretical and interpretive, each of the main contenders for being the primary unit of analysis/synthesis in research synthesis studies had connotations going both ways in the broader literature. Evidence, data, findings, facts and claims are each discussed as an embodiment of empiricism and/or as necessarily reflecting something over and above pure observation and the world readily revealing itself. Thus, it did not appear consequential whether one or another concept is used to denote the main unit of analysis/synthesis, provided one takes it to be the case that the main unit is indeed some carrier of the empirical contents of studies.

---

<sup>119</sup> The same holds for data. However, evidence is the preferred term, especially in policy documents referring to health research.

For the majority of synthesis studies, this is indeed the case – they aim to bring together empirical contents from previous research. But not necessarily. Some synthesis studies aim to integrate other types of research contents. Others end up doing so while integrating empirical contents. A wide range of synthesis studies may be having as primary or equally important units of analysis/synthesis concepts, theories, descriptions of cultural-historical contexts, methods, etc. Further to that, if we look closer into findings themselves, it becomes immediately evident that this unit of analysis/synthesis is not quite unitary. It is made of concepts, references to methods, references to samples, representations of objects, representations of features of objects, traces of theory, qualifiers of degree of certainty in the presented knowledge ... From a different perspective, it is made of words, numbers, images, graphs ... The ways in which such different components of a finding and/or study are brought together will be at least somewhat variable. A variety of descriptions will also be available.

Thus, it has been a temporary stabilisation of uncertain grounds to take a 'finding' as the main unit of analysis/synthesis in research synthesis studies. There is much to be explored about the different units of analysis/synthesis in synthesis studies and how they are brought, or come, together. The development of the Vocabulary of Elements of Findings was some movement in this direction. Relative to the vision of the future I discuss in Section 7 below, the most scientifically useful way of continuing a philosophical debate on the units of analysis/synthesis in research synthesis is by focusing on what there is in a research report, what there is in a finding, the many ways in which we can split a finding, and on specifying the matrices which shape and co-constitute findings. Probably of little interest to science, but otherwise quite intriguing philosophically and historically, is how the variety of concepts of carriers of primarily empirical knowledge (data, evidence, observations, facts, claims and findings) overlap and why certain debates on relationships with the 'non-empirical' have clustered around one rather than any other of those concepts.

As for the other main initial concern, of exploring if source study data (evidence, findings, etc.) can be typologised reliably in terms of their distance from raw

data, it became clear that the operationalisation of such an intention would be forbiddingly complex. More importantly, the use to which it was envisaged to contribute – a measure of the reliability of individual synthesis studies, in an attempt to counteract a tendency for typological judgements of reliability along quantitative/ qualitative lines – lost its initial value. In light of the case study findings, the level of reliability of synthesis studies began to appear too low across the board. Greater attention to more homogeneous studies and numerical transformations will, undoubtedly, reveal more loci of reliability in research synthesis. I also acknowledge that there are crucial differences of reliability across individual synthesis studies within a type, that it is important to explore them, and that attending to the average distance from raw data may be one such tool. But relative to the yardstick of reliability I use here, these are all ‘small’ differences. My focus is on approaches for radically enhancing the reliability of health research synthesis studies. A concern with proximity to raw data was thus integrated in a radically new vision for the future of research synthesis.

## **5.2. Transformations**

At the start of the thesis, I took transformations of findings for the purposes of research synthesis to be processes and outcomes of presumably subtle and/or well grounded changes in the form, or the contents and form, of findings that are driven by differences perceived as hampering their ready integration but also as bridgeable, as some right type of similarity is there, too. Such transformations lead to or demonstrate a reduction of relevant differences between findings which may or may not be sufficient to ensure combinability or compatibility. That is, transformations can be a test of similarity that turns out negative and not only a successful attempt at increasing similarity.

I also distinguished between processes of transformation and re-location, with ‘transformation’ seen as the generic process and ‘re-location’ as the process where evidence is moved to a rather different type of domain (e.g. theory about other kinds of phenomena, as may often be the case in highly interpretive qualitative synthesis) or related to a rather different type of entity (e.g. when new groups of entities and phenomena are formed, as may often be the case in

meta-analysis). In line with this distinction, transformations would enable/ test for a greater degree of similarity between a finding to be transformed and a reference finding, while re-locations would enable/ test for a greater degree of relevance of a finding to a reference framework.

In the process of the work, I collapsed the distinction to a distinction of degree: re-locations were seen as 'bigger' transformations. This was partly related to recognising the complexity of the unit of analysis/synthesis in research synthesis by virtue of which findings stopped being seen as the only object of transformations (see 5.1. above). As I could directly attend to the transformations of concepts, groups, types of interventions, etc., rather than seeing these as something static to which a transformed finding is applied, the re-attribution process highlighted in re-locations was no longer that important.

In combining initial ideas and observations from the case study, I suggested that the capacity of a finding *f* to transform is determined by the number of findings in the set with which it is to be made combinable or compatible; by the extent of differences and similarities between *f* and the other findings in the set; by *f*'s having characteristics that come into a 'logical contradiction' with specifications made in the synthesis question; by the degree of its multiplicity; by the extent of missing and available pieces of information in *f*; by the distance between it and the synthesis question; by the effectiveness of the tools used to enable transformations and the tools used to constrain them; by the familiarity of the researcher with the research area of the findings and that of the synthesis question; and by a combined factor of harder to capture influences, such as order of presentation of findings, recent salient information available to the researcher, his or her capacity for lateral thinking, chance juxtaposition of information, etc. In the presentation of findings from the case study, different aspects of some of those drivers of transformations were illuminated.

Many more remain to be illuminated, including some basic ones. The contents of some of the determinants needs much further specification (partly because they were not subjected to sufficient empirical exploration in the case study) or re-examination (partly because the case study showed some problems with the initial conceptualisation). For instance, how exactly should we conceptualise the

distance between a finding and a synthesis question? How can we operationalise the measurement of multiplicity? What is the 'full' set of currently employed tools for enabling and constraining transformations, what difference do they make and can we separate the two types of tool easily? Issues of relative importance and interaction between the determinants also need to be given further thought and possibly more plausible hypotheses offered. For instance, how likely is it that the extent of 'all' differences and similarities between a finding and its reference findings matters for its capacity for meaningful transformations rather than a smaller subset of these brought to the fore by the synthesis question? Or how strongly does the perception of missing/available information relate to characteristics of the researcher? Relative to a high quality descriptive account of transformations, within the framework of parameters proposed above, a large number of questions remain that require further empirical and/or conceptual exploration. I am inclined, however, to argue against giving those questions much of a priority though certainly not against addressing them.

I take this position on pragmatic grounds and relative to the concerns of this work. My interest in transformations was tied to the question of how much we can trust current research syntheses and how we can make transformations in future ones more rigorous and creative. Below, I outline a picture of the credibility of current synthesis studies. I suggest that the ways in which transformations are performed in them are reactive, rather than deliberate, responses to the pressures of insufficiency and/or excess of information. I add to this claim findings from the case study and broader observations of the research synthesis field. If the resulting picture of hugely problematic credibility of synthesis studies is valid, it will be self-evident why working out details about drivers of transformations will not be an effective way of improving those processes. It needs, however, to be taken as a provisional picture. I do not have direct evidence of data extraction from RCTs and numerical transformations in preparation for meta-analysis. I cover such studies and transformations on the basis of features I identified in this case study and consider largely generalisable (e.g. levels of missing information within a research report) and on the basis of knowledge and observations from my broader experience.

A major way in which mainstream synthesis studies resolve both the insufficiency and excess of information is by a strong leaning towards identities and 'almost identities' (of interventions, population groups, settings, finding types, etc.) and ease of judgement about similarity/ difference. This already creates a context where the need for transformations and the facing of their uncertainty are kept to the minimum. The strong leaning towards identities and 'almost identities' is a somewhat intended, somewhat unintended consequence of asking very narrow questions and setting up very specific inclusion/ exclusion criteria. As much as there is nothing devious in the approach of the mainstream systematic review – narrow questions tend to be easier to answer and specific inclusion/ exclusion criteria ensure reliability – the combination is not without disadvantages and not without alternatives. The main disadvantage is that relevant information is not taken into account. Many apparently different or irrelevant findings that can be used through transformation are left out. The main alternative to the 'narrow question – highly specific criteria' combination is to ask narrow questions by approaching them through a broad range of information. This would require, among other things, many more transformations. The reliability of the synthesis process will decrease in ways that are visible.

The leaning towards ease of judgement in the mainstream systematic review is realised through the contents of inclusion/ exclusion criteria. Theoretically, the contents of inclusion/ exclusion criteria has to reflect hypotheses of factors affecting an outcome. But judgements about correspondence with inclusion/ exclusion criteria should also be feasible, in the sense that the relevant information should be reliably available across studies. They should be generally easy, too, so as to be reliable. Thus, only widely reported (standard) parameters characterising studies tend to inform inclusion/ exclusion criteria. Also, similarity/ difference along such parameters is judged by making the comparisons easy to operationalise. Yet standard parameters correlate with outcomes only sometimes and, as the findings from the case study suggest, there is striking variability in the ways in which they are reported on and very limited information of what the labels used in such reporting stand for. Transformations in mainstream synthesis studies are thus often performed or

avoided on the basis of 'unimportant' parameters and on appearances of similarity rather than detailed background information.

In contrast, many alternative synthesis methods (with qualitative syntheses as the exemplar case) start from a variety of material where only a few dimensions of relevant similarity/ difference are specified. A much larger number of directions for exploring the combinability of findings are open, with the demands this would place on transforming findings. This excess of potential, however, needs to be curtailed. Three main approaches seem to be used. The first is the 'insight' approach. An insight flashes – a pattern is seen – findings transform towards a few common features. The possibility for and the value of justifying why those similarities, differences and transformations rather than others is denied (it is interpretive, it is a leap of thought, it is researcher dependent, it is context of discovery) against the strong valuation of transparency in research synthesis. Second, we have the approach where largely uncontroversial transformations are performed which go for the lowest common denominator of messages and limited criticism – very broad similarities are derived, often in line with well established narratives. Reliability and transparency are preserved at the cost of reaching rather generic and predictable conclusions. Finally, we have a combination of largely uncontroversial transformations and no transformations at all, where the generic and predictable conclusions are accompanied by non-synthesised descriptions of individual cases (see Pawson's criticism of narrative reviews in Pawson, 2002a on which the description of the latter two situations draws).

Exceptions notwithstanding, the following principles concerning transformations thus prevail in health research synthesis, largely for pragmatic reasons associated with a duality of too much and too little information: do not look for transformations along lines of similarity that are unusual or not immediately obvious; do not explore alternatives to the transformations you have carried out; transform on prima facie similarities rather than detailed checks of similarities and differences; transform to the lowest common denominator; transform on an insight.<sup>120</sup> This amounts to severe underuse of the richness of findings and

---

<sup>120</sup> All this can be re-formulated, equivalently, from the perspective of similarities: do not create unusual groups; do not test your grouping against alternatives; create groups on the basis of

insufficiency of rigour in studies which define themselves precisely through comprehensiveness and rigour.

My assumption at the beginning of the thesis was that having a better awareness of how processes of transformation happen and what drives and hampers them, for the right and wrong reasons, will allow us to make them more rigorous and creative – that is, that we need detailed descriptive knowledge before we can get to the normative. I no longer see the details of a descriptive account of transformations, as they are currently performed, as a particularly valuable guide towards a normative account. This is because, in response to the study findings, I no longer see the processes of transformation, *as performed within current models of health research synthesis*, as meaningfully rectifiable. One of my initial hypotheses was that if we extend the scope of certain virtues of research synthesis (e.g. transparency, comprehensiveness and critical analysis) and if we turn to a default mode of multiple coding,<sup>121</sup> we can noticeably enhance the rigour and creativity of transformations in current research synthesis, and hence the trustworthiness of its outcomes. The case study showed quite starkly that the tools I implemented in trying to extend the scope of the methodology's own virtues offered something but not enough, their reliability was low, and their cost in terms of feasibility on a larger scale would not be worth paying. It also suggested very strongly that minor modifications of such tools will not change this, and that a completely different model and infrastructure are needed.

A corollary of the above assertions is that, within the current overall model of research synthesis and the infrastructure that supports it, our methods are simply realistic in their selectivities and the inconsistency with which they apply the virtues of the enterprise. In addition, some of their outcomes are good without any relativisation to the 'best we can do'. There is value in bringing together the very closely similar and avoiding transformations. There is also

---

*prima facie* similarities; create big safe groups; create groups following an insight of what fits together. It can be reformulated from the perspective of differences, too, for that matter ...

<sup>121</sup> This need not be taken as separate from extending the scope of the methodology's own virtues – we can see multiple coding as another way of enabling comprehensiveness and systematicity. It is mentioned separately as the typical associations for ways of realising those virtues do not include multiple coding.



value in jumps of thought taking us to places that would never be reached if the path had to be walked step by step, always thinking of those who may wish to follow or of finding our way back.

But the facts that current methods are realistic in their approach, good in certain senses, and cannot be reformed in a substantial difference-making way under the same overall conditions do not mean that they are good enough or that the majority of them cannot be replaced in a very radical way. The vision for the future I have referred to on a number of occasions will be discussed shortly, immediately after the limitations of this work.

## **6. *Limitations***

The novelty and breadth of the work have resulted in compromises of precision, depth and systematicity. This is particularly visible in the literature review. As its aim was not only to contextualise the work and its key questions, but to prepare the ground for the empirical investigation, limitations of the literature review may have translated very directly into limitations of argument and findings. Many of the issues branching out of the topics of evidence, its combinability, and transformations are fundamental epistemological issues. A range of immensely rich and complex debates were thus touched upon. Broad claims were made and highly specific detail given, but the complexity of the middle level was not addressed. Representing this middle level adequately was also impeded by the fact that it was not possible to rely on a shared understanding of the basics, as issues that are self-evident to health researchers are not evident to philosophers and vice versa. Some highly relevant debates were altogether left out, such as the mixed methods, interdisciplinarity and secondary data analysis debates. Some of the debates I was most committed to bringing forth, such as those in the psychology of science, had to be given up for the sake of consistency and remained very much in the background, as considerations I was using in the analysis but at a limited depth.

A different and better informed background framing may have resulted in a different specification of questions and methodology. Many attempts were made to achieve some local systematicity – e.g. through ‘microstudies’ based on

carefully selected samples (obtained, for instance, through exemplary texts or bibliometric data), but these could only serve as a partial remedy. The state of preformation of the field of research synthesis, its disconnectedness from the philosophy of science and other meta-scientific debates, and the breadth of the issues addressed meant that this work would be falling often into traps of ignorance and obvious omissions. The price and risk of early attempts at something unusual and broad are high. On a more positive note, this is simply an approach of a very particular grounding in the meta-scientific debate, which, although limited in its scope and reliability, will provide a useful reference point to help direct further exploration.

The novelty of the methods resulted in a substantial imbalance between methods and analysis. On the one hand, many more methodological steps were prepared, envisaged and actually executed in comparison to the lines of investigation that were followed in detail – both in terms of analysis and reporting. For instance, the exploration of issues of representation was extensively prepared in the Analysis Framework and voluminous data were collected. These were not, however, analysed in depth after primary analysis showed that such findings will not explain enough of what is happening in processes of transformation. Rigour and transparency demand that the analysis is followed through as set up and that negative findings are reported more clearly, but constraints of feasibility had to be taken into account. Overall, my approach has been to do justice to the strongest hypotheses within a set (relative to current knowledge and the aims of the thesis) and report clearly positive and negative findings concerning these. Findings about weaker hypotheses were reported only if highly salient, whether positive or negative. For instance, issues of combinability arising from differences between quantitative and qualitative studies related to strong hypotheses, while issues of combinability arising from the variability in representations of the same finding did not. The reason for ‘over-designing’ the study was that, in such a new territory, many directions of exploration appeared to have equal chances of offering interesting and informative findings, and yet came with no guarantee that they lead anywhere.

Another aspect of the imbalance between methods and analyses was that many methods did not provide sufficient information for analysis. For instance, many

operationalisations for identifying and isolating elements and effects of matrices were of limited effectiveness. Expectations of a sufficiently high level of preliminary synthesis concepts and claims were not realised and comparisons between 'input' and 'output', seen as central to understanding processes of transformation, could not be carried out. Fundamental elements of the Analysis Framework were thus found to be weaker than anticipated. Although this was not unexpected – in a field of no sufficiently similar precedents failure is at least as likely as success – there may have been an approach where risks of methodological steps not working were better distributed.

The novelty of methods also meant that some of them were not pushed to their logical limits. As a result, they provided data that could support or disconfirm hypotheses, but without excluding alternatives as reliably as possible. Most typically, as patterns of difference were expected to be found closer to the surface, methods were unsuited to provide data about potentially important differences at deeper levels. For example, the composition and structure of quantitative and qualitative findings were found to be quite similar by using the Vocabulary of Elements of Findings. It was clear, however, that many avenues remained unexplored at the level of sub-elements – for instance, at the level of types of 'expressions' and 'relationships'. Although unlikely in light of existing debates, the theoretical possibility was open that differences between quantitative and qualitative findings appear at such finer levels. The directions of how to refine the methods further were clear in their general outline but were not followed. Such a process would have generated too many additional operationalisations, data and analyses. I thus settled for less secure claims.

A further key limitation of this work concerned the sample of studies. It was intentionally highly heterogeneous. The sampling framework aimed to pull together studies of a wide variety of health-related fields with their potentially incongruous fundamental assumptions and concerns; theoretical, methodological and pragmatic frameworks; concepts; values, etc. I saw this as an opportunity to allow for a broad range of transformations and challenges associated with them to emerge during the extraction-coding. Presumably, this

has opened more of the variety of transformations and their drivers. At the same time it generated evidence of the frequency and range of transformations and the force of their drivers which will not be readily applicable to the synthesis of largely homogeneous studies. In having to choose whether I address transformations in largely homogeneous or highly heterogeneous studies, where doing both simultaneously would be a contradiction in terms, I went for the option that underpins answers to much more complex health-related and social questions. While I acknowledge that focused work is needed on homogeneous studies and, even more narrowly, on homogeneous RCTs on biomedical topics where standardisation is expected to be at its highest, I do consider some of the case study findings sufficiently generalisable to such contexts. No realistic prediction based on even minimal familiarity with health research will maintain that the variety and incompleteness of information found at the level of standardised parameters will disappear without a trace in a sample of homogenous RCTs. It will be reduced for a variety of reasons, not least due to the widespread use of reporting guidelines such as the CONSORT statement. It will not disappear. The need to ascertain similarity and difference, make inclusion-exclusion judgements and create groups on the basis of such far from uniform and complete information on standardised parameters will remain. Controversial transformations will thus be made in any synthesis of RCTs, even if all other transformations involved in it are numerical conversions that are entirely innocuous, reversible and independent of problematic statistical assumptions (and we have good reasons to believe that not all numerical conversions are innocuous, reversible and independent of problematic statistical assumptions).

Very importantly and unfortunately, the approach I have used is also a classic case of the research process interfering intensely with the phenomenon researched and/or merging of method and object of study. In studying data extraction and transformations relative to demanding criteria of rigour and heuristic promise, and in approaching this through observing my own data extraction and transformations in progress, I brought to life a process which is some way, maybe a long way, away from 'normal' data extraction as executed by the majority of health researchers. Other approaches of studying

transformations in data extraction, such as comparing input and output material or performing 'cognitive interviews' with researchers while they are extracting data, are certainly valuable. But unsatisfactory trade-offs seem inevitable. If the exploratory processes used interfere little with the 'normal process' of data extraction, fewer questions can be asked. If the sample used is more representative of the health researchers population, further errors of communication are added (e.g. of communicating aims to study participants in a way that would not shape their observations). The study of transformations in research synthesis seems to be in the category of cases where many methods need to be triangulated for a fuller picture of the phenomenon of interest, with no method having a superior balance of strengths and weaknesses.

Finally, as much as I have clarified on a number of occasions that the questions asked and answered here concern the process of data extraction, which is only one of many stages of a research synthesis process, I fear that the description of what is happening at this stage has begun to sound as a description of 'the' research synthesis process, with some detail missing. It is far from it. At least as much can be said about many other elements of the process, maybe most notably those of literature searching and critical appraisal. In addition, the description of transformations is not a description of transformations happening and prepared in research synthesis, but in one of its phases only. Very interesting issues of, for instance, statistical assumptions used in quantitative transformations have thus received no attention as they tend to become relevant at later stages. Also, in accordance with the way in which the scope of the study was narrowed at the beginning (focus on 'words' and 'numbers within words'), certain types of research material received minimal attention. These include, for instance, graphs, figures and images, of which only individual cases were analysed. There is sufficient indication that processes of transformation will work differently there.

Future research will hopefully address some of these limitations, perhaps in continuity with ideas presented here. But some of this future research better break away from the assumptions and approaches of this work. I finally discuss how.

## **7. Vision for the future or an 'impossible synthesis'?**

### **7.1. Existing visions for the future of research synthesis**

Information technology has changed immensely since the beginnings of research synthesis studies within an evidence-based medicine framework. The software underpinning research synthesis has also changed immensely in line with user needs, feedback and new technological solutions. That said, the field does not engage in high visibility discussions on alternatives to the current infrastructure of research synthesis and whether technological advancements are making it possible to do research synthesis in a radically new way. I have encountered several proposals made in this direction, all of these only minimally articulated. Chalmers, Hedges and Cooper (2002), in discussing the future of research synthesis, see us as moving towards the analysis of raw data, once they become more broadly available and shared (*op. cit.*: 31-2). Computer scientists see opportunities for text enrichment, where abstracts or published papers are tagged with further information enabling much more effective automatic capture of knowledge through data mining.<sup>122</sup> This would reduce the need both for developing traditional literature searching strategies and for manual data extraction. Researchers at the Agency for Healthcare Research and Quality in the US have created a Systematic Review Data Repository where researchers can deposit data extracted for their review, in recognition of the time-consuming work involved in data extraction and the fact that data can have multiple uses. There are 104 projects in progress in the Repository as of May 2013 (Agency for Healthcare Research and Quality, 2013). Finally, much hope is being placed in big datasets, e.g. collections of electronic health records. These have been discussed as having the potential to reduce the need for randomised controlled trials (Boyle, 2012) and, respectively, for their synthesis. Circulating views on the future of research synthesis thus veer, variably, towards proximity to primary data, richer meta-data, larger volumes of information, greater automation of processes, accumulation of data enriched by critical analysis, increased connectivity between existing reviews, and increased use of routine data.

---

<sup>122</sup> Informal discussion at the *International Conference on Biomedical Ontology*, Buffalo, US, 26-30 July 2011.

## 7.2. An alternative?

I wonder if we can have our cake and eat it (and even save it for later). I wonder if this can happen by expanding on a basic vision for 1) mega-databases on 2) big questions 3) that bring together a broad range of resources, from raw data through critical analyses to non-research material, and 4) that are able to break down, in a fractal pattern, into smaller databases on narrow review topics.

This broad vision is a mirror image of limitations and strengths of three main types of enterprises – traditional systematic reviews, non-traditional reviews on broad topics, and Big Data databases. A main criticism directed at traditional systematic reviews is that they address very narrow problems, yet the precision with which they formulate these narrow problems is a great virtue. Reviews on broad issues are not as broad as the title suggests, as they tend to use a narrow range of disciplinary perspectives and resource types. Nevertheless, they make the attempt and often achieve a much deeper critical perspective in the process, as a result of the juxtaposition of highly diverse material. Large databases accumulating material on a phenomenon are a treasure trove of information, but this information is meaningless without hypotheses/research questions from which to analyse it.

Referring to a (potentially significant) number of mega-databases, rather than one mega-database, is not only out of a concern with feasibility but a reflection of a fundamental position of this thesis – it does not believe in a final, consistent, unified picture. The big questions I have in mind in constructing and probing this new vision are at the level of generality of the following examples: *How to change health behaviours? Do physical or psychological factors have more weight in prognosis in life limiting illnesses? What is the relative contribution of individual versus social and environmental factors in making health-related decisions? What are the effects of nutrition on health and how to institute change in nutrition behaviours? Are different types of communication in healthcare associated with different hard and soft outcomes? ...* For simplicity, I will revert to speaking of ‘the mega-database’, or ‘the super-synthesis database’ in specifying this vision for a possible or impossible synthesis. I will be meaning a single representative of this type of super-synthesis databases, generically construed. The features of the super-synthesis database presented below draw

on assumptions and findings of this work, broader criticisms of research synthesis, virtues and good practices in it, promises of and concerns about Big Data, and characteristics of our current research cultures. The description of these features is followed by discussion of the limits this utopia hits, too.

If, as per background claims of this work, the primary route to increasing the rigour and heuristic potential of transformations goes through increased access to information about the matrices of a study and if, as per findings from this work, this is not information we can get satisfactorily out of published papers, our super-synthesis database will need structured ways of collecting it from primary researchers and theorists/critical analysts. Building on theoretical thinking of what the matrices that shape and co-constitute findings are, the database will be prompting primary researchers to supply such information. For instance, established study designs will be associated with fields and prompts that correspond to items in current quality assessment checklists. The specifics of how certain steps were realised will be entered by researchers in a combination of structured and free text. Researchers will be prompted to articulate hesitations and concerns about rigour. The information demanded by the mega-database will be more extensive than the information demanded by either papers or data depositories. For instance, it will be asking for background, connecting information that tends to remain only in the minds and unshared notes of researchers – partly because it has become so normalised that it is not considered unknown or relevant, partly because a researcher's intuition is to withhold it as things could have been done better. The information about matrices on the super-synthesis database will also be much more consistent across studies and better structured than it is in papers and data depositories, with the benefits this has in terms of judging comparability of findings and explorations of matrix effects. In addition, research that challenges normalised matrices and often comes from distant fields will be appropriately linked. For example, empirical findings about self-management interventions will be directly connected to more theoretical research on the politics of the concept of self-management, or studies based on surveys relying on self-reporting will be connected to basic psychological research on memory and social biases.



Constructive and deconstructive work on the same topic will be placed side by side.

If, as the social studies of science, history of science and the psychology of science would suggest, we have good reasons to believe that some of the main drivers of research cannot be reconciled with a picture of science making as a fully rational enterprise, we need to find structured ways of collecting information about such drivers and relating it to the nature of research findings. For instance, this would include information on the researcher as an individual and on the social processes in his or her research team. Since, however, many scientists resist the idea of a far from perfect rationality of science and those who do not will be concerned that they will be undermining their work, relevant information may need to be collected sensitively. For instance, it may be anonymised and at least partly disconnected from the area of study ('normal' data entered on the super-synthesis database will be relatable to a traceable user and extensively connected). Alternatively, attempts may be made to foster a culture which rewards the extensive empirical exploration of purported non-rational effects on health research. Currently, the culture is of asserting or denying them rhetorically or on the basis of limited examples.

The description so far suggests that much of the information on the super-synthesis database will be supplied by primary researchers rather than, for instance, data extractors.<sup>123</sup> This would then demand the development of mechanisms through which the value of the peer review process characterising research papers and the further critical appraisal and re-interpretations coming from synthesis researchers are added and potentially improved on. A practice may be instituted, for instance, where the peers who have reviewed a paper for a journal also sanction the entries made by authors on the mega-database. At the same time, the disadvantage of a single story coming out of the authors-reviewers debate characterising research papers may be circumvented. As the text presented no longer needs to be linear and consistent, different

---

<sup>123</sup> In the context of health and medical research this practically means a default of several primary researchers supplying information on the same study. The division of labour in the field is such that different researchers have privileged knowledge of different aspects of the same project – e.g. data collectors, statisticians, analysts, etc.

perspectives on the same research decision or outcome can be left in. As for the perspective of synthesis researchers, this, too, can be incorporated as a further layer of critique and interpretation. Finally, comments of readers can also be added, with a clear indication of their source. This openness to comments of a broad range of readers will overcome the limitations of knowledge of a couple of reviewers and a single synthesis researcher or a small team. As this study made all too clear to its author, no amount of generic critical thinking, theoretical support from the meta-sciences, and internal checks and balances arising from inter-connections between entities and findings across studies can make up for the lack of narrowly specialised knowledge of a synthesis researcher. We may take this community contribution even further – if we are claiming that findings are always generated and seen within a range of matrices, we will want the super-synthesis database to ask for background information on all those community users so that its relationship to the nature of their contribution can be explored.

As a further way of getting the most out of the same piece of research, the super-synthesis database will encourage multiple classifications of concepts and findings. This will be done partly by the primary researchers, reviewers, synthesis researchers and peers. For instance, the data entry process for primary researchers will prompt them to consider where else their research fits. The synthesis researchers' and peers' commenting outlined above will naturally involve this, too (“this reminds me of ...”). In addition, database curators and contributors may proactively explore relationships between concepts and findings and propose new relationships of membership and instantiation. These could be variously approved by authors, indicated as curator/contributor suggestions, and commented on by peers.

If, as per findings from this work, rigorous multiple coding in ‘easy but many cases’ cannot be achieved proactively, as it is too cognitively demanding and emotionally destructive, then the super-synthesis database should be underpinned by extensive capacities for linking, self-population/transformation of information and simplified entry of information. For instance, as far as linking

is concerned, authors may be prompted to answer questions of how their key concept relates to a network of concepts the system already holds as connected, or, if the concept is new, what concepts the authors see as related to it and in what ways. This would be a further way of multiply coding information, while the demands on proactively coming up with 'all' related concepts, phenomena, etc. are lessened. As far as self-population of information is concerned, the super-synthesis database may, for instance, have different pre-set levels for saliency of findings, some of which context-specific, some of which generic (e.g. applies to over 80% of the population studied) which get self-populated on the basis of information provided. Or once information is provided on initial and final sample, dropouts become automatically calculated. Depending on the degree to which the change of perspective may require readjustments, the authors will be given the opportunity to correct automatically populated information.

If there are arguments both in favour of the value of processed findings and raw data, the super-synthesis database should enable the entry of both, depending on the preferences and capacity of researchers performing the data entry. It should also enable ways of automatic extraction of data from other sources – e.g. the working databases of researchers contributing data to the super-synthesis database or national statistical databases.

The super-synthesis database should also attempt to counteract one of the biggest problems in research, the bias towards positive findings.<sup>124</sup> To avoid this bias at the level of studies, where studies having positive findings are much more likely to be published, the incentives for contributing 'no difference found' data should be just as strong as those for contributing positive data. To avoid this bias at the level of findings, where positive findings are much more likely to be reported within a study and attended to in its discussion, hypotheses and expectations will be first reported in detail and findings then 'demanded' rather than the reporting being driven by findings. As researchers will no longer be

---

<sup>124</sup> I mean primarily in opposition to 'no difference found' rather than negative findings although the latter is also relevant.

constrained by journal word limits, some of the practical constraints underpinning a positivity bias will also be dealt away with.

As it is becoming more and more widely recognised in health research that there are questions on which formal research is not the only way in which 'data' are collected, the super-synthesis database will allow for other forms of knowledge to be contributed, with appropriate indication of their source. For instance, patient perspectives and experiences are being shared extensively outside of a research context such as on social media and personal pages, to the extent to which research is beginning to use these as a key data source. Lay users may be encouraged to contribute these to the super-synthesis database, too.

Finally, it will be a 'clever' system. It will scan itself for areas of a great variety of findings, or distant areas with overlap in vocabulary, or highlight gaps and over-researched areas. And it better be used on big screens<sup>125</sup> ...

### **7.3. Can it work? Will we want it to?**

It is obvious that the feasibility of a super-synthesis database with such features is hugely problematic.

This is a model as per which researchers are much more involved in, and responsible for, making research amenable to synthesis efforts and/or re-analysis – both their own and that of other researchers. Such a model requires a profound change in the research culture and in its system of awards and recognition. Contributing to such a database should become just as important as publishing a research paper. It will be a matter of positive incentives, but not only. It will also be a matter of the extent to which the disruptions and destructions following from the implementation of such an enterprise are amenable to assimilation, whether voluntary or forced by other actors. Most current practices of research synthesis will need to be dismantled, yet they have

---

<sup>125</sup> I mean this in a very literal way. Certain things need to fall in the same visual field so that they can be connected – just as in Köhler's experiments where the wooden boxes, stick and bananas had to come within the same visual field so that the chimps retrieve the bananas.

become deeply rooted in health and other applied sciences research. They are associated with a powerful infrastructure, culture and mentality; spread of knowledge and skills; and external support. The level of interconnectedness between research areas will begin to show its dangers, too – the precariousness of fundamental assumptions of certain areas of work, invisible because facts that challenge them are too far off, will become obvious. The commitment to non-essentialist ideas and the merging of scientific research with critical scientific and meta-scientific thinking will be unsettling as they break down our (easy) confidence in science. The degree of openness about one's research practices demanded can be quite uncomfortable, too. The messiness and imperfection of research will be much more visible. If all these are ways of improving research or being realistic about it, this is good. But if it discredits research relative to factors that deserve to be much more strongly discredited, it is not.

Finally, the smoothness of the descriptions above should not hide the fact that we are talking about a highly complex enterprise, both conceptually and technologically. The work needed to go in specifying the desiderata for the system and its conceptual underpinnings (e.g. classification systems) seems enormous, even before the feasibility of various technological solutions is taken into account.<sup>126</sup> This is also a system that should be accessible and editable by multiple users, the majority of whom should be fully identifiable and bear formal responsibility for the contents of their contributions. The system should have different levels of editing rights and clear lines of approval. The conceptual structure under which the data are entered is likely to be highly dynamic, at least in initial stages. That said, these are not issues without a sufficient number of precedents for effective solutions in our day and age. Feasibility is problematic, but no unique practical hindrance has been identified.

What would the benefits of such super-synthesis databases be? They will enable the gradual accumulation of ideas and findings from fields, resources and studies in numbers that are incomparably larger than those in any of the current synthesis models. A greater linkage, more varied use and extensive re-

---

<sup>126</sup> In actual practice they cannot but be developed concurrently, with each of them taking the requirements and limitations of the other into account.

use of information will become possible – the same set of findings will have the chance to contribute to a much broader range of synthesis studies and to receive support or be challenged from many more directions. Innovative or narrowly specialised work which is not visible in the mainstream or outside of its discipline will become better connected. Such increased use of findings and ideas and increased degree of interconnectedness amongst them will have important heuristic, financial and ethical implications. The super-synthesis databases will hold much more ‘complete’ information from studies than any research paper. It will also hold it in a highly structured way. This will enable automatic juxtaposition and comparison of information along a broad range of parameters which no paper collection or data depository can achieve. Added to the fact that simply ‘more’ information will be available, these features will enable transformations in bringing information together that are much more rigorous and creative than any of the current models of research synthesis can allow. Super-synthesis databases will enable a much greater degree of critical analysis of findings than that in a standard appraisal for the purposes of research synthesis (let alone a standard peer review process). They will make the meta-knowledge – criticisms, deconstructions, authors’ concerns, etc. – sit side by side with the ‘positive knowledge’. Issues raised in meta-work will thus be easier to take into account or challenge. Super-synthesis databases will also make it easy to identify where enough knowledge has accumulated for qualitatively new conclusions and where the research gaps and disconnectedness lie. The very process of using them may help researchers achieve greater reflexivity, learn about related areas (if, indeed, they are prompted to draw connections with other research) and achieve closure or clarity of new directions. This is a completely different way of performing research synthesis work, which also has profound implications for the ways in which primary research is conducted and shared.

Do I urge for movement in this direction? Yes, I do. I do not feel that any synthesis study which is underpinned by lower than this level of availability, interconnectedness and critical appraisal of information can be taken as sufficiently credible. In light of findings from this study, I see our current syntheses as having only a slightly greater likelihood of capturing ‘truth’ than

chance bestows, although for lack of anything better, I will continue using them as a preferred source of information. Do I see super-synthesis databases as a way of reaching some ultimate knowledge of a small number of fundamental laws in health and medicine and a greater confidence in our knowledge? No, this is not part of the vision.

This is a model of research which will be showing the tiniest differences between apparently similar concepts, samples, contexts, questions, etc. In every moment of bringing information together we will be aware of the large number of assumptions on the basis of which we are combining information and reaching higher-level, supposedly more generalisable, conclusions. We will be able to easily vary those assumptions and in a good number of cases our findings will vary, too. That which is currently hidden in our processes of transformation will no longer be. Seeing low-level differences and the difference making potential of differences will hardly make us more secure in our knowledge. But I doubt it will make us less secure. Retrospectively, we will see how unjustifiably confident we were before in our synthesis work and findings. But in a moment, we will continue to believe, or know, that we are doing the best we can and the assumptions on the basis of which we are bringing evidence together are the most reliable we have at the moment. Yet it should be the case that, without getting ultimate knowledge or being more justifiably secure in our knowledge, we will manage to deal more effectively with a large variety of questions in health research, as the rigour of our transformation processes and the variety of alternatives we have tried before settling for one will have increased dramatically.

How can we begin to work in this direction? The first and foremost thing to do is to obtain the support of the Cochrane Collaboration. This is to be followed by selecting pilot topics and bringing together teams of synthesis researchers, information specialists, researchers who explore issues related to the topic of interest in a variety of areas as well as researchers working in the meta-scientific fields – partly to develop the desiderata for the database, partly to map and connect their fields of study.

What should the relevant meta-scientific research be so that work on such a super-synthesis database is enabled? I would suggest that it should focus primarily on biomedical and other health-related ontologies, operationalisation of thinking about matrices, and work on breaking down studies and findings into elements that can become building blocks of the data collection so that it is valid and reliable and so that those building blocks can then be effectively linked, retrieved, compared and analysed.

How likely is it that such super-synthesis databases become a reality? It is one of those occasions when I think history weighs too heavily, systems are too strong, and resistance comes from highly cherished beliefs and statuses and is thus close to insurmountable. I am pessimistic that such type of work can get off the ground. Doubt will keep it a vision only. At the very least, a long time will pass before we can reach such a stage. In my view, the aspect of this vision we can realistically hope to achieve and which will bring about 'good' well worth bringing about is to integrate more closely the scientific with the meta-scientific in health research synthesis, both in thinking about research synthesis and in synthesising actual research.

Maybe I am wrong to be pessimistic. Maybe at this point, at the end of an investigation which should have suggested something practical and easier to do, and on this grey day of November, my trust in visions of a radically better science cannot come to life. Doubt me. Doubt me if you can.

That said, the world, including the scientific world, is full of surprises ...



\*\*\*\*\*

The window-screen sifts the blue cumulus  
From my cigarette.  
I sit and ponder the big synthesis but the answer  
I just forget.

Constellated at my feet *papers* pepper  
And salt the truth  
I taste in them nevertheless as I squirm and shift  
From the rage to the ruth

Of the somewhat helpless who is and who is not  
Quite able  
To say "Just kill the witches and you'll be a lot  
More comfortable."

And against much advice I read in the Bible  
A love-lorn column  
With wish then to use such passion and scruple without  
Getting solemn

And queasy, but can't do it, and the mixed-up  
Jig-saw  
Puzzle cut from the Garden and scattered in a world  
For the cats-paw

Mankind to put together in the pain of its truth  
Stays shining  
More dear, derelict, strict though asunder  
Than all our repining

Has learned. I must hope that solution mankindled  
May find breath

Wiser than any we know if the solving's a matter  
For life, not death.

Gray Burr. Garden Puzzle.

The original had “Constellated at my feet *newspapers* pepper/ And salt the truth” rather than “papers”. I could not resist the transformation. I believe I can back it up. And I have left a clear audit trail for you to agree or disagree with.

## Acknowledgements

*It is their work, too, so you do not thank them.* This is roughly what Umberto Eco says about thanking your supervisors in the thesis acknowledgements. It is their work, too – I could not agree more. But I do not see a valid argument leading from the premise to the conclusion. My biggest “thank you’s” are for Professor John Dupré and Dr Susan Kelly. If I had made a wish list of what my supervisors should be like, I would have chosen worse. I am very grateful for all their feedback, advice and encouragement. I am also grateful for the freedom they gave me to experiment. Precious little comments of theirs have shifted many habits of thought and style. John’s insistence on examples gave life to many pages of abstraction, the empirical grounding for which was known only to me. I also relaxed that ‘I need not be showing that I’ve read it all’, nor to have read it all. Many more sentences now begin with what is important in them rather than end with it because of a comment about reverse sentences by Susan. (Not sure about this one though!) They supported me invariably for ‘other’ things I wanted to do and apply for during the PhD years. Last but in many ways first, it has been inspiring to see how razor sharp minds can go with kindness and openness, a sense of humour, and creative responses to rules that hamper rather than help. I have been exceptionally lucky with my supervisors.

My mentor, Dr Hannah Farrimond, has given me much down-to-earth yet extremely helpful advice. She was able to see right into the many ways I could devise to make my life more difficult than unnecessary. And it became easier. Recently, we have been talking extensively about women in academia. It has been enormously motivating.

My examiners, Professor Nancy Cartwright and Professor Nicky Britten, have helped improve this work more than I thought possible at this stage of the research. Thanks to their advice, it has much more integrity than originally – it states more often what I have done rather than what I intended to do and is more transparent about its starting assumptions, including the value I had ascribed to being transparent. The stylistic advice on reducing the use of passive voice made many difficult passages more readable and dynamic. I also

liked the feeling of owning more of what I have done. It has been a privilege to have academics whose work I have enjoyed reading so much to read mine.

I would next like to thank the Economic and Social Research Council for funding me, through a grant attached to Egenis, the ESCR Centre for Genomics in Society. I am grateful to the decision makers who found promise in my unusual proposal and background. But I am also grateful to the invisible, let's say, Jane, Peter, Esther and Nathaniel – the taxpayers – who made this possible. Having come back to earning and paying taxes a few months ago, I remembered that you really want them well used.

I owe much to my friends and colleagues at Egenis and the broader department, and especially to Ann-Sophie Barwich, Pietro Berti, Louise Bezuidenhout, Daniele Carrieri, Antonio Carvallo, Jo Donaghy, Chris Elphick, Trijsje Franssen, Kate Getliffe, Jean Harrington, Anna Harris, Pierre-Olivier Méthot, Ginny Russell, Kai Wang and David Wyatt. I always felt very welcomed and have some wonderful memories of downtime together. Each has taught me something, in his or her unique way. It was also a curious law that every time my motivation would waver, I would hear one of them talk about their project. Their enthusiasm would shine through. It will catch on within hours. They would forgive me for saying that every time I saw or heard them anxious, over-stressed and over-worked, I remembered what I do not want to do. I hope I have been a help and mirror to them, too.

I have learnt much from my senior colleagues at Egenis and the broader department. The Egenis seminars have long been a place of questions to dread. I could see how but I could see the point, too. Intellectual atmosphere aside, my senior colleagues and the administrative staff at Egenis have also made me feel very welcomed and supported. I am particularly grateful to Dr Sabina Leonelli. Reading her work and attending the events on data-driven science she organised have helped me immensely with my work. We seem to have many interests in common. We also seem to have many interests, period. There was never enough time to talk 'properly'. Staffan Müller-Wille still has my admiration for a person 'who knows it all' (be it sometimes darkly!). They may not necessarily know it – I would often just listen – but I have learnt precious things from Professor Barry Barnes, Dr Christine Hauskeller, Professor Steve

Hughes, Dr Dana Wilson-Kovacs, Dr Maureen O'Malley, Dr Michael Morrison, Professor Andy Pickering and Professor Brian Rappert. Sometimes it was factual knowledge and lines of argument. Maybe more often it was how to be an academic with one's own curious idiosyncrasies. Laura Dobb, Sue Harding, Claire Packman, Natasha Simcock, Cheryl Sutton and Chee Wong, our administrative and communications staff at various points, have helped me with much paperwork, navigation of rules and random queries. Amy Rager from the Graduate School has been very helpful with squeaky administrative wheels.

In a world where quality, time and money are so inter-related, I am indebted to Georgi Iliev. He helped me achieve the quality of work I was aiming at when Jane, Peter, Esther and Nathaniel, the invisible taxpayers, could no longer do. He has been a close and truly supportive friend for many years now. He is probably also the one who would ask me the most specific questions about the PhD and listen through my overly specific answers.

I am very grateful to David Wyatt, Dr Paul Sutcliffe and Dr Dimitar Angelov who read and commented on chapters of my dissertation. All of them have a remarkable eye for detail, clarity and nuances of language. Paul has been a colleague, mentor and friend in my last couple of years as a researcher at Warwick. I owe much of my confidence to him as well as many tangible outcomes of my work. He, along with Dr Janice Koistinen, also a friend and colleague from Warwick, gave me invaluable advice when I was leaving to start my PhD. I listened. I started writing from the very beginning. Before it was too late, I stopped reading and focused on my own argument. Most of my PhD work went in a smooth flow because of their advice. Much of what I know about academic writing I owe to Professor Bill Fulford and Professor Jeremy Dale, who have directed my work at Warwick. For years they have been trying to teach me to stop apologising that this is not the best work in the world and destroying it verbally (and, in that respect, be more like a male than female medical student); have a strong argument that fits on a single page, single-spaced; cut yet more detail ... Still much to learn but they will recognise an improvement!

Many notches of the experimental spirit and sincerity of this work I owe to Michael Carhart, a friend who is in a friendship category of its own. This work

would have been braver and more direct than my previous work just by a logic of internal development. But it would never have been those things to its current extent, be it sufficient or not, if it were not for many extraordinary conversations and silences with Michael. It would have had less poetry in it, too. And just as with those conversations and silences, I do not know where the beginning and end of this thesis are.

My immediate family – Krastina Manolova, Petar Petrov and Borislav Petrov, my cousins Yolina Stoyanova-Brunner and Konstantina Stoyanova-Poix, and my closest friends – Dennitza Ivanova, Iliyana Hristova and Gloria Ayob have always trusted in me, my abilities and work much more than objectively deserved. Each of them has supported me in their own way – whether in our online coffee mornings each Sunday; by joking that I can go to their funeral only with a PhD title and that they then have many years to live; by sometimes asking about, sometimes completely ignoring my PhD; by saying many ‘well done’s’ and ‘I was telling about you to so and so’; by taking my mind off with their news and stories. I am particularly grateful to my mum for putting up with me for several months in my most intense period of writing up, 16 years after I left home. I was far from being good company. I was obsessed with my data analysis. I would leave all housework to her. I was also on the verge of deciding whether to let my heart go broken or whether the other person is funnily irrelevant to one’s love. I went for the latter. Worst of all, I made her party to a mega-project of reorganising the books in our library and ‘setting free’ the unused ones. This took us six weekends, much dust and chaos, and some drama around where Shakespeare should go since we couldn’t ‘multi-code’ him. I had forgotten it is her home now. But the smell of coffee she makes in this house is still a smell which makes me feel at home in every nice café, anywhere in the world.

I try to remind those around me that no matter how busy I am, people always come before work for me. I hope they knew that, in the bigger picture, the PhD is just a phd. I have made it wait without a speck of hesitation when I felt my help or presence were needed. But I’m afraid that my sensitivity has been dumbed down. I also suspect that they have been limiting the information flows to give me space and time to write. It is a PHD after all. I am deeply grateful to them. Now I will have the time to find out.

## Appendices

### Appendix to Chapter 2

#### 1. Strategies for the bibliometric searches in PubMed presented in Table T 2.1

The bibliometric searches I performed are only illustrative. Much more refined strategies and thorough analysis of the retrievals are needed to obtain accurate numbers. The comments accompanied by \*\*\*\*\* indicate findings which limit the drawing of clear conclusions and suggest that more in-depth investigation is required.

1. ("cost effectiveness" OR cost-effectiveness) AND Bayesian  
Limited to 'review'
2. "Bayesian meta-analysis" OR "Bayesian metaanalysis" OR "Bayesian meta analysis" OR "Bayesian synthesis"
3. "case survey" OR "quantitative case survey"  
Limited to 'review'
4. "comparative case study"  
Limited to 'review'
5. "decision modelling" OR "decision modeling"  
Limited to 'review'

\*\*\*\*\* *Combinations with "comprehensive" not found.*

6. "confidence profile"
7. "content analysis"  
Limited to 'review'

\*\*\*\*\* *"Content synthesis" retrieved primarily false positives (14 vs. 2 true positives). Added to retrieval for "content analysis".*

8. "critical interpretive synthesis"

9. "cross design synthesis" OR "cross-design synthesis"

*\*\*\*\*\* No hits for "grouped meta-analysis" OR "grouped metaanalysis" OR "grouped meta analysis".*

10. "ecological triangulation" OR "ecological sentence synthesis"

11. EPPI-approach OR "EPPI approach" OR  
EPPI-review OR "EPPI review" OR (EPPI AND review)

12. Estabrooks, CA  
Limited to 'review'

*\*\*\*\*\* The names of the three authors together were not found.*

13. "framework synthesis"

"framework analysis"  
Limited to 'review'

*\*\*\*\*\* "Framework synthesis" retrieved fewer true positives than false positives: 7 true positives out of 28 hits for all years and 6 true positives out of 18 hits for last 5 years. In Table T 2.1 these were added to the retrieval for "framework analysis" limited to 'review', which was 18 publications for all years and 12 for last 5 years.*

14. "grounded theory" OR "grounded formal theory"  
Limited to 'review'

15. hierarchical model\*  
Limited to 'review'

16. meta-ethnography OR metaethnography OR "meta ethnography"

17. meta-interpretation OR metainterpretation OR "meta-interpretation"



18. meta-narrative OR metanarrative OR “meta narrative”

19. "meta-needs assessment" OR "metaneeds assessment" OR “meta needs assesement”

20. meta-study OR metastudy OR “meta study”

21. cross-case OR “cross case”

Limited to ‘review’

*\*\*\*\*\* “Meta-matrix” and spelling variations retrieved no true positives. Miles AND Huberman, limited to ‘review’ returned only 1 publication related to thematic analysis in Miles & Huberman.*

22. "multiparameter evidence synthesis" OR "multi-parameter evidence synthesis"

23. "narrative summary"

24. “narrative synthesis”

25. "qualitative comparative analysis"

26. qualitative [ti] AND meta-analysis [ti] OR "qualitative meta-analysis"

27. "qualitative metasummary" OR "qualitative meta-summary" OR “qualitative meta summary”

*\*\*\*\*\* The last parameter – “qualitative meta summary” was subsequently excluded. The way in which the search query was translated by the database, as a way of compensating for the fact that the exact phrase was not found, led to the retrieval of a large number of false positives.*

28. meta-synthesis OR metasyntesis OR “meta synthesis”

29. “realist synthesis”

30. "textual narrative synthesis"

\*\*\*\*\* *The phrase was not found but a small number of abstracts (7 all, 4 last 5 years) were retrieved that contained all three words.*

31. 1) "thematic analysis"

Limited to 'review'

2) "thematic synthesis"

3) 1) OR 2).

## **2. Further issues with the meta-methodological discussion of units of analysis/ synthesis**

An illustration follows of difficulties into which even extended and persuasive discussions of the units of analysis/synthesis in research synthesis studies run. One such discussion is found in Greenhalgh et al., 2005. It is very helpful in articulating a range of units of information in research synthesis studies. It is also illuminating through demonstrating many of the complex relationships between such units (e.g. larger units of synthesis containing and giving meaning to smaller units or being constituted by them). Those persuasive descriptions do, however, start appearing shaky once probed.

The authors "took as [their] initial unit of analysis *the unfolding 'storyline' of a research tradition over time*":

*We took the research tradition as the initial unit of analysis for our review. In other words, before identifying, evaluating and comparing data from primary research studies, we first embarked on a systematic mapping phase to collect and compare the different over-arching storylines of the rise and fall of diffusion research that we judged relevant to our overall research question (op. cit.: 419).*

At a most basic level, "a piece of evidence" is nevertheless seen as the primary unit of analysis:

*The review process was thus somewhat laborious, since each piece of*

*evidence had to be double-handled — first for constructing the meta-narrative within its own tradition and again for contributing to the ‘rich picture’ of one of the seven dimensions of diffusion of innovations. But the double handling served a crucial purpose, since each piece of evidence was first interpreted within its own paradigm before being added to other evidence generated in a different paradigm. Conflicting findings could thus be explored in terms of contestation between incommensurable paradigms (423).*

Contradictions, thus identified, can themselves become a piece of evidence or at least datum:

*In this way, the many contradictions we were finding in our sources could be turned into data and analysed systematically ... thus allowing us to go beyond statements such as ‘the findings of primary studies were contradictory’ or that ‘more research is needed’ (423).*

Overall, the following picture seems to be emerging from Greenhalgh et al. paper when reconstruction of the meta-narrative process is attempted from the perspective of its basic units of information. Pieces of evidence, the smallest unit of information in a synthesis study, are first brought together to construct the meta-narrative of a particular research tradition, the designated “initial unit of analysis”. This meta-narrative is both a synthesis product and a building block for further synthesis. As a building block for further synthesis, it has two functions. First, (along with the other meta-narratives) it provides a set of “dimensions” of the topic of interest (in this case diffusion of innovation) that are shared across meta-narratives and around which the over-arching synthesis is constructed. Second, it provides the context of meanings within which the available evidence is re-analysed and re-synthesised for the final over-arching synthesis.

Clearly, this account is problematic, whether because of the manner of reconstruction or because of the claims on which the reconstruction is based. There is circularity in suggesting that that which is used to build a meta-narrative (“each piece of evidence”) then draws its (partly new) meaning from

the meta-narrative it helps to build, and it needs to be demonstrated whether it is a virtuous or vicious circularity. There is inconsistency in claiming incommensurability of paradigms – these are seen as incommensurable when findings are conflicting, but otherwise share key dimensions and evidence. There is an interesting suggestion about ‘contradictions’ becoming pieces of evidence, or at least some sort of data. If contradictions can turn into data, what other relationships between pieces of evidence can acquire the same status?

## Appendix to Chapter 4

### Search strategy used to generate the pool from which papers were sampled for the case study

The “brief values filter” (Petrova et al., 2012) is as follows:

1. Attitude\* (tw)
2. Perceptions (tw)
3. Qualitative (tw)
4. Coping (tw)
5. Counseling (tw)
6. Cultural (tw)
7. Ethics (tw)
8. Experiences (tw)
9. Interviews (tw)
10. Perceived (tw)
11. Personal (tw)
12. Professionals (tw)
13. QOL (tw) OR Quality of Life (MeSH)
14. Relations (tw)
15. Respondents (tw)
16. Satisfaction (tw)
17. Staff (tw)
18. Well-being (tw)
19. Adaptation, Psychological (MeSH)
20. Nurse’s Role (MeSH)
21. Social Support (MeSH)
22. OR/1-21

The Mind-Body search strategy used was:

1. Mind-Body Relations (Metaphysics) [MeSH]
2. Mind-Body Therapies [MeSH]
3. Psychosomatic Medicine [MeSH]
4. Psychophysiologic Disorders [MeSH]

5. Psychophysiology [MeSH]
6. OR/1-5

**Interface:** PubMed (over 20 million publications around the time of running the search, currently over 22 million)

**Date** search run – 22 November 2010

**Without year limits**

#1 Values search strategy retrieved 1,340,184 citations

#2 Mind-body search strategy retrieved 564,417

Together (combined with OR so as to exclude duplication), these come up to 1,794,126.

Of these, 1 229 709 (91.8% of retrieval of the Values search strategy) were unique to the Values search strategy and 453 942 were unique for the Mind-Body search strategy (80.4% of the Mind-body search strategy).

#3 Neoplasms [MeSH] retrieved 2, 177, 340

(#1 OR #2) AND #3 retrieved 98, 931

This represented 4.5% of the citations on cancer; 7.4% of the Values citations and 17.5% of the Mind-Body citations.

**Applying limits – last three years (June 07 – June 10)**

The search was run in November 2010. An earlier end date was chosen to ensure that all selected publications were indexed with Medical Subject Headings (MeSH), as some of the analysis parameters required exploration of these.

(#1 OR #2) AND #3, add year limits 18,456

(18.7% of the research of interest, according to the searches, has been done in the last three years)

#1, add year limits 233,517

(17.4% of the Values citations retrieved without year limits)

#2, add year limits 78,670

(13.9% of the Mind-Body citations retrieved without year limits)

#3, add year limits 268,937

(12.4% of the Neoplasms citations retrieved without year limits)

## Appendix to Chapter 5

### 1. Frequency of codes in terms of the number of papers from which they have “references”

“References” is the NVivo term for units of information coded under a particular code.

Number of papers	Number of codes with references from this number of papers	Percent	Cumulative Percent
0	505	33,4	33,4
1	761	50,3	83,6
2	158	10,4	94,1
3	40	2,6	96,7
4	12	,8	97,5
5	15	1,0	98,5
6	10	,7	99,1
7	6	,4	99,5
8	1	,1	99,6
10	4	,3	99,9
11	2	,1	100,0
Total	1514	100,0	



## 2. Frequency of codes in terms of the number of references within them

Number of references	Number of codes with that number of references	Percent	Cumulative Percent
0	505	33,4	33,4
1	511	33,8	67,1
2	207	13,7	80,8
3	101	6,7	87,5
4	41	2,7	90,2
5	34	2,2	92,4
6	23	1,5	93,9
7	16	1,1	95,0
8	19	1,3	96,2
9	14	,9	97,2
10	8	,5	97,7
11	5	,3	98,0
12	10	,7	98,7
14	3	,2	98,9
15	4	,3	99,1
16	2	,1	99,3
17	1	,1	99,3
18	3	,2	99,5
19	1	,1	99,6
22	1	,1	99,7
25	1	,1	99,7
28	1	,1	99,8
36	1	,1	99,9
47	1	,1	99,9
53	1	,1	100,0
Total	1514	100,0	

### **3. Sampling of findings for formalisation with the Vocabulary of Elements of Findings**

#### **For findings from NVivo**

A finding was picked from each of the branches of the coding tree, from the highest to the lowest level.

#### **For findings from the EPPI-Reviewer**

A finding was picked from each of the branches of the coding tree with the exception of the lowest sub-branches within a coding set, provided they were highly similar in topic and structure and from the same study.

This was the case relatively frequently in EPPI and more so in comparison to NVivo, as findings were coded at a greater degree of granularity in the former. In such cases sub-branches were sampled, too (see below).

#### **For finding from Excel**

Findings were sampled from each worksheet (28), with each worksheet reflecting a separate category of findings (e.g. findings about cancer type of participants, hypothesised risk factors, phenomenological experiences, etc.)

#### **Sampling of findings when there was more than one finding within a code or worksheet**

- 1. If 2 findings within a code or worksheet, use ‘ordered alteration’ to select one finding** (e.g. Branch 5 – finding 1 of 2, Branch 8 – finding 2 of 2, Branch 9 - finding 1 of 2).<sup>127</sup>
- 2. If 3 findings within a code or worksheet, use ‘ordered alteration’ to select one finding** (e.g. Branch 4 – finding 1 of 3, Branch 6 – finding 2 of 3, Branch 10 - finding 3 of 3, Branch 13 – finding 1 of 3).

---

<sup>127</sup> The branches are not necessarily consecutive to reflect the fact that there were branches with a single finding in them, to which the above procedures were not applied.

**3. If 4 or more findings within a code or worksheet, select one from each represented study.**

3.1. If the number of findings within a represented study is 2 or 3, proceed as in 1 and 2 above.

3.2. If the number of findings within a represented study is 4 or larger, query <http://www.random.org/integers/> for a single random number and select respective finding.

**Sampling from lowest branches in EPPI-Reviewer**

When, at the lowest level, branches reflected specifications of a common thematic content (e.g. “maintaining a normal life” – through going back to work, engaging in social activities one used to enjoy, moving about, etc. in a long list) and when they concerned a single study only, one finding was chosen through the ‘ordered alteration’ described above or the random numbers generator.

**Sampling of findings in cases a finding selected as above is ‘composite’ (that is, contains more than one sentence or a large number of data points):**

1. If a composite finding with more than 2 propositions or data points of identical, highly similar or already well represented structure, code 2 of those propositions or data points.
2. If a composite finding with more than 2 propositions or data points of somewhat different, but not particularly interesting structures, code 3 of those.
3. If a composite finding with more than 2 propositions or data points of different and rarely or not encountered structure, code until saturation.
4. If a composite finding where some of the elements are ‘other’ propositions, different from the specification of a ‘conventional finding’ (Chapter 5, Section 5.1.1), code up to 2 of these – to explore the specificity of the Vocabulary and also potential non-findings that can be extracted as findings in synthesis studies.

## Separating ‘conventional findings’ from ‘other’

There were a number of items that were sampled as findings (as they were found in the Findings/Results sections of studies) but were excluded as per the definition of ‘conventional finding’. These included sentences on methods; hypotheses, explanations and predictions that were formulated on the basis of or in relation to the study findings but were clearly identifiable as hypotheses, explanations and predictions and only minimally expressed a study’s empirical contents; background pragmatic information, etc. For example:

*These findings raise the emergent hypothesis of a conserved neurobehavioral symptom complex, which results from diverse triggering insults* (Study 6 – hypothesis formulated on the basis of the study findings which are not, however, described in the sentence).

*This could, at least in part, explain why chemotherapy was a significant predictor of work change* (Study 5 – explanation that only minimally mentions the study findings).

At the same time, there were untypical findings (rather than other types of claims) which were left in the sample. These included meta-findings – findings about data, findings, methods and a research field; findings from other studies as in reviews or from background and discussion sections of a primary study, and typical findings mixed with ‘other’ claims. For example:

*Although the meaning of the VAS scale was explained, all patients seemed to find it difficult to translate their shortness of breath to a VAS scale* (Study 14 – finding on method effectiveness rather than a substantive finding).

Through applying those inclusion-exclusion criteria, the final sample of findings analysed for composition and structure using the Vocabulary of Elements of Findings was reached. It consisted of 301 items.

## References

- Achinstein, P., 2001. *The book of evidence*. Oxford: Oxford University Press.
- Ades, A. and Sutton, A., 2006. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society*, 169, pp. 5-35.
- Agency for Healthcare Research and Quality, 2013. *SRDR: Systematic Review Data Repository*. [online] Available at: <<http://srdr.ahrq.gov/>> [Last accessed 31 May 2013].
- Al-Rowaili, A., Al-aeel, S. A., Al-Naim, L. S. and AL-Diab, A. I., 2009. Appropriateness of Cancer Pain Management in Saudi teaching Hospital. *The Gulf journal of oncology*, 5, pp. 37-43.
- Alley, M., 1996. *The Craft of Scientific Writing*. New York, etc.: Springer.
- Alderson, P. and Green, S. eds., 2002. *Cochrane Collaboration open learning material for reviewers*. Version 1.1. [pdf] Available at: <<http://www.cochrane-net.org/openlearning/PDF/Openlearning-full.pdf>> [Last accessed 17 May 2013].
- American Cancer Society, 2012. *Lifetime Risk of Developing or Dying From Cancer*. [online] (Last updated 29 November 2012) Available at: <<http://www.cancer.org/cancer/cancerbasics/lifetime-probability-of-developing-or-dying-from-cancer>> [Last accessed 26 May 2013].
- Anderson, L. M., Petticrew, M., Rehfuss, E., Armstrong, R., Ueffing, E., Baker, P., Francis, D., Tugwell, P., 2011. Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods*, 2, pp. 33-42.
- Andrykowski, M. A., Donovan, K. A. and Jacobsen, P. B., 2009. Magnitude and Correlates of Response Shift in Fatigue Ratings in Women Undergoing Adjuvant Therapy for Breast Cancer. *Journal of Pain and Symptom Management*, 37, pp. 341-51.
- Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F. and Chalmers, T. C., 1992. A Comparison of Results of Meta-analyses of Randomized Control Trials and Recommendations of Clinical Experts: Treatments for Myocardial Infarction. *The Journal of the American Medical Association*, 268, pp. 240-8.

- Ashcroft, R. E., 2004. Current epistemological problems in evidence based medicine. *Journal of Medical Ethics*, 30, pp. 131-5.
- Austin, J. L., 1962. *Sense and Sensibilia*. Oxford: Oxford University Press/Clarendon.
- Ayob, G., 2008. Do People Defy Generalizations? Examining the Case against Evidence-Based Medicine in Psychiatry. *Philosophy, Psychiatry, and Psychology*, 15, pp. 167-74.
- Bagshaw, S. M. and Bellomo, R., 2008. The Need to Reform Our Assessment of Evidence from Clinical Trials: A Commentary. *Philosophy, Ethics, and Humanities in Medicine*, 3, pp. 1-11.
- Banning, J. (n.d., 1993?) *Ecological Sentence Synthesis*. [pdf] Available at: <<http://mycahs.colostate.edu/James.H.Banning/PDFs/Ecological%20Sentence%20Synthesis.pdf>> [Last accessed 26 May 2013].
- Banning, J. (n.d.) *Ecological Triangulation: An Approach for Qualitative Meta-Synthesis*. [pdf] Available at: <<http://mycahs.colostate.edu/James.H.Banning/PDFs/Ecological%20Triangulation.pdf>> [Last accessed 26 May 2013].
- Barbour, R. S. and Barbour, M., 2003. Evaluating and synthesizing qualitative research: the need to develop a distinctive approach. *Journal of Evaluation in Clinical Practice*, 9, pp. 179-86.
- Barnett-Page, E. and Thomas, J., 2009. Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology*, 9.
- Bath-Hextall, F. J., Jenkinson, C., Humphreys, R. and Williams, H. C., 2012. Dietary supplements for established atopic eczema. *Cochrane Database of Systematic Reviews*, 2, Art. No.: CD005205, DOI: 10.1002/14651858.CD005205.pub3.
- Becker, L. A. and Oxman, A. D., 2011. Overviews of reviews. In: J. P. T. Higgins and S. Green, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. [online] The Cochrane Collaboration. Available at: <[www.cochrane-handbook.org](http://www.cochrane-handbook.org)> Ch.22. [Last accessed 17 May 2013].
- Beecher, H. K., 1955. The powerful placebo. *Journal of the American Medical Association*, 159, pp. 1602-6.

- Bennett, B., Friedlander, M. and Lloyd, A., 2007. The Experience of Cancer-Related Fatigue and Chronic Fatigue Syndrome: A Qualitative and Comparative Study. *Journal of Pain and Symptom Management*, 34, pp. 126-35.
- Bogen, J. and Woodward, J., 1988. Saving the Phenomena. *The Philosophical Review*, 97, pp. 303-52.
- Bohlin, I., 2012. Formalizing Syntheses of Medical Knowledge: The Rise of Meta-analysis and Systematic Reviews. *Perspectives on Science: Historical, Philosophical, Social*, 20, pp. 273-309.
- Bolton, D., 2008. The Epistemology of Randomized, Controlled Trials and Application in Psychiatry. *Philosophy, Psychiatry, and Psychology*, 15, pp. 159-65.
- Booth, A., 2011. Does “systematic” mean “comprehensive”? Issues in Sampling, Searching and Synthesis for Qualitative Systematic Reviews. [presentation] *IHSR Qualitative Research Group Seminar Series*. Veysey Building, Peninsula College of Medicine and Dentistry, Exeter, 3 October 2011.
- Booth, A., Papaioannou, D. and Sutton, A., 2012. *Systematic Approaches to a Successful Literature Review*. Los Angeles, etc.: Sage.
- Boyle, P., 2012. The ‘Big Data’ Debate. [presentation] *The ‘Big Data’ Debate – ESRC Festival of Social Science Panel Discussion*. The British Academy, London, 5 November 2012.
- Braude, H. D., 2009. Clinical Intuition versus Statistics: Different Modes of Tacit Knowledge in Clinical Epidemiology and Evidence-Based Medicine. *Theoretical Medicine and Bioethics: Philosophy of Medical Research and Practice*, 30, pp. 181-98.
- Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M. and Pill, R., 2002. Using meta ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research and Policy*, 7, pp. 209–15.
- Burr, G. (n.d.) Garden Puzzle. In: P. Leary and R. Kelly, ed. 1965. *A Controversy of Poets: An Anthology of Contemporary American Poetry*. Garden City, NY: Anchor Books, p. 39-40.
- Callebaut, W. Scientific perspectivism: A philosopher of science’s response to the challenge of big data biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, pp. 69-80.

- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. and Donovan, J., 2003. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science & Medicine*, 56, pp. 671–84.
- Cancer Research, 2011. *What cancer is*. [online] (Last updated 3 November 2011) Available at: <<http://www.cancerresearchuk.org/cancer-help/about-cancer/what-is-cancer/cells/what-cancer-is>> [Last accessed 26 May 2013].
- Cartwright, N., 1983. *How the Laws of Physics Lie?* Oxford and New York: Oxford University Press/Clarendon.
- Cartwright, N., 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Cartwright, N. and Stegenga, J., 2011. A Theory of Evidence for Evidence-Based Policy. *Proceedings of the British Academy*, 171, pp. 291-322.
- Cartwright, N., 2011. A philosopher's view of the long road from RCTs to effectiveness. *The Lancet*, 377, pp. 1400-1.
- Casarett, D., Fishman, J., O'Dwyer, P. J., Barg, F. K., Naylor, M. and Asch, D. A., 2008. How Should We Design Supportive Cancer Care? The Patient's Perspective. *Journal of Clinical Oncology*, 26, pp. 1296-301.
- Centre for Reviews and Dissemination, 2009. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. [pdf] York: CRD, University of York. Available at: <[http://www.york.ac.uk/inst/crd/pdf/Systematic\\_Reviews.pdf](http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf)> [Last accessed 17 May 2013].
- Chalmers, I., Hedges, L.V. and Cooper, H., 2002. A Brief History of Research Synthesis. *Evaluation & the Health Professions*, 25, pp. 12-37.
- Chang, H., 2011. The Philosophical Grammar of Scientific Practice. *International Studies in the Philosophy of Science*, 25, pp. 205-21.
- Charmaz, K., 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London, etc.: Sage.
- Churchland, P. M., 1979. *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.
- Churchland, P. M., 1982. The anti-realist epistemology of van Fraassen's *The Scientific Image*. *Pacific Philosophical Quarterly*, 63, pp. 226-35.



- Clancy, C. M. and Kamerow, D., 1996. Evidence-Based Medicine Meets Cost-effectiveness Analysis. *Journal of the American Medical Association*, 276, pp. 329-330.
- Clark, J. M. and Paivio, A., 1989. Observational and Theoretical Terms in Psychology: A Cognitive Perspective on Scientific Language. *American Psychologist*, 44, pp. 500-12.
- Clark-Grill, M., 2007. Questionable Gate-Keeping: Scientific Evidence for Complementary and Alternative Medicines (CAM): Response to Malcolm Parker. *Journal of Bioethical Inquiry*, 4, pp. 21-8.
- Cochrane, A. L., 1989 (reprint of 1972 ed.). *Effectiveness and efficiency: random reflections on health services*. The Nuffield Provincial Hospitals Trust/Cambridge University Press.
- Cochrane QES Register, 2013. [online] Available at: <<http://www.mendeley.com/groups/518691/cochrane-qes-register/papers/>> [Last accessed 26 May 2013].
- Colquhoun, D., 2006-2013. *DC's Improbable Science: Truth, falsehood and evidence: investigations of dubious and dishonest science*. [blog] Available at: <<http://www.dcscience.net/>> [Last accessed 26 May 2013].
- Colquhoun, D., 2011. Why philosophy is largely ignored by science. *DC's Improbable Science: Truth, falsehood and evidence: investigations of dubious and dishonest science*. [blog entry] 28 October. Available at: <<http://www.dcscience.net/?p=4799>> [Last accessed 26 May 2013].
- Creswell, J. W., 2003. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. 2nd ed. Thousand Oaks, etc.: Sage.
- Creswell, J. W. and Plano Clark, V. L., 2007. *Designing and Conducting Mixed Methods Research*. [Kindle edition] Thousand Oaks, etc.: Sage.
- Critical Appraisal Skills Programme, 2013. *10 questions to help you make sense of qualitative research*. [pdf] Available at: <<http://www.casp-uk.net/wp-content/uploads/2011/11/CASP-Qualitative-Research-Checklist-31.05.13.pdf>> [Last accessed 2 June 2013].
- Davey, M. M., Cummings, G., Newburn-Cook, C. V. and Lo, E. A., 2009. Predictors of nurse absenteeism in hospitals: a systematic review. *Journal of Nursing Management*, 17, pp. 312-30.

- Derkatch, C., 2008. Method As Argument: Boundary Work in Evidence-Based Medicine. *Social Epistemology: A Journal of Knowledge, Culture, and Policy*, 22, pp. 371-88.
- Dixey, R., Sahota, P., Atwal, S. and Turner, A., 2001. Children talking about healthy eating: data from focus groups with 300 9-11-year-olds. *British Nutrition Foundation Nutrition Bulletin*, 26, pp. 71-9.
- Dixon-Woods, M., Agarwal, S., Young, B., Jones, D. and Sutton, A., 2004. *Integrative approaches to qualitative and quantitative evidence*. [pdf] London: Health Development Agency. Available at: <[http://www.nice.org.uk/niceMedia/pdf/Integrative\\_approaches\\_evidence.pdf](http://www.nice.org.uk/niceMedia/pdf/Integrative_approaches_evidence.pdf)> [Last accessed 17 May 2013].
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. and Sutton, A. J., 2006. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, 6.
- Dixon-Woods, M., Fitzpatrick, R. and Roberts, K., 2001. Including qualitative research in systematic reviews: opportunities and problems. *Journal of Evaluation in Clinical Practice*, 7, pp. 125-33.
- Douglas, H., 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
- Dowe, D. L., 2008. Minimum Message Length and Statistically Consistent Invariant (Objective?) Bayesian Probabilistic Inference – From (Medical) 'Evidence'. *Social Epistemology: A Journal of Knowledge, Culture, and Policy*, 22, 433-60.
- Dupré, J., 1993. *The Disorder of Things*, Cambridge, MA: Harvard University Press.
- Dupré, J., 2001. *Human Nature and the Limits of Science*. Oxford: Oxford University Press/Clarendon.
- Dwivedi, R. C., Nutting, C. M., Rhys-Evans, P., Harrington, K.J. and Kazi, R., 2009. Future Perspectives for Health related Quality of life (HRQOL) studies in Head and Neck Cancer. *Giornale Italiano di Medicina del Lavoro ed Ergonomia*, 31, pp. B17-20.
- Dysart-Gale, D., 2008. Lost in Translation: Bibliotherapy and Evidence-Based Medicine. *Journal of Medical Humanities*, 29, pp. 33-43.

- Eaves, Y. D., 2001. A synthesis technique for grounded data analysis. *Journal of Advanced Nursing*, 35, pp. 654-63.
- Ebell, M. H., Siwek, J., Weiss, B. D., Woolf, S. H., Susman, J., Ewigman, B. and Bowman, M., 2004. Strength of Recommendation Taxonomy (SORT): A Patient-Centered Approach to Grading Evidence in the Medical Literature. *American Family Physician*, 69, pp. 548-56.
- Egger, M. and Smith, G. D., 1997. Meta-analysis: Potentials and promise. *British Medical Journal*, 315, pp. 1371-4.
- Egger, M., Smith, G. D. and Phillips, A. N., 1997. Meta-analysis: Principles and procedures. *British Medical Journal*, 315, pp. 1533-7.
- EPPI-Centre, 2013. *Welcome to the EPPI-Review 4 gateway*. [online] Available at: <<http://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4>> [Last accessed 26 May 2013].
- Ernst, E., Cohen, M. H. and Stone, J., 2004. Ethical Problems Arising in Evidence Based Complementary and Alternative Medicine. *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 30, pp. 156-9.
- Estabrooks, C. A., Field, P. A. and Morse, J. M., 1994. Aggregating qualitative findings: an approach to theory development. *Qualitative Health Research*, 4, pp. 503-11.
- Evans, D. and Fitzgerald, M., 2002. Reasons for physically restraining patients and residents: a systematic review and content analysis. *International Journal of Nursing Studies*, 39, pp. 735-43.
- Evidence-Based Medicine Working Group, 1992. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *Journal of the American Medical Association*, 268, pp. 2420-5.
- Falkum, E., 2008. Phronesis and Techne: The Debate on Evidence-Based Medicine in Psychiatry and Psychotherapy. *Philosophy, Psychiatry, and Psychology*, 15, pp. 141-9.
- Feigl, H., 1958. The "Mental" and the "Physical". In: H. Feigl, M. Scriven and G. Maxwell, eds. *Concepts, Theories and the Mind-Body Problem*. Volume II of the Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press, pp. 370-497.

- Feyerabend, P., 1975. *Against Method: Outline of an anarchistic\* theory of knowledge*. London: NLB.
- Finfgeld, D. L., 1999. Courage as a process of pushing beyond the struggle. *Qualitative Health Research*, 9, pp. 803-14.
- Finfgeld, D. L., 2003. Metasynthesis: The State of the Art – So Far. *Qualitative Health Research*, 13, pp. 893-904
- Fraser, J., 1997. *How to publish in biomedicine: 500 tips for success*. Abingdon: Radcliffe Medical Press.
- Fulford, K. M. W., 2004. Ten Principles of Values-Based Medicine. In: J. Radden, ed. 2004. *The Philosophy of Psychiatry: A Companion*. New York: Oxford University Press, pp. 205-234.
- Fulford, K. W. M., Thornton, T. and Graham, G., 2007. *Oxford Textbook of Philosophy and Psychiatry*. Oxford: Oxford University Press.
- Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M. and Donnelly, M., 2012. Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database of Systematic Reviews*, 2, Art. No.: CD008225, DOI: 10.1002/14651858.CD008225.pub2.
- Gallagher, J., for BBC News, 2012. *Breast screening advice updated amid controversy over tests*. [online] (Last updated 30 October 2012) Available at: <<http://www.bbc.co.uk/news/health-20121043>> [Last accessed 26 May 2013].
- Garcia, J., Bricker, L., Henderson, J., Martin, M. A., Mugford, M., Nielson, J and Roberts, T., 2002. Women's Views of Pregnancy Ultrasound: A Systematic Review. *Birth*, 29, pp. 225-50.
- Gilbert, R., Salanti, G., Harden, M. and See, S., 2005. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *The International Journal of Epidemiology*, 34, pp. 874-87.
- Glaser, B. G. and Strauss, A. L., 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine de Gruyter.
- Glass, G. V., 1976. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5, pp. 3-8.

- Glass, G. V., McGaw, B. and Smith, M. L., 1981. *Meta-Analysis in Social Research*. Beverly Hills and London: Sage.
- Glymour, C., 1980. *Theory and evidence*. Princeton: Princeton University Press.
- Goldberg, Y., Porat, R. M., Kedar, I., Shochat, C., Galinsky, D., Hamburger, T., Hubert, A., Strul, H., Kariiv, R., Ben-Avi, L., Savion, M., Pikarsky, E., Abeliovich, D., Bercovich, D., Lerer, I. and Peretz, T., 2010. An Ashkenazi founder mutation in the MSH6 gene leading to HNPCC. *Familial Cancer*, 9, pp. 141-50.
- Goldenberg, M. J., 2006. On evidence and evidence-based medicine: Lessons from the philosophy of science. *Social Science & Medicine*, 62, pp. 2621-32.
- Goodman, K. W., 2003. *Ethics and Evidence-Based Medicine: Fallibility and Responsibility in Clinical Science*. Cambridge: Cambridge University Press.
- Gøtzsche, P. C., Hróbjartsson, A., Maric, K. and Tendal, B., 2007. Data extraction errors in meta-analyses that use standardized mean differences. *Journal of the American Medical Association*, 298, pp. 430-7.
- Gough, D. and Elbourne, D., 2002. Systematic Research Synthesis to Inform Policy, Practice and Democratic Debate. *Social Policy & Society*, 1, pp. 225-36.
- Green, J. and Britten, N., 1998. Qualitative research and evidence based medicine. *British Medical Journal*, 316, pp. 1230-2.
- Green, J. and Thorogood, N., 2004. *Qualitative Methods for Health Research*. London, etc.: Sage.
- Green, S., Higgins, J. P. T, Alderson, P., Clarke, M., Mulrow, C. D. and Oxman, A. D., 2011. Introduction. In: J. P. T Higgins and S. Green, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. [online] The Cochrane Collaboration. Available at: <[www.cochrane-handbook.org](http://www.cochrane-handbook.org)> Ch.1. [Last accessed 18 May 2013].
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O. and Peacock, R., 2005. Storylines of research in diffusion of innovation: a metanarrative approach to systematic review. *Social Science & Medicine*, 61, pp. 417-30.
- Grossman, J., 2008. A Couple of the Nasties Lurking in Evidence-Based Medicine. *Social Epistemology: A Journal of Knowledge, Culture, and Policy*, 22, pp. 333-52.

- Gupta, M., 2007. Does Evidence-Based Medicine Apply to Psychiatry. *Theoretical Medicine and Bioethics: Philosophy of Medical Research and Practice*, 28, pp. 103-20.
- Guyatt, G. H., Haynes, R. B., Jaeschke, R. Z., Cook, D. J., Green, L., Naylor, C. D., Wilson, M. C. and Richardson, W. S., for the Evidence-Based Medicine Working Group, 2000. Users' Guides to The Medical Literature: XXV. Evidence-based medicine: Principles for applying the Users' Guides to patient care. *Journal of the American Medical Association*, 284, pp. 1290-6.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., Debeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P. and Schünemann, H. J., 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64, pp. 383-94.
- Guyatt, G. H., Oxman, A. D., Sultan, S., Glasziou, P., Akl, E. A., Alonso-Coello, P., Atkins, D., Kunz, R., Brozek, J., Montori, V., Jaeschke, R., Rind, D., Dahm, P., Meerpohl, J., Vist, G., Berliner, E., Norris, S., Falck-Ytter, Y., Murad, M. H. and Schünemann, H. J., for the Grade Working Group, 2011. GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, 64, pp. 1311-6.
- Hacking, I., 1983. *Representing and Intervening: Introductory topics in the philosophy of natural science*. Cambridge, etc.: Cambridge University Press.
- Hall, G. M., ed., 2013. *How to Write a Paper*. 5th ed. Chichester: Wiley-Blackwell/ BMJ Books.
- Hammersley, M., 1997. Qualitative Data Archiving: Some Reflections on its Prospects and Problems. *Sociology*, 31, pp. 131-42.
- Hanson, N. R., 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.
- Harden, A., 2011. Keynote presentation on mixed reviews. *7th Mixed Methods International Conference*. Leeds, 29 June – 5 July 2011.
- Harden, A., Garcia, J., Oliver, S., Rees, R., Shepherd, J., Brunton, G. and Oakley, A., 2004. Applying systematic review methods to studies of people's views: an example from public health research. *Journal of Epidemiology and Community Health*, 58, pp. 794-800.

- Harper, D., 2013. *Online Etymology Dictionary*. Available at:  
<<http://www.etymonline.com/>> [Last accessed 31 May 2013].
- Harris, F. M., Kendall, M., Bentley, A., Maguire, R., Worth, A., Murray, S., Boyd, K., Brown, D., Kearney, N. and Sheikh, A., 2008. Researching experiences of terminal cancer: a systematic review of methodological issues and approaches. *European Journal of Cancer Care*, 17, pp. 377-86.
- Hawkes, N., 2011. Too much information: The facts don't always speak for themselves, and the prime minister may come to regret thinking that they do. *British Medical Journal*, 343, p. 128.
- Haynes, B. R., 2006. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *Evidence-Based Medicine*, 11, pp. 162-4.
- Hempel, C. G., 1966. *Philosophy of Natural Science*, Englewood Cliffs: Prentice-Hall.
- Henry, S. G., 2006. Recognizing Tacit Knowledge in Medical Epistemology. *Theoretical Medicine and Bioethics: Philosophy of Medical Research and Practice*, 27, pp. 187-213.
- Higgins, J. and Churchill, R., 2011. New quality standards for Cochrane reviews. [presentation] 19<sup>th</sup> *Cochrane Colloquium*. Madrid, Spain, 19-22 October 2011. Cochrane Collaboration. Available at:  
<<http://www.cochrane.org/multimedia/multimedia-cochrane-colloquia-and-meetings/colloquium-madrid-2011>> [Last accessed 18 May 2013].
- Higgins, J. P. T. and Deeks, J. J., 2011. Selecting studies and collecting data. In: J. P. T Higgins and S. Green, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. [online] The Cochrane Collaboration. Available at:  
<[www.cochrane-handbook.org](http://www.cochrane-handbook.org)> Ch.7. [Last accessed 18 May 2013].
- Higgins, J. P. T and Green, S., eds., 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley-Blackwell.
- Higgins, J. P. T. and Green, S., eds., 2011. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. [online] The Cochrane Collaboration. Available at: <[www.cochrane-handbook.org](http://www.cochrane-handbook.org)> [Last accessed 18 May 2013].
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J. and Altman, D. G., 2003. Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, pp. 557-60.

- Higginson, I. J., Finlay, I., Goodwin, D. M., Cook, A. M., Hood, K., Edwards, A. G. K., Douglas, H. R. and Norman, C. E., 2002. Do Hospital-based Palliative Teams Improve Care for Patients or Families at the End of Life? *Journal of Pain and Symptom Management*, 23, pp. 96-106.
- Hök, J., Wachtler, C., Falkenberg, T. and Tishelman, C., 2007. Using narrative analysis to understand the combined use of complementary therapies and bio-medically oriented health care. *Social Science & Medicine*, 65, pp. 1642-53.
- Howell, D. C. *Fundamental Statistics for the Behavioral Sciences*. 7th ed. International Edition: Wadsworth/Cengage Learning.
- Howick, J., 2011. *The Philosophy of Evidence-based Medicine*. Chichester: Wiley-Blackwell/BMJ Books.
- Howick, J., Glasziou, P. and Aronson, J. K., 2010. Evidence-based mechanistic reasoning. *Journal of the Royal Society of Medicine*, 103, pp. 433-41.
- Howlett, P. and Morgan, M. S., eds., 2011. *How well do facts travel? The dissemination of reliable knowledge*. Cambridge: Cambridge University Press.
- Hubbard, G., Kidd, L. and Kearney, N., 2010. Disrupted lives and threats to identity: The experiences of people with colorectal cancer within the first year following diagnosis. *Health*, 14, pp. 131-47.
- Huemer, M., 2011. Sense-Data. In: E. N. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. [online] Available at: <<http://plato.stanford.edu/archives/spr2011/entries/sense-data/>> [Last accessed 18 May 2013]
- Huque, M. F., 1988. Experiences with meta-analysis in NDA submissions. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 2, pp. 28-33.
- IBM, 2013. *SPSS Statistics: Put the power of advanced statistical analysis in your hands*. [online] Available at: <<http://www-01.ibm.com/software/analytics/spss/products/statistics/>> [Last accessed 26 May 2013].
- Independent UK Panel on Breast Cancer Screening, 2012. The benefits and harms of breast cancer screening: an independent review. *The Lancet*, 380, pp. 1778-86.



- INVOLVE, 2013. *What is public involvement in research?* [online] Available at: <<http://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/>> [Last accessed 31 May 2013].
- Jaturapatporn, D., Moran, E., Obwanga, C. and Husain, A., 2010. Patients' experience of oxygen therapy and dyspnea: a qualitative study in home palliative care. *Supportive Care in Cancer*, 18, pp. 765-70.
- Jensen, L.A. and Allen, M.N., 1994. A synthesis of qualitative research in wellness–illness. *Qualitative Health Research*, 4, pp. 349-69.
- John Wiley & Sons, 2013. *The Cochrane Library*. [online] Available at: <<http://www.thecochranelibrary.com/view/0/AboutTheCochraneLibrary.html>> [Last accessed 26 May 2013].
- Jones, A. P., Remington, T., Williamson, P. R., Ashby, D. and Smyth, R. L., 2005. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. *Journal of Clinical Epidemiology*, 58, pp. 741-2.
- Jüni, P., Witschi, A., Bloch, R. and Egger, M., 1999. The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis. *The Journal of the American Medical Association*, 282, pp. 1054-60.
- Karunanathan, S., Wolfson, C., Bergman, H., Béland, F. and Hogan, D. B., 2009. A multidisciplinary systematic literature review on frailty: Overview of the methodology used by the Canadian Initiative on Frailty and Aging. *BMC Medical Research Methodology*, 9.
- Kearney, M. H., 2001. Enduring Love: A Grounded Formal Theory of Women's Experience of Domestic Violence. *Research in Nursing & Health*, 24, pp. 270-80.
- Kelly, M. and Swann, C., 2004. Foreword to Dixon-Woods et al. *Integrative approaches to qualitative and quantitative evidence*. [pdf] London: Health Development Agency. Available at: <[http://www.nice.org.uk/niceMedia/pdf/Integrative\\_approaches\\_evidence.pdf](http://www.nice.org.uk/niceMedia/pdf/Integrative_approaches_evidence.pdf)> pp. iv-vi. [Last accessed 17 May 2013].
- Kincaid, H., Dupré, J. and Wylie, A., eds., 2007. *Value-free science?: Ideals and Illusions*, Oxford, etc.: Oxford University Press.

- Kitcher, P., 2001. *Science, Truth, and Democracy*. Oxford, etc.: Oxford University Press.
- Knaapen, L. and Weisz, G., 2008. The biomedical standardization of premenstrual syndrome. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 39, pp. 120-34.
- Kuhn, T. S., 1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago and London: University of Chicago Press.
- Kulkarni, A. V., 2005. The challenges of evidence-based medicine: a philosophical perspective. *Medicine, Health Care and Philosophy*, 8, pp. 255-60.
- Kvale, S. and Brinkmann, S., 2009. *InterViews: Learning the Craft of Qualitative Research Interviewing*. Los Angeles, etc.: Sage.
- La Caze, A., 2008. Evidence-Based Medicine Can't Be. *Social Epistemology: A Journal of Knowledge, Culture, and Policy*, 22, pp. 353-70.
- La Caze, A., 2009. Evidence-Based Medicine Must Be... *Journal of Medicine and Philosophy*, 34, pp. 509-27.
- Lambert, H., 2006. Accounting for EBM: Notions of evidence in medicine. *Social Science & Medicine*, 62, pp. 2633-45.
- Latour, B., 1987. *Science in Action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
- Lenoir, T., ed., 1998. *Inscribing Science: Scientific Texts and the Materiality of Communication*. Stanford: Stanford University Press.
- Leonardi-Bee, J. and Rolfe, V., 2007. *Presenting and Interpreting Meta-Analyses*. [online] Nottingham: School of Nursing and Academic Division of Midwifery, University of Nottingham. Available at: <<http://www.nottingham.ac.uk/nmp/sonet/rlos/ebp/meta-analysis2/>> [Last accessed 18 May 2013].
- Leonelli, S., ed., 2012a. Special Section on Data-driven research in the biological and biomedical sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, pp. 1-87.
- Leonelli, S., 2012b. Introduction: Making sense of data-driven research in the biological and biomedical sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, pp. 1-3.

- Lincoln, Y. S. and Guba, E. G. 1985. *Naturalistic Inquiry*. Newbury Park, etc.: Sage.
- Lloyd Jones, M., 2005. Role development and effective practice in specialist and advanced practice roles in acute hospital settings: systematic review and meta-synthesis. *Journal of Advanced Nursing*, 49, pp. 191-209.
- Longino, H. E., 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Longino, H. E., 1996. Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy. In: L. H. Nelson and J Nelson, eds. *Feminism, Science, and the Philosophy of Science*. Dordrecht: Kluwer Academic.
- Lucas, P.J., Baird, J, Arai, L., Law, C and Roberts, H.M., 2007. Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Medical Research Methodology*, 7.
- Macbeth, F., 2011. *NICE Clinical Guidelines: new challenges and how Cochrane can help*. [presentation] 19<sup>th</sup> Cochrane Colloquium. Madrid, Spain, 19-22 October 2011. Cochrane Collaboration. Available at: <<http://www.cochrane.org/multimedia/multimedia-cochrane-colloquia-and-meetings/colloquium-madrid-2011>> [Last accessed 18 May 2013].
- Machamer, P. and Wolters, G., eds., 2004. *Science, Values and Objectivity*. Pittsburgh: University of Pittsburgh Press.
- Major, C. H. and Savin-Baden, M., 2010. *An Introduction to Qualitative Research Synthesis: Managing the information explosion in social science research*. London and New York: Routledge.
- Major, C. H. and Savin-Baden, M., 2011. Integration of qualitative evidence: towards construction of academic knowledge in social science and professional fields. *Qualitative Research*, 11, 645-63.
- Marston, C. and King, E., 2006. Factors that shape young people's sexual behaviour: a systematic review. *The Lancet*, 368, pp. 1581-6.
- MAXQDA, 2013. *Data Analysis with MAXQDA*. [online] Available at: <<http://www.maxqda.com/products/what-is-maxqda>> [Last accessed 26 May 2013].
- Maxwell, G., 1962. The Ontological Status of Theoretical Entities. In: H. Feigl and G. Maxwell, eds. *Scientific Explanation, Space, and Time*. Volume III of the

Minnesota Studies in the Philosophy of Science. Minneapolis: Minnesota University Press, pp. 3-27.

- McCullough, L. B., 2007. Constructing a Systematic Review for Argument-Based Clinical Ethics Literature: The Example of Concealed Medications. *Journal of Medicine and Philosophy*, 32, pp. 65-76.
- Meade, M. O. and Richardson, W. S., 1997. Selecting and appraising studies for a systematic review. *Annals of Internal Medicine*, 127, pp. 531-7.
- Meades, R. T., Svensson, W. E., Frank, J. W., Gada, V., Ralleigh, G., Satchithananda, K., Barrett, N. and Nijran, K. S., 2010. Carcinoma of the breast wire localisation post nuclear medicine sentinel lymph node imaging. Are radiologists receiving a significant dose? *European Radiology*, 20, pp. 529-32.
- Miles, M. B. and Huberman, A. M., 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd ed. Thousand Oaks: Sage.
- Molassiotis, A., Wengström, Y. and Kearney, N., 2010. Symptom Cluster Patterns During the First Year After Diagnosis with Cancer. *Journal of Pain and Symptom Management*, 39, pp. 847-58.
- Molewijk, B., 2004. Integrated Empirical Ethics: In Search for Clarifying Identities. *Medicine, Health Care and Philosophy: A European Journal*, 7, pp. 85-7.
- Mols, F., Thong, M. S. Y., Vreugdenhil, G. and van de Poll-Franse, L. V., 2009. Long-term cancer survivors experience work changes after diagnosis: results of a population-based study. *Psycho-Oncology*, 18, pp. 1252-60.
- Munday, D., Petrova, M. and Dale, J., 2009. Exploring preferences for place of death with terminally ill patients: a qualitative study of the experiences of GPs and community nurses. *British Medical Journal*, 339, DOI: <http://dx.doi.org/10.1136/bmj.b2391>.
- Naylor, D. C., 1995. Grey zones of clinical practice: some limits to evidence-based medicine. *The Lancet*, 345, pp. 840-2
- Noblit, G. W. and Hare, R. D., 1988. *Meta-Ethnography: Synthesizing Qualitative Studies*. Newbury Park, etc.: Sage.
- O'Connor, D., Green, S. and Higgins, J. P. T., 2011. Defining the review question and developing criteria for including studies. In: J. P. T Higgins and S. Green, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0.

- [online] The Cochrane Collaboration. Available at: <[www.cochrane-handbook.org](http://www.cochrane-handbook.org)> Ch.5. [Last accessed 18 May 2013].
- O'Malley, M. A. and Soyer, O. S. The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, pp. 58-68.
- Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J. and Oakley, A., 2005. An Emerging Framework for Including Different Types of Evidence in Systematic Reviews for Public Policy. *Evaluation*, 11, pp. 428-46.
- Oliver, S. R., Rees, R. W., Clarke-Jones, L., Milne, R., Oakley, A. R., Gabbay, J., Stein, K., Buchanan, P. and Gyte, G., 2008. A multidimensional conceptual framework for analysing public involvement in health services research involvement in health services research. *Health Expectations*, 11, pp. 72-84.
- Onwuegbuzie, A. J. and Frels, R. K., 2011. *The Review of Literature Evaluative Steps: A Mixed Methods Approach*. [handouts] Workshop presented at the 7th Mixed Methods International Conference. Leeds, 29 June – 5 July 2011.
- Oxford English Dictionary, 2013. [online] Available at: <<http://www.oed.com/>> via University of Exeter Library website <<http://as.exeter.ac.uk/library/>> [Last accessed 31 May 2013].
- Parker, M., 2002. Whither Our Art? Clinical Wisdom and Evidence-Based Medicine. *Medicine, Health Care and Philosophy: A European Journal*, 5, pp. 273-80.
- Parker, M., 2007. Two into One Won't Go: Conceptual, Clinical, Ethical and Legal Impedimenta to the Convergence of CAM and Orthodox Medicine. *Bioethical Inquiry*, 4, pp. 7-19.
- Paterson, B. L., Thorne, S. E., Canam, C. and Jillings, C., 2001. *Meta-Study of Qualitative Health Research*. [e-book] Sage Research Methods Online. Available through: University of Exeter Library website <<http://as.exeter.ac.uk/library/>> [Last accessed 18 May 2013].
- Patton, M. Q., 2002. *Qualitative Research & Evaluation Methods*. 3rd ed. Thousand Oaks, etc.: Sage.
- Paul, C. L., Clinton-McHarg, T., Sanson-Fisher, R. W., Douglas, H. and Webb, G., 2009. Are we there yet? The state of the evidence base for guidelines on breaking bad news to cancer patients. *European Journal of Cancer*, 45, pp. 2960-6.

- Pavlin, N. L., Gunn, J. M., Parker, R., Fairley, C. K. and Hocking, J., 2006. Implementing chlamydia screening: what do women think? A systematic review of the literature. *BMC Public Health*, 6.
- Pawson, R., 2002a. Evidence-based Policy: In Search of a Method. *Evaluation*, 8, pp. 157-81.
- Pawson, R., 2002b. Evidence-based Policy: The Promise of Realist Synthesis. *Evaluation*, 8, pp. 340-58.
- Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K., 2004. *Realist synthesis: an introduction*. [pdf] Manchester: University of Manchester, ESRC Research Methods Programme. Available at: <<http://www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf>> [Last accessed 18 May 2013].
- Petrova, M., Sutcliffe, P., Fulford, K. W. M. and Dale, J., 2012a. Search terms and a validated brief search filter to retrieve publications on health-related values in MEDLINE: a word frequency analysis study, *Journal of the American Medical Informatics Association*, 19, pp. 479-88.
- Petrova, M., Sutcliffe, P., Fulford, K. W. M. and Dale, J., 2012b. *VaST (Values Search Tools): A manual for searching electronic databases for publications on health-related values*. Available at: <<http://www2.warwick.ac.uk/fac/med/study/research/vbp/resources/manual.pdf>> [Last accessed 9 June 2013].
- Petticrew, M., 2001. Systematic reviews from astronomy to zoology: myths and misconceptions. *British Medical Journal*, 322, pp. 98-101.
- Pinhey, T. K. and Okinaka, A. M., 2004. Exploring the purging behavior of Asian-Pacific adolescents in Guam: Does heavy television viewing make a difference? *Deviant Behavior*, 25, pp. 27-41.
- Poincaré, H. (n.d., 1908?) *Science and Method*, London, etc.: Thomas Nelson and Sons.
- Polanyi, M., 1946. *Science, Faith and Society*. London: Geoffrey Cumberlege/ Oxford University Press.
- Popay, J., 2006. Incorporating qualitative information in systematic reviews. [presentation] 14<sup>th</sup> *Cochrane Colloquium*. Dublin, Ireland, 23-26 October 2006. Cochrane Collaboration. Available at:

<<http://www.cochrane.org/multimedia/multimedia-cochrane-colloquia-and-meetings/colloquium-dublin-2006>> [Last accessed 18 May 2013].

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., With, Roen, K. and Duffy, S., 2006. *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme*. Version 1. [pdf] Kindly provided by the first author.

Popay, J., Rogers, A. and Williams, G., 1998. Rationale and Standards for the Systematic Review of Qualitative Literature in Health Services Research. *Qualitative Health Research*, 8, pp. 341-51.

Pope, C. and Mays, N., 2006. Synthesising qualitative research. In: C. Pope and N. Mays, eds. *Qualitative research in health care*. 3rd ed. Malden, MA, etc.: Blackwell Publishing/ BMJ Books.

Pope, C., Mays, N. and Popay, J., 2007. *Synthesizing Qualitative and Quantitative Health Evidence: a Guide to Methods*. Maidenhead: Open University Press.

QSR International, 2013. *NVivo*. [online] Available at:

<[http://www.qsrinternational.com/products\\_nvivo.aspx](http://www.qsrinternational.com/products_nvivo.aspx)> [Last accessed 26 May 2013].

Quine, W. V., 1951. Two Dogmas of Empiricism. *The Philosophical Review*, 60, pp. 20-43.

Ragin, C.C., 1987. *The comparative method: moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.

Rapp, K., Klenk, J., Ulmer, H., Concin, H., Diem, G., Oberaigner, W. and Schroeder, J., 2008. Weight change and cancer risk in a cohort of more than 65 000 adults in Austria. *Annals of Oncology*, 19, pp. 641-8.

Rees, R., Oliver, K., Woodman, J. and Thomas, J., 2011. The views of young children in the UK about obesity, body size, shape and weight: a systematic review. *BMC Public Health*, 11, DOI:10.1186/1471-2458-11-188.

Roberts, K. A., Dixon-Woods, M., Fitzpatrick, R., Abrams, K. R. and Jones, D. R., 2002. Factors affecting uptake of childhood immunisation: a Bayesian synthesis of qualitative and quantitative evidence. *The Lancet*, 360, pp. 1596-99.

- Rogers, W. A., 2004. Evidence Based Medicine and Justice: A Framework for Looking at the Impact of EBM upon Vulnerable or Disadvantaged Groups. *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 30, pp. 141-5.
- Rosenberg, W. and Donald, A., 1995. Evidence based medicine: an approach to clinical problem-solving. *British Medical Journal*, 310, pp. 1122-6.
- Royle, P. and Waugh, N., 2005. A simplified search strategy for identifying randomised controlled trials for systematic reviews of health care interventions: a comparison with more exhaustive strategies. *BMC Medical Research Methodology*, 5. Available at: <[http://wrap.warwick.ac.uk/52147/1/WRAP\\_Waugh\\_filters.pdf](http://wrap.warwick.ac.uk/52147/1/WRAP_Waugh_filters.pdf)> [Last accessed 31 May 2013].
- Russell, B., 1948. *Human Knowledge: Its Scope and Limits*. London: George Allen and Unwin.
- Ryan, R. E., Kaufman, C. A. and Hill, S. J., 2009. Building blocks for meta-synthesis: data integration tables for summarising, mapping, and synthesising evidence on interventions for communicating with health consumers. *BMC Medical Research Methodology*, 9.
- Saarni, S. I. and Gylling, H. A., 2004. Evidence Based Medicine Guidelines: A Solution to Rationing or Politics Disguised As Science? *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 30, pp. 171-5.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. and Richardson, W. S., 1996. Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, p. 71-2.
- Sandelowski, M. and Barroso, J., 2002. Finding the Findings in Qualitative Studies. *Journal of Nursing Scholarship*, 34, pp. 213-9.
- Sandelowski, M. and Barroso, J., 2007. *Handbook for Synthesizing Qualitative Research*. New York: Springer.
- Sandelowski, M., Docherty, S. and Emden, C., 1997. Focus on qualitative methods. Qualitative meta-synthesis: issues and techniques. *Research in Nursing and Health*, 20, pp. 365-71.
- Schreiber, R., Crooks, D. and Stern, P.N., 1997. Qualitative Meta-Analysis. In: J. M. Morse, ed. 1997. *Completing a Qualitative Project: Details and Dialogue*. Thousand Oaks: Sage.



- Schutz, A., 1962. *Collected Papers I: The Problem of Social Reality*. The Hague: Martinus Nijhoff.
- Schwandt, T. A., 1998. The Interpretative Review of Educational Matters: Is There Any Other Kind? *Review of Educational Research*, 68, pp. 409-12.
- Sehon, S. R. and Stanley, D. E., 2003. A philosophical analysis of the evidence-based medicine debate. *BMC Health Services Research*, 3.
- Sherwood, G., 1997. Meta-Synthesis of Qualitative Analyses of Caring: Defining a Therapeutic Model of Nursing. *Advanced Practice Nursing Quarterly*, 3, pp. 32-42.
- Sherwood, G., 1999. Meta-synthesis: merging qualitative studies to develop nursing knowledge. *International Journal for Human Caring*, 3, pp. 37-42.
- Slavin, R. E., 1995. Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48, pp. 9-18.
- Smith, L. K., Pope, C. and Botha, J. I., 2005. Patients' help-seeing experiences and delay in cancer presentation: a qualitative synthesis. *The Lancet*, 366, pp. 825-31.
- Stegenga, J., 2011. Is meta-analysis the platinum standard of evidence? *Studies in History and Philosophy of Biological and Biomedical Sciences*, 42, pp. 497–507.
- Strahan, E. J., 2003. Selling thinness: How media images increase importance of weight and beauty as a basis of women's self-esteem and decrease their body satisfaction and eating. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 64, p. 996. Abstract only. Available through: PsycINFO, via University of Exeter Library website <<http://as.exeter.ac.uk/library/>> [Last accessed 18 May 2013].
- Straus, S. E., Glasziou, P., Richardson, W. S., Haynes, R. B., 2011. *Evidence-Based Medicine: How to practice and teach it*. 4th ed. Churchill Livingstone/Elsevier.
- Strech, D., Synofzik, M. and Marckmann, G., 2008. Systematic Review of Empirical Bioethics. *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 34, pp. 472-7.
- Strike, K. and Posner, G., 1983. Types of Synthesis and Their Criteria. In: S. A. Ward and L. J. Reed, eds. *Knowledge Structure and Use: Implications for Synthesis and Interpretation*. Philadelphia: Temple University Press.

- Suppe, F., ed., 1977. *The Structure of Scientific Theories*. Urbana: University of Illinois Press.
- Sutton, A. J., Abrams, K. R., Jones, D. R, Sheldon, T. A. and Song, F. 2000. *Methods for Meta-Analysis in Medical Research*. Chichester, etc.: John Wiley & Sons.
- Sutton, A. J., Cooper, N. J. and Jones, D. R., 2009. Evidence synthesis as the key to more coherent and efficient research. *BMC Medical Research Methodology*, 9.
- The Campbell Collaboration, 2013. The Campbell Collaboration: What helps? What harms? Based on what evidence? [online] Available at: <<http://www.campbellcollaboration.org/>> [Last accessed 9 June 2013].
- The Cochrane Collaboration, 2013a. *History*. [online] (Last updated 24 May) Available at: <<http://www.cochrane.org/about-us/history>> [Last accessed 26 May 2013].
- The Cochrane Collaboration, 2013b. *About us*. [online] (Last updated 31 May) Available at: <<http://www.cochrane.org/about-us>> [Last accessed 2 June 2013].
- The Cochrane Collaboration, 2013c. *Cochrane Reviews*. [online] (Last updated 24 April) Available at: <<http://www.cochrane.org/cochrane-reviews>> [Last accessed 26 May 2013].
- The Cochrane Collaboration, 2013d. *Methods Groups*. [online] Available at: <<http://www.cochrane.org/contact/methods-groups>> [Last accessed 28 May 2013].
- The Cochrane Information Management System, 2013. *RevMan*. [online] Available at: <<http://ims.cochrane.org/revman>> [Last accessed 26 May 2013].
- The Cochrane Handbook, 2008; 2011 – see Higgins, J. P. T. and Green, S., eds., 2008; 2011.
- The Joanna Briggs Institute, 2013. *Appraise Evidence*. [online] (Last updated 17 April 2013) Available at: <<http://www.joannabriggs.edu.au/SUMARI>> [Last accessed 26 May 2013].
- Thomas, J. and Harden, A., 2008. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8.
- Thomas, J., McNaught, J. and Ananiadou, S., 2011. Application of text mining within systematic reviews. *Research Synthesis Methods*, 2, pp. 1-14.

- Thomas, R. and Davies, N., 2007. Lifestyle During and After Cancer Treatment. *Clinical Oncology*, 19, pp. 616-27.
- Thombre, A., Sherman, A. C. and Simonton, S., 2010. Religious Coping and Posttraumatic Growth Among Family Caregivers of Cancer Patients in India. *Journal of Psychosocial Oncology*, 28, pp. 173-88.
- Thorne, S. and Paterson, B., 1998. Shifting Images of Chronic Illness. *IMAGE: Journal of Nursing Scholarship*, 30, pp. 173-8.
- Thornton, T., 2006. Tacit Knowledge As the Unifying Factor in Evidence Based Medicine and Clinical Judgement. *Philosophy, Ethics, and Humanities in Medicine*, 1, pp. 1-10.
- Thornton, T., 2008. Evidence-Based Medicine and Evaluativism. *Philosophy, Psychiatry, and Psychology*, 15, pp. 175-8.
- Tonelli, M. R., 1998. The philosophical limits of evidence-based medicine. *Academic Medicine*, 73, pp. 1234-40.
- Upshur, R. E. G., 1999. Priors and Prejudice. *Theoretical Medicine and Bioethics: Philosophy of Medical Research and Practice*, 20, pp. 319-27.
- Upshur, R. E. G., 2003. Argumentation and Evidence. *Theoretical Medicine and Bioethics: Philosophy of Medical Research and Practice*, 24, pp. 283-99.
- Upshur, R. E. G., 2004. Legitimacy, Authority, and Hierarchy: Critical Challenges for Evidence-Based Medicine. *Brief Treatment and Crisis Intervention* 4, pp. 197-204.
- Van Fraassen, B., 1980. *The Scientific Image*. Oxford: Oxford University Press/ Clarendon.
- Van Fraassen, B., 2002. *The Empirical Stance*, New Haven: Yale University Press.
- Van Fraassen, B., 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press/ Clarendon.
- Vanhaecht, K., De Witte, K., Depreitere, R. and Sermeus, W., 2006. Clinical pathway audit tools: a systematic review. *Journal of Nursing Management*, 14, pp. 529-37.

- Vergnes, J. N., Marchal-Sixou, C. and Nabet, C., 2010. Ethics in Systematic Reviews. *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 36, pp. 771-4.
- Vicker, G., 1981. Rationality and Intuition. In: J. Wechsler, ed. *On Aesthetics in Science*. Cambridge, MA and London, England: MIT Press.
- Vineis, P., 2004. Evidence-Based Medicine and Ethics: A Practical Approach. *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 30, pp. 126-30.
- Vlayen, J., Aertgeerts, B., Hannes, K., Sermeus, W. & Ramaekers, D., 2005. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *International Journal for Quality in Health Care*, 17, pp. 235-42.
- Walter, F. M., Emery, J., Braithwaite, D. and Marteau, T. M., 2004. Lay Understanding of Familial Risk of Common Chronic Diseases: A Systematic Review and Synthesis of Qualitative Research. *Annals of Family Medicine*, 2, pp. 583-94.
- Watine, J., 2011. What Sort of Bioethical Values Are the Evidence-Based Medicine and the GRADE Approaches Willing to Deal With? *Journal of Medical Ethics: The Journal of the Institute of Medical Ethics*, 37, pp. 184-6.
- Weed, M., 2005. "Meta Interpretation": A Method for the Interpretive Synthesis of Qualitative Research. *Forum: Qualitative Social Research (Sozialforschung)*, 6, Art 37.
- Weimer, W., 1975. The Psychology of Inference and Expectation: Some Preliminary Remarks. In: G. Maxwell and R. M. Anderson, eds. *Minnesota Studies in the Philosophy of Science, Volume VI: Induction, Probability, and Confirmation*. Minneapolis: University of Minnesota Press, pp. 430-486.
- West, S., King, V., Carey, T. S., Lohr, K. N., Mckoy, N., Sutton, S. F. and Lux, L., 2002. *Systems to Rate the Strength Of Scientific Evidence*. Report 47, prepared for the Agency for Healthcare Research and Quality. [pdf] North Carolina: Research Triangle Institute, University of North Carolina. Available at: <<http://www.thecre.com/pdf/ahrq-system-strength.pdf>> [Last accessed 18 May 2013].
- Wifstad, Å., 2008. External and Internal Evidence in Clinical Judgment: The Evidence-Based Medical Attitude. *Philosophy, Psychiatry, and Psychology*, 15, pp. 135-9.

- Wittgenstein, L., 1922. *Tractatus Logico-Philosophicus*. Translated from German by C. K. Ogden. London: Kegan Paul, Trench, Trubner and Co./ New York: Harcourt, Brace & Company. Project Gutenberg online edition available at: <<http://www.gutenberg.org/files/5740/5740-pdf.pdf>> [Last accessed 11 June 13].
- Whitaker, K. L., Brewin, C. R. and Watson, M., 2008. Intrusive cognitions and anxiety in cancer patients. *Journal of Psychosomatic Research*, 64, pp. 509-17.
- World Health Organization, 2008. The top 10 causes of death. Fact sheet No 310. [online] (Last updated June 2011) Available at: <<http://www.who.int/mediacentre/factsheets/fs310/en/index.html>> [Last accessed 26 May 2013].
- Worrall, J., 2002. What Evidence in Evidence-Based Medicine? *Philosophy of Science*, 69, pp. S316-30.
- Worrall, J., 2007a. Evidence in Medicine and Evidence-Based Medicine. *Philosophy Compass*, 2, pp. 981-1022.
- Worrall, J., 2007b. Why There's No Cause to Randomize. *The British Journal for the Philosophy of Science*, 58, pp. 451-488.
- Wylie, A., 2006. Socially Naturalized Norms of Epistemic Rationality: Aggregation and Deliberation. *The Southern Journal of Philosophy*, XLIV, pp. 43-8.
- Zarkovich, E. and Upshur, R. E. G., 2002. The Virtues of Evidence. *Theoretical Medicine*, 23, pp. 403-12.
- Zhao, S. Y., 1991. Metatheory, Metamethod, Meta-Data-Analysis – What, Why, and How? *Sociological Perspectives*, 34, pp. 377-90.

## References from Table T 2.1.

Below are references used solely in Table T 2.1 and only accessed for the purposes of obtaining the bibliometric data presented in the table:

- Ades, A. E. and Cliffe, S., 2002. Markov Chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Medical Decision Making*, 22, pp. 359-71.
- Agranoff, R. and Radin, B.A., 1991. The comparative case study approach in public administration. *Research in Public Administration*, 1, pp. 203-31.
- Arai, L., Britten, N., Popay, J., Roberts, H., Petticrew, M., Rodgers, M. and Sowden, A., 2007. Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions. *Evidence & Policy*, 3, pp. 361-83.
- Beck, C.T., 2002a. Mothering multiples: a meta-synthesis of qualitative research. *MCN, American Journal of Maternal/Child Nursing*, 27, pp. 214-21.
- Beck, C.T., 2002b. Postpartum depression: a metasynthesis. *Qualitative Health Research*, 12, pp. 453-72.
- Brunton, G., Oliver, S., Oliver, K. and Lorenc, T., 2006. *A synthesis of research addressing children's, young people's and parents' views of walking and cycling for transport*. [pdf] London: EPPI-Centre. Available at: <[http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=W4j\\_bi08vDQ%3d&tabid=942&mid=2141](http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=W4j_bi08vDQ%3d&tabid=942&mid=2141)> [Last accessed 26 May 2013].
- Clemmens, D., 2003. Adolescent motherhood: a meta-synthesis of qualitative studies. *MCN, American Journal of Maternal/Child Nursing*, 28, pp. 93-9.
- Cooper, N., Sutton, A. and Abrams, K., 2002. Decision analytic economic modelling within a Bayesian framework: application to prophylactic antibiotics' use for caesarean section. *Statistical Methods in Medical Research*, 11, pp. 491–512.
- Cress, D. and Snow, D., 2000. The outcomes of homeless mobilization: the influence of organization, disruption, political mediation, and framing. *American Journal of Sociology*, 105, pp. 1063-104.

- Dixon, D., 1996. Unifying concepts in parents' experiences with health care providers. *Journal of Family Nursing*, 2, pp. 111-32.
- Dixon-Woods, M., Kirk, D., Agarwal, S., Annandale, E., Arthur, T., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. and Sutton, A., 2005. *Vulnerable Groups and Access to Health Care: A Critical Interpretative Synthesis*. [pdf] London: London School of Hygiene & Tropical Medicine, National Co-ordinating Centre for NHS Service Delivery and Organisation R & D (NCCSDO). Available at: <[http://www.netscc.ac.uk/hsdr/files/project/SDO\\_FR\\_08-1210-025\\_V01.pdf](http://www.netscc.ac.uk/hsdr/files/project/SDO_FR_08-1210-025_V01.pdf)> [Last accessed 26 May 2013].
- Dowie, J., 2001. *Towards value-based, science-informed public health policy: conceptual framework and practical guidelines*. WHO/Department of Health Consultation on Risks to Health: Better Management for Decision-making.
- Droitcour, J., Silberman, G. and Chelimsky, E., 1993. Cross-design synthesis: a new form of meta-analysis for combining results from randomised clinical trials and medical-practice databases. *International Journal of Technology Assessment in Health Care*, 9, pp. 440-9.
- Eddy, D. M., 1989. The confidence profile method: a Bayesian method for assessing health technologies. *Operations Research*, 37, pp. 210-228.
- Eddy, D. M., Hasselblad, V. and Shachter, R., 1992. *Meta-analysis by the Confidence Profile Method*. London: Academic Press.
- Gaber, J., 2000. Meta-needs assessment. *Evaluation and Program Planning*, 23, pp. 139-47.
- Harden, A., Oakley, A. and Oliver, S., 2001. Peer-delivered health promotion for young people: a systematic review of different study designs. *Health Education Journal*, 60, pp. 1-15.
- Harden, A., Weston, R. and Oakley, A., 1999. *A review of the effectiveness and appropriateness of peer-delivered health promotion for young people*. [pdf] London: EPPI-Centre. Available at: <<http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=bCmFZQRwu-o%3d&tabid=255&mid=1071>> [Last accessed 26 May 2013].
- Jensen, J.L. and Rogers, R., 2001. Cumulating the intellectual gold of case study research. *Public Administration Review*, 61, pp. 236-46.

- Larsson, R., 1993. Case survey methodology: quantitative analysis of patterns across case studies. *Academy of Management Journal*, 36, pp. 1515-46.
- Luce, B.R. and Claxton, K., 1999. Redefining the analytical approach to pharmacoeconomics. *Health Economics*, 8, pp.187-9.
- McCormick, J., Rodney, P. and Varcoe, C., 2003. Reinterpretation across studies: an approach to meta-analysis. *Qualitative Health Research*, 13, pp. 933-44.
- McNaughton, D.B., 2000. A synthesis of qualitative home visiting research. *Public Health Nursing*, 17, pp. 405-14.
- Melinder, K. A. and Andersson, R., 2001. The impact of structural factors on the injury rate in different European countries. *European Journal of Public Health*, 11, pp. 301-8.
- Miller, D. and Reilly, J., 1995. Making an issue of food safety: the media, pressure groups and the public sphere. In: D. Maurer and J. Sobal, eds. 1995. *Food and Nutrition as Social Problems*. New York: Aldine de Gruyter.
- Neill, S.J., 2000. Acute childhood illness at home: the parents' perspective. *Journal of Advanced Nursing*, 31, pp. 821-32.
- Nelson, A.M., 2002. A Metasynthesis: Mothering Other-than-Normal Children. *Qualitative Health Research*, 12, pp. 515-30.
- Oliver, S., Oakley, L., Lumley, J. and Waters, E., 2001. Smoking cessation programmes in pregnancy: systematically addressing development, implementation, women's concerns and effectiveness. *Health Education Journal*, 60, pp. 362-70.
- Paterson, B.L., Thorne, S. and Dewis, M., 1998. Adapting to and managing diabetes. *Image – the Journal of Nursing Scholarship*, 30, pp. 57-62.
- Pawson, R. and Bellamy, J., 2006. Realist synthesis: an explanatory focus for systematic review. In: J. Popay, ed. 2006. *Putting Effectiveness into Context: Methodological Issues in the Synthesis of Evidence from Diverse Study Designs*. London: Health Development Agency.
- Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K., 2005. Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy*, 10 (Suppl 1), pp. 21-34.
- Ragin, C.C., 2000. *Fuzzy set social science*. Chicago: University of Chicago Press.



- Sandelowski, M., Barroso, J. and Voils, C.I., 2007. Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings, *Research in Nursing & Health*, 30, pp. 99-111.
- Spiegelhalter, D. J., Abrams, K. R. and Myles, J., 2004. *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. New York: Wiley.
- Suikkala, A. and Leino-Kilpi, H., 2001. Nursing student-patient relationships: a review of the literature from 1984–1998. *Journal of Advanced Nursing*, 33, pp. 42-50.
- Turner, D., Wailoo, A., Nicholson, K., Cooper, N., Sutton, A. and Abrams, K., 2003. Systematic review and economic decision modelling for the prevention and treatment of influenza A and B. *Health Technology Assessment*, 7, pp. 1-288. Available at: <<http://www.hta.ac.uk/fullmono/mon735.pdf>> [Last accessed 26 May 13].
- US General Accounting Office, 1992. *Cross-design Synthesis: A New Strategy for Medical Effectiveness Research*. Washington, DC: General Accounting Office.
- Whitehead, A., 2002. *Meta-analysis of Controlled Clinical Trials*. Chichester: Wiley.
- Yin, R.K. and Heald, K.A., 1975. Using the case survey method to analyse policy studies. *Administrative Science Quarterly* 20, pp. 371-81.