eunethta

EUnetHTA JA2
WP7 DELIVERABLE

Position paper on how to decide on the appropriate study design for primary research arising from HTA reports

The EUnetHTA JA 2 (2012-2015) has received funding from the European Union, in the framework of the Health Programme

eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

Joint Action on HTA 2012-2015

Position Paper on how to decide on the appropriate study design for primary research arising from HTA reports

June 2015

Was developed by Work Package 7 – Methodology development and evidence generation: Guidelines and pilots production

WP 7 Lead Partner: HAS

WP 7 Co-Lead Partner: IQWiG

HAS

HAUTE AUTORITÉ DE SANTÉ

IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Institute for Quality and Efficiency in Health Care

Disclaimer: EUnetHTA Joint Action 2 is supported by a grant from the European Commission. The sole responsibility for the content of this document lies with the authors and neither the European Commission nor EUnetHTA are responsible for any use that may be made of the information contained therein.

Document History and contributors

This position paper is a deliverable of EUnetHTA Joint Action 2 Work package 7 Subgroup 2.

5-7 March 2014	Andrew Cook	First Draft of Plan
April 2014	Andrew Cook	Draft 1a
13 June 2014	Andrew Cook	Draft 1b
18 June 2014	Love Linnér	Comments from TLV on Draft 1b
1 July 2014	Andrew Cook	Draft 1c
3 October 2014	Andrew Cook	Draft 1d, following input from TLV and HAS (received during 1st internal review). Remove sections on multiplicity, unusual designs. Add basic consideration of modelling.
4 November 2014	Andrew Cook	**Final version of the First draft** incorporating comments from TLV, HAS, SNHTA (received during 2nd internal review).
29 January 2015	Andrew Cook	Draft 2a, incorporating comments received during WP7, SAG and EMA/ENCePP consultation (**review period November 7th – December 15th**) Reordered in light of consultation. Much material moved to appendices.
5 March 2015	Andrew Cook	Draft 2b, updates as discussed with HAS. A chapter on dilemmas added, further to additional suggestions provided by some WP7 partners and IMS (member of ENCePP)

24 March 2015	Andrew Cook	Draft 2c, Further Revision in light of comments from SNHTA and HAS
10 April 2015	Andrew Cook	**Final version of the Second draft**
18 June 2015	Andrew Cook	Draft 3a, incorporating comments received during Public consultation **(review period April 14th – May 29th)**
30 June 2015	Andrew Cook	**Final version of the document**

Pilot team

Agency - Country		Responsibility/ Role
NETSCC - UK[1]	Andrew Cook	Lead Author
TLV – Sweden	Love Linner, Jonas Lindblom	Member of the drafting group, internal reviewer
HAS - France	Irena Guzina, François Meyer, Sophie Stamenkovic, Laura Zanetti	Coordinating agency, members of the drafting group, internal reviewers
SNHTA – Switzerland	Goedele van-Haasteren	Internal reviewer

Further contributors

Input (description of practices and/or comments on the first draft) was received from the following **WP7 members**:

AETSA, Spain	Maria Auxiliadora Castilo Munoz
AETS ISC-III, Spain	Setefilla Luengo Matos
Agenas, Italy	Antonio Migliore
AIFA, Italy	Agnese Cangini
ASSR-RER, Italy	Luciana Ballini
Avalia-t, Spain	Leonor Varela Lema
HAS, France	Irena Guzina, Sophie Stamenkovic, Laura Zanetti
IQWIG, Germany	Jorg Lauterberg, Stefan Sauerland, Jürgen Windeler
KCE, Belgium	Mattias Neyt
NICE, UK	Zoe Garrett
NOKC, Norway	Marianne Klemp, Vidgis Lauvrak
Regione del Veneto, Italy	Anna Cavazzana
SNHTA, Switzerland	Goedele van-Haasteren

[1] This report presents independent research funded in part by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

TLV, Sweden	Jonas Lindblom

Stakeholder involvement

Comments on the first draft were received from EDMA and EFPIA.

EMA and ENCePP involvement

Comments on the first draft were received from Massoud Toussi, IMS Health, and Martin Daumer, SLCMSR.

Public consultation

Comments on the second draft were received from EFPIA and Biogen.

Executive Summary

The phrase “more research is needed” is the bane of the research commissioner’s life. Following on from a piece of evidence synthesis it is so frustrating to be told that more research is needed, but not to be told exactly what that research is. This paper, along with its partner *How best to formulate research recommendations* strives to guide the systematic reviewer of a health technology in the task of constructing recommendations for Additional Evidence Generation, and focuses on the role of the reviewer in specifying the design of subsequent primary research.

We consider current practice in specifying research design across Europe and in international systematic review organisations and guideline groups.

We then discuss approaches to specifying a primary research method – when it should be done and when it should not; and consider particular dilemmas which have affected EUnetHTA partners.

We then set out the EUnetHTA position on specifying research methods in primary research:

1. Firstly, a systematic reviewer should identify and prioritise important clinical uncertainties, and set them out using the approach detailed in the companion paper to this on formulating research recommendations.
2. They should then, should he or she feel able, comment on possible research designs to address these uncertainties, if
    a. The required design is dictated by the existing evidence base.
    b. There is a substantial risk that if a design is not specified a less than ideal design is likely to be adopted (e.g. in case there would a party with a vested interest).
    c. One of the reasons discussed in section 6.2.1 can be applied.
3. Else, a study design should not be specified.

Glossary

Term	Definition
ADC	Additional Data Collection
AEG	Additional Evidence Generation
AETSA	Agencia de Evaluacion de Technologias Sanitarias de Andalucia, Andalusian Agency for Health Technology Assessment. http://www.juntadeandalucia.es/salud/servicios/aetsa/
AETS-ISCIII	Instituto de Salud Carlos III, A Health Technology Assesment agency in Madrid, Spain http://aunets.isciii.es
AIFA	Agenzia Italiana del Farmaco, the Italian Medicines Agency. http://www.agenziafarmaco.gov.it/
ASSR-RER	Agenzia Sanitaria e Sociale Regionale - Regione Emilia-Romagna, Regional Agency for Health and Social Care, Bologna, Italy. http://asr.regione.emilia-romagna.it
avalia-t	Axencia de Avaliación de Teconoloxìas Sanitarias de Galicia, Galician agency for HTA. http://avalia-t.sergas.es
CRD	Centre for Reviews and Dissemination, based in York, UK
DUETs	Database of Uncertainties about the Effects of Treatments, http://www.library.nhs.uk/duets/
Effectiveness	A measure of whether an intervention works, and how well it works, in real world clinical practice.
Efficacy	A measure of whether an intervention works, and how well it works, in ideal circumstances.
EPICOT	A way of defining a research question: Evidence, Population, Intervention, Control, Outcome, Time [1]
EUnetHTA	The European Network for Health Technology Assessment, http://www.eunethta.eu
GRADE	Grading of Recommendations Assessment, Development and Evaluation, a tool for grading the quality of healthcare

	evidence
HAS	Haute Autorité de Santé (the French National Authority for Health), a Health Technology Assessment agency in Paris, France. http://www.has-sante.fr
HTA	Health Technology Assessment
IDEAL	The IDEAL Collaboration is an initiative to improve the quality of research in surgery. http://www.ideal-collaboration.net
INAHTA	The International Network of Agencies for Health Technology Assessment, http://www.inahta.org
IQWIG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Institute for Quality and Efficiency in Health Care, Cologne, Germany. http://www.iqwig.de
KCE	Kenniscentrum – Centre d'Expertise (The Belgian Health Care Knowledge Centre). https://kce.fgov.be/
MRC	The Medical Research Council, http://www.mrc.ac.uk
NETSCC	NIHR Evaluation Trials and Studies Coordinating Centre
NICE	National Institute for Health and Care Excellence, https://www.nice.org.uk
NIHR	National Institute for Health Research, a UK research funder and coordinator, https://www.nice.org.uk
PICO	A briefer way of defining a research question than EPICOT: Population, Intervention, Control, Outcome [1]
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses, an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. http://www.prisma-statement.org
RCT	Randomised Controlled Trial
SBU	Swedish Council on Health Technology Assessment, a Health Technology Assessment Agency in Stockholm, Sweden. http://www.sbu.se
SG2	EUnetHTA 2nd Joint Action, Workpackage 7, Subgroup 2

STROBE	Strengthening the Reporting of Observational Studies in Epidemiology, international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies. http://www.strobe-statement.org/
TLV	The Swedish Dental and Pharmaceutical Benefits Agency, http://www.tlv.se
TOMADO	Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea, http://www.nets.nihr.ac.uk/projects/hta/0811003
WP7	The 7th workpackage of EUnetHTA joint action 2.

Background

Within EUnetHTA Joint action 2, Subgroup 2 of Work Package 7 is dedicated to Additional Evidence Generation (AEG) with the objective of developing and testing a methodological basis for European cooperation in this field.

Two methodological documents are associated with the work of JA2 WP7 SG2, this paper, and its partner How to best formulate research recommendations. Both strive to guide the systematic reviewer of a health technology in the task of constructing recommendations for Additional Evidence Generation.

The NIHR Evaluation Trials and Studies Coordinating Centre (NETSCC) in the UK was approached by HAS to lead the development of this paper, supported by Tandvårds-och Läkemedelsförmånsverket (TLV, The Dental and Pharmaceutical Benefits Agency) in Sweden.

1 Introduction

The phrase “more research is needed” is the bane of the research commissioner’s life. Following on from a piece of evidence synthesis it is so frustrating to be told that more research is needed, but not to be told exactly what that research is. The author advocating research is probably, at the moment he writes the words, the best placed person in the whole world to define exact evidence gaps which need to be addressed, but so often research recommendations are as an afterthought; when they should be considered an integral part of a research study.

The fundamental part of a research recommendation arising from an HTA report is a statement of the clinical uncertainties or decision problems which remain following the synthesis of evidence by the HTA authors, ideally in a prioritised order. These are vital - if they are incorrect then subsequently commissioned primary research will not be answering the most important questions.

How to formulate the question - the clinical decision problem - is the subject of another parallel WP7 SG2 paper, on *How to best formulate research recommendations*.

This paper considers how that formulated question might be best answered.

For the purpose of this paper we adopt the INAHTA definition of healthcare technology [2]:

> *Healthcare technology is defined as prevention and rehabilitation, vaccines, pharmaceuticals and devices, medical and surgical procedures, and the systems within which health is protected and maintained.*

2 Role of this document

This paper discusses the role of the reviewer in specifying the design of subsequent primary research by considering current practice in Europe and beyond.

It is not a guideline but a position paper that presents a view of EUnetHTA members as to the best approaches to specifying primary research methods to follow from systematic review type HTA reports.

It is not intended to replace an epidemiology or clinical trials textbook, but to provide consideration of various issues around designing primary research arising from a systematic review.

Similarly, it does not address the methods of evidence synthesis. We assume that appropriate evidence synthesis work (e.g. systematic review, meta-analysis, models of mixed / indirect treatment comparisons) have been undertaken before recommending further primary research. To do otherwise is wasteful [3-5].

3 Method of Production

Following initial teleconferences with HAS to discuss the remit of the work, the lead author prepared an outline, then proceeded to populate it based on publically available literature.

An initial literature search was carried out by HAS to retrieve and select most relevant articles and papers that provide a discussion on research methodology and designs.

A structured search strategy that combines Medical Subject Heading (MeSH) terms and free-text terms related to research designs in the specific context of additional evidence generation was developed (see Table 1). The Medline database was searched in March 2014 with no time limits and without restrictions on type of publication. An update was performed in October 2014.

Table 1 - Search Terms

	Terms	Number of hits
	(post-authorization[Title/Abstract] OR post authorization [Title/Abstract] OR postauthorization[Title/Abstract] OR post-marketing[Title/Abstract] OR post marketing[Title/Abstract] OR postmarketing[Title/Abstract] OR product surveillance, postmarketing[MeSH Terms] OR additional data[Title/Abstract] OR additional studies[Title/Abstract] OR additional study[Title/Abstract] OR evidence gaps[Title/Abstract] OR research gaps[Title/Abstract] OR observational[Title/Abstract] OR non interventional[Title/Abstract] OR non-interventional[Title/Abstract] OR real life[Title/Abstract] OR real world[Title/Abstract])	93
AND		
	(HTA[Title/Abstract] OR health technology assessment [Title/Abstract] OR technology assessment, biomedical[MeSH Terms] OR Pharmacoepidemiology[MeSH Terms])	
AND		

	Terms	Number of hits
	(Research/methods[Mesh:NoExp]) OR Research Design[Mesh:NoExp] OR study design[Title/Abstract] OR research design[Title/Abstract] OR study methodology[Title/Abstract]) OR research methodology[Title/Abstract] OR Clinical research[Title/Abstract])	

The search in PubMed was supplemented by a manual internet search of:

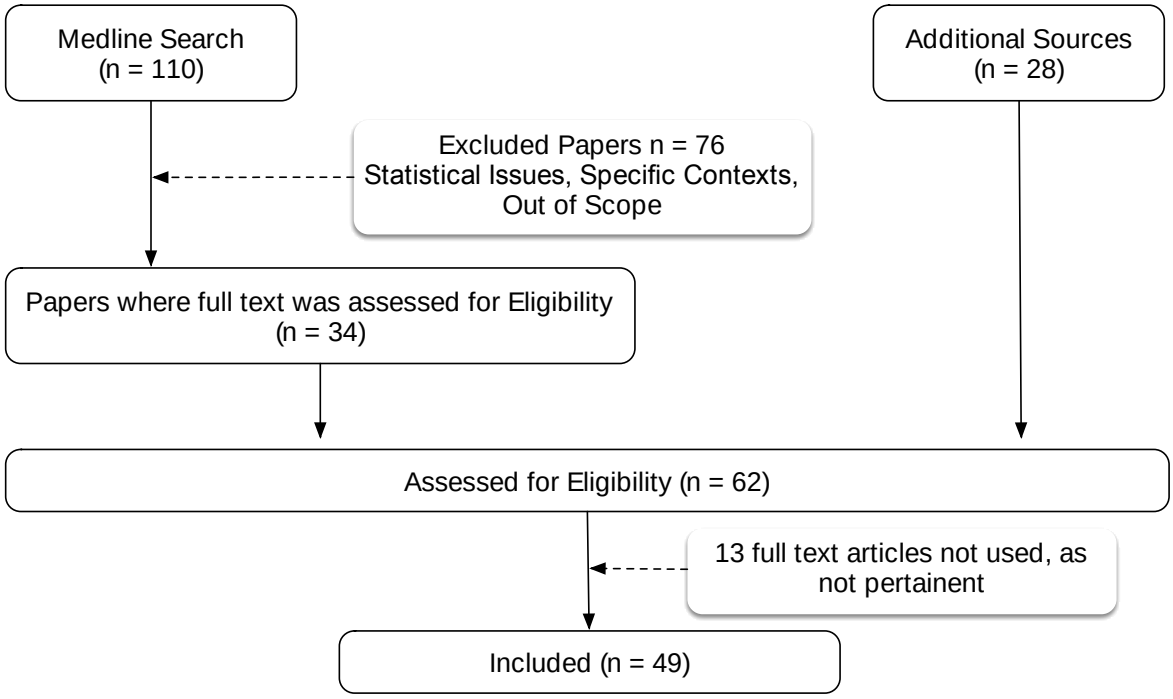
- websites of main international governmental and HTA agencies;
- other organizations of interest such as European Network of Centers for Pharmacoepidemiology and Pharmacovigilance (ENCePP)
- Google and Google scholar.

The search was performed in March 2014. Only the websites available in English and French (or with an English version of publications) were examined. First, the 'publication' tab (when available) was searched for relevant reports and documents related to formulating research recommendations or dealing with uncertainty. As a second step, a manual search in the search box was conducted using terms following keywords "design", "methodology", "gaps".

WP7 SG2 partners were also contacted in order to provide any documentation of interest from their agencies.

The synthesis process is summarised in Figure 1 on page 17.

Figure 1 - The Synthesis process for identified literature



It is important to remember that this paper does not hold itself out to be a systematic review. It presents a view of EUnetHTA members as to the best approaches to specifying primary research methods to follow from systematic review type HTA reports. That view has been informed by systematic literature searching, but also by papers and other information of which contributing EUnetHTA members were aware (either through their own reading, or as outputs of their organisations), and the experiences of those member organisations in requesting and developing primary research (used to inform section 5).

There then followed an iterative process of drafting and commenting, where EUnetHTA WP7 member organisations, SAG members and ENCePP members provided feedback and comment – on both the content and presentation of the work.

Finally, the document was sent for public consultation.

4 The Document Structure

Section 5 considers current practice and guidance in a number of European organisations, and some international, which produce systematic reviews.

Section 6 builds on this to consider when it might be appropriate for a systematic reviewer to specify a research design for subsequent primary research.

Section 7 reviews problems in specifying a primary study design which were identified by partners as part of the drafting and reviewing process for this document in early 2015.

The three appendixes provide a very high level consideration of various issues around designing primary research, but they cannot replace a text book or experienced researcher, either or both of which should be consulted before developing a real primary research project.

Appendix one illustrates considerations related to particular design questions.

Appendix two discusses whether why the reason for requesting evidence should have an influence on the selected design.

Appendix three outlines the hierarchy of evidence.

5 Current Practice in Specifying Study Designs

In this section we will review current practice in bodies with regional or national focus, and some organisations with a global scope.

5.1 Current Practice in EUnetHTA partners

The WP7 SG2 lead HAS performed in 2013 a Survey on the possibilities and conditions for performing harmonised Additional Data Collection (ADC) among EUnetHTA partners. The aim of the survey was to allow deeper understanding of HTA bodies' practices in the domain of ADC and to explore the possibilities and conditions for performing harmonized data collection in different countries.[2]

All in all, the results showed that six agencies (30%)[3] may formulate their research recommendations in the HTA report in a detailed manner, while eight (40%) provide rather limited indications for further research at the stage of production of the HTA report.

With regards to these results, EUnetHTA partners have been contacted to provide more detailed descriptions of their practices. The countries illustrated here are those for whom organisations in those countries were able to provide information in response to this request, or those for whom organisations provided necessary information during WP7 consultation on the first draft of the paper.

5.1.1 England

The two major sources of systematic review type HTA reports in England are the National Institute for Health Research (NIHR) and the National Institute for Health and Care Excellence (NICE). The Centre for Reviews Dissemination also has relevant guidance on how research recommendations should be presented.

[2] The survey has been conducted in two parts: 1° initial survey (4 questions) to identify EUnetHTA partners with a significant experience in ADC recommendations/requests; 2° more detailed survey (25 questions) for partners identified as having significant experience in the field.

[3] Twenty partners have responded to the detailed survey on ADC practices (25 questions).

5.1.1.1 NIHR Programmes

For the research programmes which publish in the NIHR Journals Library, authors are sent guidance around 6 months before their reports are due [6,7]. The NIHR does not mandate a particular format for research recommendations, but says

> *"It is perfectly acceptable and desirable to make recommendations about future research".*

In practice authors adopt a similar process to that used by NICE reports, they specify a research question (usually in PICO format[1]) and sometimes specify that they recommend a randomised controlled trial but with little further detail.

5.1.1.2 Centre for Reviews and Dissemination

The Centre for Reviews and Dissemination is funded by NIHR to provide research-based information about the effects of health and social care interventions via databases and undertake systematic reviews evaluating the research evidence on health and public health questions of national and international importance.

They have produced guidance on writing a systematic review [8]. With regards to research recommendations they state

> *"A clear statement of the implications or recommendations for future research should be made; vague statements along the lines of 'more research is needed' are not helpful and should be avoided. Specific gaps in the evidence should be highlighted to identify the research questions that need answering. Where methodological issues have been identified in existing studies, suggestions for future approaches may be made. Where possible, research recommendations should be listed in order of priority, and an indication of how rapidly the knowledge base in the area is developing should be included. This can assist in planning an update of the review and help guide commissioners when allocating funding.*
>
> *The DUETs initiative (Database of Uncertainties about the Effects of Treatments; (www.duets.nhs.uk), recommends the*

presentation of research recommendations in a structured format represented by the acronym EPICOT (Evidence, Population(s), Intervention(s), Comparison(s), Outcome(s), Time stamp). Timeliness (duration of intervention/follow-up), disease burden and suggested study design are considered as optional additional elements of a structured research recommendation. Further details and an example of how to formulate research recommendations using the EPICOT format can be found in an article published by the DUETS Working Group [1]. It is worth noting that there is some debate about the applicability of the EPICOT format for some reviews, particularly those of complex interventions [9]."

5.1.1.3 NICE

NICE has publically available process and methods guides which are updated on a regular basis.

5.1.1.3.1 Technology Appraisals

At the time of writing, NICE last reviewed its methods for technology appraisals in April 2013 [10].

For technology appraisals, when the evidence of clinical effectiveness or impact of a technology on other health outcomes is either absent, weak or uncertain, the Appraisal Committee may recommend that the technology is used only in the context of research or while the technology is recommended as an option, research is also conducted.

The committee is directed to consider

- the need for and potential value to the NHS of additional evidence that can inform the development of NICE guidance and clinical practice on the use of the technology
- the uncertainty in the analysis and what could be gained by reconsidering the decision in the light of research findings
- whether the research is feasible in circumstances when the Appraisal Committee recommends the intervention for NHS use outside the context of research

- irrecoverable costs incurred from introducing the technology
- the likely net benefits for all NHS patients of use only in a research setting during the time that the recommended research is being conducted.

If minded to make a 'research only' recommendation, the committee is directed to consider

- the likelihood that the research needed will be commissioned and successfully report
- the time it is likely to take for research findings to be available to inform subsequent NICE guidance and clinical practice
- other factors which may impact on the value of evidence generation, such as other research that is underway or likely to be commissioned and completed.

The committee is not directed to consider the methods of any resulting research.

5.1.1.3.2 Guidelines

The Guidelines Manual was updated in 2012 [11].

Identifying further research requirements is a significant part of the guideline development process. The guide however covers very little of how to do this, and instead refers users to NICE's research recommendation manual.

5.1.1.3.3 The Research Recommendations Manual

This manual was last updated in 2014 [12]. The manual discusses process extensively, in a systematic manner, mainly focusing on identifying the clinical decision problem which should be addressed, using a PICO format. The guidance discourages the specification of research methods, saying:

> *"It is seldom appropriate to specify the 'most appropriate' study design to address the proposed question as there may be a number of alternatives depending on timescale and context."*

5.1.2 Sweden

The two major sources of systematic review type HTA reports in Sweden are two government agencies under the Ministry of Health and Social Affairs: the Swedish Council on

Health Technology Assessment (SBU) and The National Board of Health and Welfare (Socialstyrelsen).

The Dental and Pharmaceutical Benefits Agency (TLV) determines whether a prescribed pharmaceutical product or medical device shall be subsidized by the state. The evaluation of the new products' cost-effectiveness can in some instances be akin to a HTA report. The agency also performs health economic evaluations of hospital medicines and medical technologies.

5.1.2.1 SBU

The Swedish Council on Health Technology Assessment (SBU) delivers systematic assessments via several different types of reports, ranging from a review of an entire therapeutic field to more limited reports covering single new interventions.

In addition to being a source of information for those who work practically in health care the assessments are also a source of information for other decision-making authorities in Sweden, such as the National Board of Health and Welfare and the Dental and Pharmaceutical Benefits Agency.

Each report also includes a section describing on-going research and also areas where the group has identified the need for further research. These suggestions are described in detail regarding patient group and intervention. Suggestions can also be made regarding appropriate methods and possible existing datasets to use. No organisation to undertake the suggested studies is specified.

Taking inspiration from UK DUET (see section 5.1.1.2), SBU also engages in identifying scientific uncertainties in health care. The findings are compiled in a database that can be accessed on the SBU website. Based on the database SBU aims to inform researchers of topics where evidence is lacking.

In an on-going pilot project, the SBU is using methods derived from the James Lind Alliance in order to prioritize areas of research, based on the knowledge gaps identified in an HTA-report on ADHD.

5.1.2.2 Socialstyrelsen

The National Board of Health and Welfare (Socialstyrelsen) is regulating several aspects of healthcare and social services in Sweden. For therapeutic areas where healthcare takes

considerable resources guidelines are published in order to support decisions concerning the allocation of resources. The guidelines are comprehensive and include recommendations for most interventions for a specific disorder but focus primarily on issues where the need for guidance is large. The guidelines are based on systematic assessments and the clinical experience of healthcare professionals (consensus methods) and include a recommendation rating for each studied intervention, from 1 (high) to 10 (low). The intervention can also be deemed as “do not”, i.e. it should be phased out from practice.

When the underlying scientific support for an intervention is weak or ambiguous, the recommendation may be set to “research only”. This recommendation is often used for newly introduced technologies and indicates that it only should be used within a structured research context where the outcome of the intervention can be followed up and evaluated. Normally it is not, within this prioritization setting, described in any further details how these interventions should be studied.

5.1.2.3 TLV

The Dental and Pharmaceutical Benefits Agency (TLV) regulates pricing of prescribed products in Sweden. The pricing policy is based on value based pricing and that the manufacturer applies for reimbursement for each new product. The application, and the cost effectiveness of the new product, is assessed by the agency using HTA methodology. The assessment is presented to the agency’s board of experts and a decision is thereafter made public.

The application for reimbursement of a new product is normally done in close conjunction with the market approval of the product. At this point in the lifecycle of the product its costs and benefits to society are often characterized by uncertainty. In these instances TLV may include a ‘coverage with evidence development’ agreement with the manufacturer in the reimbursement decision. The agreement normally implies that the manufacturer will report additional data for the product within approximately three years from the reimbursement decision. The details regarding suggestions for appropriate methods and possible data sources are defined in discussions with the manufacturer and in some instances academic expertise.

5.1.3 France

In France, the French National Authority for Health (HAS) assesses health technologies and issues recommendations on their reimbursement. HAS' Appraisal Committees may recommend reimbursement of technologies even if evidence gaps have been highlighted during the assessment. In that case, HAS' report includes also a recommendation on additional data that should be submitted by the manufacturer at the time of the reassessment of the technology (usually 5 years later, or sometimes with shorter timeline if earlier reassessment is required).

The Committees' recommendations specify specific clinical data required (outcomes), preferably in a prioritised manner, the population, comparators (if needed) and timelines, but rarely mandate a specific design. The best source of data can sometimes be suggested, for example if information could be extracted from French health assurance database or gathered through institutional cohorts or registries financed by the Ministry of Health.

It is the manufacturer that then submits the protocol for research, usually 6 months after the Committees' recommendation. HAS ensures that the proposed protocol will adequately address to the research question, and recommends amendments to the protocol, if necessary. These amendments can include changes in study design.

HAS has also published guidance intended to help manufacturers or clinical research organisations [13,14]. It includes suggestions on design options and methodological choices, along with a discussion of advantages and drawbacks of each option. Some recommendations on how to best address the research question have been formulated, based on the HAS' experience, but no specific design is mandated.

5.1.4 Spain

In the Spanish system the bulk of Health Technology Assessment is undertaken by organisations with a regional rather than national remit. This section discusses current practice in those regions which have provided information.

5.1.4.1 Avalia-T, Galicia

The recommendations delivered by Avalia-T are based on uncertainties that are derived from the literature.

The recommendation will depend on the whether it is determined that the evidence is lacking or insufficient to draw up firm conclusions on clinical/cost-effectiveness or safety (GRADE assessment [15]) or whether the uncertainties are much related to the application in real practice scenarios or long term safety issues (modes of delivery, optimal service design, patient defined outcomes, adequacy of use, economic impact, etc.).

In the first case, a general recommendation can be made regarding the need for further research to ascertain efficacy or safety, including considerations on the appropriate study design (for example if a clinical trial is considered essential) and outcome variables. When the technology is considered to be of high value but important uncertainty gaps remain, the technology can be recommended for reimbursement under special follow up conditions (subject to specific indications, application and evaluation protocols, authorized only in designated centres and for a limited period of time). If the technology is considered effective and safe but there are important doubts as to the applicability of the results when the technology is used in wider populations or important uncertainty regarding the diffusion or application of the technology in real life practice (off label use, cost deviations) a recommendation is made regarding the convenience of a post-introduction observation. A prioritisation tool that allows for scoring the technologies to ascertain the level of uncertainty (high, intermediate, low) would be applied.

The recommendations made in the reports specify the uncertainties and research scenario (special follow up conditions, post-introduction observation) but the specific design AND outcome variables are discussed and decided upon in a further stage with all relevant stakeholders (HTA doers, health care professionals, representatives from health authority).

5.1.4.2 AETS-ISCIII

Reports performed in AETS-ISCIII sometimes include recommendations for future research.

They are usually general recommendations. AETS-ISCIII does not have a standardised process. Specific gaps in the evidence are usually highlighted.

Some authors include the most convenient research design depending on the HTA results.

These are general research recommendations by the authors, without a consensus with other agents involved in the technology.

5.1.4.3 AETSA

The Andalusian Health Technology Assessment Agency delivers information through different types of products: systematic reviews, rapid assessment reviews, evidence-based recommendations for the selection and management on the use of new pharmaceuticals (known as Drug use recommendations), and clinical practice guidelines.

Although AETSA´s methodological guidance for the development of HTA appraisals does not specify that in the cases where there are gaps of evidence, in terms of clinical efficacy, effectiveness or safety, HTA doers should make recommendations for future research, specific gaps in the available evidence are highlighted in the reports.

On the other hand, clinical practice guidelines and recommendations for drugs therapies produced in AETSA include a section about recommendations for future research, following the indications of the methodological guidance developed by the Spanish National Health Service (Elaboración de guías de práctica clínica en el Sistema Nacional de Salud. Manual metodológico [16])

The guidance states:

> *"It is desirable that the proposals for future research are outlined with certain level of concision in order to be useful and comprehensible for a future research project. If the evidence available is limited, the proposal could be a general one. On the contrary, when a theme has been studied in depth, the proposal could be more indicated into detail, for example designating a specific subgroup or defining the period of follow up needed.*
>
> *If it is possible, the proposals should follow the PICO structure (suggesting the population, potential interventions and comparators, and the outcomes of interest). Ideally, if the research question required this, the design of the study and the time required for long term follow up should be specified."*

5.1.5 Norway

In Norway, the Norwegian Knowledge Centre for the Health Services commissions systematic reviews. In their guidance to reviewers they encourage research recommendations to be presented in EPICOT format [1], and also to provide answers to

- Which study designs are most optimal, or possible?
- What time framework should the studies have?
- What evaluation elements should be included (ethical, organizational)?
- What professional groups are the most relevant?
- Is international collaboration relevant?
- What is the expected value of perfect information if examined.

5.1.6 Italy

In Italy, the HTA field is covered by several regional agencies in addition to the national level.

5.1.6.1 AIFA

AIFA is the national level Italian Medicines Agency. It uses two different approaches for requesting and supporting additional data collection.

For reimbursement purposes, the AIFA Scientific Committee (CTS) assesses available evidence and, after having identified gaps in the data (e.g. on effectiveness, safety, cost-effectiveness, condition of use), recommends to collect additional evidence mainly through registries or clinical studies.

Should a registry be required, AIFA Units formulates the forms, including data to be collected and eligibility criteria for prescribing medicines, which will be submitted to CTS for final approval. The forms will be filled in by physicians and pharmacists in order to prescribe and to obtain revenue in case of Managed Entry Agreements. Since 2013, new registries have been implemented with a new structure which enables a better quality of data collection and of information which will be analysable.

Moreover, another way to collect additional data is in place in Italy and it is called “Independent research”. The general aim of the program is to support clinical research on

medicines in areas of interest for the National Health Service (NHS) and where commercial support is normally insufficient. For the identification of research areas a scientific committee (Committee for research and development, R&D) supports AIFA. In 2012, two main areas of medicines research were identified including:

Area 1. Head to head comparison of medicines and therapeutic strategies applied in health conditions with high impact on public health and National Health Service

Area 2. Pharmacoepidemiology studies on benefit/risk of treatments and studies assessing the efficacy of strategies for the improvement of the appropriateness of medicines use.

For independent research program, the researchers submit the “letter of intent” (i.e. a synthesis of the study protocol) which is evaluated by the R&D Committee for funding decision purposes. Therefore, there is not a discussion about the study design.

5.1.6.2 Regione del Veneto

The HTA reports performed by Regione del Veneto do not include a specific or standardized section about further needed research, but if any evidence gaps are identified by the assessment, these can be discussed in the report. They rarely suggest a specific research design.

The recommendations performed are usually addressed to the appropriate use of the technology: whether this should be used without particular limits or limited to specific conditions (specific indications, only in authorized specific centers, monitored use etc).

5.1.6.3 ASSR-RER

In the only two reports in which recommendations for research were formulated, ASSR has expressed the indication on study designs. In the HTA report on robotic-assisted surgery they stated the need for RCTs in specific clinical indications for robotic surgery, outlining population, comparators and outcomes. When these recommendations for research were taken up by surgeons, they strenuously argued in order to be allowed to carry out a non randomized prospective controlled study, because of the difficulties (reported also in the IDEAL paper) that randomization poses in surgery. There was concern that had the RCT design not been specified, surgeons would have opted for a case series study. In the other HTA report on innovative radiation therapy (tomotherapy) ASSR specified the PICOs and designs for additional evidence generation and RCTs were carried out.

5.1.7 Germany

Infrequently (depending on commission, context and assessment results) IQWiG reviewers outline the main features of additional studies which would answer the respective research question of the REA (Relative Effectiveness Assessment), and help to reduce the identified uncertainties in regard to clinical effectiveness and safety.

5.1.8 Switzerland

Switzerland has several institutions that provide HTA reports, which generally evaluate evidence generated from finalized studies. If significant evidence gaps at the level of efficacy, safety or cost-effectiveness are being discovered the Federal Office of Public Health (FOPH) may recommend a technology for conditional reimbursement. Additional evidence generation is demanded. A set of criteria have been defined for technologies to be considered for this form of reimbursement. One of the criteria regards the evaluation concept, including study design.

5.1.9 Belgium

The Belgian Health Care Knowledge Centre (KCE) is a semi-governmental institution with the mission to advise policymakers on decisions relating to health care and health insurance, on the basis of scientific analysis and research. Their activities include Health Technology Assessment, developing clinical practice guidelines, investigating the optimal means of organizing and funding health care and developing effective research instruments.

KCE can recommend a specific design for further research to policy makers. KCE is an independent institute and is not involved in the decision-making or implementation process.

In Belgium, Coverage with evidence development mechanisms exist for drugs and devices, in the context of either temporary reimbursement contracts that are linked to data gathering and an evaluation, or in the context of a restricted clinical application. The temporary reimbursement contracts are set up by RIZIV-INAMI[4] and a working group composed of

[4] National Institute for Health and Disability Insurance (‘Rijksinstituut voor Ziekte- en Invaliditeitsverzekering’/’Institut National d’Assurance Maladie-Invalidité’)

various stakeholders (some having advisory role only) : representatives of the applicant ; committee members (CTG-CRM[5] for drugs and CTIIMH-CRIDMI[6] for devices) ; representative of the Minister of Social Affairs ; representative of the Minister of Budget ; representative of the Minister of Economics (for drugs only) ; representatives of the sickness funds ; representatives of the professional association of the pharmaceutical or device industry.

For new promising technologies with good indications (but not sufficient evidence) that they are safe and (cost-)effective, coverage with evidence development (CED) in research is suggested while HTA should be the standard for the assessment of most procedures.

[5] Drug Reimbursement Committee ('Commissie voor Tegemoetkoming Geneesmiddelen'/'Commission de Remboursement des Médicaments')

[6] Committee for the Reimbursement of Invasive Medical Devices and Implants ('Commissie Tegemoetkoming Implantaten en Invasieve Medische Hulpmiddelen'/'Commission de Remboursements des Implants et des Dispositifs Médicaux Invasifs')

5.2 International

Some groups do not have an obvious national home, and are discussed in this international section.

5.2.1 The Cochrane Collaboration

The Cochrane Handbook for Systematic Reviews of Interventions lays out a structure for research recommendations in section 12.7.3. [17]

It dwells on how the research question should be defined, recommending the EPICOT format of Brown et al[1].

It suggests that

> *"…it is also helpful to note the study types, as well as any particular design features, that would best address the research question…"*

but does not go into any detail about how the reviewer should consider the design options.

5.2.2 PRISMA

PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. It is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses [18].

Item 26 of the PRISMA statement says

> *"Provide a general interpretation of the results in the context of other evidence, and implications for future research".*

The example given dwells on the clinical decision problem, and recommends both a trial (though no further detail is given) and the need to consider different designs to answer a particularly cogent question [18].

The PRISMA group advise authors to make explicit research recommendations, and suggest that systematic reviews have great potential to guide primary studies. They have nothing to say about choosing particular designs.

5.2.3 STROBE

The STROBE statement - The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement – provides guidelines for reporting observational studies. [19,20]

It makes no comment on the design of subsequent primary research.

6 About Study Designs

6.1 Is it feasible to answer an identified research question?

Our belief is that the feasibility of the recommended primary research question is not an issue which the reviewer needs to consider. The job of the reviewer is to identify important clinical decision problems which remain after completing a review. It is important that primary researchers and funders are informed of all the important questions to be answered, not filtered through the reviewer's belief of what might be feasible. Should a primary researcher engage with the question, they may be able to adopt approaches which the reviewer did not consider. There may be underlying work to be conducted first to maximise the chances of coming to a useful conclusion. Uncertainties around the ability to deliver a definitive answer from primary research – eg is it possible to recruit centres and patients, will the clinicians deliver the allocated interventions, will the patients accept the allocated interventions and return for follow-up assessment – may best be managed initially with feasibility or pilot work rather than a full scale attempt to derive efficacy data. [21] [22-24]

To emphasise – the most important thing that a reviewer can do when identifying research questions following a review is to enumerate the important questions, ideally in priority order.

6.2 Should a study design be specified?

While there are reasons why a reviewer should suggest an approach to conduct primary research (discussed in section 6.2.1) there are also reasons that a systematic reviewer should not specify a particular primary research design, but should constrain their role to identifying the clinically important research question discussed in section 6.2.2.

It is often appropriate for those undertaking primary research to discuss their proposed designs with reviewers, assessment agencies or appraisal agencies. Chalmers et al suggest

that all research should be informed by the existing evidence, and that implies that primary researcher should ensure as much as possible that their findings can be incorporated into the evidence base, in either meta-analysis, systematic reviews or guidelines. [3]

6.2.1 Reasons for Specifying a Study Design

- Practical reasons
    - The organisation[7] has a mechanism for funding or delivering particular designs and not others.
    - The customer for the research wants to see particular designs (eg will only accept RCTs to influence its guidance)
    - The organisation is required to specify particular designs (eg may only fund registry studies)
- Waste minimisation
    - There is already a large volume of lower hierarchy designs, or low bias designs delivered poorly, and the reviewer wishes to acquire evidence using designs which are less subject to bias – and to make it clear to primary research funders and researchers that further small studies at risk of bias will be of no use in changing policy or planning services.
    - The reviewer wishes to address more than one question, and has established that it is practical to do so e.g. Two drugs for a disease with no plausibility of interaction would suggest using a factorial trial design.
    - There are uncertainties which preclude moving directly to a definitive randomised trial, which could be addressed through feasibility or pilot work, or by following a framework such as the UK Medical Research Council's Complex Intervention evaluation framework[25] or the IDEAL recommendations [26].
    - There is a need to fit the new evidence into an existing evidence base – for example the confidence interval of an existing meta-analysis needs to be narrowed.

[7] The 'organisation' means the group specifying the design, such as an HTA agency.

- To ensure scientific quality or usefulness of research
    - If the main uncertainties are around safety, and especially long-term safety, there may be a need to steer primary researchers towards cohort and registry studies, and away from trials.
    - There may be a requirement to ensure that entry to cohort and registry studies is based on indication, not treatment received, to ensure population coverage and a view of natural history. While device manufacturers (and to a lesser extent pharmaceutical companies) may maintain registries, these are often (for commercial reasons) gated by the use of the manufacturers product and so less useful than they could be for comparing different approaches to management of a condition.
- To ensure appropriate comparisons are made.
    - A manufacturer will often wish to to compare its offerings to placebo rather than active treatment. This may be for regulatory purposes (eg to obtain marketing authorisation); or because the standard of care differs in various countries, and so a single study with an active comparator would only be relevant and pragmatic in a small number of countries. Health services funders will of course want the most relevant information for their country. This can be mitigated by specifying a robust study design which may include a direct head-to-head comparison of relevance to clinical practice in one or more countries.
    - To provide information to precisely fill the gaps identified in a systematic review
    - If there is a pre-existing meta-analysis it may be most appropriate to power a subsequent trial based on the information required to move the meta-analysis, rather than as a stand-alone trial.
- Social and ethical reasons
    - When a particular design is not socially or ethically acceptable. However, these two justifications should be viewed with caution - the phrase “that’s not

ethical" is often used when what the speaker means is "I don't like that". Similarly, the appropriate response to a study design being seen as socially unacceptable may be education rather than finding another less rigorous design.

6.2.2 Reasons for not Specifying a Study Design

The reviewer should particularly avoid specifying a design if

- Reviewers lack primary research expertise across a range of possible designs. While skilled in analysis of primary research it is unlikely that any given systematic reviewer is an expert in primary research design: skill in interpreting deficiency in reporting of primary research does not necessarily translate to design expertise.
- Reviewers lack time or resources to fully explore the implications of the design they are planning to recommend, and whether the proposed study is feasible.

To identify the cogent primary research questions arising from a review is a vital skill - what is the gap that a clinician, commissioner or patient faces? Part of this is identifying the reasons for a question - for example if the uncertainty around a new surgical procedure related to safety, efficacy, or both?

The underlying uncertainty will help guide a primary researcher to delivering an appropriate design to reduce identified uncertainties. Sometimes the uncertainty might best be addressed through a programme of related studies alongside the technology lifecycle such as a mixture of developmental and effectiveness studies. An approach such as the IDEAL framework might be adopted [26].

It would be unfortunate if that researcher delivered a randomised controlled trial specified by a reviewer, when the uncertainties would have been better addressed by a cohort study or other observational design.

It is better not to specify a design, than to specify the wrong design.

7 Dilemmas in specifying a study design

This section will consider problems in specifying a primary research study design arising from evidence synthesis, which have been identified in consultation with EUnetHTA partners and stakeholders. It cannot hope to be exhaustive or authoritative, but will hopefully provide guidance.

This section may also be quite time sensitive – in that problems expressed during the consultation in the first half of 2015 may not be topical a couple of years later. More than any other, this is probably the section of the document most in need of regular review and update.

7.1 What is the most appropriate design for addressing particular classes of evidence gap?

The table below sets out a general starting point for design selection. It may be necessary, or even desirable, to deviate from these suggestions depending on a local situation. A pragmatic approach should be adopted.

In all cases, appropriate methodologists – such as epidemiologists, statisticians and health economists – should be consulted.

The hierarchy of evidence, which may help with the selection of a design, is discussed further in appendix 3.

Class of Gap	Starting Design	Why
Effectiveness	A randomised controlled trial, based in the real practice with an active comparator if possible (eg a pragmatic randomised controlled trial). For a superiority question analysis should be by intention to treat. If the question is around equivalence then a per protocol analysis will also be required.	This experimental design, if well conducted, minimised bias and ensures confounders are evenly distributed across study arms. The active control allows direct comparison against current practice – thus allowing a decision to be made on which intervention in the study is superior.
Efficacy	A randomised controlled trial, usually with a placebo or sham control arm.	The experimental approach minimises bias. The placebo or sham allows the isolation of the effect of the intervention under test.
Safety	Observational study. For ongoing monitoring of potentially serious but rare events a registry is probably most appropriate.	For adverse events, especially rare but serious adverse events, large numbers of patients may be required to characterise the total risk. Example are joint replacement registers, where every implanted prosthetic joint is registered; and drug adverse event registers where the intention is to capture all adverse events which may be attributable to a drug.

7.2 What to do about multiple evidence gaps?

It is not unusual that a review identifies more than one important clinical uncertainty. Appropriate steps in this case could be [27] [28-30]

1. Rank the uncertainties in order of importance. Importance might be determined by the number of affected patients, the cost of managing the disease either to the health system or society, or the DALY burden which might be alleviated were research to take place. A formal value of information analysis might be conducted to help with this ranking. Any organisation undertaking reviews should have their own guidance to help with this prioritisation.
2. Consider whether some of the uncertainties might be grouped together with an appropriate study design. For example should there be issues of both duration of effect and safety then an observational study might conveniently assess both. Knowing this might effect the value for money of any given research project.
3. The reviewer can then apply the steps in section 6.2 to determine whether any further recommendations should be made on design.

7.3 When is it acceptable to specify a design other than an RCT?

Randomisation is often recognised as the gold standard for addressing questions of effectiveness or efficacy. Randomisation serves to ensure than on average the groups involved in a study are sufficiently similar that the only important difference between them are the interventions which are under the control of the investigator. Therefore, any difference in outcomes can reasonably be attributed to the interventions – allowing the inference of causation.

However there are circumstances where a randomised approach is not possible – and this often relates to the lifecycle of the technology under investigation, and where it sits in the spectrum of possible technologies to use for a patient.

7.3.1 New Technologies

Some technologies are new – and hence have little existing evidence – but are viewed as having great potential for impact for reasons such as knowledge of pathophysiology, biological plausibility, and action of similar technologies. For example, transcutaneous aortic valve replacement for aortic stenosis in the mid 2000s, or prostatic artery embolization for benign prostatic hypertrophy in the 2010s. New technologies, especially non-pharmaceuticals) may have two characteristics

- They are not in widespread use. For technologies which are operator dependent (eg surgical interventions) this offers the problem of lack of completely proficient practitioners to implement a trial.
- Their use is not standard. The problem here is a lack of agreement among clinicians as to what exactly should be included in a trial – each practitioner uses the technology is a slightly different way, and if their way isn't included as a trial arm they won't believe the results.

A solution to these issues is presented by the IDEAL framework [26].

For interventions with few practitioners, but with promise, the framework recommends prospective development studies to explore the technology with patients and to develop expertise. For the next stage – with many practitioners but little consensus on the detail of use – exploration studies are suggested. In these studies the experience of practitioners is brought together so the community can come to a consensus on use, and then proceed to a randomised efficacy or effectiveness trial if appropriate.

7.3.2 Too Many Technologies

Sometimes there are just too many competing technologies to handle neatly in a single trial – but no obvious front-runners to compare. Consider for instance the multiple ways in which liver metastases can be managed – surgery, external radiotherapy, brachytherapy, selective internal radiotherapy (SIRT), cryotherapy, high intensity ultrasound (HIFU), etc

There are two reasonable approaches to consider.

The first is an IDEAL 2b study, using a non-randomised approach with developing consensus to narrow down the exact research question to be asked [26].

The second is a multiple arm RCT, with adaptive allocation ratios as the trial proceeds [31-33]. This relies on an outcome measure being available for some patients within the recruitment period of the study. The allocation of the next patient is then weighted towards the interventions to which allocation would provide the most information. The requirement of the results of individuals being available within the recruitment time period is often not met, especially in pragmatic research – so adaptive designs are the exception rather then the rule. Adaptive designs are particularly methodologically challenging, and appropriate advice should be sought before recommending one.

7.3.3 Equipoise for the community, but not individuals

Imagine a well-established set of competing technologies with differential use on a national basis. Possibly each hospital has chosen their technology, and the physicians believe in it. There is community level equipoise – as a whole the community has substantial uncertainty as to which technology is superior. But there is no individual level equipoise – no individual physician has uncertainty as to what the appropriate technology to use is.

In this circumstance, a randomised trial can be difficult, if not impossible. Even if clinicians agree to enter patients to a trial they will often find excuses not to do so, so they can apply their preferred treatment. For some technologies, randomisation may mean a patient having to travel to a different hospital (if the required equipment is expensive or difficult to transport) and so the patient will not agree to take part in a trial.

To some extent for equipment-based technologies, if hospital management will agree, this can be address with a cluster randomised design – the equipment for one intervention is removed from some hospitals and made available in others. This will not work if the equipment is not unique to the procedures – and patients who prefer a procedure not available in their local hospital will present to one where it is available thus introducing selection bias.

More likely approaches would be an IDEAL 2b study to attempt to introduce uncertainty to practitioners [26], or take advantage of the distribution of the technologies to undertake a natural experiment [34,35].

7.4 Linking a design to the question

A not uncommon approach to a perceived need for further data is to set up a registry.

While registries can be useful, they are limited by the data they collect, and potentially by the lack of a clear analysis plan.

It is essential that any primary data collection in response to an identified evidence need is capable of addressing the identified question.

There are two common failings of registries, which can be avoided with a little thought.

1. **Entry restricted to patients receiving a specific technology**. This may seem to be the right approach if the driving questions are around safety, or the natural history of patients once they receive the technology in question. However, this approach precludes comparison. If registration is disease based rather than treatment based it becomes possible to make comparisons to a contemporaneous group who received alternative treatments, or no treatment. While necessarily more subject to bias than an experimental design, such an approach may exhibit more external validity.
2. **Lack of a clear analysis plan**. Registries are often open ended. In some this is appropriate – for the questions of interest aren't known at the inception of the registry. For example joint registries are set up to examine the safety and longevity of joint prostheses, but the exact comparisons to be made aren't known initially. Where a question is known then an analysis plan should be specified prospectively, with the time of analysis determined and the comparisons to be made described. Deciding on analysis based on what looks interesting in the data will lead to bias.

It would not be unreasonable to conclude from these two points that in general a prospective cohort study might be preferred to a registry.

7.4.1 Feasibility Studies

Feasibility Studies are pieces of research done before a main study in order to answer the question "Can this study be done?". They are used to estimate important parameters that are needed to design the main study. For instance:

- standard deviation of the outcome measure, which is needed in some cases to estimate sample size

- willingness of participants to be randomised
- willingness of clinicians to recruit participants
- number of eligible patients; carers or other appropriate participants
- characteristics of the proposed outcome measure and in some cases feasibility studies might involve designing a suitable outcome measure follow-up rates, response rates to questionnaires, adherence/compliance rates, ICCs in cluster trials, etc
- availability of data needed or the usefulness and limitations of a particular database
- time needed to collect and analyse data

Feasibility studies for randomised controlled trials may not themselves be randomised. Crucially, feasibility studies do not evaluate the outcome of interest; that is left to the main study.

If a feasibility study is a small randomised controlled trial, it need not have a primary outcome and the usual sort of power calculation is not normally undertaken. Instead, the sample size should be adequate to estimate the critical parameters (e.g. recruitment rate) to the necessary degree of precision. [21]

7.4.2 Pilot studies

Pilot studies are a version of the main study that is run in miniature to test whether the components of the main study can all work together. It is focused on the processes of the main study, for example to ensure recruitment, randomisation, treatment, and follow-up assessments all run smoothly. It will therefore resemble the main study in many respects, including an assessment of the primary outcome. In some cases this will be the first phase of the substantive study and data from the pilot phase may contribute to the final analysis; this can be referred to as an internal pilot. Or at the end of the pilot study the data may be analysed and set aside, a so-called external pilot. [21]

8 Summary

This position paper considers what a systematic reviewer might say about the desired methods of subsequent primary research arising from a systematic review. It does not touch on specifying the clinical question to be addressed, but on how that question might best be answered.

We have considered current practice from some national and a few global independent HTA bodies, and discovered that few organisations have issued guidance or recommendations on this topic. An example of typical guidance is:

> *"If appropriate consider suggesting what might be the most appropriate study design to address the proposed question" [36]*

It appears that HTA bodies generally confine themselves to identifying clinical uncertainties that require further primary research, but that they refrain from defining a specific research study design.

We then moved on to consider when a design should be specified. We came to the conclusion that reviewers could suggest an approach to conduct primary research if one of the following reasons can be applied:

- Practical considerations relating to the reviewing organisation's potential involvement in the subsequent research
- Research Waste Minimisation
- To ensure the usefulness of subsequent research, in the context of existing knowledge
- To ensure the most appropriate comparisons are made
- When there are overriding social or ethical issues.

On the other hand, we considered that reviewers should particularly avoid recommending a primary study design when they lacked expertise across a range of primary research methodologies, or lacked the resource to consider the more subtle implications of a recommended design or to investigate its feasibility.

We then rehearsed practical issues which had arisen from EUnetHTA partners during the production of this document. We thought about the relationship between types of research gap and the most appropriate study designs to address those gaps. We considered the situation of multiple evidence gaps and how to prioritise them. We considered when non-experimental evidence might be acceptable, and how to address problems with new or numerous technologies, or the distribution of uncertainty in the clinical community; and the challenges they present to conducting definitive randomised controlled trials.

We finally thought about the role of feasibility and pilot work ahead of definitive trials.

9 Conclusions

The phrase “more research is needed” is the bane of the research commissioner’s life, and as a stand-alone phrase should be abolished – it should always be accompanied by a description of the evidence gap which needs to be addressed.

The most difficult part of a research project is defining the appropriate, important underlying research question. Next comes selecting the correct study design.

It follows that the most important thing a systematic reviewer can do to facilitate future primary research is to specify, as exactly as possible, the important research questions arising from a review [37].

Specifying a research design is optional. It should involve consultation with guidance producers, trialists and other primary researchers, and if available clinical trials units [38] and research design services [39] (or their local equivalents).

What happens next will vary greatly with context. Some countries have robust and comprehensive processes for taking suggestions for pragmatic research, evaluating them, and funding those most likely to be of benefit. Others have systems which can address certain technology areas (e.g. pharmaceuticals) but not others. There are also efforts at an EU level to bring some level of community coordination to pragmatic primary research [40], which while they are not active currently, may be consumers of these recommendations in future.

Therefore, we conclude with the EUnetHTA position on how to decide on the appropriate study design to recommend:

1. Firstly, a systematic reviewer should identify and prioritise important clinical uncertainties, and set them out using the approach detailed in the companion paper to this on formulating research recommendations.
2. They could then, should he or she feel able, comment on possible research designs to address these uncertainties, if
    a. The required design is dictated by the existing evidence base.

b. There is a substantial risk that if a design is not specified a less than ideal design is likely to be adopted (e.g. in case there would be a party with a vested interest)
c. One of the reasons discussed in section 6.2.1 can be applied.

3. Else, a study design should not be specified.

10 Appendix 1 – Considerations related to particular design questions

10.1 Comparative Designs… or not?

To determine which intervention works best for which patients and which pose the greatest benefits and harms, primary research usually needs to include a comparator. The appropriate choice of comparative study design is influenced by the level of concern about selection bias versus the applicability of the results in the general population as well as pragmatic issues of data availability and quality. Randomized trials and observational studies are the most prominent study designs used to produce evidence in comparative effectiveness research. The different study designs have merits and drawbacks each, and should therefore be viewed not as alternatives but as complementary approaches. Reviewers should evaluate the available evidence and recommend new comparative research that adequately addresses the research question, regardless of study design.

For some questions, a comparator may not be needed – but these tend to be descriptive epidemiological questions such as 'how many people are taking this drug', or 'how many cases of this disease do we see each year'. Such questions are useful for service planning – but are unlikely to tell us how well things work.

10.2 Experimental v Observational Study Designs

Randomised Experimental designs – those where the investigator determines what happens to each patient through a non-predictable method – serve to evenly spread confounding factors, both known and unknown, between the groups in the study. The investigator can therefore be confident that any difference between the groups is caused either by chance, or by the differences he has imposed on them. Consequently, RCTs are usually preferred to evaluate effectives, as the design allows the attribution of causation. However, RCTs can also be expensive and slow, and subject to much regulation.

Observational studies are often convenient - different implementation of health care interventions in different places open an opportunity for a natural experiment[34,41]. This is a common approach in public health, but less commonly used in health services research.

The main methodological challenge of a natural experiment is confounding by indication - the risk that the two intervention groups differ for reasons associated with the choice of intervention.

However, it may be that randomisation is not possible, or is inappropriate.

For example, in the IDEAL framework [26] non-randomised designs are advocated in order to tightly agree specify the interventions to be assessed in advance of undertaking an experimental design.

Similarly, the MRC Complex intervention framework sets out when it is appropriate to undertake non-randomised studies in the development of a complex intervention [25,41].

There are also circumstances in which it is impossible to randomise who gets the interventions. These commonly occur in public health, where a health authority decides to, for example, fluoridate a water supply. In health services research we might envisage a situation where the surgeons in one hospital have a different practice to those in another, and neither group has sufficient uncertainty to participate in a randomised controlled trial. In such circumstances cohort designs of the underlying natural experiments should be considered [35].

One of the attractions of an indication based (ie including people based on a disease) is the possibility of complete data coverage of a population. This can be undermined depending on national views on data protection – in some countries, it may be illegal to collect data without the individuals' explicit consent, in others this is completely legitimate.

10.3 When to choose Prospective or Retrospective Study Designs?

Overall, study designs should be plausible, unbiased, affordable, provide value for money and study outcomes should correspond with the evidence gap. Above all, the outputs must be of use to the decision maker that the research is intended to influence.

Prospective research must be plausible, that is it must be practical to deliver. In most cases for example, requiring most patients with a condition to join a trial is not possible.

Prospective study designs help to remove some sources of bias, but tend to be more expensive and are necessarily slower to deliver than retrospective designs[42].

Retrospective designs suffer from recall or record bias[42,43]. They can be comparatively quick to undertake, but are limited by the quality of data already collected (or not collected at all) - which was usually not collected with research in mind. However, it has been suggested that in some cases such as drug utilisation studies this can be advantageous as non-research data may be less encumbered by social desirability bias. In the case of rare events, a retrospective case-control approach may be the only practical approach to obtaining any information in a timely way.

Value for money can only be determined by an individual funder (or a group of funders acting together), but approaches such as Value of Information analysis (see section 10.4.3 on Health Economic Modelling) may help inform this decision.

10.4 Touching on Modelling

In many cases, clinical or economic modelling is undertaken as part of an initial evidence synthesis, alongside a systematic review. Alternatively, questions resulting from systematic reviews may be best addressed by modelling study designs.

10.4.1 Limitations of Modelling

A model represents a best guess about the world. It attempts to synthesise the real world.

George Box tells us

all models are wrong, but some are useful [44,45]

So the main question for us is to consider to what extent are models of health care useful?

10.4.2 Clinical Modelling

Clinical Modelling is used to predict what would happen in situations outside directly observed evidence - for example beyond the duration of follow up.

Primary research is limited by

1. The expense of long term follow up.
2. The time required for long term follow up.
3. The willingness of participants to provide data in the long term.

4. The ability of researchers to keep track of participants beyond the duration of the original study.

The longer follow up gets, the lower the proportion of participants who can be followed up, and hence the more biased the responding sample becomes.

Clinical models can therefore be used to extrapolate beyond what is known from primary data, or to experiment with changes in treatment thresholds, disease severity etc.

The model is limited by the underlying data, and especially by uncertainty about how long observed behaviours in primary data will persist beyond the observation period

Before recommending primary research, a systematic review could consider whether the identified evidence gaps could be best served through modelling rather than new data collection.

10.4.3 Economic Modelling

The role of economic modelling is to address issues such as value for money and affordability. In contrast to clinical modelling, the stories that economic analysis tells could rarely be measured directly in primary research - though the models rely greatly (but not exclusively) on primary research – experimental, observational, and descriptive – to deliver their results.

Economic Models have two roles:

1. To help commissioners and reimbursement agencies decide whether to cover a particular technology
2. To determine if further research on a topic is economically sensible

10.4.3.1 Should a technology be provided within a national healthcare budget?

An agency deciding how to spend a national health care budget needs to reassure itself that individual funding or reimbursement decisions represent value for money.

This is commonly done by undertaking a cost-effectiveness analysis, informed by an underlying systematic review.

The cost effectiveness analysis allows the expected returns from the use of a health care intervention to be expressed in terms of a universally applicable measure – commonly the

Quality Adjusted Life Year (QALY), though other measures are used such as the DALY or more natural measures such as life years gained or lost – and therefore the intervention to be characterised as delivering a certain amount of benefit for a given cost. That is, the societal costs and benefits can be express as how many exta QALYs are returned compared to an alternative treatment for each extra unit of currency invested. This is known as the incremental cost-effectiveness ratio, or ICER.

This then allows different interventions and disease states to be compared in terms of the expected health gain returned for a given investment. For example, if intervention A for disease Y costs €10/QALY[8], and intervention B for disease Z costs €20/QALY, we might reasonably expect a national health care system to prefer to invest in the first intervention rather than the second should resources be limited.

Some agencies which advise on health care spending use some form of threshold for the ICER to guide their decision making. . For example NICE in England present in their methods guide the factors that the Committee will take into consideration when the most plausible ICERs that are less than £20,000, between £20,000 and £30,000, and above £30,000 per QALY gained.[46]

This is not the only criteria which should be applied to the decision of where to invest public money, but it provides a useful guide.

10.4.3.2 Is further research economically sensible?

Torgerson and Byford [47] discuss the role of economic analysis in advance of clinical trials i.e. In association with a systematic review; in order to decide whether a trial might be sufficient value for money.

Claxton et al have developed these ideas, and have formulated the concept of Value of Information Analysis, with the aim of establishing before a trial is undertaken how much a funder should be willing to pay for the results of the trial.

This has been tested with both NICE and the United Kingdom's Health Technology Assessment programme [48] with limited effect.

[8] QALY, or any other unit of benefit which the health care system is interested in purchasing

11 Appendix 2 - Does why we want evidence matter?

Different study designs lend themselves to answering different questions. In this section we will consider two issues which may drive the selection of a study design.

11.1 Effectiveness versus Safety

A decision maker may want further clinical information about a technology in a number of different aspects[9]. Here we will consider.

1. Effectiveness - how well does a technology work? For the purposes of this document, this includes both efficacy (whether it works at all in idea circumstances) and effectiveness (whether it works in real world use).
2. Safety - what are the adverse and unexpected effects of a technology.

The limitations of different design affect these two types of uncertainties in different ways.

While we suggest below that effectiveness issues for interventions are best addressed by randomised designs, if possible, and safety issues by observational studies this does not preclude collecting data in any way practical.

For assessing technologies where the course of disease is not necessarily changed by the technology under test, e.g. diagnostics, there may be better designs where all technologies can be applied to each patient.

Sometimes the only way to collect effectiveness data is by observation (e.g. in a natural experiment [35]), and safety data should always be collected in an experimental design even if it is unlikely that the small number of patients in an experimental design will be able to demonstrate anything but common safety outcomes.

11.1.1 Efficacy and Effectiveness

In order to determine how well a health technology works, it is usually necessary to compare it to the counterfactual situation, that is: to what would happen without that technology in place.

[9] Others might include natural history, resource use, utility states, etc.

To address questions like this for interventions the most powerful design may be a crossover trial, where the patient is exposed to each technology under consideration in sequence. Unfortunately, this design can often not be used - as it relies on the disease state returning to a constant level between the alternative treatments. An example of where this approach has been effective is the TOMADO trial of mandibular advancement devices in sleep apnoea[49,50].

The more usual approach is a randomised design where each participant experiences one of the available treatment options. If the fundamental question is of efficacy then the control condition is likely to be no intervention or a placebo. If the consideration if one of effectiveness then the control is more likely to be treatment as usual. The more related to effectiveness a question is, the more pragmatic the required design.

If the considerations are purely scientific, then only if an experimental design is not practical would an observational design be considered. However, medical research must be conducted in the real world, and sometimes real world considerations – such as politics, funding, acceptability, or inability to persuade clinicians or patients of the benefits of a randomised approach – mandates that an observational approach is required. In this case, there must be careful discussion on the limitations of observational data, and an understanding before funding a project as to whether it can actually deliver an answer which can be acted upon – else the observational research project can only be a waste of resources.

There is also a challenge that the more pragmatic a study is – the more it is embedded in and predicated on the setting of health care delivery – the less generalizable it may be to other settings. Therefore, it may be challenging to take the learning from effectiveness trials in one country and apply them to another.

11.1.2 Safety

Safety questions are often best served by a different approach. It is important, when assessing safety, to be sure that the occurrence of rare but serious outcomes is appropriately characterised.

To this end, the size of the study population may be more important than randomising patients to different intervention groups Therefore cohort and registry designs are often appropriate.

This can be seen in drug surveillance, where many countries have national surveillance systems where clinicians report suspected adverse drug reactions - in essence the entire population of a country is the cohort collecting adverse drug reaction data. Similarly, national joint registers attempt to capture information on all patients receiving joint prostheses, to allow faulty designs to be spotted early. These approaches have their flaws. They are dependent on the reporting of events, but events are not as actively sought out as for patients enrolled in trials. Reporting is also subject to bias, with more serious events being more likely to be reported, indeed for well-established drugs physicians are often directed not to submit minor or expected adverse reactions.

For drugs or devices where there is a specific cause for concern then a condition of licencing may be the establishment of an appropriate register with more pro-active follow-up of patients in order to accurately characterised the risks of an intervention.

Some researcher advocate using experimental designs when it is necessary to establish a causal link between an intervention and an adverse event – but this is an unusual occurrence in pragmatic research.

11.2 Technology Development

The appropriate approach to assessing a technology varies depending where in its lifecycle the technology is.

For example the IDEAL framework developed in Oxford and adopted internationally describes an approach to developing surgical interventions - but could plausibly be adopted outside the operating theatre. [26]

	1 Idea	2a Development	2b Exploration	3 Assessment	4 Long-term study
Purpose	Proof of concept	Development	Learning	Assessment	Surveillance
Number and types of patients	Single digit; highly selected	Few; selected	Many; may expand to mixed; broadening indication	Many; expanded indications (well defined)	All eligible
Number and types of surgeons	Very few; innovators	Few; innovators and some early adopters	Many; innovators, early adopters, early majority	Many; early majority	All eligible
Output	Description	Description	Measurement; comparison	Comparison; complete information for non-RCT participants	Description; audit, regional variation; quality assurance; risk adjustment
Intervention	Evolving; procedure inception	Evolving; procedure development	Evolving; procedure refinement; community learning	Stable	Stable
Method	Structured case reports	Prospective development studies	Research database; explanatory or feasibility RCT (efficacy trial); diseased based (diagnostic)	RCT with or without additions/ modifications; alternative designs	Registry; routine database (eg, SCOAP, STS, NSQIP); rare-case reports
Outcomes	Proof of concept; technical achievement; disasters; dramatic successes	Mainly safety; technical and procedural success	Safety; clinical outcomes (specific and graded); short-term outcomes; patient-centred (reported) outcomes; feasibility outcomes	Clinical outcomes (specific and graded); middle-term and long-term outcomes; patient-centred (reported) outcomes; cost-effectiveness	Rare events; long-term outcomes; quality assurance
Ethical approval	Sometimes	Yes	Yes	Yes	No
Examples	NOTES video[6]	Tissue engineered vessels[7]	Italian D2 gastrectomy study[8]	Swedish obese patients study[9]	UK national adult cardiac surgical database[10]

RCT=randomised controlled trial. SCOAP=Surgical Clinical Outcomes Assessment Programme. STS=Society of Thoracic Surgeons. NSQIP=National Surgical Quality Improvement Program. NOTES=natural orifice translumenal endoscopic surgery.

Table: **Stages of surgical innovation**

IDEAL describes how different study designs are appropriate at the various stages of technology development, ranging from the original study idea to long-term follow-up registries for adverse events reporting.

Similarly, the MRC complex intervention guidance [25,41] considers the development of any complex intervention (i.e. an intervention with potentially multiple interacting components) in phases akin to the tradition drug development pathway.

The key activity is to identify what the important unknowns are about a technology, and select an appropriate study design to address them.

12 Appendix 3 – The Hierarchy of Evidence

It is generally accepted that for any given question, a hierarchy of designs exists such that all other things being equal a design higher in the hierarchy would be preferred over one lower. These hierarchies are usually constructed such that designs with lower bias (ie a lower risk of systematic error) are ranked higher.

Considering prospective research to answer effective questions we might expect to see a hierarchy such as

1. Randomised Controlled Trials
2. Cohort Studies
3. Case-Control Studies
4. Cross-sectional studies
5. Case Series
6. Case Reports
7. Expert Opinion

References

1 Brown P, Brunnhuber K, Chalkidou K, *et al.* How to formulate research recommendations. 2006;**333**:804–6. doi:10.1136/bmj.38987.492014.94

2 International network of agencies for HTA. International network of agencies for HTA. inahta.net. http://www.inahta.net/ (accessed 26 Oct2012).

3 Chalmers I, Glasziou PP. Avoidable waste in the production and reporting of research evidence. 2009;**374**:86–9. doi:10.1016/S0140-6736(09)60329-9

4 Chalmers I. How to increase value and reduce waste when research priorities are set. *The Lancet*

5 Turner S, Wright D, Maeso R, *et al.* Avoiding waste of research resource: cohort study of publication rate for funded studies from a major UK research funder. *Trials* Published Online First: November 2013. doi:10.1186/1745-6215-14-S1-P95

6 National Institute for Health Research,. NIHR Journals Library. http://www.journalslibrary.nihr.ac.uk/ (accessed 10 Mar2014).

7 NIHR Journals Library. Information for Authors. journalslibrary.nihr.ac.uk. 2014.http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0006/25476/Information-for-Authors.pdf (accessed 10 Mar2014).

8 Centre for Reviews and. *Systematic Reviews*. York: : CRD, University of York 2009.

9 Greenhalgh TT. How to formulate research recommendations: The pie or the slice? 2006;**333**:917–7. doi:10.1136/bmj.333.7574.917-a

10 NICE. *Guide to the methods of technology appraisal 2013*. National Institute for Health and Care Excellence 2013.

11 NICE. *The Guidelines Manual 2012*. National Institute for Health and Care Excellence 2012. http://www.nice.org.uk/guidelinesmanual

12 NICE. *Research Recommendations Interim Process and methods guide December 2014 https://www.nice.org.uk/Media/Default/About/what-we-do/Research-and-development/Research%20Recommendations%20Interim%20Process%20and%20methods%20guide.pdf*

13 Stamenkovic S, Solesse A, Zanetti L, *et al.* [French Health Authority (HAS) Guidance: The Post-registration Studies on Health Technologies (Drugs, Medical Devices and Instruments): Principles and Methods]. *Therapie* 2012;**67**:409–21.

doi:10.2515/therapie/2012065

14 Haute Authorité de Santé, (null). Les études post-inscription sur les technologies de santé. 2012;:1–42.

15 Guyatt GH, Oxman AD, Schünemann HJ, *et al.* GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of clinical epidemiology* 2011;**64**:380–2. doi:10.1016/j.jclinepi.2010.09.011

16 de trabajo sobre GPC G. *Manual Metodológico de elaboración de Guías de Práctica Clínica en el Sistema Nacional de Salud*. Instituto Aragonés de Ciencias de la Salud (I+CS). Avda. Gómez Laguna 25 - 3ª planta, 50009 Zaragoza España. e-mail: iacs@guiasalud.es <mailto:iacs@guiasalud.es>

17 The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions*. 5 ed. The Cochrane Collaboration 2011. http://www.cochrane-handbook.org

18 Liberati A, Altman DG, Tetzlaff J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. 2009. b2700.

19 Elm von E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. 2007;**370**:1453–7. doi:10.1016/S0140-6736(07)61602-X

20 *strobe-statement.org*. http://www.strobe-statement.org (accessed 4 Nov2014).

21 NIHR Evaluation, Trials and Studies Coordinating Centre. Feasibility and Pilot Studies - NIHR Definitions. nihr.ac.uk. http://www.nihr.ac.uk/CCF/RfPB/FAQs/Feasibility_and_pilot_studies.pdf (accessed 5 Mar2015).

22 Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *Journal of evaluation in clinical practice* 2004;**10**:307–12. doi:10.1111/j.2002.384.doc.x

23 Kasenda B, Elm von E, You J, *et al.* Prevalence, Characteristics, and Publication of Discontinued Randomized Trials. *JAMA* 2014;**311**:1045–52. doi:10.1001/jama.2014.1361

24 Campbell MK, Entwistle VA, Cuthbertson B, *et al.* Developing a placebo-controlled trial in surgery: Issues of design, acceptability and feasibility. *Trials* 2011;**12**:50. doi:10.1186/1745-6215-12-50

25 Craig P, Dieppe P, Macintyre S, *et al. Developing and evaluating complex interventions: new guidance*. Glasgow: : Medical Research Council 2008.

26 McCulloch P, Altman DG, Campbell WB, *et al.* No surgical innovation without

evaluation: the IDEAL recommendations. 2009;**374**:1105–12. doi:10.1016/S0140-6736(09)61116-8

27 Cowan K, Oliver S. The James Lind Alliance Guidebook. *James Lind Alliance* 2012.

28 Burt CG, Cima RR, Koltun WA, *et al.* Developing a research agenda for the American Society of Colon and Rectal Surgeons: results of a delphi approach. *Dis Colon Rectum* 2009;**52**:898–905. doi:10.1007/DCR.0b013e3181a0b358

29 Tiernan J, Cook A, Geh I, *et al.* Use of a modified Delphi approach to develop research priorities for the Association of Coloproctology of Great Britain and Ireland. *Colorectal Dis* 2014;**16**:965–70. doi:10.1111/codi.12790

30 Whear R, Thompson-Coon J, Boddy K, *et al.* Establishing local priorities for a health research agenda. *Health expectations : an international journal of public participation in health care and health policy* Published Online First: 6 December 2012. doi:10.1111/hex.12029

31 Coffey CS, Levin B, Clark C, *et al.* Overview, hurdles, and future work in adaptive designs: perspectives from a National Institutes of Health-funded workshop. *Clinical trials (London, England)* 2012;**9**:671–80.

32 Orloff JJ, Stanski D. Innovative approaches to clinical development and trial design. *Ann Ist Super Sanita* 2011;**47**:8–13.

33 Bornkamp B, Bretz F, Dmitrienko A, *et al.* Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of biopharmaceutical statistics* 2007;**17**:965–95.

34 Craig P, Cooper C, Gunnell D. Using Natural Experiments to Evaluate Population Health Interventions. *Glasgow, UK: Medical Research Council* 2011.

35 Craig P, Cooper C, Gunnell D, *et al.* Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *J Epidemiol Community Health* 2012;**66**:1182–6. doi:10.1136/jech-2011-200375

36 NICE. Research Recommendations: Process and Methods Guide. 2011;:1–24.http://www.nice.org.uk/media/FC2/5E/ResearchRecommendationManual.pdf

37 Cook A. Subclinical hypothyroidism: Let's identify research questions. *BMJ* 2008;**337**:a1259–9. doi:10.1136/bmj.a1259

38 UKCRC CTU Network. ukcrc-ctu.org.uk. http://www.ukcrc-ctu.org.uk/ (accessed 4 Nov2014).

39 Research Design Service. rds.nihr.ac.uk. http://www.rds.nihr.ac.uk/ (accessed 4 Nov2014).

40 Commission E. Horizon 202 Synchronised Call initiatives. 2020-horizon.com.

http://www.2020-horizon.com/Synchronised-Call-initiatives-i1798.html (accessed 21 Jan2015).

41 Craig P, Dieppe P, Macintyre S, *et al.* Developing and evaluating complex interventions: the new Medical Research Council guidance. *Int J Nurs Stud* 2013;**50**:587–92. doi:10.1016/j.ijnurstu.2012.09.010

42 Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004;**58**:635–41. doi:10.1136/jech.2003.008466

43 Guyatt GH, Oxman AD, Vist G, *et al.* GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *Journal of clinical epidemiology* 2011;**64**:407–15. doi:10.1016/j.jclinepi.2010.07.017

44 Box G. Science and statistics. *Journal of the American Statistical Association* 1976;**71**:791–9. doi:10.1080/01621459.1976.10480949

45 Box G, Wilson KB. *Empirical Model-Building and Response Surfaces*. New York: : John Wiley and Son 1987.

46 Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. 2007;**335**:358–9. doi:10.1136/bmj.39308.560069.BE

47 Torgerson DJ, Byford S. Economic modelling before clinical trials. 2002;**325**:98. doi:10.1136/bmj.325.7355.98

48 Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics* 2006;**24**:1055–68.

49 S1 TOMADO: A crossover randomised controlled trial of oral mandibular advancement devices for obstructive sleep apnoea-hypopnoea. 2013;**68**:A4–A4. doi:10.1136/thoraxjnl-2013-204457.7

50 Quinnell T. TOMADO: crossover randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea. nets.nihr.ac.uk. http://www.nets.nihr.ac.uk/projects/hta/0811003 (accessed 17 Jul2014).

