

EXPEDITED SCIENTIFIC RESEARCH AND REPORTING (ESRO) AT KCE



EXPEDITED SCIENTIFIC RESEARCH AND REPORTING (ESRO) AT KCE

DOMINIQUE ROBERFROID, DIEGO CASTANARES-ZAPATERO, PATRICE CHALON, ELENA COSTA, MARIE DAUVIRIN, CHRISTOPHE DE MEESTER, ANJA DESOMER, NICOLAS FAIRON, LUC HOURLAY, PASCALE JONCKHEER, LAURENCE KOHN, MÉLANIE LEFÈVRE, MURIEL LEVY, SIEN OMBELET, LORENA SAN MIGUEL, CARINE VAN DE VOORDE, KOEN VAN DEN HEEDE



Title:	Expedited Scientific Research and Reporting (ESRO) at KCE
Authors:	Dominique Roberfroid (KCE), Diego Castanares-Zapatero (KCE), Patrice Chalon (KCE), Elena Costa (KCE), Marie Dauvrin (KCE), Christophe de Meester (KCE), Anja Desomer (KCE), Nicolas Fairon (KCE), Luc Hourlay (KCE), Pascale Jonckheer (KCE), Laurence Kohn (KCE), Mélanie Lefèvre (KCE), Muriel Levy (KCE), Sien Ombelet (KCE), Lorena San Miguel (KCE), Carine Van de Voorde (KCE), Koen Van den Heede (KCE)
Project facilitator:	Nathalie Swartenbroekx (KCE), Els Van Bruystegem (KCE)
Reviewers:	Gudrun Briat (KCE), Cécile Camberlin (KCE), Irina Cleemput (KCE), Justien Cornelis (KCE), Fabian Desimpel (KCE), Marijke Eyssen (KCE), Frank Hulstaert (KCE), Christophe Janssens (KCE), Vicky Jespers (KCE), Roos Leroy (KCE), Céline Ricour (KCE), Claudia Schönborn (KCE), Emily Van Coolput (KCE), Ann Van den Bruel (KCE), Leen Verleye (KCE)
External experts:	Sophie Blanchard (HAS - Haute Autorité de Santé, France), Nicole Mittmann (CDA-AMC - Canada's Drug Agency, Canada)
Acknowledgements:	Wendy Christiaens (KCE), Jens Detollenaere (KCE), Mathieu Laurent (Spade.be), Ine Verhulst (KCE)
Reported interests KCE members:	KCE is a federal institution funded by INAMI/RIZIV, by the Federal Public Service of Health, Food chain Safety and Environment, and by the Federal Public Service of Social Security. KCE's mission is to advise policymakers on decisions relating to health care and health insurance on the basis of scientific and objective research. It is expected to identify and shed light on the best possible solutions, in the context of an accessible, high-quality health care system with due regard for growing demand and budgetary constraints. KCE has no interest in companies (commercial or non-commercial, i.e., hospitals and universities), associations (e.g. professional associations, unions), individuals or organisations (e.g. lobby groups) that could be positively or negatively affected (financially or in any other way) by the implementation of the recommendations. All experts involved in the writing of the report or the peer-review process completed a declaration of interest form. Information on potential conflicts of interest is published in the colophon of this report. All members of the KCE Expert Team make yearly declarations of interest and further details of these are available upon request.
Reported interests:	All experts consulted within this report were selected because of their involvement in the topic of expedited research and reporting. Therefore, by definition, each of them might have a certain degree of conflict of interest to the main topic of this report.
Layout:	Ine Verhulst, Joyce Grijseels



Disclaimer:

- The external experts were consulted about a (preliminary) version of the scientific report. Their comments were discussed during meetings. They did not co-author the scientific report and did not necessarily agree with its content.
- Finally, this report has been approved by common assent by the Executive Board.
- Only the KCE is responsible for errors or omissions that could persist. The policy recommendations are also under the full responsibility of the KCE.

Publication date: 09 July 2024

Project number: 2023-70

Domain: Method

MeSH: Methods; Information Storage and Retrieval; Operations research

NLM Classification: [W 26.55.14](#)

Language: English

Format: Adobe® PDF™ (A4)

Legal depot: D/2024/10.273/41

ISSN: 2466-6459

Copyright: KCE reports are published under a “by/nc/nd” Creative Commons Licence
<http://kce.fgov.be/content/about-copyrights-for-kce-publications>.



How to refer to this document?

Roberfroid D, Castanares-Zapatero D, Chalon P, Costa E, Dauvrin M, de Meester C, Desomer A, Fairon N, Hourlay L, Jonckheer P, Kohn L, Lefèvre M, Levy M, Ombelet S, San Miguel L, Van de Voorde C, Van den Heede K. Expedited Scientific Research and Reporting (ESRO) at KCE. Method Brussels: Belgian Health Care Knowledge Centre (KCE). 2024. KCE Reports 386. D/2024/10.273/41. <https://doi.org/10.57598/R386C>

This document is available on the website of the Belgian Health Care Knowledge Centre.



■ TABLE OF CONTENTS

LIST OF FIGURES.....	6
LIST OF TABLES	7
LIST OF ABBREVIATIONS	8
■ SCIENTIFIC REPORT	9
1 RATIONALE	9
2 OBJECTIVES & METHODS.....	10
3 GENERAL ESRO FRAMEWORK AT KCE.....	11
3.1 INTRODUCTION	11
3.2 SCREENING AND RANKING RESEARCH REQUESTS.....	13
3.2.1 Regular annual call	14
3.2.2 Urgent requests	14
3.3 DESIGNING ESRO METHODS IN THE RESEARCH PROTOCOL.....	18
3.3.1 Project scoping phase.....	19
3.3.2 Protocol challenge phase.....	20
3.4 HAVING HUMAN RESOURCES READILY AVAILABLE FOR <i>AD HOC</i> REQUESTS	20
3.5 WHEN IS IT APPROPRIATE TO ABRIDGE OR SKIP THE EXPERT AND STAKEHOLDER MEETINGS?.....	21
3.6 WHEN IS IT APPROPRIATE TO SKIP PROPOSING CLINICAL OR POLICY RECOMMENDATIONS?.....	21
3.7 EXPEDITING THE PEER REVIEW OF SCIENTIFIC REPORTS.....	22
3.8 EXPEDITING THE PEER REVIEW OF REPORT SYNTHESSES	23
3.9 REPORTING	23
3.9.1 Synthesis and translation	23
3.9.2 Labelling ESRO reports	23



4	EXPEDITED LITERATURE REVIEWS	24
4.1	WHEN IS IT APPROPRIATE TO SEARCH ONLY MEDLINE?	24
4.1.1	Expedited reviews of RCTs	25
4.1.2	Expedited reviews of non-randomised studies	25
4.2	WHEN IS IT APPROPRIATE TO LIMIT THE SEARCH TO CORE JOURNALS?	26
4.3	WHEN IS IT APPROPRIATE TO OPTIMISE THE PRECISION OF THE SEARCH STRATEGY?	26
4.3.1	Constructing the optimised search strategy	26
4.3.2	Quality assurance	27
4.4	WHEN IS IT APPROPRIATE TO SKIP THE RETRIEVAL OF GREY EVIDENCE?	27
4.5	WHEN IS IT APPROPRIATE TO RELY ON ONE SINGLE REVIEWER?	29
4.6	WHEN IS IT APPROPRIATE TO INCLUDE (PARTLY OR MOSTLY) SURROGATE OR MODELLING STUDIES?	29
4.6.1	Surrogate-based studies	29
4.6.2	Modelling studies	31
4.7	WHEN IS IT APPROPRIATE TO BASE THE REVIEW (PARTLY OR MOSTLY) ON INTERIM RESULTS?	31
4.8	WHEN IS IT APPROPRIATE TO RELY ON A PUBLISHED SYSTEMATIC REVIEW, WITH OR WITHOUT UPDATE?	33
4.8.1	Background	33
4.8.2	Deciding when and whether to conduct an update of an existing systematic review	34
4.8.3	Planning and conducting the update of an existing systematic review	39
4.8.4	Conclusions	40
4.9	WHEN IS IT APPROPRIATE TO ABRIDGE OR SKIP THE QUALITY APPRAISAL OF STUDIES? .	41
4.9.1	Strategies to decrease time investment for quality appraisal	42
4.9.2	Description of commonly used appraisal tools for different study types	42



4.10	WHEN IS IT APPROPRIATE TO ABRIDGE OR SKIP THE GRADE APPRAISAL?	44
5	EXPEDITED INTERNATIONAL COMPARISON.....	46
5.1	CAN AN INTERNATIONAL COMPARISON BE A STAND-ALONE KCE PRODUCT?	46
5.2	HOW TO LIMIT THE SCOPE OF THE COMPARISON?	47
5.3	IS THERE A MINIMUM OF COUNTRIES/REGIONS TO BE STUDIED? IS THERE A MAXIMUM NUMBER?.....	48
5.4	WHICH ARE THE PREFERRED METHODS TO COLLECT DATA IN CASE OF AN EXPEDITED INTERNATIONAL COMPARISON?	48
5.5	DO WE NEED TO REPORT INFORMATION COUNTRY-BY-COUNTRY OR CAN WE ONLY REPORT A TRANSVERSAL ANALYSIS?.....	49
6	RAPID QUALITATIVE RESEARCH	52
6.1	WHAT IS MEANT BY RQR?	54
6.1.1	Definition	54
6.1.2	Methods used in RQR.....	55
6.2	CAN A RQR PROJECT BE A STAND-ALONE KCE PRODUCT?	57
6.3	HOW TO BETTER FOCUS THE SCOPE OF RQR?	57
6.3.1	Scoping the topic	57
6.3.2	Scoping the methods	57
6.4	ARE THERE PREEXISTING DATA AVAILABLE FOR SECONDARY ANALYSIS?	58
6.4.1	Availability of relevant scientific literature?.....	58
6.4.2	Availability of relevant and robust preexisting data?	58
6.5	IS IT POSSIBLE TO EXPEDITE THE COLLECTION OF PRIMARY DATA?	59
6.5.1	Is there a-priori theoretical framework applicable for the data collection and the analysis?	59
6.5.2	Can obtaining the ethical clearance be expedited?.....	59
6.5.3	Is there a possibility to collect data through proxies?	60



6.5.4	Is it possible to reduce the number of study participants?	61
6.5.5	Is it possible to accelerate the recruitment process?.....	61
6.5.6	Is there a possibility to restrict the languages used?	62
6.5.7	Which are the most preferred data collection techniques in case of RQR?	63
6.5.8	Is it possible to expedite the design and validation of instruments for data collection?.....	64
6.5.9	Is it possible to skip the full transcriptions of interviews and focus groups?	65
6.5.10	Is a validation of the results by study participants needed?.....	66
6.6	IS IT POSSIBLE TO EXPEDITE THE ANALYTICAL PROCESS?	66
6.6.1	Which level of analysis is expected at minimal?	66
6.6.2	Which analytical approach is the most expeditive / efficient?	67
6.6.3	Is it possible to rely on one single researcher for the analytic process?	67
6.6.4	Can the use of software help speed up the analysis?	68
6.6.5	Is it possible to skip or abridge the triangulation of sources?.....	68
6.6.6	Is it possible to skip or abridge the process of reflexivity?	68
6.7	DISCUSSION	69
6.7.1	Decisional tree	69
6.7.2	Conditions for making RQR effective.....	71
7	AUTOMATION	72
7.1	BACKGROUND	72
7.2	AIM AND OBJECTIVES.....	74
7.3	METHODS	74
7.3.1	Methods to provide practical advice on the use of Generative AI	74
7.3.2	Methods to identify tools and applications for specific research purposes	74
7.4	RESULTS.....	78



7.4.1	Generative AI.....	78
7.4.2	Tools and applications for specific research purposes	83
7.5	DISCUSSION	89
7.6	CONCLUSIONS	93
■	BIBLIOGRAPHY	94
■	APPENDIX.....	103
APPENDIX 1.	COMPARISON OF QUALITY APPRAISAL TOOLS.....	103
APPENDIX 2.	PROPOSAL FOR AN APPRAISAL TOOL OF <i>IN VITRO</i> STUDIES	109
APPENDIX 3.	TIME/RESOURCE MATRIX OF NUNNS	113
APPENDIX 4.	AUTOMATION	114
APPENDIX 5.	OBSOLETE TOOLS FOR AUTOMATION.....	125



LIST OF FIGURES

Figure 1 – Standard KCE report production process	12
Figure 2 – Fast track production process for urgent research requests.....	16
Figure 3– Algorithm for the decision on when and how to update a systematic review.....	41
Figure 4 – Decisional tree for Rapid Qualitative Research at KCE.....	70
Figure 5 – Relationship between different concepts in Artificial Intelligence (AI)	73
Figure 6 – Potential inclusion of automated tools in the lifecycle of a KCE project	76
Figure 7 – Suggestions of automated tools that can be used in the lifecycle of a KCE project.....	91
Figure 8 – Time/resource Matrix ¹¹⁸	113



LIST OF TABLES

Table 1 – Decision framework to assess the need for updating SRs.....	34
Table 2– Assessment of the state of the existing literature	35
Table 3 – Application of step 1 (assess currency) suggested by Garner et al. to the context of KCE research	35
Table 4 – Application of step 2 (assess relevance) suggested by Garner et al. to the context of KCE research.....	36
Table 5 – Tools for predicting how new studies can affect the new review conclusions.....	37
Table 6 – The adapted Ottawa method and rationale.....	38
Table 7 – Application of step 3 (assess the effect of updating the review) to the context of KCE research.....	39
Table 8 – Steps for planning and conducting the update of a review from the Cochrane Handbook for Systematic Reviews of Interventions.....	40
Table 9 – Recommendations for rating the certainty of evidence in ESRO	45
Table 10 – Steps for an accelerated international comparison.....	50
Table 11 – Overview of a selection of RQR approaches	56
Table 12 – Dimensions used to evaluate tools and applications for research automation	77
Table 13 – Comparative table of the retrieved tools for management and streamlining of literature review	85
Table 14 – Comparison of the features of different automatic transcription software.....	88
Table 15 – Comparison of the items included in the different assessment tools, by type of bias	104
Table 16 – Comparison of advantages & disadvantages of the different appraisal tools	107
Table 17 – Evaluated tools for retrieval and analysis of pertinent literature and scientific writing.....	122
Table 18 – Evaluated tools for qualitative data collection and analysis	123



LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITION
AI	Artificial Intelligence
AIHTA	Austria HTA Agency
AMSTAR	A MeaSurement Tool to Assess systematic Reviews
CDA	Canada's Drug Agency
DMC	Data Monitoring Committee
DSMB	Data Safety Monitoring Board
EMA	European Medicine Agency
ESRO	Expedited Scientific Research and Reporting
GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
HAS	Haute Autorité de Santé
HTA	Health Technology Assessment
ICTRP	International Clinical Trials Registry Platform
LLM	Large language model
NITAG	National Immunization Technical Advisory Group
PPF	Pre-project fiche
PF	Project fiche
RAP	Rapid Appraisal
REAM	Rapid Evaluation and Appraisal Methods
ROBINS	Risk Of Bias In Non-randomised Studies
ROBIS	Risk of Bias in Systematic reviews
RQR	Rapid Qualitative Research
RRMG	Rapid Reviews Methods Group
SR	Systematic Review
TPF	Topic Proposal Form



■ SCIENTIFIC REPORT

1. RATIONALE

Under the pressure of the COVID-19 pandemic and the needs of urgent decision-making, the KCE delivered faster evidence syntheses and recommendations. Many examples are accessible on the [KCE website](#). Beyond the exceptional circumstances of the pandemic, the demand from various decision makers for accelerated evidence-informed guidance is on the rise, and responding to these needs in due time constitutes a challenge and requires adapted methodologies. Many other agencies involved in enhancing evidence-informed decision making face the same challenge worldwide.¹ This acceleration process also affects the way we write and publish reports. The rapid evaluation of vaccines against RSV infections or GLP-1 receptor agonists are recent examples. However, there is so far no general KCE guidance on how to proceed in such cases.

Expedited Scientific Research and Reporting (ESRO) is defined as the use of accelerated scientific methods in comparison with more traditional and generally used approaches, but whose validity remains assured. The ESRO approach aims toward the best possible balance between expert-time involvement and the output in terms of decision-making support, while monitoring output quality. Expedited methods may be required in specific situations, e.g. supporting urgent decision-making, assessing new or emerging technologies or interventions with potential clinical applications, providing an up-to-date synthesis in rapidly evolving fields.¹ ESRO also concerns efficient and timely publication of results and recommendations.

ESRO is particularly necessary when:

1. The decision-making concerns urgent, high priority health issues
2. An expedited approach brings enough information to respond to decision-makers' needs
3. A qualitative trade-off has to be found between numerous priority projects and human resources at KCE



The ESRO approach is thus not limited to rapid response to urgent requests, but aims at maximising the efficiency in all KCE projects beyond situations of urgency or work overload. A distinction has to be made between ESRO products (e.g. an expedited literature review) and the ESRO approach which is valid to all aspects of KCE activities. In other words, ESRO considerations will be present in all KCE projects but will not necessarily result each time in ESRO products, and some chapters of a report can be ESRO-based and others not.

2. OBJECTIVES & METHODS

The objectives of this process notes are as follows:

1. To generate a framework defining when and how the ESRO approach should be applied at KCE
2. To define a range of methods & procedures available to implement ESRO
3. To publish a user-friendly process note and to update/cross-fertilize other KCE process notes if relevant (e.g. the 2017 process note on rapid reviews)

To grasp such a complex multifaceted topic, we applied the following complementary methods:

1. A narrative literature review including the most recent sources established by authoritative authors (e.g. the Cochrane Rapid Reviews Methods Group).
2. A consultation of international experts from key institutions in the field of evidence-informed health policy recommendations, i.e. the Haute Autorité de Santé (HAS) in France and the Canada's Drug Agency (CDA-AMC).^a
3. An internal consultation of KCE experts with previous experiences or interest in expedited methods.
4. The development of a general KCE framework for implementation of ESRO.

^a CDA-AMC was named Canada's Drug and Health Technology Agency (CADTH) until May 1, 2024. The National Institute for Health and Care

Excellence (NICE) in the UK and the National Health Care Institute (ZIN) in the Netherlands agreed to contribute but eventually did not.



Several KCE researchers contributed to this guideline. The tasks were shared in three main work packages: WP1 developed a general framework for ESRO at KCE; WP2 covered methods to expedite literature reviews, qualitative researches and international comparisons; and WP3 explored the possibility of expediting the research process by means of automation of key steps such as literature review, data collection and data analysis.

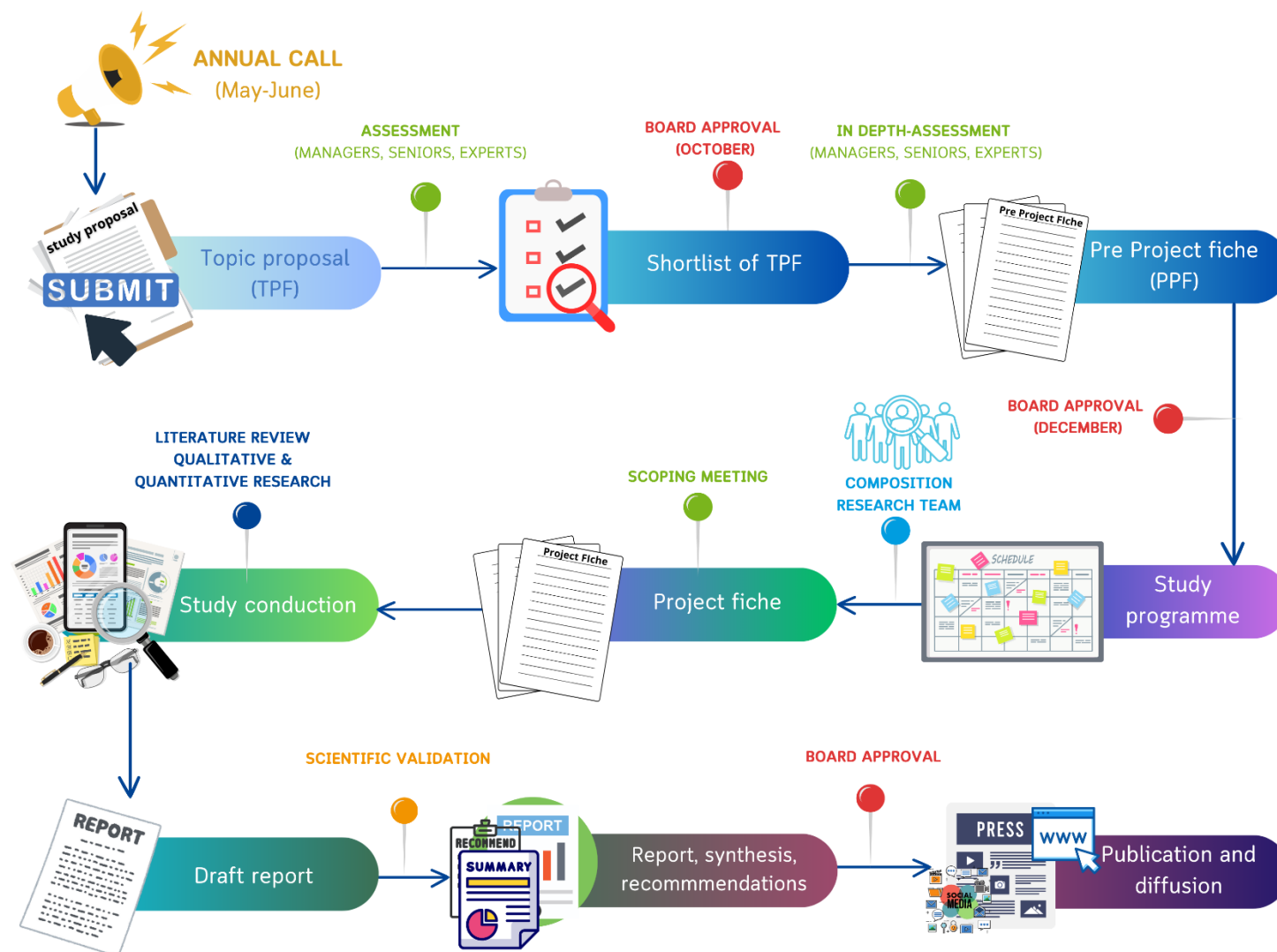
3. GENERAL ESRO FRAMEWORK AT KCE

3.1. Introduction

There are multiple steps between the submission of an information request by decision makers or commissioners and the publication of KCE reports, as pictured in Figure 1. These steps are described on KCE website ([Topic selection for the annual programme](#) & [Establishment of the definitive study programme](#)). This screen-to-publish research chain with different quality control stations has contributed significantly to building KCE's reputation for scientific rigour and quality. However, it takes usually at least eighteen months between research request submission and study completion, which means that the one-year study program is not finished when the next annual call is organized, and this results in the progressive accumulation of delayed studies. Therefore, for reasons of responsiveness to the increasing number of urgent requests from specific stakeholders but also to possibly improve the general efficiency of the organisation, revisiting the merits of current procedures is timely. In this chapter, we discuss the possible streamlining and shortening of relevant procedures and organisational aspects, while scientific methodological aspects of ESRO products are addressed in chapters 4, 5, and 6.



Figure 1 – Standard KCE report production process





3.2. Screening and ranking research requests

Key messages

- Urgent research requests from commissioners (the so-called *ad hoc* projects) are on the rise, and it is at the heart of the KCE's mission to assist in timely decision making. However, the current selection procedures must be adapted to assess thoroughly and rapidly the relevance and urgency of these requests, and in case of acceptance, start the research as soon as possible.
 - The relevance (and feasibility within the time constraint imposed by urgency) can be assessed through a fast-track screening and ranking process comprising two main steps:
 - A research protocol (the equivalent of the current project fiche (PF)), is drafted in one round by two KCE experts in close collaboration with the commissioners and other experts. The recommendations for an appropriate project scoping apply;
 - The research protocol is evaluated according to the usual KCE criteria, but by fewer reviewers (e.g. one KCE senior expert + one KCE expert in the specific domain of the request + two managers). The challenge phase of the proposal occurs at this stage.
 - The urgency can be approached through complementary angles based on:
 - The consequences of rescheduling/postponing the decision-making
 - The consequences of rescheduling/postponing/refusing the request
 - The identification of the specific locus of urgency, if any, possibly resulting in a fragmentation of the request over time
- The overall score given to the proposal by the reviewers justifies, in a structured and transparent manner, why an urgent request is accepted, rejected or postponed until the next regular annual call.
 - This ESRO procedure lasts no more than four weeks. In case of health emergency, this delay will be as short as possible.
 - The approval of the decision by the Board can be done through a written procedure to avoid the waiting time until the next Board meeting
 - In some cases, the research proposal can already result in an informative document for the commissioners, and thus be shared with them.
 - Submitters of *ad hoc* applications should be well informed in advance of the conditions to be met for their application to be eligible for a fast track procedure.



3.2.1. Regular annual call

The usual procedures for selecting topic proposals to be included in the following year's programme are well established. Each September KCE ranks the topics based on several criteria: frequency, severity, and room for improvement of the health problem related to the proposed topic; policy relevance of the proposed topic; and feasibility of the proposed study.^b As these procedures extend over several months, we assume that proposals following this path should *de facto* not be considered very urgent.^c However, specific procedures could possibly be re-examined with a view to increasing the fluidity of the process while maintaining the neutrality of the critical evaluation, for example:

1. **Reduce the number of persons involved in the evaluation** of topic proposals (currently eight individuals are involved: three senior experts, two experts, and three 3 managers)
2. **Re-analyse the procedures for approvals by the KCE's Executive Board.** Currently, the Board issues three approvals: on a short list of topic proposals in October, on the study programme in December, and on the work programme in March of the following year. As the dates are fixed, *de facto* new projects can only begin in January at the earliest, i.e., at least 6 months after their submission by commissioners. The possibility of the Board **examining project proposals in a single round could be explored** (see the proposal for ESRO projects in section 3.8). Moreover, **applying a written procedure** could avoid the waiting time linked to fixed dates of Board meetings.

These possible revisions relate to organisational and strategic decisions, and are not further developed here.

3.2.2. Urgent requests

Since the foundation of KCE, policy makers and their administrations could introduce so-called *ad hoc* projects, outside the annual call. Indeed, it has always been acknowledged that some health policy questions requiring urgent decisions might benefit from a scientific underpinning. In recent years, as already mentioned, we have observed an increasing number of *ad hoc* research requests submitted by various stakeholders. It is at the heart of the KCE's mission to assist in timely decision making, and it demonstrates that the KCE has become an essential scientific player also when rapid health policy decisions are required. But obviously, this trend poses challenges on the one hand in terms of organisation of the work programme, and on the other hand in terms of scientific research methodology.

The fast track of the *ad hoc* requests, which each time requires a reshuffle of the annual programme and teams, should be used wisely, i.e. the relevance and urgency of the commissioners' request must be assessed in an organised manner, and submitters of such applications should be well informed in advance of the conditions to be met for their application to be eligible for a fast track procedure. The methodology described here applies in normal circumstances and will be adapted in times of health crisis if necessary.

^b [Topic selection for the annual programme | KCE \(fgov.be\)](#)

^c Of note, for some proposals submitted following the annual call the degree of urgency may change over time



3.2.2.1. *Assessment of request relevance*

KCE has acquired over the years a strong experience in prioritising most relevant research proposals. A description of the processes is given elsewhere (see section 3.2.1). Obviously, this sound approach of ranking should apply to any proposal submitted to KCE, even the *ad hoc* requests. This is a matter of transparency and fairness towards all actors of the health system. However, we propose to apply an expedited procedure for urgent requests, so that the KCE can quickly position itself and start the research as soon as possible in case the research proposal is accepted (see Figure 2). The ESRO procedure should not take longer than four weeks and will have the following two main steps:

1. Drafting a research proposal

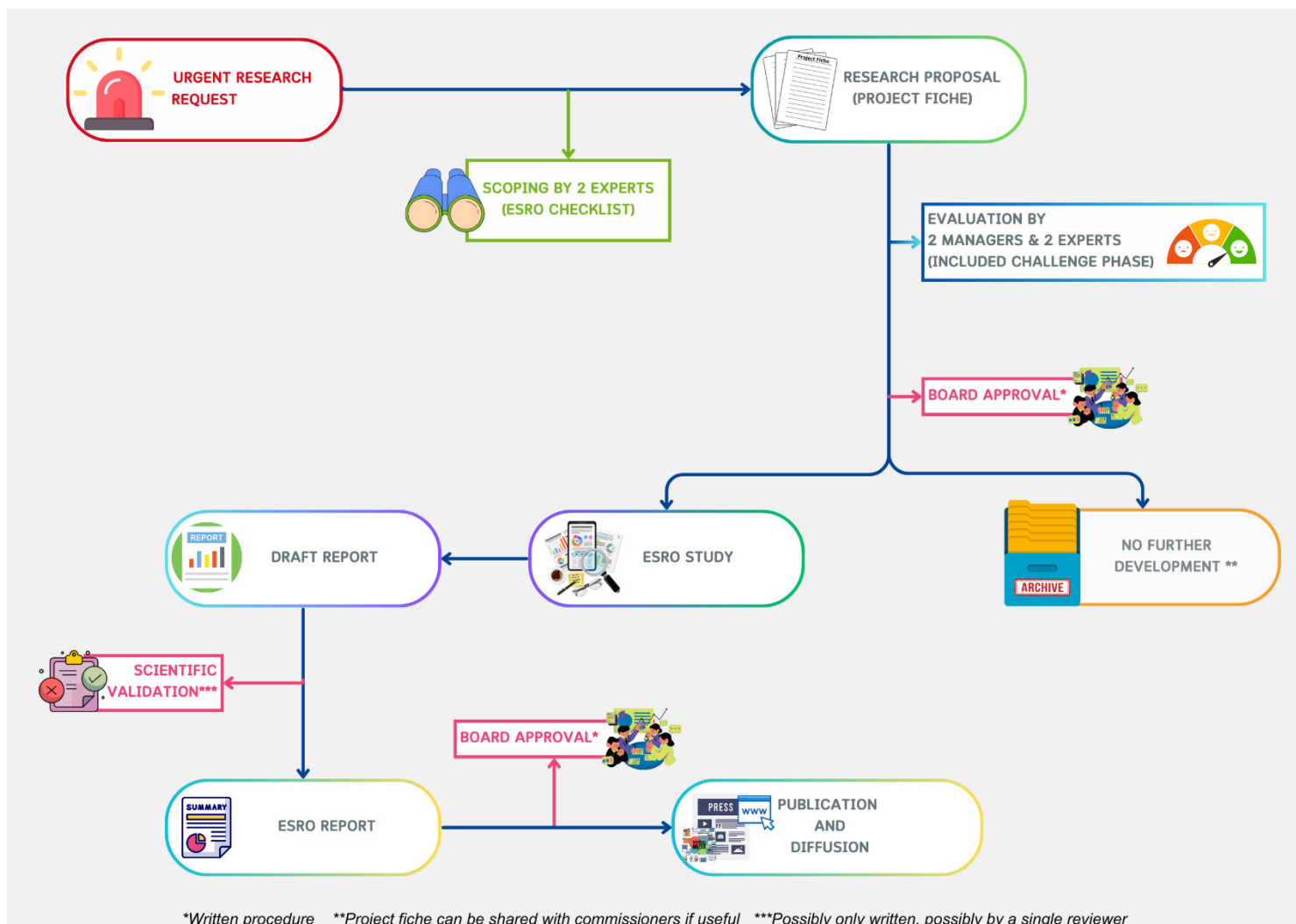
The current project fiche (PF) corresponds in KCE terminology to a short research protocol, and we propose to call it as such from now on. **It is drafted directly by two KCE experts (one senior expert and one expert in the specific domain of the request), i.e. the research request is screened in one unique round overarching the stages of the initial assessment (resulting in a shortlist of topic proposals, TPF) and in-depth evaluation of the topic (pre-project fiche, PPF).** The recommendations for an appropriate scoping, as explained in section 3.3.1, apply. Relying directly on an abbreviated research protocol is important as assessing feasibility within the time constraint imposed by urgency and available human resources is key to decide on acceptance or rejection of the request. The drafting of the research protocol is obviously done in close collaboration with the commissioners, while clarifying that it does not amount to a promise to carry out the project, and with well-informed experts from KCE and other relevant institutions. Other aspects pertaining to research protocol are discussed in section 3.3.

2. Evaluation of the research proposal

The research proposal is **evaluated according to the usual KCE criteria**. We can possibly **reduce the number of reviewers** for a lower impact on KCE regular activities (e.g. two senior experts + two managers). The question of project feasibility is obviously crucial. Nevertheless, questions of severity, political relevance, room for improvement remain important. One can imagine that political relevance will score high, but this will not automatically be the case. The true level of urgency must be assessed, as well as the policy decision that is expected (see section 3.2.2.2). The critical appraisal (challenge) of the research protocol also occurs at this stage (see section 3.3.2).



Figure 2 – Fast track production process for urgent research requests





3.2.2.2. Assessment of request urgency

The true urgency of the request must be examined at the same time as its relevance. Although less practiced so far, this assessment is essential not only to determine the most appropriate research methods, but also to possibly amend the overall KCE work plan. In an attempt to establish the level of urgency of a commissioner request, three main elements can be considered.

1. Consequences of rescheduling/postponing the decision making:

One way of assessing the urgency of a request is examining if the decision making can be rescheduled/postponed and the plausible consequences of doing it. This aspect must be discussed in depth with the commissioner. For example, it may happen that a request is considered urgent by the commissioner for circumstantial reasons, e.g. parliamentary questions on the government programme, or the approaching end of the government legislature. The validity of these circumstantial reasons must be evaluated on a case-by-case basis. Although this kind of consequence analysis may be tricky, it is important to try as much as possible to avoid that a circumstantial political burden translates into a higher programmatic burden at KCE and a perceived loss of political independence by the other stakeholders. In other words, it must be established that the request corresponds to an emergency at the health policy level.

2. Consequences of rescheduling/postponing/refusing the request:

In case the decision making cannot be postponed, it is relevant to examine how serious would be the **consequences of not delivering KCE recommendations or not on time**. That would usually result in a **decision making which is not sufficiently evidence-informed**. Whether this bears serious consequences will depend on the topic. For example, delayed recommendations on transferring the management of obstructive apnoea syndrome to the home setting will have much less public health

consequences than delayed recommendations linked to the toxicity of a new drug or vaccine. Another example is the urgent request to propose quality indicators for the management of severe child obesity in reference centres. As the date of the signature of the conventions between the centres and RIZIV-INAMI was approaching, delaying the request would have resulted in incomplete RIZIV-INAMI conventions.^d

An alternative to rescheduling/postponing/refusing the request would be to look at the problem from the other end and decide what useful output can be delivered in the allotted time, i.e. whether an acceptable trade-off between time and scientific rigour can be defined. For situation-tailored reports, which do not concern only urgent requests, please see the content of section 4. Areas of uncertainty should be clearly announced in advance and emphasised again in the report.

3. Determining the locus of urgency:

It may happen that within an *ad hoc* request, **some parts are more urgent than others to support the decision making. The request can be fragmented**. As an example, RSV infection in infants represents an important public health problem. Since early 2023, a new vaccine is available, whose efficacy and safety seem to be high. The EMA already issued a marketing authorisation, although only interim results were available at the time. Performing a full-fledged HTA would have required one to two years because Belgian clinical data on RSV infections are not easy to gather. It would not have been acceptable for parents and clinicians to wait two more RSV transmission seasons before getting a recommendation on using the vaccine, and a rapid systematic review of clinical efficacy and safety was needed and performed in one month in November 2023 to feed into the Belgian Superior Health Council's advice. The cost-effectiveness analysis can be conducted in a second stage.

Alongside the fragmentation of the request, it is necessary to **examine whether the decision-making itself cannot be fragmented**, i.e. there may

^d KCE website: <https://kce.fgov.be/en/developing-a-set-of-quality-indicators-to-monitor-the-quality-of-care-delivered-by-paediatric>



be a different timeframe for different parts of the decision-making. Again, this should be examined in close collaboration with the commissioners.

3.2.2.3. Overall scoring

The overall score accounting for relevance and urgency can provide a structured and transparent way to justify why an urgent request is accepted, rejected or postponed to the next regular annual call. This ESRO procedure should **not take longer than four weeks**. In case of health emergency, the delay should be as short as possible.

The approval of the decision by the Board can be done through a written procedure to avoid the waiting time until the next Board meeting (see section 3.8). This score is also a strong argument for discussing with the commissioner the decision to continue the project or not. **In some cases, the research proposal can already result in an informative document for the commissioner.**

3.3. Designing ESRO methods in the research protocol

Key messages

- ESRO does not apply only to urgent research request but aims at defining the best possible balance between research inputs and outputs in all KCE projects.
- A careful scoping of research questions must occur in close collaboration with the commissioners who define as precisely as possible their information needs and the nature of the anticipated recommendations.
- Relevant research methods are described with precision in a full-fledged research protocol (replacing the current Project Fiche in KCE terminology). They aim at generating the evidence strictly necessary to formulate the recommendations identified as key for decision making. In every KCE project there is an explicit consideration during the scoping phase whether ESRO can be applied at least to some parts of the study.

- The research protocol is critically appraised (challenge phase) by two experts (one senior expert and one expert with recognised expertise in the topic and/or the proposed methodology) and two managers. Usefulness and efficiency in (in-)validating the expected recommendations are central criteria.
- Substantial amendments to the protocol along the way call for a new challenge phase.

ESRO does not apply only to urgent research request but aims at defining the best possible balance between research inputs and outputs in all KCE projects. It is therefore crucial to describe in detail what the objectives of the project are and what methods and resources will be used to reach them. Currently, these are described in broad terms in a document called a pre-project fiche (PPF), and the annual program is established on the basis of these documents, which are still preliminary in most cases. It is only at the start of the project, when the proposal has already been approved by the Board and the research team has already been formed, that the principal investigator brainstorms, together with the project coordinator, about a further delineation of the scope and about the recommendations that could/should result from the project. These last changes, usually done with no formal guidance, result in a project fiche (PF). We plead for a full-fledged research protocol to be developed in a single step (therefore encompassing the current PPF and PF), and for this protocol to contain all the details on the methods that will be implemented. The drafting of the protocol is done in two stages: a scoping stage and a research proposal challenging step.



3.3.1. Project scoping phase

3.3.1.1. Defining the expected outputs

The ultimate output(s) critical for decision making should be clarified through a stringent project scoping in close collaboration with the commissioners and, if possible, knowledge users.^e **An efficient project scoping corresponds to defining from the project onset what could be the main possible recommendations useful for the decision making, with the goal of underlining what data and analysis are necessary to (in-)validate the recommendations foreseen.** For example, a systematic literature review of background rates of herpes zoster will take time and resources, while a review limited to recent publications in neighbouring countries is a sufficient source of information to benchmark rates in Belgium and compute clinical and budgetary impacts of a vaccination (KCE report 360). Similarly, it is not necessary to carry out a systematic review on the incidence of all infectious and mental diseases in incarcerated individuals to yield recommendations on the re-organisation of health care in prison. When the research project is about developing a clinical guideline, the number of outcomes should be kept minimal with respect to those that are likely to help the decision-making and not to those that, in the guideline group's opinion, are likely to have been reported in the extant literature. When the evidence base is known to be large, other restrictions may be imposed to enhance the directness of the evidence (e.g. race, geographic/economic region of where research was conducted, study design, settings etc.). It must also be noted that in some ESRO projects no policy or clinical recommendations are proposed (see section 3.6). Whatsoever the outputs they must be formally verified with the knowledge users if they correspond to their information needs.^{2, 3}

3.3.1.2. Drafting the research protocol

The scoping stage must result in a **detailed research protocol** defining the methods necessary to apply to deliver the expected outputs, e.g. the search strategy in bibliographic databases, the statistical analysis plan, or the minimal number of patients interviewed.² For sake of consistency, transparency and quality, a formal KCE checklist to streamline the reflection on most appropriate ESRO methods to consider when drafting a research protocol was developed (see sections 4, 5, 6, 7). Moreover, the structure of the research protocol should follow the international recommendations for reporting, e.g. the SRQR list (Standards for Reporting Qualitative Research) for qualitative research or the PRISMA-P statement (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) for review protocols.⁴

ESRO does not apply only to urgent research requests, but aims at finding the best possible balance between research inputs and outputs in all KCE projects. Therefore, In every KCE project, the scoping phase of every KCE projects should include explicit consideration of whether the ESRO approach can be applied, to the entire or at least to some parts of the study.

^e We define the commissioners as the person or group requesting a KCE report, and knowledge users as individuals or groups responsible for, or affected by, health and healthcare related decisions that KCE reports can inform.



3.3.2. Protocol challenge phase

The protocol challenge phase is a thorough review of the relevance of the project scope and proposed research methods. Usefulness and efficiency of methods in (in-)validating the expected recommendations are central criteria. A protocol challenge will also be organised during the project every time the principal investigator proposes substantial amendments to the research protocol.

The methodology of the ESRO project scoping and challenge phases must still be translated into a KCE process note. The general principles would be as follows:

1. The project objectives and expected outputs are defined, as well as the research strictly necessary to deliver these outputs (not necessarily policy or clinical recommendations).
2. The methods that will be used to achieve the objectives are also described precisely.
3. The objectives, methods, resources and timeline are explained in a well-structured full-fledged research protocol, taking into account the international recommendations on reporting.
4. The research protocol is critically reviewed by two KCE experts (one senior expert and one expert with recognised expertise in the topic and/or the proposed methodology), and by two managers.
5. Consequently, the research protocol is amended, when appropriate, by consensus between the principal investigator and the reviewers. When a consensus is not reached, a member of the management team is asked to enhance a final decision during a challenge meeting.
6. If the principal investigator considers that the research protocol must be substantially revised along the way, a new challenge phase is organised.

3.4. Having human resources readily available for *ad hoc* requests

We focus here only on the human resources to be mobilised to ensure that *ad hoc* requests are processed in a timely manner. Other aspects of the organisation of human resources, such as outsourcing part of projects, internally shifting staff from one project to another in the event of time constraints or insufficient staff, and the prevention of bottlenecks partly due to insufficient staff in key positions, are not discussed here.

To manage *ad hoc* requests, human resources must be quickly available, that is to say a sort of rapid response team is needed. This rapid response team acts at two levels, and different experts may be involved at each level:

1. A core group of five to ten experts are identified and trained to form an ESRO team. All *ad hoc* requests are sent to them, for example through a generic ESROteam@kce.fgov.be mailbox. Based on availability and expertise, two experts organise the assessment of relevance and urgency of the request as described in section 3.2.2. These persons can of course request input from any relevant expert, whether internally or externally, and are in close contact with the research commissioners.
2. From the moment the decision to positively respond to the *ad hoc* request is taken, a team of KCE experts can be appointed to conduct the research, including from the core group mentioned above.



3.5. When is it appropriate to abridge or skip the expert and stakeholder meetings?

Key messages

Expert and stakeholder meetings can be skipped in specific circumstances where their added value is limited, e.g. a rapid review of the efficacy and safety of a new drug. Alternatively they can be replaced by written feedback on specific scientific questions.

Involving experts, stakeholders, and in the broad sense knowledge users is important to ensure on the one hand that the research question pursued remains relevant for decision making, and on the other hand to benefit as much as possible of existing expertise. However, formally organising stakeholder consultation can be a time and resource consuming undertaking. Its added value will depend very much on the topic covered, both in terms of its content and complexity, as well as the policy recommendations envisaged. In some circumstances this approach can be accelerated or even avoided:

- **Technical and focused research question.** When it comes to a project with a very specific and technical focus, for example a review of the benefits and side effects of a new drug, for which all the useful information is expected to be found in peer-reviewed articles.
- **No policy recommendations.** When it is a project that is not intended to result in policy recommendations, for example a literature review of knowns and unknowns on GPL-1 agonist receptors.
- **Existing expert group.** When the project itself is developed within a group of experts already established, for example the Therapeutics Viral Diseases Task Force or the NITAG.

^f KCE website: <https://kce.fgov.be/sites/default/files/2023-09/KCE%20rapport%20paxlovid.pdf>

^g KCE website: <https://kce.fgov.be/en/publications/covid-19-contribution/risk-benefit-model-for-tts-events-applied-to-the-johnson-johnson-vaccine>

If it is decided to maintain a consultation of experts and/or stakeholders, one way to speed up the process may be to request written feedback. The feedback requested could relate to different parts of the project depending on the expertise of the people approached. Whatever methodology is used, it is important that it is reported transparently.

3.6. When is it appropriate to skip proposing clinical or policy recommendations?

Key messages

- It is possible that an ESRO report proposes no recommendation when the objective is to inform knowledge users, or to inform decision-makers taking full responsibility for the issuing of recommendations.
- An ESRO report can propose recommendations, but these are straightforward, do not concern complex issues, and are limited in number.

For some kind of ESRO reports, no recommendations are foreseen or necessary because the main aim is simply to inform knowledge users, either clinicians or patients (recent examples at KCE: a systematic literature review of the efficacy and safety of Paxlovid to treat COVID-19,^f and a risk-benefit model for thrombocytopenia syndrome events applied to the Johnson & Johnson vaccine against COVID-19^g) or to inform decision makers taking full responsibility of issuing recommendations (recent examples at KCE include literature reviews of the efficacy and safety of maternal vaccination^h and nirsevimabⁱ against RSV infection in infants to feed in the advice of the Belgian Superior Health Council on that topic).

^h KCE website: https://kce.fgov.be/sites/default/files/2023-12/2023-60RSV_Abrysvo_Maternal_Vaccine.pdf

ⁱ KCE website: https://kce.fgov.be/sites/default/files/2023-12/2023-60RSV_Nirsevimab_Infection_Infants.pdf



Sometimes, the ESRO report can propose recommendations, but these are straightforward, do not concern complex issues, and are limited in numbers. A recent example at KCE is the decision to stop using Evusheld in Belgium because its efficacy against omicron variants of COVID-19 had vanished and no evidence of the efficacy of doubling the dosage, as suggested by the company and the FDA, could be retrieved.^j

3.7. Expediting the peer review of scientific reports

Key messages

- The peer review by three external independent scientists (validation in KCE vocabulary) applies to all KCE reports, including those in which ESRO methods are applied.
- For ESRO projects, the validation can take place in writing, as for the peer review of scientific articles. A meeting is organised only in the event of a major disagreement between the reviewers, or between the reviewers and the research team of the project.
- In exceptional situations, the validation of ESRO projects can rely on one single reviewer, preferably an external expert. What makes the situation exceptional must be described in the research protocol and in the report.

At KCE, we call a validation the peer review of our scientific reports by three external independent scientists. External reviewers send their written comments and responses from KCE experts are discussed during a face-to-face meeting. The external review, an important step to guarantee the scientific quality of our reports, should be maintained for all reports, including those in which ESRO methods are applied. However, the organisation of the meeting can be time consuming, and sometimes finding a date that suits everyone delays validation. Therefore, **for ESRO reports, validation can happen entirely in writing**, like the peer review of scientific articles. A

meeting focusing on the remaining disagreements could be organised only if the reviewers are not satisfied with the adjustments made or responses provided, and some major comments remain unsolved.

In exceptional situations, the external validation process as described above may appear disproportionate. This will for example be the case for stand-alone expedited reviews of a new drug or vaccine including only 1-2 pivotal trials (e.g. the rapid literature reviews on Abrysvo® maternal vaccine or monoclonal antibody nirsevimab to protect infants from RSV infection are recent examples at KCE) or stand-alone international comparisons where key national experts were involved in data collection and had been given the opportunity to comment on KCE analysis and results. In another situations, the deadlines may be too short to apply the regular validation process, as it was the case for advices on drugs against COVID-19. What makes the situation exceptional must be described in the research protocol and in the report (e.g. very tight timeline, report co-developed by recognized experts in the field, for example members of the Taskforce Therapeutics Viral Diseases or the NITAG). In such situations, in addition to the written procedure described above, the number of validators could be reduced to one, preferably an external expert (e.g. this strategy was applied in the report on quality indicators for the management of paediatric obesity^k), but possibly a KCE expert who was not involved in the project.

^j KCE website: https://kce.fgov.be/sites/default/files/2022-11/TFvd_Evusheld%20prophylaxis_23112022.pdf

^k KCE website: <https://kce.fgov.be/en/developing-a-set-of-quality-indicators-to-monitor-the-quality-of-care-delivered-by-paediatric>



3.8. Expediting the peer review of report syntheses

At KCE, we call a synthesis a short document written in everyday language covering the main results of the scientific report as well as the recommendations intended for the various knowledge users and decision makers, and we call an internal review a process of critical review of the synthesis by two KCE experts. This review focuses mainly on the coherence and readability of the text, not on the scientific methodology and results of the report. Although an internal review may take some additional time to organise (find one or two reviewers who have to dedicate some of their working time to a critical review, and possible further revisions by the main authors), it is important because the synthesis includes the policy recommendations and is probably the reference document for a large proportion of knowledge users.

Under ESRO approach, the internal review can be expedited by:

1. **Involving only one reviewer**, for critical review and pointing to e.g. omissions in rationale, lack of clarity, inconsistencies.
2. **Using only a written format** mirroring the peer review process of scientific articles.

For the sake of time saving, the revision can be done without a formal meeting. In some cases, it may be indicated to discuss some remaining issues during a short face-to-face or online meeting with one or both reviewers.

3.9. Reporting

3.9.1. Synthesis and translation

In usual KCE reports, the synthesis may be up to 35 pages. For ESRO reports, the synthesis should be restricted to 5-10 pages or even less for very short reports. For some ESRO reports (e.g. a stand-alone literature review), the synthesis is limited to the main results, discussion and recommendations, if any, and presented as a 2-pager executive summary. This can be in English, or also translated in French and Dutch, depending on the target audience and the time available.

3.9.2. Labelling ESRO reports

Not all reports including some ESRO aspects or methods necessarily require a different layout than the regular KCE reports, as long as the methodology used is transparently reported and any possible limitations resulting from the expedited nature of the process are clearly explained.

In contrast, a clear and explicit signage must be created for ESRO reports, e.g. a rapid literature review.

This specific lay-out is under development, and will include the following elements:

1. Easily identifiable within a dedicated series (e.g. R³ Rapid Response Report)
2. Portrait layout, with or without picture
3. Stand-alone executive summary
 - a. Key points in the forefront
 - b. Translated in French and Dutch if necessary. The decision depends on the target audience and the time available.
 - c. Including a hyperlink towards the ESRO report (and specific chapters)
 - d. Very short colophon at the bottom of the document
4. Scientific report in English
5. Colophon at the end of the document

3.9.3. Approval by the Board

As a rule, the proposal by the management to include an ESRO project in the study programme, or its rejection, should be approved by the Board. The same rule applies for an ESRO draft report before its publication, as for any other KCE report, even if it does not contain recommendations or when the recommendations based on KCE work are endorsed by a group or institution which has the legal status to do so (e.g. reviews on RSV vaccines that were inserted in NITAG reports).



However, an important modification to the current procedures is proposed. **When indicated, a formal approval of the ESRO report by written (e-mail or online platform) can be delivered. This is possible throughout the year to minimise the waiting time (no pre-established blocking yearly schedule).** The current so-called “written procedure” of the Board applies¹. The usual rules for voting apply. In case of a justified opposition, the report may be amended and resubmitted for a written approval, or the approval may be postponed for discussion during the next regular Board meeting.

¹ Whenever a Board member within a set time period (usually 5 working days) formally declares not to agree with the conclusions/decisions (e.g. on budget), the point has to be discussed during the next Board meeting. No comment means approval (opt-out approach)

4. EXPEDITED LITERATURE REVIEWS

Expedited reviews use modified systematic review methods to accelerate the review process while maintaining systematic, transparent and reproducible methods.⁵ There is no one-size-fits-all solution for expedited literature reviews.⁵ On the one hand, there are various methodological adaptations possible, and on the other hand there may be specific circumstances where expedited reviews may be particularly appropriate, particularly when scientists are willing to accept some less certainty and a small risk for opposite conclusions to be able to deliver in due time for decision making.⁶ For example, during the COVID-19 pandemic KCE has produced “rapid reports” with less certainty about results, but with nevertheless enough certainty according to the KCE authors to be worthwhile publishing (be it sometimes with a disclaimer). These two dimensions (methods and circumstances) are considered for each of the questions below.

4.1. When is it appropriate to search only MEDLINE?

Key messages

- At least two electronic bibliographic databases should be searched.
- In expedited reviews of RCTs, two strategies can be implemented: either combining MEDLINE with Embase or CENTRAL, or searching MEDLINE with a supplementary search method (e.g. searching also ClinicalTrials.gov).
- For retrieving other study designs, the validity of searching MEDLINE only is not established.



4.1.1. Expedited reviews of RCTs

According to the Cochrane Rapid Reviews Methods Group (RRMG), limiting the number of databases searched can be a viable option in ESRO projects. MEDLINE has been found to have high coverage for studies included in systematic reviews (SRs).⁵ Some authors reported that for review of RCTs, searching only MEDLINE changes the results in only a limited proportion of cases.^{7, 8} In spite of these findings, the Cochrane RRMG considers that searching MEDLINE alone is not a viable option because coverage varies depending on topics and relevant study designs, and search sensitivity is lower than database coverage. They propose two alternative strategies:

1. Using a combination of two MECIR-required^m databases (i.e. MEDLINE, Embase, CENTRALⁿ), which is likely to be sufficient for retrieving published RCTs and to arrive at the same review conclusions as a comprehensive search.^{5, 6}
2. Combining MEDLINE with an appropriate supplementary search method:
 - A *PubMed similar articles* search is viable for updating clinically focused SRs of RCTs and holds potential for *de novo* searches for these designs.⁹ There is some indication that this approach also works for newly created reviews on clinically focused topics.⁹
 - A combination of ClinicalTrials.gov or the International Clinical Trials Registry Platform (ICTRP) for RCTs on newly approved drug^o
 - Combining CENTRAL or Embase with reference list checking.⁶ This can be done with the help of software for forward and

^m MECIR stands for Methodological Expectations of Cochrane Intervention Reviews, and are standards that should guide the conduct and report of Cochrane Intervention Reviews. They are drawn from the Cochrane Handbook for Systematic Reviews of Interventions (the 'Handbook').

ⁿ CENTRAL is a database for RCTs and quasi-RCTs. This means no search filter is necessary. Unlike MEDLINE and Embase, CENTRAL also contains grey literature, in particular study register entries and conference abstracts. It

backward citation chasing, such as citationchaser (see Chapter on automation).

4.1.2. Expedited reviews of non-randomised studies

Rapid reviews of non-RCT studies have been less studied. Subsequently, recommendations are less precise.⁵ For example, MEDLINE coverage varies depending on topics and study designs. Therefore, it is particularly important to define during the scoping stage what level of certainty is needed for the decision making, and to anticipate the consequences of a (very) inaccurate result.⁶

has a very high coverage of RCTs but is updated less frequently than databases like MEDLINE or Embase.⁵

- RCTs on non-pharmacological interventions seem to be less consistently registered in ClinicalTrials.gov. And combining MEDLINE and reference list checking might be less effective than combining the method with Embase or CENTRAL search strategies.⁶



4.2. When is it appropriate to limit the search to core journals?

Key messages

It is not appropriate to limit the search to core journals.

Other HTA agencies, such as the Austrian HTA agency (AIHTA) or the Canada's Drug Agency (CDA-AMC) limit their literature searches in those situations in which it is decided that a rapid review should be considered (given time constraints and urgency).^{p,q} For example, AIHTA mentions the option to limit their searches to core journals published in MEDLINE. A limit to Core Clinical Journals restricts retrieval to those journals considered of high impact. This list was available until 2020. Following that date, the Medical library association published in 2023 an adapted list of journals available in MEDLINE, known as Clinically Useful Journals (CUJ).¹⁰ Although this may be a way to greatly limit the results of searches, it is rather restrictive and thus AIHTA recommends adding to this list of CUJ, journals that are considered essential in the respective subject of research not listed in the "MEDLINE core journals". Expert surveys or consultations are recommended as a way to identify such journals. Given the fact that the list of CUJ comprise only around 5% of the journals captured in MEDLINE, **this is an approach that should at present not be prioritised for KCE reviews.**

4.3. When is it appropriate to optimise the precision of the search strategy?

Key messages

- In ESRO approach, it is appropriate to optimise the precision of the search strategy provided that an acceptable level of sensitivity is maintained.
- Various options exist to expedite the construction of an optimised search strategy, as presented in the text .
- For quality assurance purpose, an information specialist should be involved, at least to ensure the peer review and the validation of the search strategies prepared by experts.

4.3.1. Constructing the optimised search strategy

The construction of a search strategy to optimise the balance between sensitivity and specificity according to the intended purpose of the search is critical to all evidence synthesis products. Achieving this balance is even more important in the context of ESRO. An important way to reduce the number of records retrieved by a database search is to increase search precision without sacrificing too much sensitivity. According to the Cochrane RRMG, the following modifications can be envisaged for an expedited review:⁵

1. Dividing search terms for each element into "definite" (most commonly used to describe a particular element in the literature), "probable" (less commonly used) and "possibles" (rarely used) allows to calibrate the strategy for relevance and numbers of results, i.e. omitting such a

^p CADTH; Health Technology Review Rapid Review Process Version: 2.0. February 2022. Available at: https://www.cadth.ca/sites/default/files/attachments/2022-02/Rapid%20Review%20Process_0.pdf

^q Gartlehner, G. (2009): [Internal manual. Part 2 (2nd Ed.)]. HTA-Projektbericht 06. 2009 (upd. July 2020). Available at: [Internal manual. Part 2 (2nd Ed.)] - Repository of AIHTA GmbH



possible search term from the search strategy is likely to reduce the size of the search result without losing relevant studies.⁵

2. Using proximity/adjacency operators (which are offered in many database interfaces, including PubMed since 2022^r) allow to set a maximum distance between two search terms. They are more precise than AND-combinations but less restrictive than phrases.
3. Restricting subject headings to records where they are the focus of the article is another strategy.
4. Limiting free-text terms to occurrence in the title also can improve search precision, as the most important concepts of a study are often mentioned in the title.
5. Accounting for frequency of word occurrence. For example, Ovid interfaces offers a frequency operator, which allows a searcher to define a minimum number of times a term has to appear in a certain field, with the assumption that a recurrent term is more likely related to the topic of the article.⁵
6. Using limits and restrictions appropriately and with caution (languages, publication dates, settings, etc...). Validated search filters can be used.⁵ Clinical or methodological justification for restrictions must be provided.²

Other means to expedite the search strategy development are:

1. Reusing or adapting complete or partial search strategies of published systematic reviews (see also section 4.7.2.2)
2. Using the findings of preliminary searches to assess whether grey literature may be relevant for a topic and what types (e.g. clinical trial registrations, preprints, theses).
3. Using automated syntax translation can save time and reduce errors when translating a primary search strategy to different databases.⁵

However, manual adjustments will usually be necessary (see also section 7).

4.3.2. Quality assurance

For defining the best possible search strategy corresponding to the objective of an ESRO project, the support of an information specialist is required.² At a minimum, he/she should ensure the peer review and the validation of the search strategies elaborated by KCE researchers, for example using the Peer Review of Electronic Search Strategy (PRESS).¹¹ PRESS recommends a maximum turnaround time of five working days for feedback. If a full peer review of the search strategy is not possible, review team member with search experience should check the search strategy for spelling errors and correct use of Boolean operators at a minimum.⁵ The sensitivity of the optimised search strategy must be tested, for example by retrieving eligible studies found through preliminary searching.⁵

4.4. When is it appropriate to skip the retrieval of grey evidence?

Key messages

- In ESRO approach, it is appropriate to skip the retrieval of grey evidence most of the time.
- This may be less the case however for emerging pathologies or treatments, or understudied topics.
- Evaluation of the balance between the risk (lower certainty of results) and the benefit (less time- and resource-consuming) is topic-dependent and should be explained in the research protocol and report.

^r Information on proximity search capabilities in PubMed can be found at https://www.nlm.nih.gov/pubs/techbull/nd22/nd22_pubmed_proximity_search_available.html



Grey evidence is understood here as any evidence accessible outside peer-reviewed journals. The main grey literature resources are listed by the Cornell University (<https://guides.library.cornell.edu/evidence-synthesis/grey-literature>; accessed 14/03/2024), including links to preprint repositories. For many review topics, most evidence is found in peer-reviewed journal articles, making bibliographic databases the main resource of systematic searching. This may be less the case however for emerging pathologies or treatments, or understudied topics. Finding grey literature and searching it systematically is challenging and may be time-consuming. The relevance of searching such evidence is topic-dependent and must be carefully established based on the following considerations:

1. Does the risk of skipping the retrieval of grey literature (missing relevant information sources, lower certainty of results) outweigh the benefits (less time- and resource-consuming)?
 - a. **Adverse effects of an intervention.** When the review includes a question relating to the safety of an intervention (e.g. a new drug, a screening programme), grey literature should be searched, including the vigilance databases of regulators. This might be less crucial for a review on behavioural changes following a health promotion program, for example. On the contrary, one could argue that the evaluation of public health interventions may be less often published in peer-reviewed journals and it is therefore more important to screen the grey literature, but this argument is more difficult to support in the frame of an expedited research. Such considerations should be explained in the research protocol.
 - b. **Interrupted or unpublished studies.** Retrieving interrupted or yet unpublished randomised controlled trials (RCTs) requires searching clinical trials registries, such as [ClinicalTrials.gov](https://clinicaltrials.gov) or the WHO-led International Clinical Trials Registry Platform ([ICTRP](https://www.ictRP.org)). Of note, RCTs on interventions other than new pharmaceuticals appear to be less systematically registered in these databases, and it may be appropriate to search other registries, for example the Trials Register of Promoting Health Interventions (TRoPHI)

covering over 12 000 randomised and non-randomised trials ([website](#) visited on 14/03/2024)

- c. **Studies published without peer-review.** Retrieving preprint studies is required when the publication of new results in classical peer-reviewed journals might be too slow to allow urgent and adaptative decision making, as was, for example, the case during the COVID-19 pandemic. In such circumstances, preprint registries must be screened, e.g. medRxiv.¹² The peer-review and critical appraisal can be carried out internally.
2. Is the search for grey literature facilitated by the availability of the following elements:
 - a. **Study registries.** There are well organised registries in which most of the grey literature is concentrated, e.g. [ClinicalTrials.gov](https://clinicaltrials.gov) for interrupted or unpublished RCTs. Studies confirm that the omission of grey literature searches rarely affects the conclusions of RCTs reviews.^{6, 13} Of the different types of grey literature, excluding conference abstracts seems to have the least impact.¹⁴
 - b. **Topic experts.** There is a group of dedicated experts in the area of investigation who are involved in the search for information and can reliably point to sources of evidence not (yet) published in a peer-reviewed journal. If relevant records are identified outside of the usual bibliographic databases, searching further the grey literature might be important to answer the research question.⁵
 - c. **Search engines.** For example, the CADTH's Research Information Services team has developed and made available a grey literature resource, the [CADTH Grey Matters tool](#).^s

^s This search engine was not tested



4.5. When is it appropriate to rely on one single reviewer?

The Cochrane RRMG has recently proposed that for rapid reviews, one person can extract the data while a second person verifies accuracy and completeness only of critical data that can affect the results or conclusions.² The second reviewer will also need to check the full texts. Any disagreements should be resolved through discussion and consensus between reviewers, involving a third reviewer if agreement cannot be reached.²

4.6. When is it appropriate to include (partly or mostly) surrogate or modelling studies?

Key messages

- When clinical studies using hard clinical endpoints are **not available** or **not feasible**, studies using surrogate endpoints can be used to inform decisions.
- These surrogate endpoints can be either **biomarkers** or **intermediate clinical outcomes**; both types of endpoints should be thoroughly validated, i.e. there should be sufficient evidence that they are reliable predictors of the target clinical outcome.
- Quality assessment of studies using surrogate outcomes can be done using risk of bias tools available for the different study types, bearing in mind that the use of a surrogate outcome already adds a potential bias.
- Modelling studies are sometimes used to inform public health decisions; quality assessment tools are available for this study type. It is important to assess whether the model's assumptions (inputs and dynamics) are realistic and/or backed by evidence.

In normal circumstances, treatment and policy recommendations should be based on clinical studies with important clinical outcomes. However, in some situations, no clinical studies are available yet or may ever become available, and recommendations or conclusions have to be based on other study types or on surrogate endpoints.

4.6.1. Surrogate-based studies

For example, in the case of vaccines or antibody treatment for rapidly changing pathogens, often yearly or biyearly adaptations to the (resulting) antibodies have to be made to accommodate the genetic drift of these pathogens. Such adaptations can usually not be tested in randomized controlled trials with clinical endpoints because of time constraints. Randomized clinical trials are also not feasible for vaccination studies of rare infections (e.g. meningococcal disease), as sample sizes would be too large. It is therefore widely accepted within the field of vaccination studies to make use of “immunobridging” or **correlates of protection**, provided these are well validated. For example, FDA approval for new influenza vaccines is based on randomized controlled trials which typically cannot demonstrate clinical efficacy of the vaccine, but in which immunogenicity criteria and safety are considered sufficient for approval.^t

For COVID-19 as well, presence of high titres of neutralization antibodies after vaccination was shown to be a good (although imperfect) correlate of protection.^{15, 16} Reliable correlates of protection have also been described for other infectious diseases.¹⁷ Typically, vaccine studies using this surrogate endpoint will attempt to demonstrate a significant protective **antibody titre increase** (above a certain pre-defined and validated threshold) after vaccination.

Of note, for vaccines targeting multiple serotypes (e.g. anti-pneumococcal vaccines), different thresholds may apply for the different serotypes to assure adequate protection.¹⁸ These differences in clinically validated thresholds are however not always taken into account when approval is

^t Exact criteria for yearly approval of new versions of known vaccines could not readily be found on the FDA website.



granted; often a “universal” threshold is applied for all serotype-specific antibodies, leaving some serotypes inadequately covered by the vaccine.

Another example of an often used (clinical) surrogate endpoint is a decrease of $\geq 5\%$ or $\geq 10\%$ of body weight to approximate the protective effect of interventions on the risk of cardiovascular events.¹⁹

Assessing the validity of the suggested surrogate is thus the central question. This entails the availability of a good surrogate candidate and good quality studies to evaluate it. If these two conditions are met, basing the review and recommendations on current evidence seems reasonable, even in the absence of clinical RCTS.

4.6.1.1. Good quality surrogates

Surrogate endpoints can be roughly divided into two types: **biomarkers** and **intermediate clinical outcomes**.²⁰ Biomarkers can be defined as measurable biological molecules whose presence or level is predictive for the target outcome of interest; examples are neutralizing antibodies in the case of infectious diseases, or LDL cholesterol in the case of cardiovascular mortality. Intermediate outcomes are clinical outcomes which are predictive of the target outcome of “ultimate interest”; examples are weight loss or angina frequency for cardiovascular mortality. A good surrogate outcome is ideally a “mediator” of the effect of the intervention under study on the target outcome, i.e., it is not just merely associated with it, but it is (at least partly) responsible for the intervention’s effect on the target outcome. Surrogate outcomes which are associated with the target outcome but are not part of the causal pathway of the intervention on the outcome, can generate misleading conclusions. Even when the surrogate outcome is hypothesized to be on the causal pathway, effects of an intervention on the surrogate outcome may not be reflected on the target outcome. An example of this is the repeated failure of HDL-increasing therapies to provide a mortality benefit, despite the strong association between low HDL cholesterol and cardiovascular mortality.^{21, 22} Therefore, before surrogate endpoints can be used, there should be good-quality trials available which demonstrate that changes in the surrogate endpoints are reflected by proportional changes in the target outcomes (preferably with a clear dose-response relationship). The surrogate endpoint framework suggested by Ciani et al. also mentions that surrogate endpoints should be **validated**, i.e. they should reliably

predict benefit or harm of the treatment, based on scientific evidence.²⁰ It is important to note that many randomized clinical trials make use of surrogate endpoints, often without explicitly mentioning this. A review from 2010 found that 17% of RCTs published in 2005 and 2006 used surrogate endpoints, of which 57% clearly reported this outcome was a surrogate, and only 35% discussed the validity of this surrogate.²³ Another review found that no less than 62% of RCTs published in high impact journals in 2013 and 2014 used surrogate outcomes as the primary outcome.²⁴

4.6.1.2. Good quality evaluation studies

The usual risk of bias tools for interventional studies can be used when assessing study quality of studies using surrogate endpoints, keeping in mind that using a surrogate outcome already introduces bias.

In the case of emerging or rapidly changing epidemic illnesses, however, large randomized trials are not always available, and the treatment’s effect on the surrogate outcome is sometimes assessed by small *in vivo* studies (evaluation of safety and neutralizing antibody production in small patient groups, not always randomized) or even combinations of *in vitro* and pharmacokinetic studies (e.g. monoclonal antibodies showing virus neutralization in a cell-based assay, in combination with a pharmacokinetic analysis and safety assessment in small patient groups). In such cases, quality assessment can be challenging as quality assessment tools for *in vitro* or phase I/II studies are less well established (see section 4.8).

Real-world evidence (i.e. post-licensure observational studies) is also often used to provide evidence of effectiveness for these types of interventions. For influenza vaccination, for example, the Centre for Disease Control (CDC) of the US assesses vaccine effectiveness (VE) yearly through case control studies in the US Flu VE Network. For rare infectious diseases, such as meningococcal infections, case-control studies or before-after incidence comparisons are also often used to demonstrate vaccine effectiveness.^{25, 26} Quality of such observational studies can be assessed using the tools described below (see section 4.8).



4.6.2. Modelling studies

Another example of non-clinical studies used to inform public health decisions are modelling studies. In this type of study, (efficacy) data from other study types (ideally clinical studies, but often also laboratory or epidemiological) are used and modelled onto a population of interest. Possible reasons of doing this are to assess the impact of an intervention on the population of interest, to predict future health outcomes based on current data or to predict efficacy of treatment.

This practice is mostly a theoretical exercise, but it can offer very valuable perspectives for public health decisions. However, results of these modelling studies should be interpreted with caution. Conclusions are based on the assumptions such a model makes, and care should be taken to take these at face value.

Moreover, a model is by definition a simplification of underlying real-world dynamic processes. Predictions based on a model, even a very good one with correct assumptions, can thus still differ considerably from reality. For example, the FDA gave emergency approval to a higher dose of tixagevimab/cilgavimab (Evusheld) for prevention of SARS-CoV-2 infection in 2022, as pharmacokinetic models showed that a higher dose would overcome the decreased affinity of Evusheld for the emerging Omicron variant.²⁷ However, benefit of using this increased dose of Evusheld was not confirmed in clinical studies, leading to withdrawal of the authorization in 2023.²⁸

Some templates and tools for quality assessment of modelling studies have been developed.²⁹ These will be further discussed in section 4.8.

For a rapid appraisal by policy makers, the framework of modelling studies, designed by Thompson et al, recommends evaluating the following elements of model utility:³⁰

- **Instrumental utility:** is the model “fit for purpose”? The authors recommend checking whether reasonable real-world inputs are incorporated, whether real-world mechanisms drive the model and whether the outputs of the model are measurable. Questions include: are data inputs robust? Does the model include ranges of uncertainty? Does the model incorporate differences between individuals and

groups? Are the mechanics of the model formally defined and explained? For all questions regarding instrumental utility, see Thompson et al.³⁰

- **Conceptual utility:** can the model be effectively communicated to the public?
- **Political utility:** is the instrumental and conceptual utility of the model strong enough to support defensible policy actions?

Especially the questions regarding the instrumental utility of the model are interesting to consider when drafting recommendations; a model will only have political utility when it has conceptual utility, and it will only have conceptual utility when it has instrumental utility.³⁰

4.7. When is it appropriate to base the review (partly or mostly) on interim results?

Key messages

- It has been demonstrated that interim analysis of results tends to be biased towards a greater intervention efficacy than reported in final analysis.
- There are circumstances when interim results are crucial, particularly in epidemic situation. The use of interim results should be restricted to such situations.
- Certain precautions must be respected in order to guarantee as much as possible the validity of the results:
 - The registered research protocol should describe anticipated timing of interim analyses
 - Statistical methodology to control type-1 error should be duly described and applied
 - The registered research protocol should specify rules or thresholds to guide decisions



- The contribution of independent members of the Data Safety Monitoring Board (DSMB) or the Data Monitoring Committee (DMC) to ensure the validity of interim analysis should be duly described.
- If these precautions are not met, the use of interim results should be discarded.

Interim analysis of randomised controlled trials (RCTs) consists of the early looks of the data, usually by an independent Data Monitoring Committee (DMC) or Data Safety Monitoring Board (DSMB). The first aims are generally to examine safety and guide decisions on clinical modifications (e.g. study sample size).³¹ In some cases, interim results are published in peer-reviewed journals before the final analysis and interim results in efficacy are used to formulate clinical recommendations.

However, several authors emphasise that interim results should be taken with caution.^{32,33,34,31} In particular, two systematic reviews^{33,34} assessed the reliability of interim results in comparison with the final ones regarding treatment effect.

The first review searched RCTs in Medline (1990-2014) and in 8 journals (such as the BMJ, NEJM, Lancet) from 1996 until 2014.³³ A total of 63 trials were identified, in which the same quantitative comparison was reported in 58 trials. More than half of the trials were in oncology and 30% of trials were still recruiting when the interim report was published. The analysis showed that the final results were generally of a smaller magnitude than the interim ones, with a smaller effect size observed in 39 (67%) trials and the same or a larger effect size in 19 (33%). There was a marked reduction, defined as a $\geq 20\%$ decrease in the size of the treatment effect from interim to final analysis, in 11 (19%) trials compared to a marked increase in 3 (5%), $p = 0.057$. The magnitude of change in effect size from interim to final results was larger in trials with commercial company support and increases as the proportion of final events at the interim report decreased. There was no material association with other factors such as type of disease, use of blinding, pre-specifying the interim analysis, proportion of final sample size at the interim report, year of interim publication, year of final publication, time

between interim and final report, impact factor of the interim publication journal or impact factor of the final publication journal. **The authors concluded that while examining interim results is an important part of trial monitoring, publishing them need to be considered with caution, or better still avoided, particularly while patients are still being recruited.**³³

The other review compared matched interim analysis publications and final analysis publications in oncology clinical trials.³⁴ The authors focused on cancer progression (as interim analysis) and overall survival cancer (final analysis) in 33 RCTs retrieved via PubMed from 2005 to 2015. They found an important difference between interim (statistically significant progression-free survival benefit in 31 RCTs) and final analysis (statistically significant overall survival benefit in 8 RCTs). Among the 33 RCTs, only 3 showed that progression-free survival strongly correlated with overall survival. It is interesting also to notice that among matched pairs ($n=23$), interim analyses were published in more prominent general medicine journals such as the NEJM or The Lancet compared with final analyses (rather published in oncology journals (e.g. Journal of Clinical Oncology, The Lancet: Oncology). *"The median impact factor was 44 (IQR 24–72) for journals publishing interim analyses versus 24 (IQR 11–34) for journals publishing final analyses."* The authors concluded that **publishing interim analysis with progression-free survival apart from the final mature overall survival is not just unreliable but likely to generate unsupported and inappropriate hype.**³⁴ They also suggested to ask journals that published interim analyses to provide timely links to the final publication, even if it is published in a different journal.³⁴

In a crisis period as the COVID-19 pandemic was, investigators' pressure to publish interim results of treatment effects was high. The case of Molnupiravir, the first antiviral against COVID-19, illustrates well the problem of overoptimistic interim results. Initially, a planned interim analysis from Merck showed an efficacy of approximately 50% with respect to the primary outcome of hospitalization for any cause or death through day 29. The efficacy decreased to approximately 30% by the end of the trial, and the discrepancy between the two results could not be explained.³⁵ However, despite their well-demonstrated risk of bias, interim analysis is still used outside urgent situations. A very recent example is the EMA granting a



marketing authorisation for a vaccine against RSV in pregnant women whereas the RCT was still ongoing.³⁶ It is difficult to justify such an approach, and KCE should not apply it.

Several authors highlight some safeguards when using results of interim analysis. The role of the Data Safety Monitoring Board (DSMB) or the Data Monitoring Committee (DMC) in protecting participant safety, data quality, and interim analyses in RCTs is underlined.^{37, 38} DSMBs are crucial in maintaining the scientific integrity of RCTs by allowing an interim review of data that is independent of the sponsor but also by avoiding potential bias into the study before data have “matured”— i.e., *that is, accumulated sufficient power to ensure that the results reflect the true outcomes of the completed study*.³⁸ Interim results publication should provide acknowledgement to the DMC/DSMB members with the scope of their work to verify members’ expertise in relation to the trial subject matter.^{37, 38} It is important to mention that this information appears to be underreported.³⁷

More recently, a guidance on interim analysis method emphasizes that some methods should be used in interim analysis for efficacy (the group sequential methods and alpha-spending functions) to maintain some control over type I error (i.e. finding a statistically significant result when there is no underlying intervention effect in reality).³¹ Prespecifying the interim analytic plan to the extent possible is paramount. *“Detailed documentation should describe anticipated timing of interim analyses, proposed statistical methodology, and any prespecified rules or thresholds to guide decisions”*.³¹ Due to the instability of test statistics, especially with small sample size, the same authors recommend to reserve efficacy interim analyses for later phase, confirmatory studies with large sample sizes and lengthy follow-up. Evaluation of the interim results should be interpreted in the context of internal and external study factors such as other information on the relevant disease areas in other studies. The importance of an independent and unbiased expertise from the members of the DSMB is also highlighted.³¹

4.8. When is it appropriate to rely on a published systematic review, with or without update?

Key messages

- There is no standardised approach nor unified methods for conducting an update of an existing systematic review/HTA.
- The decision-making process for implementing an update should consider several key-points:
 1. Assessment of the currency of the research question, and the overall quality of the existing SR/HTA.
 2. Identification of relevant new methods, new studies, new validated methods, or other information
 3. Assessment of the effect of updating the existing SR/HTA (likelihood of the inclusion of new studies or new methods to modify the findings or the credibility).
- The Ottawa method (and its adapted version) may simplify and accelerate the process of assessing the existence of new information/studies and assess the impact of new studies in the updating process.
- An practical algorithm has been developed and is provided in this section.

4.8.1. Background

Systematic reviews (SRs) and Health technology assessments (HTAs) summarise scientific evidence on the effects of interventions or treatments in an efficient and reproducible way. As performing *de novo* SRs/HTAs is resource- and time-intensive, there is a significant interest in re-using an existing SR/HTA, if appropriate.³⁹ The ability to use such an approach depends on whether an existing SR/HTA is appropriate to answer the current research question and, whether the retrieved report needs to be updated. Updating an existing SR/HTA (rather than conducting a new research) is indeed recognised as an effective process, that can save



time.^{39, 40} However, there is no standard approach to assessing whether an existing SR/HTA can be used without an update, nor unified methods to use for updating a SR/HTA.

Since SR/HTA can become outdated quite rapidly,⁴¹ the Cochrane Handbook for Systematic Reviews of Interventions identified key-factors and considerations for determining whether an update is needed.⁴² In addition, the Cochrane document suggests several methodological factors for conducting the update. In a complementary way, other authors have recently identified key factors for guiding the fulfilment on the update of SRs (guidance on how to conduct an update and on which type of systematic reports they should ideally be conducted).^{43, 44} They have simultaneously developed a decision tool to help with that process. These documents share similar key points and many of them naturally overlap.

Here, a summary and synthesis of the various factors to be considered is provided, with the aim of developing a methodological framework for conducting expedited reviews at KCE.

4.8.2. *Deciding when and whether to conduct an update of an existing systematic review*

With regard to the decision-making process of updating an existing SR, the Cochrane working group's approach is largely based on the international consensus from Garner *et al.*, published in 2016 (an international panel of authors, editors, clinicians, statisticians, information specialists, methodologists, and guideline developers were invited to develop a common approach to updating systematic reviews).³⁹ The three steps of the process are presented in Table 1.

Table 1 – Decision framework to assess the need for updating SRs

Steps	Consideration
Step 1: Assess currency (i.e., relevance of topic at the time the update is being considered)	Does the published review (and the need for a new review) still address a current question? Has the review had good access or use? Did the review use valid methods and was it well conducted?
Step 2: Check the relevance of methods, studies included and other information	Are there any relevant new studies or other information? e.g. updating full strategy or using abbreviated search strategy, tracking studies in registers, ... Are there any new relevant, validated methods used now that were not used in the SR? e.g. low quality SR not considering the assessment of the risk of bias, without synthesis of the evidence (GRADE),...
Step 3: Assess the effect of updating the review	Will the adoption of new methods, the inclusion of new studies or new data change the findings or credibility?

Source: Garner *et al.*, BMJ 2016³⁹

4.8.2.1. *Step 1 and Step 2 - Assessing currency and relevance of an existing Systematic Review*

Several factors that should be considered for integrating existing SRs into new reviews have been proposed. Three of those proposed by Robinson *et al.* are useful when assessing the relevance of a SR/HTA and include⁴⁴:

1. The assessment of the PICO elements, which should be appropriate to the current research question.
2. The timeframe of the search (currency)
3. The review methods.

Likewise, the Cochrane workgroup mentions that the first step to be considered is to assess whether it is still relevant for policy makers.⁴² This may require knowledge of the area to answer the question and may need the involvement of experts and/or stakeholders.



Other points suggested by the Cochrane workgroup are:

1. To check whether the existing review is frequently accessed or cited: reviews that are frequently cited may be important to be updated, whereas reviews rarely cited are unlikely to address a highly relevant question and may therefore not be a source for an update.⁴²
2. To assess the quality of the existing review and whether appropriate and valid methods have been used. For this purpose, the quality should be appraised using the AMSTAR 2 tool (or ROBIS).

If the SR has been conducted with insufficient quality, or if its research question is no longer current, it should be disregarded and no update should be performed.

A quite similar approach has been proposed by Kim *et al.* who have identified the key factors for addressing the state of literature and thus, determining if an existing SR is suitable for updating (See Table 2 for more information).⁴³ In their algorithm, if there are several pre-existing SRs identified in the scoping phase, with discordant results^u for unknown reasons, a *de novo* SR should be performed. On the contrary, if results are concordant, the best quality SR found among the pre-existing SRs should be used, as such, only if it is up-to-date and no additional relevant studies have been published after its search date. Otherwise, an update of this pre-existing good quality SR should be performed.

An adapted and simplified algorithm was developed, based on the present publications.^{39, 43, 44} It should be noted here that if the planned review has a wider scope in one or multiple PICO elements^v, an update of an existing SR would no longer be suitable and instead, a systematic review of existing reviews (an umbrella review) may be more appropriate.⁴³

^u Potential reasons for discordance include: timing (leading to differences in studies included), extracted data, synthesis of results or their interpretation.

Table 2– Assessment of the state of the existing literature

Steps and questions

Relevance: are there one or more relevant SRs available?

Methodological quality: are there relevant SRs of sufficiently high quality in methodology?

Reporting quality: are the relevant SRs of sufficiently high quality in reporting?

Comprehensiveness: are the relevant SRs comprehensive?

Currency: are the relevant SRs up to date?

Discordance: are the relevant SRs discordant for unknown reasons?

Source: Kim *et al.*, *Systematic Reviews*, 2022 (extract from Table 1 – Key factors and considerations for selecting to conduct one review type over another)⁴³

Table 3 – Application of step 1 (assess currency) suggested by Garner *et al.* to the context of KCE research

Does the published review still address a current question?

At KCE: The relevance of the research question is first addressed during the selection of topics for the yearly KCE working programme, and then, during a scoping meeting. Only when a research question is of interest, a decision will be made to go ahead with a KCE study/review. Thus, if one or more previous SRs are identified, covering the same research question and similar PICO, their relevance would already have been indirectly assessed. Moreover, when considering using available SRs as the starting point of a new KCE report, the search for SRs is limited to recent SRs (search date considered).

Has the review had good access or use?

At KCE: This is not something checked at KCE and although it could be easily added to our quality appraisal of existing SRs for additional

^v i.e. different interventions used for the same condition (population); a same intervention is used for different populations; different outcomes or timepoints are studied



information, the quality of the SR and how recent it is should be given more weight in the decision to select a specific SR, given that frequently “cited” SRs may not necessarily be of high quality.

Did the review use valid methods and was it well conducted?

At KCE: This is something we always do before using a SR as the starting point of a new research project (i.e., quality appraisal using AMSTAR 2).

4.8.2.2. Step 2 - Identification of relevant new methods, studies and other information.

If the quality of the targeted SR is satisfactory and if the research question still is relevant (currency of research question), the next step is to consider whether new information, or new methods not considered in the SR are available. On the one hand, the Cochrane group advocates considering whether newly recommended methods could be integrated into the review (e.g. methods to assess bias, to synthesise evidence or to include different study designs such as economic evaluation, qualitative research, cluster randomised trials). Some methods have been suggested to simplify the process of assessing the existence of new information/studies and ensuring a rapid assessment of the need for updating a SR. A particularly interesting approach is the one presented in the OTTAWA method.⁴⁵ Some of the shortcuts proposed by such method could be considered for our rapid reviews, while some others may be more difficult to justify.

First, the OTTAWA method suggests to perform the literature search purely in PubMed. In addition to this, it uses journal filters limiting the search to five general interest journals, as well as a few specialized journals (depending on the topic), most frequently cited in the original SR.

A limit to pivotal trials based on sample size and journal impact is also suggested.

Once this is carried out, specific criteria are used to determine whether a SR is out of date. These criteria are described in section 4.7.2.3.

It is important to mention that once the identification of the relevant literature is performed, the final selection and data extraction for the “qualitative” signals in the original OTTAWA method is performed based on title and abstract only. No review of full articles is carried out.

Table 4 – Application of step 2 (assess relevance) suggested by Garner et al. to the context of KCE research

Are there any relevant new studies or other information?

At KCE: Any rapid approaches to come up with specific abbreviated search strategies to check for relevant new evidence (e.g. the Ottawa method) should be explored. At the moment, most often a full regular search strategy is used.

For example, limiting the search to one database (e.g. MEDLINE) could be a way to save time, with a limited risk to miss the most important studies (see section 4.1). Particularly if experts in the field are consulted and grey literature is checked for key references. On the other hand, the limitation to some journals may be more difficult to implement and offer less reliable results.

Are there any new relevant, validated methods used now that were not used in the SR ? e.g. GRADE

At KCE: This is something currently considered during the quality appraisal of any SR. Validated, recent methods are generally used at KCE when updating an existing SR.



4.8.2.3. Step 3 - Assessment of the effect of updating a systematic review

How new data might change the conclusions of a review is a crucial consideration in deciding whether or not to undertake an update. If relevant studies not captured in the original SR are identified (by running the search strategy of the original SR^w), it is necessary to assess whether the addition of this new information might influence or modify the conclusions. The Cochrane working group mentions that there are several possible scenarios: reviews having results with a high level of certainty are less likely to change after the addition of new information, while new studies may show unexpected effects or provide insight into new details about the observed effects.⁴² Likewise, Kim *et al.* mention that if other SRs show discordant results without any clear reasons^x, a *de novo* SR may be better, rather than using one or more conflicting SRs.⁴³

Several formal tools for predicting the impact of new studies in the updating process are available. Among the most pragmatic and convenient are the OTTAWA, or the RAND methods (See Table 5). The RAND and the OTTAWA methods were compared and tested in a pilot study published in 2012 with the aim to assess their consistency (or lack of) in signaling the need to update SRs.⁴⁶ The pilot study showed that in general, they appear to offer similar signals. Although the OTTAWA method was recognised as being more transparent and simpler, the authors of this pilot study proposed a hybrid approach to respond to some of its limitations. In particular, they appreciated the importance of including references provided by experts (as per the RAND method), and they suggested not to focus only on pivotal trials (which is particularly important in areas where the quality of the existent evidence is low, with variable study designs).⁴⁶

An “adapted” OTTAWA method allowing the assessment of the need to update topic areas in clinical practice guidelines was presented by Goossen

et al. in 2022.⁴⁷ Although updating guidelines is not the topic of interest of our document (focused on SRs), this 2022 paper highlights interesting points for consideration regarding the limitations of the OTTAWA method and how these may be addressed via adaptations. A table summarising the differences between the original and the adapted OTTAWA methods is presented in an appendix. Table 6 shows how the adapted OTTAWA method suggested by Goossen *et al.*⁴⁷ is applied in practice, its main risks as well as considerations to make in the context of KCE studies.

Table 5 – Tools for predicting how new studies can affect the new review conclusions

Steps	Description
Ottawa method	A simple PubMed search (using the three largest and three most recent trials from the original review) to identify new studies. If new studies are found, then the method uses quantitative signals (e.g. change in significance, effect size) to assess the likelihood that the new studies will change the review conclusion, thus triggering an update.
Adapted Ottawa method	This method includes certain adaptations for addressing the main critiques and limitations of the original OTTAWA method. An explanation of the weaknesses of the original method and the justification for the adaptations is offered in the Table 6.
RAND method	An abbreviated search of five major journals to identify new studies, and a search of the US Food and Drug Administration website and external expert judgment to determine the currency of the report findings.

Source: Garner *et al.*, *BMJ* 2016³⁹

^w An abbreviated search strategy is a possibility suggested by the Cochrane group.

^x e.g. differences in the methods, or different timeframes



Table 6 – The adapted Ottawa method and rationale

Adapted Ottawa method*	Gains and risks	To be considered for KCE
Focused literature searches		
In a single database: Ovid-Medline	- Option allowing to save time - Risk to miss evidence published elsewhere, but unlikely to miss the most important studies, unless studying a field in which evidence is very scarce.	- If one database is consulted, experts in the field should always be consulted to ensure that all key-references are identified and considered. - Grey literature should be searched with key-terms to further limit the risk to miss important references.
With journal filter: general interest medical journals with IF ≥ 6, specialty journals with IF ≥ 2	- Advantage to rapidly identify highly relevant evidence in the topic - The possibility to limit the search to a selection of journals need to be discussed with experts in the field to ensure in which cases this may be appropriate.	Journals of interest should be defined with experts in the field (and to be considered even if IF < 2 for some medical/economical specialities).
Study selection		
Using predefined criteria, abridged from criteria selections from existing systematic review	Efficiency gained by using filters (timeframe of publication, language, similar PICO items)	For a majority of topics, language limitation to English is generally sufficient. All other aspects are generally already considered at present for KCE projects.
Eligibility assessed for all relevant new studies	- Potentially relevant signals are not missed - High sensitivity can increase the number of studies selected	Discussion with experts could allow an appropriate interpretation of new relevant studies, particularly if they have discordant conclusions.
Comparison of study results with results from existing systematic review		
Qualitative updating signals only	Estimation if new results are in line or in contradiction with existing review. If in line, no need for updating the SR.	Instead, a short description with references of the studies can be considered. Alternatively a table with the references of the more recent studies can be provided for completeness.
Experts solicited to decide on updates		
Based on the proposed need to update topic area	Identification of relevant evidence, directly applicable to healthcare professionals involved in the topic (minimize the risk of missing key information and sources)	/
Invited to submit additional PICO questions	- Way to be appropriately in line with the potential changes of the studied topic (refinement) - Risk to enlarge the scope of the research	It is necessary to be discussed with experts before searching for new evidence (some characteristics of a topic can evolve within a discipline).

Legend: * This approach is only recommended for consideration for updating an existing systematic review of high-quality.



It should be noted that the recommendations of the RAND method are USA-centered and focus on RCTs of drugs. Searching the 5 major journals is likely to be inappropriate in most cases (more relevant for new pharmaceuticals for which a limited number of large trials are available). The RAND approach is therefore not recommended for routine KCE reviews. From all these methods, the most specific that could be considered in the context of KCE's work is the adapted OTTAWA method. It is worth noting that the focused literature searches recommended in the adapted OTTAWA method should only be used once a good-quality systematic review has been found (and not to search for it).

Table 7 – Application of step 3 (assess the effect of updating the review) to the context of KCE research

Garner et al. propose a checklist to decide on when and how to update systematic reviews.³⁹

At KCE: This is a consideration that has already been made in previous KCE reports to justify not to update a MA, but generally, a narrative summary for any new papers was still provided for completeness.

The Cochrane working group highlights that reviews presenting a high level of certainty in their conclusions are less likely to change after the addition of new information, while new studies may show unexpected effects or provide insight into new details about the observed effects.⁴² The Cochrane working group suggests several formal tools for predicting the impact of new studies in the updating process. Among the most pragmatic and convenient is the “modified” OTTAWA method. KCE should explore the suggestions of this method to rapidly check if any relevant important studies have been published since the search date of the original SR. Even in the presence of new studies, if their conclusions are in line with those of the original high-quality SR identified, a decision may be made not to update the SR and simply summarise its results. Nevertheless, references for the new studies should be provided for the interested reader.

4.8.3. Planning and conducting the update of an existing systematic review

Although the updating of a review should be contextualised for each case,⁴³ the Cochrane working group highlighted several key factors for planning and carrying out the update.⁴² They are summarised in Table 7 and an algorithm is proposed in Figure 2.

**Table 8 – Steps for planning and conducting the update of a review from the Cochrane Handbook for Systematic Reviews of Interventions**

Steps	Considerations
Reconsidering review questions and eligibility criteria	All elements of the review question (PICO) should be reconsidered: • P: new understanding of definition of patient populations • I: new interventions (or older that are no more used) • C: new comparators (new standard of care) • O: standardised core outcomes sets, or new extended outcomes (e.g. economic aspects) • D: consider to add new study designs to potentially widen the evidence
Splitting and merging reviews	Since evidence is accumulating, splitting the review into two or more (more manageable) reviews with more narrowly defined questions could be an option. It is also a way to accelerate the process of updating. Similarly, merging two or more narrow reviews to give a more complete view on a topic of interest could also be considered.
Planning the search strategy for an update	• The search strategy can be re-run as specified in the existing review, with the addition of date limits set to the period following the most recent search. • An information specialist or healthcare librarian should be consulted to ensure the strategy remains appropriate (changes in databases require also the help of an information expert).
Planning the methods for an update	Authors are encouraged to consult current guidance on review methods and compare these with the methods used in the existing review to identify important changes.
Incorporating feedback and comments from reviewers	Comments may represent valid concerns and can usefully identify additional studies that were overlooked by the review authors
Conducting and reporting the update	• Authors should clearly alert readers that this is an update of an earlier version • The updated search should be presented in a PRISMA-type flow diagram.

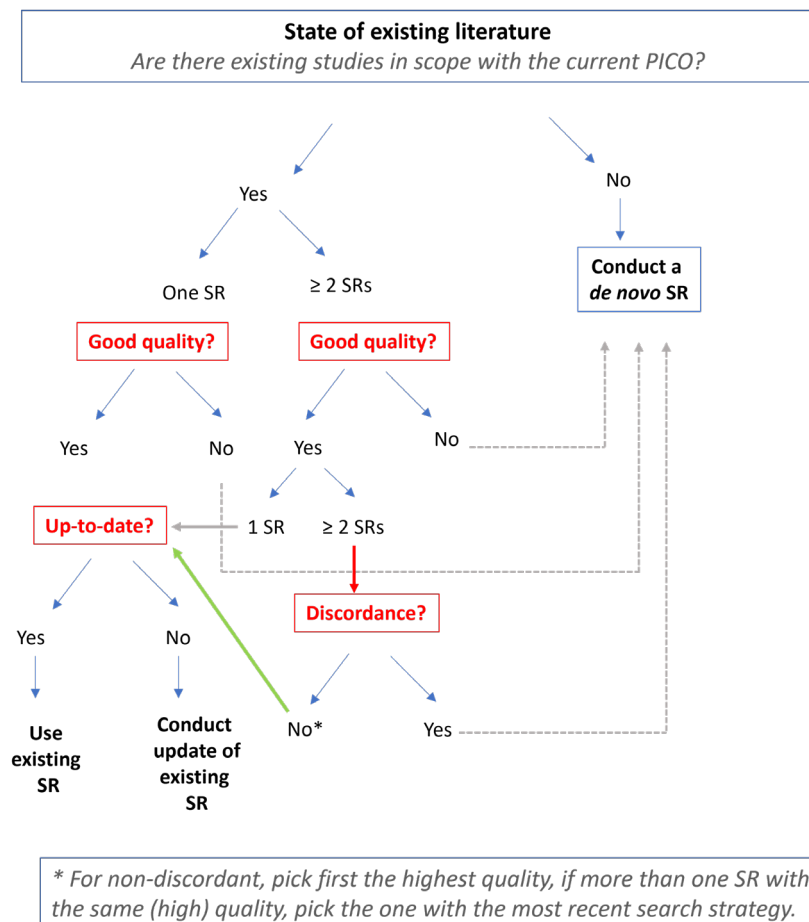
Source: Garner et al., *BMJ* 2016³⁹ and Cumpston et al. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023

4.8.4. Conclusions

An important number of the steps suggested in the literature and described in this section have already been used at KCE. However, some additional approaches could still be considered. In particular, the adapted OTTAWA method for assessing the need to update an existing SR under certain circumstances, should be discussed and explored internally (see Table 6., Table 8, and Figure 3 for more details)



Figure 3– Algorithm for the decision on when and how to update a systematic review



4.9. When is it appropriate to abridge or skip the quality appraisal of studies?

Key messages

- Appraising the quality of studies is a crucial element of any KCE projects, and standard tools to achieve this assessment are available for most study designs (see text).
- For expedited reviews, the following strategies are deemed appropriate:
 - One unique experienced reviewer does the critical appraisal, a second person verify the judgments.
 - The appraisal of a study stops when the risk of bias is considered “critical” in one of the assessed domains, even if the appraisal tool (e.g. ROBINS-I) is not complete.
 - For appraising systematic reviews, AMSTAR 2 is preferred to ROBIS.
 - For appraising observational studies, ROBINS-I is a complete but complex appraisal tool. A KCE-tailored grid, attempting to balance completeness with ease-of-use, is in development.

First of all, there is a conceptual difference between methodological **quality appraisal** (was the study well designed and executed, and thus likely to obtain reliable results?) and **risk of bias assessment** (what is the risk that the effect size of this study is biased?), although both concepts are often used interchangeably. Criteria for a good quality study and one at low risk of bias will often overlap, but not always. Depending on the goal of the appraisal, slightly different strategies could be employed. For example, for risk of bias assessment, each separate outcome must be assessed separately, as some outcomes may be at higher risk of bias than others (e.g. subjective versus objective outcome). If this is needed in the context of an expedited review, it is therefore important to focus only on critical outcomes or those most likely to inform public health decisions.²



4.9.1. *Strategies to decrease time investment for quality appraisal*

When performing an expedited review, quality and/or risk of bias appraisal of the included studies can be very time-consuming, especially for non-randomized studies. Some regulatory or advisory agencies make recommendations without performing a formal quality appraisal of all the included studies. Examples of this are the European Medicines Agency (marketing authorisation of drugs) and (in Belgium) the Superior Health Council. However, for KCE recommendations, we do not advise to skip all forms of quality appraisal, as basing recommendations on low-quality or biased evidence can lead to wrong decisions. Rather, a more time-efficient way of quality appraisal should be considered.

For a formal systematic review, two reviewers should score the quality of studies using the same appraisal tool.^{48, 49} One way of decreasing the time necessary to perform a review is to **limit the quality appraisal to one (experienced) reviewer**. The drawbacks associated with only one reviewer performing the literature search and the quality appraisal (increased risk of omitting important studies and of errors in the quality appraisal, respectively) are probably offset by the reduced time frame in which this is possible and should be considered acceptable when expedited review is called for. The Cochrane Methods Rapid Reviews group advise to have one person perform the assessment and a second person to verify the judgments (based on expert opinion).²

Another principle that could be applied when performing quality appraisal to decrease the time investment, is to stop the appraisal of a study when the risk of bias is considered “critical” or “high” (depending on the tool) in one of the assessed domains. When a Cochrane appraisal tool is used, an assessment of “critical risk” or “high risk” for one domain renders the whole study at high risk of bias, therefore appraisal of further domains will not change the overall appraisal. Such an approach is especially useful when all studies assessed at critical/high risk of bias will be excluded from the (meta-)analysis or when the primary aim is to provide a general quality assessment, without going into detail for each study. However, when the goal is to interpret for each individual study how bias could have skewed the parameter estimate one way or another, it is preferable to perform the assessment across all domains, as different biases may influence each

other in opposing directions. Furthermore, other appraisal tools often do not apply the concept of “critical risk” or do not consider the whole study at high risk when one domain is considered at risk, and this strategy can thus not be used when employing these tools.

Another strategy could be to pick an **easy-to-use and quick tool for assessment**. In general, shorter and easier tools will be less comprehensive than lengthier and elaborate tools; a compromise between ease-of-use and comprehensiveness should be aimed for when selecting an appraisal tool.

4.9.2. *Description of commonly used appraisal tools for different study types*

For **randomized controlled trials**, the Cochrane Risk of Bias 2 tool (RoB2) is widely used. This tool focusses on risk of bias in the randomization process, deviations from the intended interventions, missing outcome data, measurement of outcome and selection of the reported result.

For other study types, however, there is less consensus on the ideal tool to use.

For **meta-analyses and systematic reviews**, ROBIS (Risk Of Bias In Systematic reviews) and AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews) are commonly used.⁵⁰ In 2020, the Cochrane Collaboration launched a pilot version of the ROB-ME tool (Risk of Bias due to Missing Evidence), which was adapted to its current version in October 2023, aimed at assessment of meta-analyses. Of note, as the name suggests, only the possibility of missing evidence is considered as a bias in this tool.

One study reported similar inter-rater reliability of AMSTAR and ROBIS (slightly higher for AMSTAR) and similar reliability, but with a substantial higher time investment and expertise needed for ROBIS as compared to AMSTAR.⁵⁰ This finding would favour use of AMSTAR(2) for an expedited review. On the other hand, another study suggested AMSTAR and ROBIS should be used as complementary tools, as AMSTAR focuses on methodological quality whereas ROBIS focuses on risk of bias, which are related but not identical concepts.⁵¹ However, in the context of an expedited review, this is not considered feasible.



For **observational studies** as well, a multitude of tools have been proposed, including the Cochrane's ROBINS-I tool (Risk of Bias in non-randomized studies – interventional). ROBINS-I is intended for all non-randomized study designs, including quasi-experimental studies, cohort studies and case-control studies.⁵² However, ROBINS-I is, in line with other Cochrane bias assessment tools, very elaborate, lengthy and difficult to interpret for non-experienced reviewers. It is also worth noting that ROBINS-I is very strict in its assessment, as the score obtained by ROBINS-I is intended to be directly comparable to scores obtained by RoB2 of randomized controlled trials (i.e. an observational study assessed at low risk of bias with the ROBINS-I can be considered to offer equivalent level of evidence as a RCT assessed at low risk of bias).⁵³ In practice, this means the large majority of observational studies is at moderate risk of bias at best.

For these reasons, many systematic reviews/meta-analyses have chosen to use other tools to assess risk of bias in observational studies, such as the Newcastle-Ottawa scale. Of note, Cochrane collaboration also endorses the use of this scale when expertise in using ROBINS-I is lacking, thereby acknowledging the difficulties in using their own assessment tool.⁵² In the KCE Process Book^y, checklists for critical appraisal of observational studies (both for cohort and case-control studies) can also be found. A comparison of some of the most commonly used tools, including this KCE checklist, can be found in Appendix 1.1. An adapted KCE tool for quality appraisal, based on the most important bias domains in the identified appraisal checklists and using simple signalling questions, is currently being developed.

For other study types, assessment tools are also available, although there is often a lack of consensus on which tool to use and for some study designs no appropriate instruments for quality assessment are available at all. Below you can find a short discussion of available assessment tools based on the study type:

- **Clinical phase I studies:** these studies are often on small numbers of subjects and are non-randomized and non-blinded. As safety is a more important endpoint than efficacy, such studies should be interpreted to

only give information on dosing and safety. No specific quality appraisal tools for such studies could be retrieved.

- **Clinical phase II studies:** as these are usually randomized studies, the RoB-2 could be used. Alternatively (and faster), a so-called Key Methodological Score has been developed for such studies, using only three questions to assess study quality.^{54, 55} The score consists of the following items: the clear definition of a **primary endpoint**, the clear justification of the **number of patients (sample size)** included and the definition, a priori, of the **population** on which those criteria were evaluated.⁵⁵
- **Animal studies:** for animal studies, several instruments have been developed, including the STAIR, the CAMARADES and the SYRCLE instrument.⁵⁶ The SYRCLE instrument was established in 2008 and is based on the Cochrane RoB tool. It discusses selection bias, performance bias, detection bias, attrition bias and reporting bias, based on a 10-item checklist. This has become the most recommended tool for animal studies.^{56, 57}
- **Cell culture studies:** for *in vitro* studies on cell culture lines, there is a lack of recognized and validated instruments for quality assessment or risk of bias assessment.⁵⁸ A recent review identified the tools used for cell culture studies in 82 systematic & scoping reviews and meta-analyses; they found that 32 different tools had been used to assess quality. Some tools originated from another field but were adapted to apply to cell culture experimental studies. Most frequently, the SYRCLE tool (designed for animal studies) was adapted for use. Some authors developed their own tool, specifically for this purpose.⁵⁸ Many biases inherent to clinical studies do not pertain to *in vitro* studies (e.g. selection of participants); on the other hand, some biases are typical for *in vitro* and not clinical studies. For example, variations in the experimental conditions will cause bias, and must be interrogated in an assessment tool of cell culture studies. Based on the review of Barbosa

^y KCE website: <https://processbook.kce.be/>



et al, we propose a tool for assessment of in vitro studies in Appendix 1.2.

- **Diagnostic studies:** the QUADAS-2 checklist and QUADAS-C extension are recommended tools for evaluating diagnostic studies. Other instruments, such as the Cochrane risk of bias tool or JBI critical appraisal checklist for diagnostic test accuracy are based on the QUADAS tools.⁵⁶ No tools for expedited review were found.
- **Qualitative studies:** checklists for qualitative studies have been developed by the Critical Appraisal Skill Programme (CASP), Joanna Brigs Institute (JBI) and the Cabinet Office. Among these, the CASP checklist is the most frequently recommended tool.⁵⁶
- **Modelling studies:** assessment tools differ according to the type of modelling study. For diagnostic and prognostic prediction models, the PROBAST tool (Prediction model Risk of Bias Assessment Tool) can be used;^{29, 56} this tool includes the domains of participants, predictors, outcome and analysis. For each domain, signalling questions are provided. Also for prediction models, the CHARMS checklist (Checklist for Critical Appraisal and data extraction for systematic Reviews of Modelling studies) was developed. Of note, to facilitate use of these two tools, Fernandez-Felix et al. developed an Excel template to easily extract the information for both tools from included studies and automatically calculate the risk of bias.²⁹ The CASP clinical prediction rule checklist is suitable for studies using modelling to develop clinical prediction rules.⁵⁶

4.10. When is it appropriate to abridge or skip the GRADE appraisal?

Key messages

- Appraising the quality of evidence is a crucial element of any KCE review of interventions, and GRADE applies to each of them
- In ESRO projects, a number of adaptations can be considered:
 - The rating of certainty of evidence can be limited to the main intervention and comparator, and with a limited number of critical benefit and harm outcomes, for example the two most important outcomes for each.
 - The rating of quality can be done by only one reviewer who provides a clear rationale for each decision to up- or downrate the evidence quality. A second experienced reviewer should then check the rating decisions and their rationales.
 - If the ESRO review is based on the effect estimates of an existing well-conducted systematic review, the GRADE of that review can be used or be added if missing.

We consider that GRADE should always be performed as assessing the certainty (or quality) of evidence constitutes an essential piece of information to elaborate recommendations for decision-making. Even for documents with no recommendations, appraising the quality of evidence is relevant. However, when ESRO is required, Gartlehner *et al.* proposed the adaptations summarised in Table 5.⁵⁹

**Table 9 – Recommendations for rating the certainty of evidence in ESRO**

Recommendations to accelerate GRADE application	GRADE recommends a literature review or a Delphi-like approach involving knowledge users and people with the condition to rate outcomes that are most important for decision making. If a literature review or a formal Delphi approach are not feasible, use informal judgements of knowledge users, topic experts (including people who live with the condition), or internal team members to prioritise the outcomes to be graded.
	GRADE recommends rating the quality of evidence of the seven most important outcomes representing benefits and harms. To accelerate the process, consider rating fewer than seven outcomes .
	GRADE guidance recommends that two reviewers independently rate the quality of evidence and then agree on a final rating. To accelerate the process, consider using a single reviewer to rate the certainty of evidence, and verify all decisions (and footnoted rationales) by a second reviewer .
Recommendations when using an existing systematic review or network meta-analysis	If effect estimates of a well-conducted systematic review, meta-analysis, or network meta-analysis are incorporated to address parts of a key question of the ESRO review, consider using existing quality of evidence grades from such review .
	For network meta-analyses, GRADE recommends rating the quality of evidence for direct and indirect estimates separately. To accelerate the process, rate only the evidence quality of the direct estimate . If there is incoherence with the indirect estimate, further downgrade.
	If a network meta-analysis presents only indirect estimates, rate the quality of evidence and then further downgrade for indirectness.

Source: Gartlehner et al. 2023⁵⁹

- Restricting the number of outcomes. Because not all comparisons and outcomes are equally important for decision making and rating the quality of evidence can be labour-intensive, prioritising which comparisons and outcomes are most important for decision making is crucial for the efficiency of an ESRO review.⁵⁹ The rating of certainty of evidence can be limited to the main intervention and comparator, and with a limited number of critical benefits and harms outcomes. For example, if appropriate, outcomes for rating could be limited to the two most important outcomes reflecting benefits and the two most important outcomes reflecting harms. Another way that can help restrict the number of outcomes is to include only outcomes critical for decision making, and not outcomes rated as simply important for decision making. If a literature review or a Delphi approach to rate the importance of outcomes is not feasible, informal judgements of knowledge users, topic experts or other team members can be used.⁵⁹
- Limiting to one reviewer. GRADE guidance recommends that two reviewers independently assess the quality of evidence of each relevant outcome. In case of differences in judgements regarding the uprating or



downrating of individual domains that determine the quality of evidence, investigators need to achieve consensus by discussion or involve a third reviewer to obtain a consensus. For ESRO reviews, the rating of certainty can be done by only one reviewer who provides a clear rationale for each decision to up- or downrate the evidence quality. A second, senior reviewer with experience or formal training in rating evidence quality, should then check the rating decisions and their rationales.⁵⁹

- Recycling GRADE of existing systematic reviews. One approach to ESRO production is to include and, if necessary, update existing systematic reviews that address parts of a key research question (see section 4.7). If such a review is methodologically robust and used GRADE, results and evidence quality can be considered for use. If the existing well-conducted review did not use GRADE, GRADE can be implemented based on the reported results. If literature searches of an existing review need to be updated, grades of evidence quality might need to be revised.⁵⁹ Special considerations on how to deal with direct and indirect results of network meta-analysis are explained in the paper by Gartlehner *et al.* (see Table 6).⁵⁹

5. EXPEDITED INTERNATIONAL COMPARISON

Key messages

- A stand-alone international comparison should be reserved for interventions for which previous research has already shown that they work.
- The difference between an expedited international comparison and a “standard” international comparison can be related to the scope of the comparison, the number of countries included in the comparison, the method(s) to collect information or in reporting the results.

In this section we describe some questions that are relevant in the context of an expedited international comparison. Nevertheless, the distinction between a well-scoped international comparison and an expedited international comparison is somewhat vague. Therefore, it might be worthwhile to apply the principles that are described below when a classic international comparison is conducted as well (cf. KCE process notes^z).

5.1. Can an international comparison be a stand-alone KCE product?

In principle, an expedited international comparison can be part of a KCE product/report which also includes other methods to answer the research question(s), or it can be a stand-alone product.

^z https://processbook.kce.be/sites/default/files/2021-11/2015-72%28Method%29_HSR%20Process%20Notes%20Literature%20review%20and%20international%20comparison_Report_0.pdf



Under which conditions can an expedited international comparison be a stand-alone product?

The difference between an expedited international comparison and a “standard” international comparison can be related to the scope of the comparison (section 5.2), the number of countries included in the comparison (section 5.3), the method(s) to collect information (section 5.4) or in reporting the results (section 5.5). However, whatever the answer to the questions asked in the next paragraphs, it has to be decided under which conditions an expedited international comparison can be a stand-alone product.

An international comparison is typically one of the research methods in health services research, alongside other methods such as a literature review, data analysis or qualitative methods. In such a mixed-methods context, the research question(s) and policy recommendations are considered from various complementary perspectives. A stand-alone international comparison does not allow such multi-faceted perspective. Therefore, a stand-alone international comparison should be reserved for interventions for which previous research has already shown that they work. This previous research can, for example, be a Cochrane systematic review or a KCE report. The international comparison could then focus on, for instance, implementation issues. We can illustrate this with an example. Several KCE reports recommended the concentration of complex surgical procedures for cancers because of a volume-outcome relationship. Given the available evidence provided in these reports, a stand-alone expedited international comparison could be conducted to decide on implementation conditions (e.g. minimum composition of multidisciplinary teams in reference centres; geographical accessibility). Consequently, recommendations can also be formulated in such case.

In the absence of such review or report, an expedited international comparison is in itself insufficient to formulate recommendations. The reason is obviously that health care system organisation differs between countries, and that conclusions arising from evaluation of foreign health care systems are therefore not necessarily transferable to Belgium. It can, however, be valuable to provide policymakers with an overview of international practice examples.

5.2. How to limit the scope of the comparison?

In comparison to a classical international comparison, the scope of an expedited comparison should be even more focussed and limited to the main points than it is the case in a classical international comparison. In order to do so, the following approaches could be followed:

- Organising a scoping meeting before the start of the international comparison with main experts in the field, preferably experts with international contacts.
- Try to identify in the scoping meeting the (organisational) aspects and questions that are central to the research questions, those that (could) have a clear impact on the studied disease or health service, and those that possibly differ between the countries included in the international comparison.
- It should be made explicit that with the international comparison we will not be able to address all possible interesting questions about the subject and that for sake of time, only a limited number of questions will be addressed.
- A distinction could be made between “nice to have” and “must have” questions. If information is easy to find and process, the “nice to have” questions can also be addressed, but they should not take up excessive amounts of time.
- If only few countries are studied (see section 5.3), the choice of those countries is important in order to give the maximum amount of comparative information. At least one of the chosen countries should have a similar healthcare structure to Belgium but with differences regarding the organisation of care that is studied (as those organisational aspects would in theory be relatively easy to implement in a Belgian setting). Neighbouring countries are usually interesting as they have similar population health profiles and comparable access to healthcare.

The scoping rationale and process must be clearly explained in the research protocol (see section 3.3).



5.3. Is there a minimum of countries/regions to be studied? Is there a maximum number?

The decision about the number of countries is secondary to the decision about the scope and research questions of the international comparison. It is based on pre-defined criteria and should include at least one country with a similar healthcare structure as Belgium (see section 5.2). As a rule of thumb an international comparison includes four or five countries and the number should be limited (e.g. not more than 8-10 countries). It is important to have two countries (or regions) as absolute minimum to enable a meaningful comparison with the Belgian situation.

5.4. Which are the preferred methods to collect data in case of an expedited international comparison?

For an expedited international comparison, less time-consuming sources (e.g. reports, key informants) are preferred over more time-consuming sources of information such as questionnaires, surveys, data registry, etc.

An adapted process could combine a (non-systematic) review of scientific literature, grey literature, laws and regulations, and relevant websites with implication of key informants. The following sequence is suggested:

- Possibly, first contact with key informants: to obtain first information or documents before starting a literature review
- Review of scientific literature, grey literature, laws and regulations, and relevant websites
- Contact with key informants:
 - provide additional information or documents

- provide further explanation on information retrieved from existing documents (sometimes in their own language)
- ‘Validation’ by key informants who:
 - verify that no important information was missed
 - ‘validate’ the final country chapter or transversal analysis and comparative tables

Depending on the topic and the access to international key informants, variants of the above approach need to be considered. The data gathering, for instance, can be focused on contacting key informants. In that case it is recommended to ask at least 2 key informants per country to fill out a questionnaire (e.g. on organisational aspects, implementation strategy payment models). The results of this first round can be compared to identify differences and similarities in responses that need to be assessed in a second round with the key informants. As an example, this approach was implemented for the KCE report on the home-based management of obstructive sleep apnoea.^{aa}

Site visits can be considered if they allow to accelerate the process (e.g. so that key informants can demonstrate how an organisational model works in practice and at the same time being interviewed to gather additional data about, for instance, facilitators and barriers of the model, how it was implemented elsewhere, etc.). This depends on the topic (i.e. if there is something to ‘demonstrate’).

Participation in international conferences can also be an option to accelerate the process of data gathering (i.e. several key informants at the same place) but only if the timing is consistent with the project.

^{aa} KCE website: <https://kce.fgov.be/en/publications/all-reports/organisation-of-diagnosis-and-treatment-of-obstructive-sleep-apnoea-syndrome-an-international>



5.5. Do we need to report information country-by-country or can we only report a transversal analysis?

An expedited international comparison will require an efficient data extraction and reporting method. During the project planning phase it is important to make a choice and assess if it fits the team composition and task distribution.

- In case the international comparison is carried out by many different KCE experts (e.g. one country per expert) it is important to define a template that describes which results will be presented and how they will be presented. Ideally, the template is built when the first country is analysed, but a theoretical template could be built in advance and adapted after the first countries are analysed. Each country report can be validated by a country expert. Based on the country reports a transversal analysis can be made.
- In case one or two KCE experts carry out the entire international comparison it might be more efficient to limit the chapter in the report to a transversal analysis only. This approach will require a data extraction template that includes the same themes that will be treated in the transversal analysis chapter. Yet, in contrast with the 'country-by-country' approach the data extraction templates per country will not include written text. It is limited to bullet points based on contacts with key informants or publications. This approach can only be followed when the KCE experts work on the international comparison in a consecutive period (i.e., periods of interruptions will cause recall bias and include a risk that experts have to go back to the initial sources of information). A transversal analysis can be submitted to country experts for validation with the request to review if the information in this chapter for the respective country is correct.

Starting from the process note on international comparisons, where could the process be accelerated?

**Table 10 – Steps for an accelerated international comparison**

Step	Comment
Preparatory work	
Explore the topic by performing a horizon scan	It is possible to rely on a consultation of Belgian experts with a broad international network within the domain to gather the necessary information for the horizon scan.
Defining the topic should be integrated in this exploration phase	For expedited international comparison, the topic should be well defined at the beginning, in collaboration with the commissioner. It should be made explicit from the start on which 'must have' questions the international comparison will focus.
Evaluate the anticipated amount of information	
Discuss which aspects of the phenomenon you want to analyse	Be precise on which questions you will try to answer and stick to it
Determine at which level the comparison should be done	For an expedited international comparison, it is unlikely that the focus will be on an entire health system or an entire sector. It is more likely to focus on a domain-based comparison or an intervention/technology-based comparison.
Assess if there is a need to (re)phrase the research question(s).	
Reflect on other (broader) aspects related to the primary question	Could be skipped – focus only on one (well defined) RQ
Evaluate how many and which (type of) countries	The rule of thumb for an expedited international comparison is to select between 2 and 5 countries or regions with at least one country with similar characteristics as Belgium.
Determine which sources of information you want to use and which method(s)	Preferrable methods that are not time consuming (e.g. reports, key informants)
Selection of countries	
Define criteria for selection	At least one of the chosen countries should have a similar healthcare structure to Belgium but with differences regarding the organisation of care that is studied
Draft a longlist of countries	When the consultation of Belgian experts results in 2-5 suggested countries, it can be considered to immediately define the shortlist of countries that will be studied.
Reduce to a shortlist	
Collecting the evidence	
Peer-reviewed scientific literature	
Grey literature	Check for available information in previous KCE reports



Law and regulations	Check for available information in previous KCE reports
Websites	
Key informants	Questionnaires/surveys should, by default, not be considered for expedited international comparisons. However, interviews or open-ended questionnaires with key informants could be a key to accelerate the process. Another option is that the KCE experts complete the country profiles to the best of their knowledge and then send it out to country representatives to amend, and respond to specific questions, etc.
Data	Should not be considered for expedited international comparisons.
Site visits	Should not be considered for expedited international comparisons unless it will clearly speed up the process (e.g. opportunity to interview several key informants and learn how the system in a country works by seeing a 'best-practice' example in real life)
International conferences	Could be an option to accelerate the process (gathers all key informants at the same place) but only if timing is consistent.
Data extraction and presentation	
Analysis matrix on a couple of countries	Consider starting with a theoretical matrix
Adapt analysis matrix	
Transversal analysis (comparative tables)	Can be considered as the reporting method when there are a limited number of experts working on the international comparison.
Validation	
With key informants	Could replace full validation process



6. RAPID QUALITATIVE RESEARCH

RQR are qualitative study conducted over a short timeframe (from a few weeks to a few months); with a participatory design (representatives of local populations or institutions are involved in the research), an early involvement of stakeholders in the process to create ownership; a combination of multiple research methods and triangulated data during data analysis; a parallel data collection and analysis, are team-based (i.e., researchers work collaboratively on all aspects of the research process); and need a clear focus (i.e., time is invested in a thorough scoping phase, to ensure the whole team agrees on the RQ's, the study design and dissemination plan).

Can a RQR project be a stand-alone KCE product?

RQR should not be a stand-alone KCE product. Mixed methods approach should be preferred to obtain robust and acceptable recommendations.

How to better focus the scope of RQR?

Close collaboration with stakeholders helps to delineate the scope of the study and then contribute to RQR. Scoping on the topic and on the methods is an essential first step in RQR, this is already in use in (qualitative) studies at KCE.

Are there preexisting data available for secondary analysis?

Secondary data can help, if there are sufficient information about their data collection process, their quality, and the possible biases.

Is it possible to expedite the collection of primary data?

Is there a-priori theoretical framework applicable for the data collection and the analysis?

An a-priori framework can help to expedite the research as it frames the data collection and analysis. Including experts knowledgeable with the topics in the research team can help to expedite the data collection and

analysis but the experts should question themselves about their knowledge and remain open to alternative / new perspectives on the topic.

Can obtaining the ethical clearance be expedited?

Ethical approvals remain under the control of the ethics committees. One can only expedite this process by selecting an ethical committee reputed for its rapidity / with several meetings per month according to an agenda accessible to the researchers.

Is there a possibility to collect data through proxies?

Proxies can be used if they reflect the diversity of the investigated groups and are sufficiently aware of the major challenges occurring on the field.

Is it possible to reduce the number of study participants?

The number of participants can be reduced if the study uses purposive sampling and if a progressive analysis is done during data collection to ensure that the data collected reach saturation as much as possible (without aiming at perfect saturation).

Is it possible to accelerate the recruitment process?

The recruitment process can be accelerated by using natural groups and informal focus groups if the researchers have the possibility to visit study settings: this is therefore limited to some topics (e.g. experiences in attending a specific health service or working behaviours in a chosen health care unit). It can also be helped through the collaboration with the stakeholders who can act as intermediary between the researchers and the participants. This can therefore lead to a selection bias, an influence bias or social desirability bias.

Is there a possibility to restrict the languages used?

No, selecting a single language for the data collection is not recommended and can lead to biased outcomes as the spoken language is also a proxy for relevant contextual variables.

***Which are the most preferred data collection techniques in case of RQR?***

Focus groups are longer and require more participants than interviews. Opting for the second one could accelerate data collection process but to the detriment of the richness due to interaction between participants. Online data collection in general population could be a useful mean to accelerate data collection, but it is important to keep the risk of selection bias in mind, particularly as regards less digitalized stakeholders.

Is it possible to expedite the design and validation of instruments for data collection?

This can be achieved if the KCE develops a common database of validated interview guides/typical questions that can be used in an interview guide/focus group guide (in Dutch and French to reduce the need for translation). Guides tend to follow the same structure and some questions are not specific to a topic (e.g. ice-breaker questions, questions reflecting a SWOT, conclusion questions...).

Is it possible to skip the full transcriptions of interviews and focus groups?

Transcription can partly be skipped for rapid research, and it should be possible to work from field notes, RAP sheets, or partial transcription depending on the team, the subject and the type of results that will be disseminated. Direct transcript should be avoided in RQR and be done only if the researcher has a specific expectation of sharing this "raw" material.

Is a validation of the results by study participants needed?

Validation of results per participant can be skipped as it is time consuming and other efficient triangulation techniques can be preferred.

Is it possible to expedite the analytical process?***Which level of analysis is expected at minimal?***

Due to the resources dedicated to the qualitative part of a research project for the KCE studies, the expected level of analysis is at minimum descriptive.

Is it possible to rely on one single researcher for the analytic process?

One single researcher could be sufficient if there are possibilities for the team to give their feedback on the interpretation during regular team meetings, to ensure the validity of the analysis. It is also wise to allow the qualitative researchers to concentrate on this study during data collection, analysis, and reporting. It is also advised to endorse a team member with a cross-checker role (responsible for going through the data and to identify inconsistencies and bringing them up in team meetings) to avoid this bias.

Can the use of software help speed up the analysis?

Analysis could be fastened using a software but cannot be replaced by AI-performed analysis.

Is it possible to skip or abridge the triangulation of sources?

Triangulation is always required as its part of the validity of the qualitative research and should be achieved by triangulation of sources.

Is it possible to skip or abridge the process of reflexivity?

Reflexivity is a validity criterion of qualitative research and should be always integrated in a RQR. Participation to the meetings of the qualitative cell can help the researchers involved in RQR but they need to have sufficient time to attend it.



Rapid Qualitative Research (RQR) is not new but has been developing over the past decades. Some scholars see it as a response to the criticism on social scientists (and by extension all researchers relying on qualitative approach) for running behind the facts and producing research results too slow (especially in emergency situations).

A 2021 literature review highlighted the following reasons for the use of rapid qualitative methods: time savings, cost reduction, enhanced data quantity, improved efficiency, heightened accuracy, and the generation of a more nuanced interpretation of participants' realities.⁶⁰

In other words, RQR aims at producing information in a brief time span without losing rigor, nor quality, as publications seem to confirm this: quick, but not dirty! Indeed, the first concern when it comes to rapid (qualitative) research is often about quality and rigour.⁶¹

The following sections review the process of qualitative research and identify how, and under which conditions the process can be accelerated, while not sacrificing the quality and the rigour. Therefore, we reviewed the literature (Medline, Embase and Sociological Abstract) and the websites of several research institutes (and related journals) known for doing qualitative research to identify criteria to decide when RQR is feasible, and how it can be performed.^{bb} Twenty-six papers were identified (see references).

For further details on qualitative methods at KCE, readers can consult the KCE process book on qualitative methods (KCE report 187).^{cc}

^{bb} Association pour la Recherche Qualitative (ARQ), International Institute for Qualitative Methodology, International Journal of Qualitative Methods, National Centre for Social Research, Rapid Research Evaluation and

6.1. What is meant by RQR?

6.1.1. Definition

There is no consensus on what is meant by 'rapid' qualitative research. There is considerable debate in the field concerning the definition of RQR, and this is mostly related to the timeframe, ranging from a few days to a few months and the applied methods.⁶¹⁻⁶³ Vindrola-Padros mentions the definition of Beebe (2014) who defines it as "*research designed to address the need for cost-effective and timely results in rapidly changing situations*" (Beebe 2014, cited by Vindrola-Padros, p 3).⁶¹ Furthermore, while maintaining citation of Beebe (2014), Vindrola-Padros reports that a rapid qualitative inquiry is an "*intensive, team-based qualitative inquiry with a focus on the insider's or emic perspective, using multiple sources and triangulation and using iterative data analysis and additional data collection to quickly develop a preliminary understanding of a situation*" (Vindrola-Padros, 2021, p.2).⁶¹ She also provides her own definition, stressing that "*rapid qualitative research is considered as qualitative research that lasts anywhere from a few days to six months or might have a longer timeframe, but include multiple short and intensive stages of data collection and analysis, as in the case of rapid feedback evaluations and rapid cycle evaluations*" (Vindrola-Padros et al., 2021, p. 3).⁶⁰

Referring to the work of McNall on Rapid Evaluation and Appraisal Methods (REAM), Vindrola-Padros pointed out the following common features in RQR:⁶⁴

- the study is conducted over a **short timeframe** (from a few weeks to a few months);
- the study design tends to be **participatory** (representatives of local populations or institutions are involved in the research) and stakeholders are involved early in the process to create ownership;

Appraisal Lab (RREAL), National Institute for Health and Care Excellence (NICE), Haute Autorité de Santé (HAS)

^{cc} See more: <https://processbook.kce.be/qualitative-data-gcp-hsr-hta>



- the studies combine **multiple research methods and triangulated data** during data analysis;
- **data collection and analysis** tend to be carried out **in parallel** and preliminary findings are used to decide on additional data collection;
- the studies are **team-based** (i.e., researchers work collaboratively on all aspects of the research process);
- the studies need a **clear focus** (i.e., time is invested in a thorough scoping phase, to ensure the whole team agrees on the RQ's, the study design and dissemination plan).

Often rapid methods are characterised by rapid feedback loops so that newly emerging insights can quickly inform decision makers. This means that often several research cycles succeed one another.

6.1.2. Methods used in RQR

From our literature review, it appears that frequently rapid approaches use both quantitative and qualitative data collection methods, while this is probably not an intrinsic feature of RQR but to be linked to contexts where they apply most often, e.g. development or medical-oriented projects/evaluation. Assessment/appraisal approaches typically focus on assessing a situation (e.g. to explore or to design an intervention), while evaluation approaches tend to focus on projects, programs, or systems.

The field of RQR is diverse and various design approaches have been developed. Over the last two decades, several methods of rapid evaluation, assessment, and appraisal have emerged. There has been cross-fertilisation between approaches, making them almost indistinguishable. It is more useful to look at the set of core principles and methods than to get lost in different approaches and names. Although the origins and contexts of practice of these approaches might be different, they share the idea of making available trustworthy and actionable information for decision makers in a brief time delay. They try to find a balance between speed and trustworthiness.

Rapid evaluation and assessment methods require an appropriate context of application. Typically, there is an urgent demand for information based on which action will be taken. Hence the context is characterized by a certain sense of urgency. Rapid evaluation and assessment methods do not aim at contributing to an existing body of theory, but to provide decision makers with information of sufficient quality to improve their decision making and subsequently the effectiveness of actions taken. Hence rapid evaluation and assessment methods are oriented towards action.

Table 10 presents an overview of a selection of RQR approaches, including advice for the KCE. It should be noted that additional methods exist but need to be further investigated in the context of qualitative research at KCE (e.g. rapid ethnographies or Rapid Assessment, Response and Evaluation (RARE)).



Table 11 – Overview of a selection of RQR approaches

Method	Description	Characteristics of team	Data collection	Notes for KCE
Real-time evaluations (RTE) ⁶⁵⁻⁶⁷	Method to evaluate rapidly the effectiveness and impact of responses to emergencies. Observations and recommendations are shared on an ongoing basis with decision makers.	Evaluators might be part of the crisis response cell.	Systematic collection and revision of data as the crisis unfolds. Use of a mixed-methods approach, (e.g. semi structured interviews, site-visits, in-depth interview, focus groups, use of quantitative databases such as patient registries, reviews of documents).	Can be used for understanding behaviours of patients regarding a new public health measure (e.g. vaccinations) leading to adaptation in the communication by competent authorities. Like the methods used in the Taskforce Therapeutics or the approach for <i>ad hoc</i> questions.
Rapid-feedback evaluation (RFE) ⁶⁸⁻⁷⁰	Aim at providing quick answers to focused questions, e.g. about the performance of a certain program or intervention. Feedback needed to decide how to correct a problem within the program or the intervention.	Can be used by a mixed method team (not only qualitative).	Consists of collecting, analysing, and evaluating data on the performance of a program or intervention.	Appears not limited to qualitative methods. Like the current studies at KCE.
Rapid assessment (RA) ^{71, 72}	Wants to quickly generate information to inform decision making. Is pragmatic and action-oriented.	Mixed method team can be useful.	Use of small samples of key informants to gather information in a rapid and pragmatic way.	Example of KCE report on CBRNe where scoping interviews were used. ⁷³
Rapid ethnographic assessment (REA) ⁷⁴⁻⁷⁶	Aim at providing rapid assessments of local conditions to inform the design of interventions.	Mixed method team can be useful.	Use a limited range of research methods (e.g. key informant interviews, questionnaires, and existing data sets).	Example at KCE: Site visits in HSR studies.
Rapid evaluation methods (REM) ⁷⁷	Aim at providing quick, accurate and economical methods of health services and patient satisfaction, developed by the WHO. ⁷⁷	Mixed method team can be useful.	Use of a wide range of data collection techniques, e.g. clinic exit interviews, health staff interviews, observations focus group discussions, review of patient records	Used to have baseline data needed for an analysis of a defined situation ⁷⁷ To date this method has never been used at KCE

This list is non-exhaustive, other methods exist and can be used in rapid qualitative research.



6.2. Can a RQR project be a stand-alone KCE product?

Referring to the literature on RQR, Vindrola-Padros identified that RQR can be used: “(1) as a preliminary study used to inform longer-term research (Coreil et al., 1989), (2) as a short study carried out to explore new questions that might have remained or emerged after the longer study ended (Munoz-Plaza et al. 2016, (3) in parallel to a longer study, potentially using rapid feedback loops or cycles to inform the development of the longer study (Shrank 2013), (4) as a study on its own (Beebe 2014)” (Vindrola-Padros, 2021, p. 5).⁶¹ The last option is the one we will develop here.

In their systematic review on RQR methods during complex health emergencies, Johnson *et al.* retrieved 13 articles about qualitative studies and 9 about mixed-methods design.⁷⁸ This shows that a stand-alone rapid qualitative study is possible. Among the stand-alone qualitative studies, Johnson *et al.* found a study on the systematic assessment of the national health system capacity to respond to pandemic influenza⁷⁹, a study on the assessment of Ebola virus disease survivor needs in Sierra-Leone⁸⁰ or a study aiming at identifying the sociocultural determinants related to community resistance in Guinea after Ebola outbreak.⁸¹

Until now, qualitative research at KCE has always been part of mixed-method studies. Qualitative part of a KCE study contributes to the triangulation of other data by adding elements of understanding of the perceptions, meanings, and experiences of the stakeholders. If little is known about the subject, or if it is an emerging phenomenon such as new health behaviours, for example, a single exploratory qualitative study may be envisaged. However, to carry out a rapid study, it is important to have a good knowledge of the subject and to be able to refer to a theoretical framework (see section 1.5.1). Furthermore, since qualitative approaches are already easily accused of subjectivity, the risk of basing recommendations solely on a qualitative approach is that they are only partially founded.

RQR should not be a stand-alone KCE product. Mixed methods approach should be preferred to obtain robust and acceptable recommendations.

6.3. How to better focus the scope of RQR?

6.3.1. Scoping the topic

Vindrola-Padros *et al.* and Horwood *et al.* suggested actively engaging with stakeholders to better scope the study topic, a practice already in use at KCE. A close collaboration helps the research team to be flexible and refine the research questions while considering the dynamics of the context. Further information about the stakeholder involvement can be found in the [KCE report 174C](#).^{dd}

Animation techniques to facilitate and expedite the discussions with stakeholders can be found in the supplement of the KCE report 340 bis.^{ee}

Close collaboration with stakeholders helps to delineate the scope of the study and then contribute to Rapid Qualitative Research.

6.3.2. Scoping the methods

The methodology applied can be focused, for example by limiting the scope of the interview to focus on a few topics. Targeted interviews are normally short in duration, which allows researchers to cover more participants or to shorten the data collection phase.⁶¹

Most of the qualitative data collection concern the patient experiences regarding a specific issue (via interviews or group discussions). Other frequent qualitative data collection at KCE concern group discussions and focus groups with professionals and stakeholders around a specific issue. The willingness to better involve patients as part of the research process (although not limited to the qualitative dimension) also contributes to the

^{dd} For KCE internal use: an up-to-date list of stakeholders is provided via the project guide available on the [S:\Project Guide](#)

^{ee} https://kce.fgov.be/sites/default/files/2021-11/KCE_340bis_Process_Note_animation_techniques_0.pdf



growing need for the development of skills inherent to qualitative methods such as active listening, moderation, and group animation.

Like other research methods applied at KCE, KCE qualitative studies tend to be pragmatic, relying on a brief period for data collection, favouring thematic analysis rather than in-depth analysis as in grounded theory^{ff}. In that sense, comparing to existing practices in academic research, qualitative parts of KCE studies already satisfy to the principles of RQR.

Scoping on the topic and on the methods is an essential first step in RQR, this is already in use in (qualitative) studies at KCE.

6.4. Are there preexisting data available for secondary analysis?

6.4.1. Availability of relevant scientific literature?

If there is literature available and of good quality, it could be sufficient to analyse it to answer qualitative questions raised in the study (see chapter 4 about how to expedite a literature review).

Like other literature reviews, one always needs to question the possible extrapolation to the Belgian context: it requires to select qualitative studies with sufficient quality and with enough elements to understand the context and the target groups. One may think about selecting qualitative studies conducted in similar cultural, religious, or socioeconomic contexts (e.g. limiting to neighbouring countries of Belgium). Checklists for the quality appraisal are available in the KCE report 187.⁸²

6.4.2. Availability of relevant and robust preexisting data?

Obviously, using existing data to perform analyses will result in a precious saving of time. It is always advisable to first look for existing data and analyse it before undertaking any new data collection. Previous KCE studies can be a source of data (e.g. transcript of interviews conducted with

patients). It can be useful to search in online data sources such as blogs, social networks (Instagram, Facebook...) or online forum.⁸³ These are public data and easy to access. This strategy was previously used in the KCE report 364 on Post-intensive Care Syndrome (PICS) for general practitioners to collect patient perspectives.⁸⁴

It is worth noting that secondary data, whether collected for a specific research question or available through sources like media testimonials or newspapers, may not perfectly address the actual research question. Moreover, even if it is not specific to secondary data or RQR, it requires a careful appraisal of the quality, the validity, the trustworthiness and the relevance of the data before using.⁸⁵ This will be further developed in an update of the KCE report 187. Researchers also have a limited control of the design, and the data collection methods that were used. It must also require informing the participants, when signing informed consent, that the data can be reused in other studies. This aspect should be further discussed as part of the ongoing discussion on ethics at KCE.

On the practical side, reusing previous interviews or focus groups conducted by KCE researchers requires that the data need to be fully transcribed to facilitate the identification of the relevant material. It also implies that the researchers complete each transcript with their field notes to provide information about the context of the interview / focus group, non-verbal clues, observations, etc.

The recommendation to first search for existing data applies to all research, for pragmatic and ethical reasons (i.e., unnecessary solicitation of people), regardless of the urgency of the study. Depending on what exists and its applicability to the Belgian context, an original data collection should be carried out.

Secondary data can help, if there are sufficient information about their data collection process, their quality and the possible biases.

^{ff} See also the KCE report 187: <https://processbook.kce.be/qualitative-data-gcp-hsr-hta>



6.5. Is it possible to expedite the collection of primary data?

When it is not possible to restrict the response to qualitative research questions using literature or secondary data collection, it will be necessary to collect new data. We will go through the whole process to identify the steps that can be suppressed or alleviated to accelerate the research.

6.5.1. *Is there a-priori theoretical framework applicable for the data collection and the analysis?*

Not all qualitative researches start from a theoretical framework. However, at the KCE, as in many rapid qualitative studies, relying on a theoretical framework linked to the topic can be useful because it saves time. A good knowledge and understanding of the topics are also an asset at this stage.

Indeed, a theoretical framework provides a structure for conceptualizing and conducting the study, guiding every step from the formulation of research questions to the interpretation of findings. The framework helps ensure that the research is grounded in relevant theory and contributes to the advancement of knowledge in the field, in constructing the interview guide and in analysing and interpreting the results. For example, in the KCE study on the prevention of Post-Intensive Care Syndrome, the study mobilised the Theoretical Domains Framework (TDF) that synthesizes constructs from 33 behaviour change theories into 14 “construct domains,” or clusters of related constructs that may explain practice change or the absence of change.⁸⁶

Vindrola-Padros therefore stressed the need for devoting sufficient time for the theoretical grounding: “lack of theoretical grounding of rapid qualitative research due to time pressures and the provision of more actionable findings, leads to the production of research that is instrumental, atheoretical and without a critical angle” (Vindrola-Padros, 2021, p. 7).⁶¹

An a-priori framework can help to expedite the research as it frames the data collection and analysis.

Including experts knowledgeable with the topics in the research team can help to expedite the data collection and analysis but the experts should question themselves about their knowledge and remain open to alternative / new perspectives on the topic.

6.5.2. *Can obtaining the ethical clearance be expedited?*

The approval from an ethics committee is compulsory when the study addresses health-related information (e.g. diagnosis, treatment, symptoms), individual and personal experience as patient or user of the health care system.⁹⁹

This phase can be particularly time-consuming due to the extensive preparation required for submission files and the subsequent delays in the review process by the committee. It frequently poses challenges for (KCE) researchers especially regarding the appropriateness of submitting qualitative research protocols to current ethical committees. This uncertainty arises because the procedures in place often do not align with the specific needs of qualitative research, as they are typically modelled after criteria more suited to clinical research assessments. If a research team is aiming for rapid research, the procedure for evaluating the study protocol by an ethics committee seems to be the main bottleneck.⁸⁷

Ethical approvals remain under the control of the ethics committees. One can only expedite this process by selecting an ethical committee reputed for its rapidity / with several meetings per month according to an agenda accessible to the researchers.

6.5.2.1. *Possibilities of fast tracks approval*

During the COVID crisis, some countries introduced a fast-track system for approving studies relating to COVID-19, as was the case in the UK, for example. This procedure reduces the approval to a few weeks instead of a few months, but ethical review still represented an important bureaucratic burden because the documentation required for the applications is not

⁹⁹ The scope of application varies from country to country, e.g. in Canada, you need one for surveys with professionals.



reduced.⁸⁸ A fast-track process does not compromise the depth of the evaluation. Instead, it enabled ethics committees to prioritise applications with a distinct focus on addressing challenges arising from the pandemic, giving priority to such studies over those centred on different subjects.⁸⁹ This was the case for Belgium during COVID crisis.

Vindrola-Padros *et al.* reported that some researchers suggest establishing separate ethical review processes for research that is deemed to be time-sensitive, e.g. for research on emergencies.⁸⁷ This ethics committee requires a combination of speed, depth, and proportionality. *Médecins Sans Frontières* (MSF) for example has established its own ethical review board with an expedited review process if the research poses minimal risks to participants in the human category. This includes descriptive studies that involve monitoring and evaluation to test a new approach, social science research in health and health systems, and prevalence and incidence studies.⁹⁰

A fast track system of approval should be discussed to allow RQR with individual patients but this is not limited to the KCE. Additionally, when a contact point exists (before the secretary of the ethics committee), the research team can explain them the degree of urgency.

6.5.2.2. Possibilities to shorten the preparation of the submission

Independently from the delays imposed by the ethics committee, it should be possible to gain some time in the preparation of the file to submit building a database of consent forms/information forms to start with a base from similar type of research and just customise it according to the study and to make the messages more user-friendly for the participants when necessary. Copies of all KCE dossiers submitted to ethics committees, including the comments received and the revisions done, can be centralised to have multiple examples available and to prepare the ethical applications as quickly and accurately as possible.

^{hh} Sensitive data according to European Commission: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en

6.5.3. Is there a possibility to collect data through proxies?

To facilitate prompt data collection, one may explore obtaining data from patient representatives or healthcare practitioners on “non personal” topics or on “non sensitive data”^{hh} (e.g. experiences in each service or perceptions about problems encountered in general in the health care system). Not all studies require collecting data reflecting individual patient experiences. This will be determined by the topic of the study (see, e.g. the KCE study 2022-02 on palliative middle care – still ongoing now of writing). The expected outcome also differ: patient representatives will provide experiential expertise and collective knowledge through their own experiences and exchanges with multiple other patients with similar experiences. One should not forget that patient associations cannot provide information about all issues experienced by those using the health care system (e.g. there is no patient association of prisoners).

Researchers should pay attention to the bias that can occurⁱⁱ. As recalled by Vindrola-Padros, “due to time pressures, researchers will recruit participants or observe situations that are easily accessible. This might mean that views of some participants will feature more prominently or that a multiplicity of voices will be lost. Rapid qualitative researchers have tried to address this potential challenge, by designing sampling guides where researchers map the different groups of participants they would like to include in the study and set out sampling targets for each” (Vindrola-Padros, 2021, p. 7).⁶¹ Therefore researchers should ensure that these representatives be likely to reflect the diversity of the professional profiles and still have a connection with those active on the field.

Ethical concerns may be less apparent in these cases since individuals may be less inclined to disclose information about their own health, and time could also be gained because asking for an ethical clearance would not be necessary. However, this needs to be clarified by the ongoing KCE working group on ethical issues.

ⁱⁱ See also the KCE report 187: <https://processbook.kce.be/qualitative-data-gcp-hsr-hta>



Proxies can be used if they reflect the diversity of the investigated groups and are sufficiently aware of the major challenges occurring on the field.

6.5.4. *Is it possible to reduce the number of study participants?*

Sample size in qualitative research is determined by the principle of saturation. In practice, it implies that conducting an additional interview or focus group will not generate additional relevant data. Among the different RQR found, the number of interviews, observations and focus groups highly vary across the different studies: from 11 interviews⁹¹ to more than 300⁶³. These variations can have several explanations: the complexity of the study topics, the available material and human resources, and the time dedicated to the data collection. In theory, therefore, it is difficult to predict the number of interviews required.

One way of speeding up the research process can be to limit the number of interviews. This can be done when the data collection instruments and the characteristics of interest of the participants are carefully planned. Additional details on purposive sampling are provided in the KCE report 187.⁸²

Data should also be progressively analysed during the data collection process to ensure that the diversity and the saturation are gradually reached.

Similarly, to interviews, saturation also determines the number of focus groups expected in a project. When it comes to the number of participants per group, it is usually recommended to have 6 to 12 participants, the ideal number being 8 to 10).⁸² In case of an online group, a smaller number of participants is recommended: 3 to 4, maximum six persons.⁹²

The conclusion is that there is no golden rule about the number of interviews or focus groups.

In KCE studies, because of the limited time resources, the number of the interviews is usually fixed *a priori*, and the building of the sample is based on maximum 2 or 3 priority criteria (next to the language spoken, see further) to reach maximum diversity within the sample size decided. Moreover, KCE qualitative studies never try to reach perfect saturation.

The number of participants can be reduced if the study uses purposive sampling and if a progressive analysis is done during data collection to

ensure that the data collected reach saturation as much as possible (without aiming at perfect saturation).

6.5.5. *Is it possible to accelerate the recruitment process?*

Some facilitators were identified that can be transposed in all studies:

- Having pre-existing collaboration with potential participants (e.g. research team belonging to an academic hospital): this is not likely to occur for KCE.
- Offering incentives for participation: incentives need therefore to be sufficiently attractive and reflect the nature of the participation. Nevertheless, ensuring participants understand they won't receive tangible benefits is crucial. This is why the issue of offering incentives can be contentious. However, options such as reimbursing travel expenses or providing a modest compensation or a small gift are available.
- Being known as a research centre: when recruiting in a large audience (e.g. via social networks), being clearly identified as an official research centre with a strong expertise can contribute to an increased awareness among the population which, in turn, can be more willing to participate in studies.
- Working closely with stakeholders that can help for the recruitment: they can disseminate information or directly contact participants. Here, collaboration with patient (platform) organisations is very valuable.
- Having an inventory of key representative structures in and out the healthcare sector (e.g. the Belgian Disability Forum, the Observatory of Chronic Diseases, the Council of Elders...): relying on these structures can be useful when the topic is well defined and when a "expert" perspective is needed (assuming that those in the structures are more informed and aware than the general population).

For enhancing focus group recruitment, Vindrola-Padros suggests using natural groups and informal focus groups.⁶¹ Natural groups are groups conducted with participants that were not created for the study but are therefore existing).⁶¹ These groups can consist of individuals who come



together unexpectedly (e.g. people waiting in line outside the bank) or existing groups who come regularly together to a specific (non-research related) purpose (e.g. parents who join baby playgroups). An example of the implication of a natural group is the use of “open stalls”, these are stalls set at key locations around the community where individuals could be asked about their views (related to the study purpose) (see the study of Rowa-Dewar *et al.*, 2008, cited by Vindrola-Padros, 2021, p. 59).⁶¹ In the KCE context, one can, among others, rely on patient associations, support groups organised in health services, existing working group of stakeholders on a defined topic or team meetings in a hospital ward.

Informal focus groups are quite like natural groups: it consists in opportunistic discussion based on an existing social group. The content of the discussion and its duration are determined by the participants. It can include other activities (if it remains sedentary and not disruptive). As the discussion is open and flexible, participants may change over the time of the discussion (people arriving, people leaving).^{jj 93}

The formation of groups can also be combined with other data collection methods. For example, Vindrola-Padros refers to a study where the researchers conducted informal focus groups while they carried out observations during staff tours in a community health centre.⁶¹ The unscheduled focus groups gave the researchers the flexibility to incorporate these group discussions into other activities at the health centre. It can be envisioned to do such unscheduled focus groups when planning site visits as done in several KCE projects.

The recruitment process can be accelerate by using natural groups and informal focus groups if the researchers have the possibility to visit study settings: this is therefore limited to some topics (e.g. experiences in attending a specific health service or working behaviours in a chosen health care unit). It can also be helped through the collaboration with the stakeholders who can act as intermediary between the researchers and the participants. This can therefore lead to a selection bias, an influence bias or social desirability bias.

6.5.6. *Is there a possibility to restrict the languages used?*

Because Belgium has three official languages (Dutch, French and German), in KCE studies it is expected of having a multilingual sample, leading to a theoretically tripling of the recruitment material such as information documents, invitations, mailing lists... For practical reasons, data are only collected in French and Dutch, the country's two major languages. However, this is very time- and resource-consuming. Reflecting on how to accelerate the research, we might wonder whether to focus on one language only should be asked.

Multilingual sample makes sense if there are theoretical expectations related to the linguistic and cultural affiliation. In other words, speaking French or Dutch is considered as influencing the thematic under study. For example, based on statistics, we can hypothesise that euthanasia or transgender identity are topics where cultural differences between Flemish and French patients are likely to occur. But the spoken language can also refer to other aspects such as the place of living or the influence of the regional policies on the participants' responses. Another example is the investigation on access to cervix cancer screening: it needs to be considered by region because of the different regional policies and thus requires a bilingual recruitment. For other topics, the region or the province play a key role and then require recruiting in both communities and across the country: investigating access to emergency departments only in the Brussels region, i.e., can lead to biased recommendations.

The literature does not provide advice on how to accelerate the recruitment in multilingual settings. Only two studies reported RQR conducted in multilingual settings: a cross-national study in South America⁸⁹ and an international study relying on eight different languages⁹⁴. In fact, conducting RQR in multilingual settings appears to be the exception rather than the rule.

No, selecting a single language for the data collection is not recommended and can lead to biased outcomes as the spoken language is also a proxy for relevant contextual variables.

jj The authors explained, i.e., visiting a household while prepping a meal.



6.5.7. Which are the most preferred data collection techniques in case of RQR?

A description of some qualitative research data collection techniques can be retrieved in the KCE report 187.⁸² This section focuses on aspects when the principles of RQR are applicable.

Across the different qualitative research data collection methods, a common strategy can be identified to obtain a more rapid research process, i.e., to carry out data collection and analysis in parallel.⁶¹

6.5.7.1. Rapid interviewing process

Vindrola-Padros suggests adapting the interviewing methodology as following to turn it into a more expedited process⁶¹:

- During the interview: each interviewer audio records the interview but also takes notes documenting the conversation. The notes are helpful for guiding the discussion, can act as safety net in case audio record fails and allows to have a summary of the main findings at the end of the interview.
- After the interview: the summary of the findings should be entered in a Rapid interviewing Process (RAP) sheet (which should be developed and agreed by the team). An example is provided in the Appendix 1.3. At the end of each data collection day, the team has a summary of the main findings from all the interviews carried out to date. If data are being collected from different data sources, these can also be added to the same RAP sheet. The RAP sheet can then also act as a triangulation tool.
- As interviews are ongoing: RAP sheets can be used to cross-check data collection and ensure all researchers are collecting data in the same way. RAP sheets are modifiable throughout the data collections stage, removing or adding topics. New questions can be added to the interview topic guide or changes made to the sampling strategy. RAP sheets can be used for different aims of the study, such as comparison of data across different team members (with one RAP sheet per researcher), or to compare data across different research sites or cases (one RAP sheet per study site). Even if data are collected from multiple sources,

these would be added to the same RAP sheet of the specific case or site. At the end of the data collection phase, the RAP sheets can facilitate cross-case comparisons. For comparisons across populations (e.g. patients' groups), then a RAP sheet per sample group is more appropriate. RAP sheets can also be used to capture changes over time, with a RAP sheet at different stages in the studies. For instance, if the study aims to capture changes before and after implementation of an intervention, the team can then compare findings across the different RAP sheets, filled in at various stages of the study.

- After data collection is complete: the team can make the choice of carrying out full or selected transcription during or after data collection. RAP sheets can also be used to identify salient topics and then go back to the transcripts for more details.

For focus group, the same process can be followed, i.e., use of detailed notes to identify the main findings.

Focus groups are longer and require more participants than interviews. Opting for the second one could accelerate data collection process but to the detriment of the richness due to interaction between participants.

6.5.7.2. Observations

During observations in RQR, the researcher will not have enough time to collect detailed notes and might rely on the use of jottings or brief notes or a strategic selection of words that can act as memory aide for the development of more expanded notes afterwards.

The manual of RAP, developed by Scrimshaw and Hurtado⁹⁵, cited by Vindrola-Padros⁶¹, reported three types of information records that each researcher needs to maintain during a study:

- *Brief diary*: notes on what was done each day to act as chronological record.
- *Brief fieldnotes*: a few words on what was seen or heard.
- *Expanded fieldnotes*: on the same day, the researcher should expand on the brief fieldnotes and add comments and impressions in parentheses. The development of these expanded notes can already



identify gaps in data collection, so that the researcher can collect this lacking data during the next interview / focus group.

It is worth recalling that qualitative researchers are expected to use memos as part of the data collection and analytic process but also as part of the (self) reflexivity⁹⁶ (see also section 1.6.6).

6.5.7.3. *Visual methods*

Visual methods can be used as both alternative or complementary methods to collect and generate data. In RQR the use of video or photography can facilitate the intensive periods of data collection.⁶¹ For example, Singh *et al.* relied on visual summary maps of the individual interviews but also maps representing similar settings and sites, that were combined to form “meta-maps” representing group’s experience.⁹⁷

6.5.7.4. *Videoconferencing or online data collection?*

Online research was used in RQR we retrieved from the literature. It allows to gain time by avoiding people, interviewees and researchers, to travel to collect data. However, the online participation leads to a possible selection bias favouring people with a good digital literacy, i.e., youngest, or higher educated participants. Online forum can also be used as a complementary method to other data collection methods but, as experienced at KCE, cannot replace interviews or focus groups.^{kk} Previous KCE studies used online data collection tools: interviews via Zoom (KCE report 349 on decisional capacity⁹⁸, KCE report 361 on access to health care for people with intellectual disabilities⁹⁹), online focus groups (KCE report 353 on ICU staffing¹⁰⁰), Miro assisted group discussions (KCE report 326 on antenatal care¹⁰¹) and sometimes through a combination of online forum and interviews via Zoom (KCE report 344 on Long COVID).

It should therefore recall that it is often easier to build a trust relationship when participants are sitting live together, rather than each on one side of a screen. This could be of more importance in interviews on sensitive topics. Also helping the interviewee in case of emotional distress is much more difficult when the researchers are not physically present.

Online data collection in general population could be a useful mean to accelerate data collection, but it is important to keep the risk of selection bias in mind, particularly as regards less digitalized stakeholders.

6.5.8. *Is it possible to expedite the design and validation of instruments for data collection?*

To date, studies at KCE always consider collecting data in Dutch and French (see also section 1.5.6). This bilingual aspect is not specific to qualitative research. Conducting a study in a multilingual setting requires a careful process for validating the translation of the tool for data collection. At KCE, the tool for data collection is usually first developed in one language, validated by peers in the language of development, then translated, validated by peers and, if necessary, the original version is adapted. Ideally, a reverse translation should be done to ensure the adequacy of both versions.¹⁰²

This can be achieved if the KCE develops a common database of validated interview guides/typical questions that can be used in an interview guide/focus group guide (in Dutch and French to reduce the need for translation) (see also section 1.5.2. on ethics). Guides tend to follow the same structure and some questions are not specific to a topic (e.g. ice-breaker questions, questions reflecting a SWOT, conclusion questions...).

^{kk} For more information, see the KCE report 348: <https://kce.fgov.be/en/identifying-patient-needs-methodological-approach-and-application>



6.5.9. *Is it possible to skip the full transcriptions of interviews and focus groups?*

Transcription involves converting the spoken content of an interview/group discussion into a written form. It has traditionally been regarded as an indispensable aspect of qualitative interview research, making it challenging to question or deviate from this practice.¹⁰³ Transcripts present several advantages:

As highlighted in the KCE report 187, transcribing is essential as it furnishes a comprehensive and verbatim account of the interview.⁸² However, it is important to note that in the KCE context, this should be construed as a word-for-word transfer from audio to text. This means excluding nuances such as tone, inflection, and the timing of responses, as incorporating these elements, while enhancing transcription quality, may be deemed excessive in the specific research context of KCE.

Transcripts allow to check for internal validity credibility of the results because they make member checks possible, i.e., informants may be asked to read transcripts of dialogues in which they have participated in checking if their words matched what they intended.¹⁰⁴

However, transcribing takes time, and not every interview may require a full transcription.

Wanat *et al.* recall in their review that some researchers emphasize the significance of acknowledging the diversity within qualitative traditions and schools.¹⁰³ They argue that, while transcription holds significant value for certain qualitative approaches (e.g. grounded theory), it may not be deemed essential for others. Tape and notebook analysis, involving notetaking from a playback of the recorded interview, can be an alternative.⁸²

In RQR, it can be decided to forego the transcription of data, after a careful evaluation of the necessity of having (or not) a full transcription in the context of the study. Under this rationale, in rapid approach, some researchers choose to bypass the full transcriptions and decide to analyse the data directly from the audios¹⁰⁵ or from field notes. Others opt for a partial in-house transcription and postpone full transcription to have access to more complete material, useful for more in-depth analyses later, and to be able to document the results by using quotes.¹⁰⁶ To note that if the interviews are

audio-recorded, quotes are not lost, it just takes time to locate them to reproduce in the report or a scientific paper.⁹⁷ To have rapid access to illustrative verbatims, data can be directly reported in a structured matrix (one line per data source and one column per theme) to summarize key information and short illustrative quotations from the interviews¹⁰⁶ (see also section 1.6).

Another option is to select only the interviews they consider most important and to transcribe only those in full.¹⁰³ Several authors have, however, made all the interviews transcribed *verbatim*.^{88, 91, 94, 107-109}

The decision to use (full) transcripts or not requires careful consideration regarding the subsequent implications, including workload and the nature of analysis feasible in these circumstances. For instance, working without transcripts is only appropriate when the researchers are immersed in the subject (no interruption in the analytical work) and when there is a guarantee of stability in the research team (the same researcher collects and analyses the data). It is difficult for someone new to a project to immerse herself/himself in the analysis without having heard the interviews, taken part in them or had access to the transcripts.¹⁰³ In addition, inexperienced qualitative researchers attempting to analyse partial notes may introduce bias. It is then recommended to produce complete transcripts for the initial interviews to familiarize the researcher with key messages before considering audio or field notes analysis.

Directly analysing interview recordings instead of working with transcripts can be an interesting option for speeding up the process and present also the advantage that it can reduce the risk of misinterpreting dialogue as the researchers can hear the participant's tone, inflection, and pauses¹¹⁰, what was lost in a 'simple' transcription.⁹⁷

To expedite the transcription process, the research team may choose either professional transcription services or an automatic method, such as utilizing an online tool during online interviews. However, it is essential to note that even with the latter option, a substantial amount of time will still be necessary for thorough checks, edits, and the anonymization of the transcripts.⁸⁸ If a software can transcribe instantly the discussion, this feedback phase could be shortened but at the moment we do not have identified such transcription tool compatible with EU GDPR.



Transcription can partly be skipped for rapid research, and it should be possible to work from field notes, RAP sheets (see section 1.6.2), or partial transcription depending on the team, the subject and the type of results that will be disseminated. Direct transcript should be avoided in RQR and be done only if the researcher has a specific expectation of sharing this “raw” material.

6.5.10. *Is a validation of the results by study participants needed?*

The member validation – that is sharing data with participants to gain their feedback - is one dimension of the triangulation, a key aspect of the quality of the qualitative research – expedited or not. Triangulation can be reached by other methods but can never be skipped.⁸² Consequently, validation per participant is not always used as part of an RQR (but the same applies to qualitative studies in general). As an example, Gilchrist *et al.* relied on member checking in their RQR¹¹¹ while Khayyat *et al.* mentioned that they did not return the transcripts to participants for feedback or correction.⁹¹

This validation needs to be done if no other triangulation method is used. Yet, sharing the direct transcript of the interview can generate adverse reactions: participants may not recall their discourse or find it confronting to read it. Consequently, some participants may ask to modify their contribution or to cancel it. If a validation per participant is needed, it can be more suitable to provide a synthesis of the findings. Based on the KCE experience, sharing the synthesis can be accompanied by specific questions to guide the revision of the document by the participants (e.g. getting feedback to clarify unclear aspects or helping to prioritise the recommendations). The feedback of the participants is then framed and can be easier to integrate for the research team.

Validation of results per participants can be skipped as it is time consuming and other efficient triangulation techniques can be preferred.

6.6. Is it possible to expedite the analytical process?

6.6.1. *Which level of analysis is expected at minimal?*

As explained in the process note on qualitative research, the analytic approach will depend on the design and the aims of the research. Some designs and/or research questions require an inductive, others a deductive approach.⁸²

In a rapid process, as we recommend starting from an existing framework where possible (see 1.5.2), mainly a deductive approach will be applied to the data. We did not find information on the minimal expected level of analysis. In qualitative research, the level of the analysis may “(...)seek simply to describe people’s views or behaviours or move beyond this to provide explanation that can take the form of classifications, typologies, patterns, models and theories” (Pope and Mays¹¹², 2006, p. 67). Next to a basic level, i.e., a descriptive account of what was said (by whom) related to topics and questions (also called the “manifest level”), a higher level of analysis could be interpretative: this is the level of identifying the “meanings” of the responses, also sometimes called the latent level of analysis. This second level of analysis can to a large degree be inspired by theories.⁸²

Most often used in rapid ethnographies (in healthcare) is the thematic analysis that could remain at a ‘relatively’ basic level, with some minor modifications to the methodology to increase the speed of analysis, such as: mapping of the code structure during data collection (this reduces the amount of time needed to develop the initial codes), dividing the interview transcripts and fieldnotes between different members of the team and no full double coding of the data but with a limited degree of cross-checking between researchers.⁶¹

Due to the resources dedicated to the qualitative part of a research project for the KCE studies, the expected level of analysis is at minimum descriptive.



6.6.2. Which analytical approach is the most expeditive / efficient?

Across the analytical approaches, such as table-based methods or thematic analysis, Vindrola-Padros⁶¹ identified following elements which can enhance the analytical process in RQR:

- Primary field analysis: during data collection, brief summaries of data are already made. Also, frequent meetings within the research team to discuss emerging findings enhance the data analysis in parallel with data collection.
- Reduction of data: tables and matrices are often used to reduce data to brief overviews such as the RAP sheets (see Appendix 1.3).
- Development of a codebook in advance or during the initial stages of data collection: this allows the researchers to start the data analysis while data collections is ongoing. However, the authors warned that setting pre-established categories might blind the researchers from findings they had not anticipated. An in-between approach has been suggested with some codes decided on a priori and adding codes during and after data collection, but the time-saving benefits of this approach is limited.

Other approaches can be retrieved in the handbook of Vindrola-Padros⁶¹, such as visualisation methods (e.g. programme theory, causal networks, etc) but their applicability for KCE reports should be further investigated within the frame of the KCE report 187, rather than already suggesting them as an option to facilitate RQR in KCE studies.

We need therefore to bear in mind the warning of Vindrola-Padros:

“The tension between breadth and depth of data is also problematic in rapid qualitative research: it will not be able to capture all the relevant social, cultural, political, and economic factors at stake and will not be able to engage with conflicting views or contradictions. As a result, the depth and focus on complexity that makes qualitative research unique, will be lost. A counter argument is the need for focus of the scope of the study. This tension between breadth and depth, leading to gaps in data collection, could be a matter of study design, regardless of the length of the study (long studies also deal with this tensions)” (Vindrola-Padros, 2021, p. 7).⁶¹

6.6.3. Is it possible to rely on one single researcher for the analytic process?

From the retrieved articles, it emerged that the number of researchers involved in the analytic process varied from one team to another. It is not always possible to distinguish if all researchers were involved at the same time, with the same percentage of working time. However, the analytical process is shared by at least two researchers. It is also worth noting that, in the retrieved studies, the researchers can rely on an interdisciplinary team – allowing for different interpretations of the data and consequently contributing to a richer understanding of the findings. In the retrieved studies, we also identified that the RQR research teams included experienced researchers in RQR (not all researchers were experimented, but a duo of researchers involved in analysis should include at least an experimented researcher). Moreover, some research teams can rely on students or less experienced researchers to manage certain tasks (e.g. contacting the participants), allowing them to focus on their analysis.

Vindrola-Padros⁶¹ mentioned also a challenge regarding the process of data analysis. In RQR there is sometimes not enough time for cross-checking between research members, which could affect the consistency in data collection and analysis. The use of manuals, tables, or other types of pre-established forms for data collection is recommended. Endorsing a team member with a cross-checker role (responsible for going through the data and to identify inconsistencies and bringing them up in team meetings) can also be a solution to avoid this bias.

Finally, it also appeared that the researchers were able to concentrate their efforts on their study (contrary to KCE researchers often involved in several projects at the same time). Being able to focus on the analysis can contribute to accelerate the analysis process.



One single researcher could be sufficient if there are possibilities for the team to give their feedback on the interpretation during regular team meetings, to ensure the validity of the analysis. It is also wise to allow the qualitative researchers to concentrate on this study during data collection, analysis and reporting. It is also advised to endorse a team member with a cross-checker role (responsible for going through the data and to identify inconsistencies and bringing them up in team meetings) to avoid this bias.

6.6.4. *Can the use of software help speed up the analysis?*

Across the studies we retrieved from the literature, the use of software was inconsistent. This can be explained by the fact that some software requires a certain level of training to be used at their best potential. In teams with various levels of mastering these resources, it can be easier and faster to use fewer complex tools such as an Excel spreadsheet. At KCE, some experts are already familiar with Nvivo: it may contribute to fasten the analysis if, when there are several researchers involved in the project, all of them can use it in a collaborative way. It also allows for integrating various sources of data such as audio files, transcripts of interviews, visual material, or literature.

Analysis could be fasten by the use of a software but cannot be replaced by AI-performed analysis.

6.6.5. *Is it possible to skip or abridge the triangulation of sources?*

Triangulation is considered as a quality criterion of qualitative research and can be achieved via different methods.⁸² All qualitative studies (whether expedited or not) need at least one form of triangulation to be considered as valid: it also serves to prevent accusations of biased findings because of a single method or source or researchers' biases.⁸²

When comparing the different forms of triangulation, triangulation of sources consist in comparing data obtained by different methods or derived at different times.⁸² Methods triangulation, consisting in comparing data obtained by different methods, appears to be more efficient and is consistent to the existing practices at KCE where studies often rely on different data sources (e.g. quantitative data, literature search, analysis of the context...).⁸² The "different times" can be difficult to achieve in RQR.

Triangulation is always required as it part of the validity of the qualitative research and should be achieved by triangulation of sources.

6.6.6. *Is it possible to skip or abridge the process of reflexivity?*

Similarly to triangulation, reflexivity is essential to qualitative research, independently of its expedited character.⁸² Reflexivity implies that the researchers reflect on how the research have changed themselves and how it affected the process of the research.¹¹³ It can even appear as more essential to stress as one can share the concerns raised by Singh *et al.*: that the RQR may be accused of lacking of rigour and in-depth analysis.⁹⁷ In our literature review, most of the studies reported reflexivity. As an example, Rankl *et al.* detailed their reflexivity process in a context of a RQR.¹¹⁴

Kohn and Christiaens reported four good practices are advised to ensure reflexivity: theoretical openness, methodological openness, awareness of the social setting of the research itself and awareness of the wider social context.⁸²

Regarding the theoretical openness, several studies incorporated theoretical frameworks, as a starting point of their process. For example, the Behavioural Wheel Change was used by Bielinska *et al.* to investigate advance care planning in later life.¹¹⁵ Higham *et al.*⁸⁸ mobilised the Framework for Theorising and Evaluating Non-adoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies (NASSS) developed by Greenhalgh *et al.*¹¹⁶ and Normalisation Process Theory (NPT) developed by May and Finch to study infection prevention in an UK hospital during the COVID-19 pandemic.(see section 1.5.1)

In terms of methodological openness, it implies a clear and detailed reporting of all methodological steps in the study.

Reflexivity also implies to provide a detailed description of the context and setting. This is, in KCE studies, addressed in the context description of the issue at stake. It is thus important to devote time to describing the context, even if it is considered well-known by the audience.



Finally, qualitative research (expedited or not) also implies a form of self-reflexivity of the researchers. Borsca *et al.* define it as following: *“self-reflexivity, which is essential in qualitative research, goes hand in hand with an emancipatory attitude, trying to enrich the social discourses on mental health with wisdom “from within” and “bottom-up,” instead of “top-down,” in its double sense: theory-driven as well as hierarchical”* (Borsca *et al.* , 2022, p. 7).¹¹⁷ (Self) Reflexivity can be achieved by regular meetings with peer-researchers, involved or not in the study.

Reflexivity is a validity criteria of qualitative research and should be always integrated in a RQR. Participation to the meetings of the qualitative cell can help the researchers involved in RQR but they need to have sufficient time to attend it.

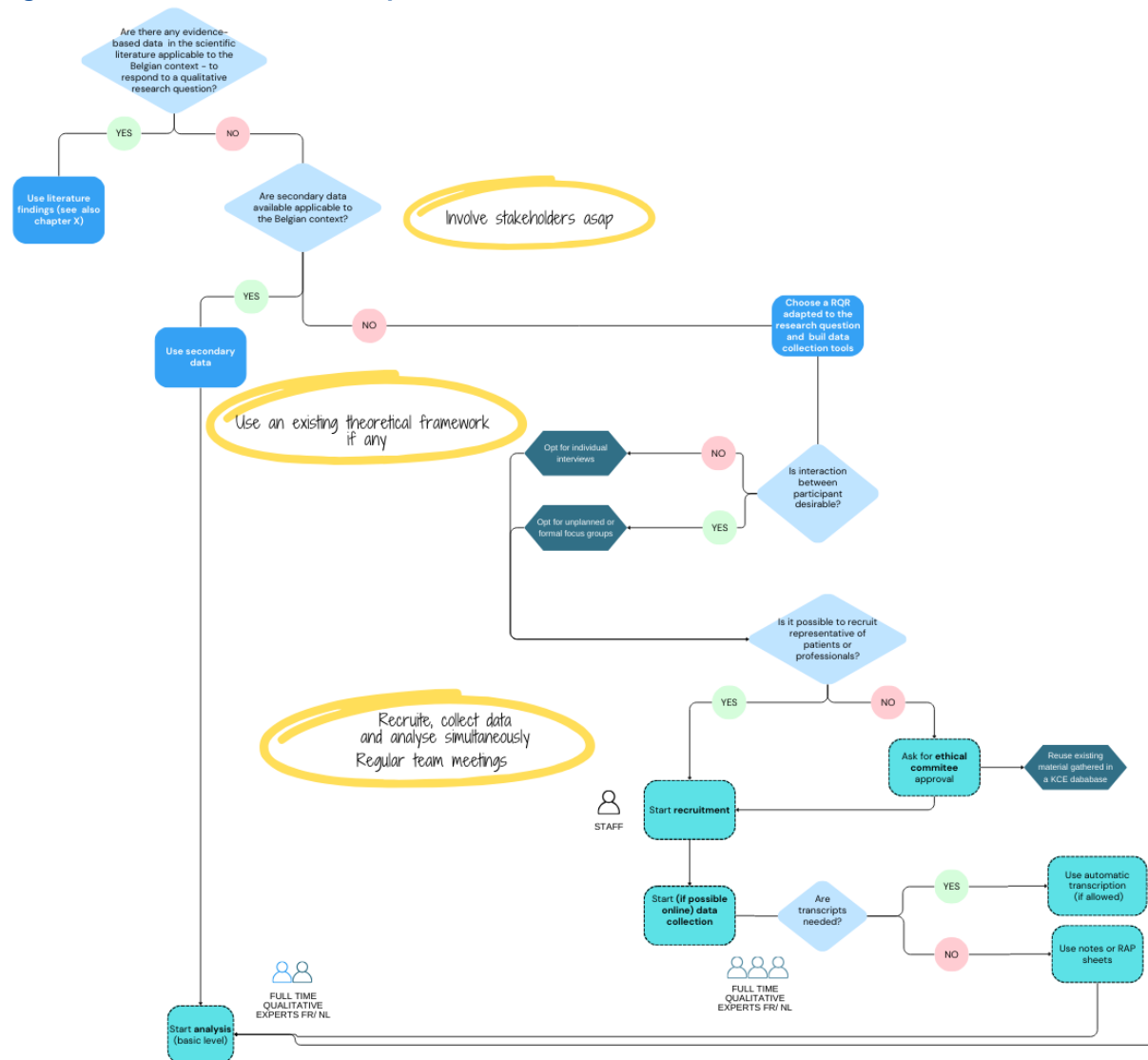
6.7. Discussion

6.7.1. Decisional tree

Vindrola-Padros recalls that RQR needs to be flexible and needs to take into account the contextual changes over time.⁶¹ But, this also implies that RQR may be not adapted to all research contexts and topics. To decide when to conduct a RQR, she refers to the resource-time matrix developed by Nunns (see Appendix 1.4).¹¹⁸ This matrix mainly focuses on the human resource and the time constraints and insufficiently takes into account the different steps of the qualitative research. Figure 4 presents a decisional tree, adapted to the KCE context, which aims at helping to plan an RQR, considering the different steps of a qualitative research. It also includes elements related to the team and the management of the project. Boxes include the section numbers where additional information is available. All principles of qualitative research as reported in the KCE report 187 are complementary to this decisional tree.



Figure 4 – Decisional tree for Rapid Qualitative Research at KCE





6.7.2. Conditions for making RQR effective

As highlighted through this chapter, it is necessary to have a certain number of human resources to conduct research in a rapid time frame. It requires researchers and support staff. Some time-consuming tasks can be delegated to the staff: e.g. contacting participants per phone to recruit them. Therefore, it should not be done at the expenses of the other missions of the staff and require that staff receives adequate information to be comfortable by doing this.

Regarding the researchers needed to conduct an RQR, the following aspects can be stressed:

- Each researcher needs continuous training in evolving qualitative research methods. In RQR methods, often the research teams are enlarged, but the involvement of inexperienced researchers might cause a delay in study set-up and start of data collection, because they need more time for training.⁶¹
- The team needs to include experienced researchers able to conduct an interview and taking notes simultaneously. The golden standard is usually conducting interviews with two researchers: one leading the discussion, the other observing and taking notes.
- The team needs to include researchers experienced in their contacts with the stakeholders. Ensuring that some of the team members had previous (positive) collaboration with the key stakeholders of the topic can help to accelerate the process. Indeed, they do not need to first learn how to function together, allowing them to work together directly.
- The team needs to include experienced researchers in conducting analysis. Therefore, one can only develop its expertise by practicing regularly. Outside “expedited” research, qualitative experts need to have regular opportunities to analyse data, with the support of more experienced KCE researchers.
- The team needs to have regular meetings to discuss the data collection as it goes and contribute to the different interpretations of the data, including reflexivity (see also sections 1.5.4, 1.6.5, 1.6.6).
- Conducting interviews is already starting the analysis: this can be seen as an argument in favour of researchers collecting and analysing the data, rather than having researchers doing the interviews and others doing the data analysis.
- Having the possibility of being immersed in the data collection and analysis may contribute to a faster process: it requires limiting the number of simultaneous projects.

These conditions, if not fulfilled, can jeopardise the entire process of an expedited research. Vindrola-Padros⁶¹ pointed out several situations where it might be best to avoid carrying out rapid qualitative studies (only the situations applicable for KCE research were retained):

- *“When the time and resources available for the study are severely constrained and there is a lack of existing evidence to help inform the study design.*
- *When there are clear signs that the team will not be able to get access to the people or spaces [settings] required to carry out data collection for the study.*
- *When the research questions cannot be answered with a rapid study design as they require long-term engagement with a specific topic or area.*
- *When researchers are being asked to ‘cut corners’ or carry out research in a way that might violate quality of ethical standards of research”. (Vindrola-Padros, 2021,p.6).⁶¹*



7. AUTOMATION

Key messages

- Automation has the potential of expediting several steps in the research process, this can be achieved via the use of tools specifically designed to complete a research task or via the use of generative Artificial Intelligence (AI) models such as Large Language Models (LLMs).
- Generative AI can be used for a variety of tasks such as the production of text summaries, translation, data extraction from text, code generation and exploratory knowledge. The user should be aware that it is subject to potential biases and may produce factually wrong (hallucinatory) output. The use of carefully crafted prompts, following prompt engineering techniques has the potential to increase the accuracy and relevance of the Generative AI's output. A library of prompts of proven efficacy should be developed within the KCE.
- When using generative AI, a responsible use is recommended, both from a legal and an environmental standpoint. The user should always check the terms and conditions relative to the use of the chosen model.
- We investigated and selected some tools that have the potential to be beneficial in expediting specific steps of the KCE research process, but these tools need to be tested in a real-life project setting to be fully validated. We selected tools to manage and streamline literature reviews, in which some degree of AI assistance can be found in the screening phase, but still need to undergo human verification. AI-based semantic search engines may be interesting in an exploratory phase of literature retrieval, but the reproducibility of their results is variable. The transcription software Wisper.ai is currently being tested in-house.
- This work is a preliminary step to a regular *technological watch*, that should take the form of an online living document.

7.1. Background

Automation refers to the use of technology to perform tasks or processes with minimal or no human intervention. It involves the application of machines, computer software, and other technologies to operate and control a process or system. An ideally reliable automation process could take over repetitive and time-consuming tasks, and free time for other, more intellectual activities.

Any part of a research project has some tasks that are at least partly amenable to automation. For instance, classically carried out health or medical-related systematic review is completed in over thirteen months¹¹⁹, and is potentially obsolete within two years⁴¹.

The term “Synthesis gap” describes how policymaking requests go faster than the speed at which it can be produced by using traditional methods¹²⁰. Automation could assist with reducing the *synthesis gap*, provided its assistance produces reliable and reproducible results.

At the 2023 ISPOR (International Society of Pharmacoeconomics and Outcome Research) edition, Mangat et al.¹²¹ reported how the use of AI tools is not the subject of any specific guidance across Health Technology Assessment (HTA) bodies, with the exception of NICE, which makes some recommendations in the use of AI in the screening of articles for Systematic Literature Review (SLR), demonstrating how research institutions are only currently embracing the potential of AI in their processes.

Box 1 – Definition of concepts

Automation: Automation involves using technology to perform tasks with minimal human intervention. This can range from simple mechanical systems to advanced AI systems. The goal is to increase efficiency, reduce errors, and free up human time for more complex tasks.

Artificial intelligence (AI): Artificial intelligence is a technology that enables computers and machines to simulate human intelligence and problem-solving capabilities.



Machine learning (ML): Machine learning is a branch of AI and computer science that focuses on the use of data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.

Natural Language Processing (NLP): Natural language processing combines computational linguistics—rule-based modelling of human language—with statistical and machine learning models to enable computers and digital devices to recognise, understand and generate text and speech.

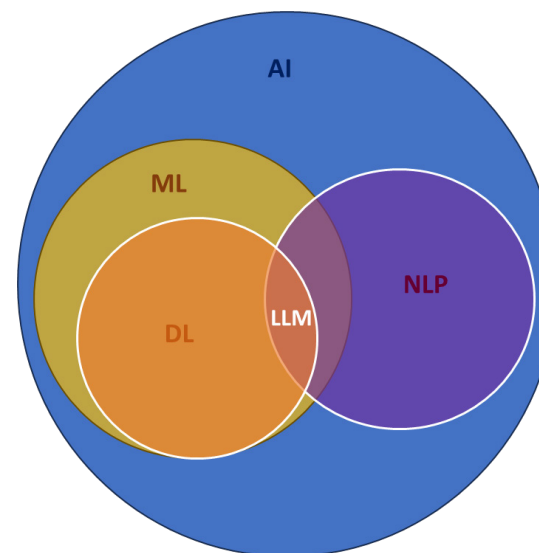
Deep learning (DL): Deep learning is a subset of machine learning that uses multi-layered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain.

Large language models (LLMs): Large language models are a category of foundation models trained on immense amounts of data making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks.

Generative AI: Generative AI refers to deep-learning models that can take raw data and “learn” to generate statistically probable outputs when prompted. At a high level, generative models encode a simplified representation of their training data and draw from it to create a new work that is similar, but not identical, to the original data.^{ll}

Application programming interface (API): An application programming interface, is a set of rules or protocols that let software applications communicate with each other to exchange data, features and functiona^{mm}lity.ⁿⁿ

Figure 5 – Relationship between different concepts in Artificial Intelligence (AI)



Legend: AI: Artificial Intelligence; DL: Deep Learning; LLM: Large Language Models; ML: Machine Learning; NLP: Natural Language Processing.

Source: adapted from <https://medium.com/@deepthi.sudharsan/decoding-the-relationship-language-models-and-natural-language-processing-bbc0cd6754e2>.

^{ll} <https://research.ibm.com/blog/what-is-generative-AI>

^{mm} Source: <https://www.ibm.com>

ⁿⁿ <https://www.gartner.com/en/information-technology/glossary/application-programming-interface>



7.2. Aim and Objectives

The aim of this chapter is to investigate potential automation tools that could be integrated in the KCE project workflow to expedite the research process.

The field of automation is currently in full expansion, so this work is to be considered as preliminary to a regular *technological watch*, that should take the form of a living document.

This chapter has as main objectives:

- To provide basic practical advice on **how to use Generative AI** in the KCE project lifecycle
- To identify tools and applications that may have a **specific purpose** in expediting phases of the KCE project lifecycle

7.3. Methods

The general methods used to produce this chapter were pragmatic. We investigated the experiences and resources of each individual member of the research group and we proceeded to perform:

- Exploratory pragmatic search of the literature via Medline, snowballing of references on relevant articles, search of grey literature and of commercial websites offering automation tools, from Novembre 2023 till March 2024.
- Information collection at relevant national and international research networks or organisations, including the Cochrane Collaboration, the European Association for Health Information and Libraries (EAHIL), Health Technology Assessment international (HTAi), the Institute for Quality and Efficiency in Health Care (IQWiG), the International Network of Agencies for Health Technology Assessment (INAHTA), the International Society for Pharmacoeconomics and Outcomes Research

(ISPOR), and the Sheffield Centre for Health and Related Research (SCHARR).

- The distribution of a short survey to medical librarians and information specialists within Belgian universities and health research organisations.
- Participation to seminars, conferences and online courses related to the use of AI.
- Participation of the research team in a shared technological watch^{oo}.

7.3.1. *Methods to provide practical advice on the use of Generative AI*

A guidance on prompt engineering was developed based on the material covered during an online course on prompt engineering attended by one member of the working group, and a pragmatic review of the literature on this topic. Advice on how LLMs can support code generation was derived from first-hand experience and a pragmatic literature review. University websites were consulted to create a guidance note on the responsible use of Generative AI. The opinion of an external expert was also sought.

7.3.2. *Methods to identify tools and applications for specific research purposes*

7.3.2.1. *Methods for the retrieval of tools and applications for specific research purposes*

To retrieve a representative sample of available tools for automation, the working group examined separately the different phases of a KCE project lifecycle (Figure 2), and divided the process into two categories:

- **Retrieval and analysis of pertinent literature:** for this task automation has the potential to expedite all the phases of a classical systematic

^{oo} The searching and sharing of information about developments in a specific technological field



literature review (retrieval, screening, evaluation and synthesis of the literature) and has an added value in getting access in a short timeframe to relevant background information in the scoping phase and while evaluating submitted projects.

- **Data collection and analysis:** in this phase automation has the potential to avoid the unnecessary use of external collaborators for the transcription of collected qualitative data (e.g. interviews, focus groups) or translation, and provide a base for expediting the analysis of transcripts. Automation has also the potential to expedite the execution of certain quantitative data collection (e.g. survey design) and analysis tasks. It may not directly be linked to the analysis itself, but also include coding, sharing analysis, etc.

Box 2 – Perception of the needs of data collection and analysis tools

To better understand the needs of KCE experts for expediting qualitative data analysis, we organised a meeting with a senior expert in qualitative research and another one with KCE's qualitative research cell (n=5). For quantitative data analysis, an email was sent to all data analysts, inquiring about their perceived usefulness of data analysis tools, and their relation with using tools to automate their daily tasks. In the process of inquiry on KCE experts' needs for automation we received the following response:

Qualitative data collection and analysis

During the two meetings with KCE experts in qualitative research, two needs emerged: the requirement for automatic audio transcription tools for interviews and focus groups, and the need for assisted qualitative data analysis tools. Concerning the automatic transcription tools, it was deemed crucial that the tools are able to transcribe in both French and Dutch, accommodating various accents in these languages and distinguishing between different speakers' voices. For the assisted qualitative data analysis tools, emphasis was placed on their capacity to

code in both French and Dutch, using both inductive and deductive coding methods. Members from the qualitative research cell were generally satisfied with the use of Nvivo to analyse qualitative data.

Quantitative data collection and analysis

Discussions with expert data analysts on automation in quantitative data analysis resulted in the decision not to pursue tools for such a task. The current opinion is that each project is different, so not amenable to a "one size fits all" procedure and that automation will not save time in the analyst's need to "feel" the data.

The most efficient way of saving time would be to **share macros and codes developed in-house**. We can think of code-sharing platforms (e.g. Github) where everyone could 'draw' the codes or macros that are of interest for a particular analysis.

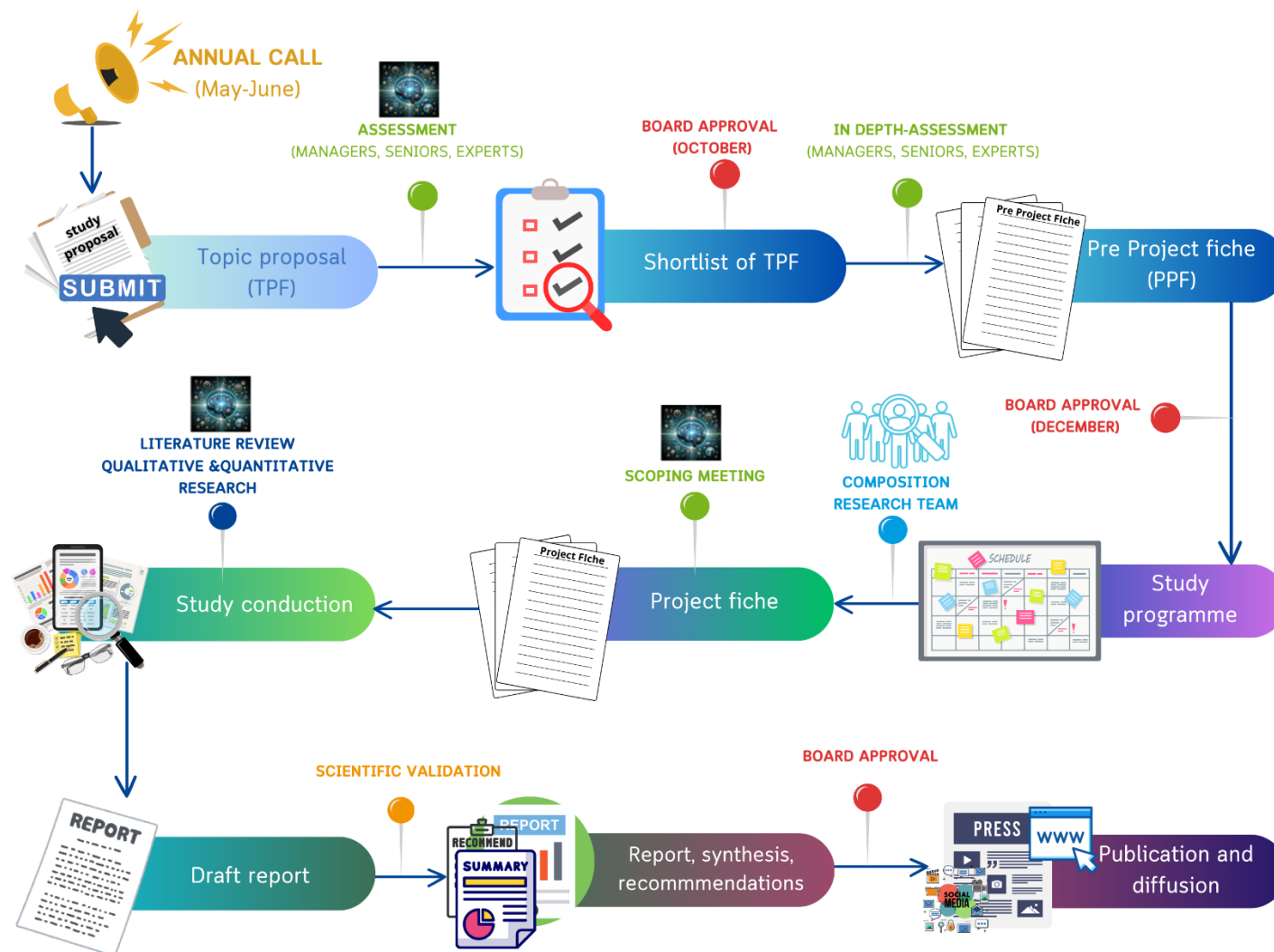
The use of generative AI in data analysis is something new, and appears to be interesting to explore.

General remarks:

1. When we talk about automation, it is difficult to separate methods relative to qualitative and quantitative data analysis, since in AI models, both kind of data (whether structured or unstructured at the moment of collection) will undergo the same steps of structuring and analysis.
2. We proposed in this report some specific tools, but we are not exhaustive. Lots of tools are built by developers to respond to a specific need, but are not commercialised. They often use AI models developed by large corporations such as Google, Amazon and Microsoft.
3. Some investigated tools have become obsolete in the course of the redaction of this document. To signal to the reader that the current tool may be obsolete, the relative sections have been moved to the appendix.



Figure 6 – Potential inclusion of automated tools in the lifecycle of a KCE project





7.3.2.2. *Methods for the evaluation of tools and applications for specific research purposes*

To evaluate retrieved tools and applications, the research group (n= 5) established by consensus during three iterative discussions (two face-to face and one online), the evaluation process for candidate tools. It was decided that the evaluation process should ideally include two phases. The first phase, that is reported in this document, is based on a rapid evaluation by a single researcher of a set of predefined items that have been considered relevant attributes for a tool to be useful in the KCE research setting, the second phase should involve, for the tools that have been considered “of interest” in phase one, the evaluation of such items in a real-life testing of the tool in the context of a study.

We discussed the most important items on which each potential tool should be evaluated to determine each tool's suitability in expediting research in the framework of a KCE project. We came to an unanimous consensus on the definition of seven items: resource saving, reliability, liability, learning curve, user friendliness, cost, frequency of use (Table 11).

We decided that each item should be evaluated using a 5-point Likert scale, where 1 is consistently the worst possible score and 5 is the best possible score (i.e., unlikely/somewhat unlikely/neutral/somewhat likely/likely). Some points of attention for the evaluation of certain dimensions were added to enhance consistency in the evaluation process (Table 11). We agreed that a score of 1 or 2 in the dimension “Liability” in the first round of evaluation for a tool that was used to process sensitive data, would result in the exclusion of the tool for further evaluation.

Table 12 – Dimensions used to evaluate tools and applications for research automation

Dimension	Definition	Points of attention
A Resource saving	The tool/ application is likely to save (human) resources in the execution of the task compared with standard methods	
B Reliability	The tool/application is likely to NOT introduce flaws in the results of the research process	
C Liability	The tool/application is likely to NOT expose KCE to legal liability	Respect of GDPR* Server located within the EU Secondary use of data Data Owner(s)
D Learning curve	The tool/application does not require extensive training in competencies that cannot be found in-house	
E User friendliness	The tool/application is easy to use	
F Cost	The cost of the tool/application is likely to be proportionate to its usefulness	Unique purchase License type (annual fee /Monthly fee / institutional fee/ individual fee / ...) Frequency of tool update (impact cost)
G Frequency of use	The tool/ application is likely to be used regularly/frequently at the KCE	

*GDPR: General Data Protection Regulation



7.4. Results

7.4.1. Generative AI

Generative AI tools, such as LLM based chatbots, are being developed and continuously improved. The most popular one is ChatGPT, a variant of the GPT language model developed by OpenAI. ChatGPT is designed for conversational applications, that is based on its “understanding” and “production” of natural language.

It is important to note that while LLMs are proficient in language generation and understanding, they do not possess true comprehension in the human sense. Its responses are based on patterns and associations learned from the data on which they are trained.

The LLM “understands” the input text or “**prompt**”, then gets the first few words from rearranging the input (question) into a response, then it finds the most probable next word, where every next word is reweighed on the basis of the preceding word. The reweighing step is called “transformer” and the principle of re-evaluating the weights based on the salience given to previous parts of the phrase is called “attention”^{PP}. The chatbot continues to select the word with the highest probability, and then repeating this process for each subsequent word until a full response is generated.

To maximize the benefits of an LLM chatbot, users need to learn effective prompting techniques. This involves being clear, specific, and direct in their queries and instructions (see section 7.4.1.1).

Beyond generating text, such models are able to generate images, audio and video and are used in coding (see section X). The latest version of ChatGPT-4 can handle diverse forms of texts, such as databases, and can even perform analysis and construct graphs based on the data. However,

this kind of process may expose to liability since it is unclear how the data fed into the chatbot are used.

One critical constraint of LLMs lies in the knowledge cut-off date of training data. Frequently used LLMs such as ChatGPT and Gemini are trained on historical data up to a specific point in time. For instance, ChatGPT’s training data only extends until September 2021, which means it lacks awareness of events or developments beyond that date. This limitation impacts its ability to provide real-time information or context. To offset this drawback some models are internet-connected, allowing them to access up-to-date information and quote sources. For example, an AI language model integrated with web browsing capabilities can pull recent data from online articles, scientific journals, or databases. This does not necessarily mean that the output of such models is accurate, as the output of an LLM always needs to be checked for factual errors or **hallucinations**. The term hallucinations describe situations in which the content produced by LLMs is not based on reality or their training data, resulting in outputs that may be inaccurate or misleading¹²².

7.4.1.1. Prompt engineering

A prompt is a natural language instruction (input) that is given to a LLM in order to produce the desired response (output). Prompts make it possible for us to interact with an AI without using code, but simply relying on the language we use in every day conversations.

Prompt engineering involves the careful crafting and fine-tuning of input prompts to steer the responses of LLMs, aiming for outputs that are accurate, relevant, and coherent¹²³.

The use of prompt engineering techniques has shown a significant increase in the output’s quality¹²⁴. Although it involves established strategies, it currently still involves a significant amount of trial and error, **lacking a deep**

^{PP} For a clear visual explanation of this process, please see https://www.theguardian.com/technology/ng-interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer?CMP=Share_iOSApp_Other



theoretical foundation that explains why one way of framing a task is more effective, beyond the fact that it produces superior outcomes. Prompt engineering techniques may be particularly useful to avoid LLM's **hallucinations**.

To interact with LLMs, a general rule is to write in a clear and simple manner, describe precisely the task to execute, avoid stating what NOT to do, but instead concentrate on what needs to be done and finally give the necessary context¹²⁵.

We provide below some examples of prompt engineering principles and techniques, trying to concentrate on fundamental and robust concepts. However, the field is in full expansion so new, more efficient techniques may already be available at the time of publication of this document. While we are keeping the conceptual discussion of the prompting techniques and some prompt examples in the main text, most prompt examples and their output can be found in Appendix 1.5.1.

Prompt elements

To present a LLM with a general question without providing specific instructions or context is called **naïve prompting**. This is a simple technique, useful for quick interactions and for getting general information, but may result in generic or ambiguous responses¹²⁶.

To elicit a more precise response to execute a specific task, the structure of a prompt can include up to four essential elements⁹⁹. These elements—each critical in its own right—can be utilised in full or in selected combinations, depending on the specific requirements of the task:

Instructions that guide the task,

Context that provides background information or additional framing of the task, to steer the model in the desired direction

Input data that serves as the basis for the response

Output indicator that specifies the desired format or type of response.

For example, if the task is to explore what are the advantages for the patients of the available gestational diabetes screening strategies in the Belgian context, a naïve prompt could be as follows (*Prompt 1*) *What are the advantages of different gestational diabetes screening strategies?*

A prompt crafted to get a more precise response, and that includes all four elements could be as follows: (*Prompt 2*) *There are two types of gestational diabetes screening in Belgium, one is a one-step strategy, the other is a two-step strategy. Analyse advantages of both strategies and give the two main advantages for each strategy with three bullet points, from a patient's point of view*

The right contextualisation is crucial to get a relevant response. A technique to steer the LLM in the desired direction is giving the LLM, in the prompt, a “role” to play (**role prompting**)¹²³. When the goal is to obtain a response within a specialised field, we may request the LLM to adopt the persona of an expert in the corresponding field (**expert prompting**)¹²⁷. For instance, when the model is prompted to assume the role of a public health expert, it is inclined to offer a thorough and contextually precise response when questioned about matters pertaining to public health.

The different parts of the prompt can be demarcated by using **delimiters**, for instance you may use triple quotations to differentiate input data from the instructions.

A prompt example that incorporate all the aforementioned elements may be the following:

(*Prompt 3*) *List two main advantages and two main disadvantages for the following gestational diabetes screening strategies available in Belgium. One-step screening strategy and two step screening strategy. Impersonate a concerned patient. Put the results in a table.*

Zero-shot prompting and few-shots prompting

⁹⁹ <https://www.promptingguide.ai/introduction/elements>



The examples of prompts from the previous section are examples of **zero-shot prompting**^{rr}: in this technique we rely on the training data of the LLM to perform the task, without giving any example or additional training. When zero-shot prompting fails to produce desired results, it is advisable to supplement the prompt with demonstrations or examples, transitioning to a **few-shots prompting** approach. Few-shot prompting involves providing a language model with a small number of examples (shots) to illustrate the task at hand. Each example typically consists of an input followed by the desired output.

Chain of thought prompting

In **chain of thought** prompting, we actively promote factual accuracy within the model by guiding it to adhere to a structured sequence of steps in its "reasoning"¹²⁷, which is a step beyond the few-shots prompting method described in the previous section, since we do not limit ourselves only to providing examples of the desired output to a certain input, but we **describe the logical steps** that link input and output.

Another technique that is used to elicit chain of thought "reasoning" from a LLM is zero-shot chain of thought prompting¹²⁸ (that basically only involves adding a sentence akin to "Let's think it step by step" to the original prompt).

Prompt chaining

In the **prompt chaining** approach we need to break the task that we want to complete into subtasks. After identifying those subtasks, the LLM is prompted with one at a time, and its response becomes input for the next prompt in the sequence. This technique may be useful to complete complex tasks that a LLM could find challenging if faced with only one extremely detailed prompt^{ss}. For example, if we wish to summarise a long or complex text we might want to first let the LLM give us the main points of the text in bullet points, and after take the bullet points as an input and ask the LLM to make a fluent text.

It also is useful for mathematical reasoning, data analysis and visualisation, as shown in the example in the Appendix X, where the output was produced using ChatGPT 4.

Some further reflections on prompt engineering

We have explored some of the most current prompt engineering techniques for text generation, but prompts may be used also for image, sound and video generation. Besides, many more techniques are available.

Ultimately, crafting the right prompt to perform a task is an iterative process, that requires extensive trial and error. There are a number of tools that help design efficient prompts, and prompts for specific domains are stored in a number of prompts libraries. Examples of both tools and libraries can be found at the following website: <https://www.promptingguide.ai/tools>.

A specific prompt library in our institution could be useful to share prompts that have been proven useful in the expedition of certain tasks in the research process.

7.4.1.2. LLMs to help in code generation

LLMs offer powerful code generation capabilities. These models have been trained on vast amounts of text, including code from various programming languages, documentation, and forums, enabling them to understand and generate code effectively. LLMs can assist programmers in several ways, making the development process more efficient and accessible.

- **Code completion and suggestion:** One of the primary ways LLMs help in programming is through code completion and suggestion. As a programmer types, the LLM can predict the next line of code or complete the current line, based on the context provided by the existing code. This feature is incredibly useful for speeding up the coding process and reducing syntactical errors. For example, when a developer starts typing a function in Python, the LLM can suggest the complete function signature and even the body based on common

^{rr} <https://www.promptingguide.ai/techniques/zeroshot>

^{ss} https://www.promptingguide.ai/techniques/prompt_chaining



patterns it has learned. Those tools are already part of code editors, natively or as plugins, such as Github Copilot and JetBrains AI assistant. Examples of prompts and code generated by Github copilot can be found here (<https://github.blog/2023-06-20-how-to-write-better-prompts-for-github-copilot/>). LLM can help generate a single line of code such as a regular expression, a function or a whole class. Models such as chatGPT have a memory span big enough to help building a set of classes interacting with each other, remembering the methods implemented by each class.

- **Debugging and error correction:** LLMs are also useful at identifying errors in code and suggesting corrections. By understanding the syntax and semantics of the programming language, as well as the intent behind the code, these models can highlight syntactical errors, logical mistakes, or even suggest better coding practices. For instance, if a developer mistakenly uses a variable that is not defined, the LLM can point out the error and suggest a fix, such as defining the variable or correcting its name to match an existing one.
- **Code documentation and explanation:** Another significant advantage of using LLMs in programming is their ability to generate basic documentation and explanations for code. By simply providing a piece of code to an LLM, it can generate comments explaining what the code does, how it works, and, to a certain extent, even why certain decisions were made. This not only helps in maintaining and understanding codebases but also assists in onboarding new developers to a project. Note that the output will always need to be adapted and contextualised by the programmer to document methods and decisions, especially when alternatives were not chosen.
- **Example generation and learning assistance:** LLMs can generate examples of code based on specified requirements or functionalities. This is particularly helpful for learning new programming concepts or languages. The LLM can provide a complete, annotated example, which serves as both a practical guide and a learning tool.

While LLMs have dramatically improved the efficiency and creativity in coding tasks, their suggestions are not infallible. An experienced programmer's expertise becomes crucial in identifying and correcting errors that might be introduced by an LLM. These models, despite their advanced capabilities, can sometimes generate code that is syntactically correct but logically flawed or not optimal for a given context or that poses security threats. For instance, an LLM might suggest an algorithm that solves the problem at hand but is inefficient in terms of computational complexity.

Moreover, understanding the underlying principles and having a deep knowledge of the programming language and its ecosystems allows developers to critically evaluate the suggestions provided by LLMs. This critical evaluation ensures that the integrated code aligns with best practices and project-specific guidelines which might not be apparent to the LLM. Additionally, experienced programmers can spot subtle errors in logic or implementation that could lead to bugs or vulnerabilities in the software.

These considerations apply to programming in general. In our institution, programming languages (ex. R, SAS, etc...) are - not only but mostly - used for health care data manipulation and analysis (as opposed to web engineering or applications development for example). The knowledge of these underlying data is obviously paramount, which adds to the computational complexity and makes the LLM output validation crucial.

How to responsibly use Generative AI – Guidance note

This guidance note is not intended as an exhaustive set of best practices and should not be viewed as a final policy. As the AI field is evolving rapidly, it is anticipated that this document should be updated regularly.

Content authenticity: AI-generated content can be inaccurate, misleading, or entirely fabricated ("hallucinated") or may contain copyrighted material. The user is responsible for any content that is published that includes AI-generated material. It should be kept in mind that Generative AI tools are often trained on large, unmoderated bodies of text, such as text posted to the internet. This can result in the production of biased and other unintended content. Information checking of LLM's output is therefore mandatory



Copyright: AI tools can use content available on the Web (text, images, etc.). Not all the content provided on the Web is flagged with copyrights mentions^{tt}. And even if so, it is unclear if, and how, the AI tool takes this into account. Therefore, the probability that the AI-generated content contains copyrighted material is not negligible and results need to be checked to identify potential copyright infringement. Generally, any new content automatically receives copyright protection, granting the creator both moral rights (such as the right to be recognised as the creator) and economic rights. However, the situation regarding AI-generated content remains complex and unresolved. Typically, providers of AI tools do not claim ownership of the content created using their tools, potentially as a precaution to avoid assuming legal responsibility. Computer programs are subject to copyrights, therefore prompts could be considered as part of a computer program, so people would be tempted to consider the resulting creation as copyrighted. However some courts have stated that a creation generated by a computer (software) is not subject to copyright since copyright can only concern a human being, and cannot be attributed to the owner of the computer (software creator) having generated this creation^{uu}. The prompt is considered as instructions to a commissioned artist, but the machine “interprets” those instructions to deliver a creation^{vv}. Re using content created with AI tools exposes therefore to copyrights issues, so checking the latest version of the terms and conditions of the LLM of choice, both regarding the use and storage of the input data (see data security below) and on the rights of use of output data.

Why cite LLM chatbots: The use of LLM chatbots should be cited for different reasons, including:

- To uphold the principle of academic integrity by giving due credit to your sources.

- To provide transparency about the origin of the information, which can have a different bias profile than human-generated content.
- To contribute to the important task of tracking the influence of AI in our collective knowledge, highlighting its role in various domains.

Most academic publishers agree that LLM chatbots should not be given journal authorship but accept their responsible use when they are cited correctly. This implies that the nature and limitations of LLM chatbots should be taken into account and that they should be used for explorative purposes and text refinement rather than content creation. When submitting a manuscript to a scientific journal, we recommend checking the author guidelines of the selected journal.

How to cite LLM chatbots: Currently, there are no specific guidelines established for citing LLM chatbots. If you are allowed to use LLM chatbots to help you write a manuscript, then we recommend to check the style guides to obtain information on the latest recommended format to use to cite LLM chatbots in your work. Citation rules for LLM chatbots vary depending on the style guide and are evolving over time.

Data security: Confidential or sensitive information should not be provided to AI tools until the location of storage and computation is clearly defined (within EU)^{ww}. Also, the AI tool should clearly state if the content provided for analysis is then added to its “database” to enrich its language model, or not. You should not enter data classified as confidential (e.g. non-public research data, patient data) or (privacy) sensitive into publicly-available Generative AI tools. Information shared with Generative AI tools using default settings is not private and could expose proprietary or sensitive information to unauthorised parties.

^{tt} <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/license/>

^{uu} <https://findfx.thomsonreuters.com/gfx/legaldocs/lbvgooeoqvq/AI%20COPYRIGHT%20LAWSUIT%20thalerdecision.pdf>

^{vv} <https://www.theartnewspaper.com/2023/09/02/artificial-intelligence-lawsuit-decision-us-copyright-law>

^{ww} <https://www.wired.com/story/how-to-use-ai-tools-protect-privacy/>,
<https://oecd.ai/en/work/the-ai-data-challenge-how-do-we-protect-privacy-and-other-fundamental-rights-in-an-ai-driven-world>



7.4.2. Tools and applications for specific research purposes

We focused on retrieving and evaluating tools that were designed and responded to the need to expedite specific steps in the research process and were sufficiently limited in scope for us to be able to apply our evaluation items.

Thirty one specific tools were retrieved. Among these tools 14 were related to the retrieval and analysis of pertinent literature and writing assistance category, of these 13 were considered as candidates for real-life testing but 3 were already obsolete at the time of publication. Seventeen were related to qualitative data collection and analysis, of these 8 were considered to be candidates for real-life testing. Tools **specifically** designed to enable automation in quantitative data analysis were not specifically explored, as no need for such tools was requested by KCE experts (Box 2).

Tools that were considered “of interest” were selected for the second phase of evaluation in real-life conditions are described below. The complete list of evaluated tools can be found in the Appendix 1.5.5 .

7.4.2.1. Expedited retrieval and analysis of pertinent literature and scientific writing

The range of tools designed to aid in literature reviews is constantly growing. Although AI-assisted automation is not widely implemented in these tools, it is typically used during the screening stage to help prioritise more relevant papers. However, this type of assistance is still viewed as requiring human verification to ensure accuracy¹²⁹. Despite this limited use of AI, the user interfaces of these tools have been enhanced to allow easier and more efficient navigation through all stages of the literature review process, potentially speeding up the workflow.

Tools for management and streamlining of literature review

Tools for management and streamlining of literature review are specialised software applications designed to assist researchers in efficiently organising, conducting and analysing literature. These tools typically offer features like reference management, data extraction, annotation capabilities, and collaboration functionalities, streamlining the process of sifting through vast amounts of literature, synthesizing findings, and managing citations and references. They claim to enhance the accuracy and speed of literature reviews while reducing the manual workload involved.

In the SR toolbox^{xx} there are more than 240 tools indexed, of which 224 support health care literature reviews. We chose to evaluate three tools (Nested knowledge, EPPI reviewer, Distiller SR) on the basis of the work of Cowie et al.¹³⁰ to which we added the tool Covidence on the basis of internal discussion.

Covidence is a comprehensive tool designed specifically for systematic review management. It functions as a tool for title and abstract screening, full text screening, it assists in full text data extraction and quality appraisal for conducting systematic reviews (SR). Automation is limited to the title and abstract screening, where studies are sorted using machine learning. The manufacturers claim that the use of this tool saves an average 35% time per review. It has a new Randomised Controlled Trial (RCT) filter feature that removes non-RCT references. Additionally, Covidence automates the generation of PRISMA flow diagrams and allows for the export of quality assessments, risk of bias evaluations, and study data directly to RevMan Web, RevMan5, CSV, and Excel for comprehensive report generation. There is the possibility of trying the tool with one review with up to 500 titles for free, otherwise the price is 264 Euros/year for one review with unlimited collaborators. The price for organisations needs to be requested. The interface is user friendly and completely intuitive¹³¹.

^{xx} <http://www.systematicreviewtools.com/>



Nested Knowledge is a platform for SR and meta-analysis. The software is composed of two parts: “AutoLit” that is used to search, screen, extract data and complete critical appraisal and “Synthesis” that is used to visualise and analyse data, publish and share insights by generating automatic outputs. The cost is particularly elevated, between 295 and 695 dollars per user per month (approximately between 314 and 647 euros) for companies, although less costly subscriptions with limited features are also available. A free trial with limited features is possible. It is uncertain whether this tool is worth it compared with other less expensive alternatives¹³⁰.

EPPI reviewer is a web-based software developed by the Evidence for Policy and Practice Information and Co-ordinating Center (EPPI-Center) in London, which caters to all types of systematic reviews, including meta-analysis, framework synthesis, and thematic synthesis. This tool manages references (importing, identifying duplicates), stores PDF files, and incorporates text mining technology. It uses AI-assisted technology in the abstract screening phase, it includes study type classifiers, it incorporates the RobotReviewer automatic quality appraisal tool, it can perform automated review updates using Microsoft academics. The pricing structure includes a user fee per month and a sharable review fee. Effective use requires thorough training and a well-defined research objective. It is possible to have a limited trial user account, the fees for a shareable review is 40 Euros/month and the user fee is 12 Euros/user/month.

DistillerSR is a systematic review and literature review software that automates the management of literature collection, screening, and assessment using in part AI workflows. The platform is designed to handle dual reviewer screening, conflict resolution, capturing exclusion reasons, risk of bias assessments, duplicate detection, and reporting templates such as PRISMA. It offers features such as machine learning classifiers, NLP for information processing, and AI-powered double-checking of exclusion decisions. Like in the other streamlining tools, AI assistance is concentrated in the screening phase of the review. The software owners claim that it reduces the screening burden by 60%. The price is consequential, for

professional use it is 227 Euros/user/month and for company use it is 277 Euros/user/month.

Systematic Review Accelerator: The Systematic Review Accelerator (SRA) is a suite of automation tools designed to speed up various steps in the systematic review process. The tools that are collected in this suite include:

- Polyglot Search Translator: Automatically translates search strategies across multiple databases
- Deduplicator: Identifies and removes duplicate studies from search results,
- Screenatron: Enhances the speed of title and abstract screening by allowing reviewers to quickly include or exclude articles through customizable hotkeys and a streamlined user interface.
- Disputatron: Facilitates the resolution of disagreements between reviewers by detecting screening decision disagreements.
- Spidercite: Automatically finds and downloads citing and cited articles from an uploaded set of studies
- Revman Replicant: Automatically displays the results of the systematic review from meta-analyses (forest plots)
- Word Frequency Analyser: Assists with designing search strings to find relevant literature by analysing the frequency of words and phrases in a set of search results, aiding in the development of a refined search strategy.
- Search Refinery: Refines the search strategy to be maximally efficient in finding only relevant literature

All the tools that are included in the suite are free, but they do not automatically interconnect with one another.


Table 13 – Comparative table of the retrieved tools for management and streamlining of literature review

Feature	Covidence	DistillerSR	EPPI Reviewer	Nested Knowledge	Systematic Accelerator	Review
Target Use Case	Systematic management review	Systematic management review	All types of literature review (systematic reviews, meta-analyses, 'narrative' reviews, meta-ethnographies)	Systematic literature review and meta-analysis	All types of evidence reviews or synthesis	
Key Functions	- Dual reviewer screening - Duplicate records detection - Conflict resolution - Data extraction - Quality assessment - Configurable reporting (e.g. PRISMA) - Literature review project management	- Literature searches - Dual reviewer screening - Duplicate records detection - Conflict resolution - Data extraction - Risk of bias assessment - Configurable reporting (e.g. PRISMA) - Literature review project management	- Literature searches - Dual reviewer screening - Duplicate records detection - Dual reviewer screening - Conflict resolution - Data extraction - Risk of bias assessment - Configurable reporting (e.g. PRISMA) - Literature review project management	"AutoLit:" Search, screen, tagging, data extraction "Synthesis": Qualitative and quantitative analysis, manuscript creation	- Designs and writes review methods section - Designs, refines and translates (for different databases) search strategy - Removes duplicate records - Screens records - Forward and backward citations	
Most relevant Automation Features	- Machine learning for title and abstract screening - Assistance in the creation of evidence tables	- AI-powered pre-screening: machine learning classifiers - Integrations with data providers for full-text documents (e.g. PubMed) - AI-powered double-checking of exclusion decisions and validation categorisation of records	- Deduplication - Automatic clustering of studies 'Priority screening' for screening efficiency - Five study type classifiers (RCT, systematic review, economic evaluation, COVID-19 categories, and Long COVI - Assistance in quality appraisal - Automated review updates using Microsoft Academic	- Automation in abstract screening phase - Automation in output production after data extraction - Automation on a creation of concept hierarchy via tagging.	- Polyglot Search Translator - The Deduplicator - The Screenatron - The Disputatron - SearchRefinery - SpiderCite - RevMan Replicant - Word Frequency Analyser	



Price	264 Euros/year for one review with unlimited collaborators for one year. The price for organisations needs to be requested.	Professional use: 227 Euros/user/month Company use: 277 Euros/user/month.	Shareable review: 40 Euros/month. User fee is 12 Euros/user/month.	Company: approximately between 314 and 647 Euros/user/monthly	Free
--------------	---	--	---	---	------

Tools for forward and backward citation

CitationChaser is an open source, easy-to-use tool for quick backward and forward citation chasing, developed by Haddaway et al. It generates standardised output files that can be quickly and effectively combined with the results of bibliographic database searches to reduce duplication and increase the breadth of the pool of potentially pertinent records that can be screened within evidence synthesis¹³². The tool uses the Lens.org bibliographic database aggregator, which compiles content from five bibliographic resources: Microsoft Academic Graph (MAG), CrossRef, CORE, PubMed, and PubMed Central¹³³. It may also be useful in cases where a few core papers are identified by external experts (study evaluation phase and scoping) in order to gain rapid access to relevant related literature. The use is free of charge.

ResearchRabbit is a tool that enables to identify relevant literature, create alerts for newly-published papers and share collections with others. The starting point is one or more selected papers ("seed papers") identified by the researcher or provided by external experts. The tool provides, on the basis of the seed papers a network of relevant articles. The database contains 100s of millions of academic articles and covers more than 90% of materials that can be found in major databases used by academic institutions (such as Scopus, Web of Science, and others), and also has materials not found in these academic databases. ResearchRabbit does not create its own search engine algorithms. Instead, it uses search algorithms that have been developed by the National Institutes of Health (NIH) and Semantic Scholar. The use is free of charge.

Semantic search engines

Semantic search engines are AI-powered search engines that, in the scientific domain, are designed to perform literature search in a way that goes beyond keyword search, aiming to "understand" the context and intent behind a user's query. They might find relevant studies that might be missed by traditional keyword-based searches but they may also reflect the biases of the data the AI is trained upon. Semantic search engines are not as reproducible as keyword-based search engines in the retrieval of medical literature¹³⁴. They are often paired with other AI-assisted features such as summary of information.

Consensus.app, also known simply as Consensus, is an AI-powered search engine designed to find and summarise scientific research papers. It aims to provide users with evidence-based answers to their queries by extracting and distilling data directly from scientific research. Consensus is capable of searching through over 200 million scientific papers to provide users with relevant studies and key insights. The reliability and reproducibility of its results could not be thoroughly assessed. All features are available for a limited amount of time free of charge, a premium access with unlimited access costs USD 8.99 per month.

Perplexity.ai is an AI-powered search and answer engine (based on ChatGPT technology). Perplexity's main strengths are that it cites sources, it uses a knowledge base including webpages, documents and datasets, and it tailors results to user's interests and preferences over time (i.e., personalisation). Similar to other semantic search tools, its reliability and reproducibility could not be thoroughly assessed^{135, 136}. A limited set of features is available for free, a Pro subscription costs USD 20 per month.



Tools for PDF data extraction

As far as tools for data extraction from PDFs is concerned, it needs to be noted that ChatGPT4 has recently integrated an upload feature that allows to upload and analyse PDFs, this feature is also available free of charge using MS copilot (that uses ChatGPT4 APIs) in the web browser Edge, so tools like ChatPDF and Ask Your PDF may already be obsolete. A brief description can be found in the Appendix.

Tools for automated quality appraisal

The only tool retrieved was the free application RobotReviewer, but the website seems no longer functional and a request for further information has not currently been answered. Robot Reviewer seems not to be supported anymore as a free application. Its technology is however part of the EPPI Reviewer tool. A brief description of RobotReviewer can be found in the Appendix.

Tools for assistance in scientific writing

Jenni.ai is a tool that is designed to assist in scientific writing. It uses NLP to provide style and content suggestions while redacting. It has features such as *autocomplete*: which provides suggestions for text content based on information that is given about the field and the type of text requires. It may also retrieve citations, or manage existing ones, and place them in-text with a variety of styles. It can paraphrase and rewrite, avoiding unwanted plagiarism. the tool has also an inbuilt AI research assistant (AskJenni) that may serve as semantic search engine or can answer questions about the document that is being written. A limited set of features, including 200 AI words per day, is available free of charge. The whole range of features is available for USD 20 per month.

Apart from specific tools designed to assist in scientific writing, LLMs in general have the potential of assisting in tasks such as grammar and language checks, translation, drafting paragraphs, summarizing research papers, and generating clear and concise scientific writing.

7.4.2.2. *Expedited data collection and analysis*

Tools for audio transcription

There exist a large number of automatic transcription software that convert audio or speech to text and are using automated speech recognition (ASR) and NLP. In general, the quality of the audio recording influences the accuracy rate of the automatic transcription. Six were shortlisted as they were GDPR compliant and had their data servers located in the EU. The features of the six different software are compared in Table 9.

All automatic transcription software supported Dutch, French and English among other languages. Audio files of any size and length can be uploaded to the automatic transcription software. Most of the selected software provide a text or transcription editor to manually edit the transcribed text, including bookmarking, commenting and highlighting functions. The software can transcribe multi-speaker files and can also distinguish different speakers in the transcript. Different speakers are indicated as “speaker 1”, “speaker 2”, etc. and can be renamed in the online editor (e.g. Amberscript). Some of the software automatically add timestamps, which can also be edited in the online editor. Whisper AI^{yy} is the only system that is free and open-source, which means it can be implemented and run on local computers. In addition, Whisper AI is a multitasking model which can perform multilingual speech recognition, speech translation, and language identification.

^{yy} <https://github.com/openai/whisper> (accessed 04/12/2023)



Table 14 – Comparison of the features of different automatic transcription software

Features	Amberscript	Happyscribe	Trint	SmartScribe	Konch	Whisper AI
Free trial	Yes	Yes (10 minutes)	Yes	Yes (15 min)	Yes (15 min)	N/A
Accuracy rate (under optimal condition)*	85%	85%	99%	95%	90%	89-95% for FR, NL, EN
Online editor access	Yes	Yes	Yes	Yes	Yes	No
Server based in	Western Europe	Ireland	EU (choice)	France	EU	Local computer
Security standards	GDPR compliant and ISO certified ISO 27001 & ISO9001 certified	GDPR compliant and Tier IV, PCI DSS, and ISO 27001 compliant facility	ISO 27001:2013 certified and GDPR compliant	GDPR compliant	GDPR compliant	N/A
Use data to train AI	No	Yes	No	No	No	No
Collaboration	Multi-user account and access	Can share transcripts in view-only or edit mode	Can share transcripts with read-only access or editing and sharing permissions.	?	Multi-user account for reviewing	
Export formats	Word, json, txt	Txt, docx, pdf, html, json, maxqda, vtt, srt, stl, final cut, premiere, avid	.docx, .srt, .vtt, .txt, .stl, .edl, .html, .xml, .csv	Word, txt, json, odt	Word docx, MS Excel, csv, srt, vtt	Txt
Price	€15/hour or €40/month. (5 hours of audio/video uploaded)	Basic: €14/month (120 minutes). Business: €39/month (600 minutes)	Starter: €48 per user/month (7 files)	€10/hour	€56/year (10 hours)	Free

*Accuracy rates are based on information provided by the manufacturers of the software.



Tools for qualitative content analysis

Atlas.ti is a tool that claims fully automated code generation for interviews, survey data, or focus group material, whether it involves transcriptions, raw audio material, or video recordings. The claim is that it reduces analysis time by up to 90% increasing accuracy. It is also able to transform analysis results into meaningful graphs, networks, and diagrams. The servers are located in Berlin. An overview of the AI features (i.e., AI coding, included in Atlas.ti) is provided in [this video](#).

MAXQDA is a software package for qualitative and mixed methods research. The tool has a transcription function and can: summarise selected passages, analyse coded segments and suggest subcodes, create automatic summaries of coded segments, and customise results output. The cost is not disclosed.

NVivo which is currently being used at KCE, is also at the early stages of integrating Generative AI into their software. For instance, NVivo claims to leverage machine learning to automate data coding and transcription of audio and video recordings (90% accuracy from quality recordings). The availability of these options depend on the NVivo license.

A note on web scraping

While web scraping is not a specific tool and its application within the KCE research process remains somewhat ambiguous, we believe it is important to elucidate the concept, as it has emerged frequently in our technological watch. Web scraping refers to the automated extraction of vast quantities of data from websites. This technique automates what would otherwise be a manual process of collecting information, akin to copying and pasting data from a website. Through web scraping, data is gathered and then parsed for further processing or analysis.

Python is the preferred language for web scraping, although R, with certain packages (like rvest, RSelenium, and phantomjs), serves as an alternative. In Python, various modules are employed based on the scraping needs:

- **Beautiful Soup:** This module is designed for extracting data from structured documents (HTML or XML), such as retrieving numbers from a table on a webpage. Its use requires familiarity with the webpage's structure but allows for quick data retrieval once set up, making it ideal for pages that are updated regularly.
- **Selenium:** This module identifies the desired information regardless of the website's structure, catering to sites with dynamic content written in JavaScript. Its implementation is more complex.

Users must be aware that extracting data from websites without the owner's consent may constitute a legal offense. Furthermore, to maintain effectiveness, web scraping tools often require updates, especially when the target websites are updated or redesigned. Websites may also employ measures such as IP tracking or CAPTCHA to prevent scraping.

Web scraping is a potent method for data collection from websites, with applications spanning business, research, and other fields. It demands specific programming skills and awareness of legal and ethical boundaries. Given these considerations, we have chosen not to explore this topic further.

7.5. Discussion

Within the lifecycle of a KCE research project the options for automation are multiple, and in continuous expansion (see Figure 6 and Box 3).

A recent paper from the Cochrane Rapid Reviews Method Group by Affenburger et al.¹³⁷ has issued some recommendations in the production of rapid reviews, these recommendations may be summarised as follows:

- Utilise collaborative online platforms that enable parallel work, real-time project management, and centralized review details.
- Implement automation software to support, but not replace, human reviewers and their judgement.
- Ensure transparency in reporting methodologies and potential biases due to supportive software.



Specific considerations are given for the use of emerging software, such as LLMs. The use of emerging software should be preceded by validation studies using real-world data, and its reliability needs to be ascertained.

Our observations on the use of existing and emerging software in literature synthesis are largely in accord with the conclusions of Affengruber et al., although in our analysis we have not limited ourselves in the use of tools in the production of literature reviews, but we have expanded the scope to other aspects of the KCE-project lifecycle (See Figure 6)., therefore we have included in our work the possibility of using a wider range of software.

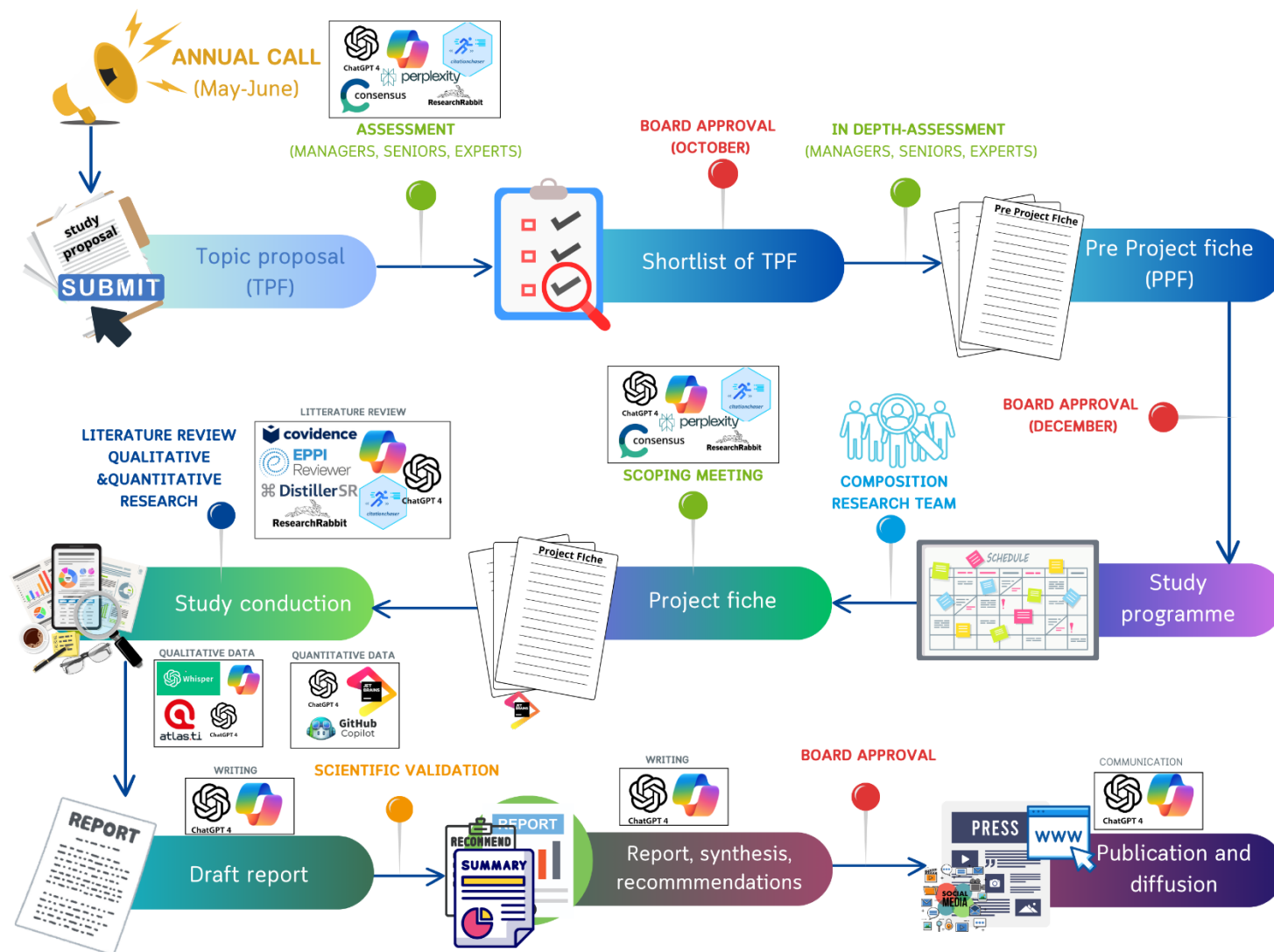
Tools for streamlining systematic literature reviews can save time by consolidating tasks typically performed across multiple platforms into a single environment, which also facilitates real-time collaboration. Currently, the use of AI-assisted tasks is mainly limited to the abstract screening phase, although some tools are being implemented in other phases as well. Among these tools, Covidence may soon be tested in a real-life KCE project.

Tools for **forward and backward citation** and **semantic search engines** may be integrated in a systematic literature review process by completing article retrieval with papers that may be missed using a keyword-based search. However, their characteristics seem more appropriate to speed up and diversify key literature retrieval when evaluating TPFs and during the scoping phase of a project. Although an AI-assisted semantic search may retrieve more relevant literature than keyword search, it currently still suffers from a lack of reproducibility. A recent paper on AI-assisted literature retrieval, still suggests the use of keyword-based search engines such as PubMed for reproducible literature screening¹³⁴.

Automated quality appraisal of the literature is technically feasible and promising, but the only standalone tool that has been evaluated for this purpose seem no longer supported by its development team, and is instead integrated in a paying tool for streamlining literature reviews. The use of such a tool could be tested in a real-life KCE project in order to evaluate its actual added value.



Figure 7 – Suggestions of automated tools that can be used in the lifecycle of a KCE project





Box 3 – Examples of how automated tools can be used in the lifecycle of a KCE project

Evaluation of TPF: to search for/check relevant articles

Scoping: to search for/check relevant articles and obtain summaries of key articles, to help with brainstorming

Literature review: to identify relevant articles and extract key information, to accelerate the process of systematic literature review

Qualitative data: to provide suggestions for design and content of interview guide/protocols, to transcribe interviews, to support with data analysis.

Quantitative data: to provide suggestions for survey design, support code generation, data analysis and visualization.

Writing: to improve sentences, to summarise scientific reports, to assist in translation tasks, to support the generation of synthesis in French and Dutch.

Communication: to develop text for social media posts, to generate pictures for the scientific report's cover page and social media posts.

Generative AI can be used in a variety of tasks: from assistance to writing (help with rephrasing, increase readability), to assistance with compiling text summaries, to more complex tasks such as assistance to coding and data analysis. Royer et al. (ISPOR Europe congress proceedings) used ChatGPT 3.5 to assess abstracts by designing a set of clear and concise questions for each inclusion and exclusion criterion, where ChatGPT 3.5 responses were converted into binary outputs, that computed a score, based on which researchers were able to exclude 25% of abstracts with a model recall rate (sensitivity) of 95.8%, without requiring any kind of training whatsoever on behalf of the AI¹³⁸. A sensitivity of 95.9% for title and abstract screening and of 97.1% for full text screening were achieved with GTP4¹³⁹. This suggests that with refined and precise prompting, LLM models may be able to achieve a certain degree of accuracy in retrieving data from texts, potentially expediting the processing and data extraction from documents. Carefully

crafted prompts used with AI-assisted interaction with PDFs may be useful for this purpose, but need to be developed internally, and could be published in the appendices of the report's Method section.

Some **questions remain open about the use of generative AI**, including concerns about its accuracy due to the potential for generating unreliable or non-factual output. There are also issues surrounding research integrity, as, for instance, there is no consensus on the acceptable degree of LLM assistance in writing. Additionally, explainability is a concern, given that LLMs are opaque "black box" tools that can be difficult to integrate into the scientific method. Moreover, bias in the training data may lead to reflected biases in the produced output. Finally, there is a question about how overuse or overreliance on LLMs may lead to deskilling and a lack of critical thinking in scientific writing¹⁴⁰.

Therefore the use of generative AI needs to be responsible. This includes being mindful of the fact that content produced by LLMs may be biased, fabricated ("hallucinated") or subject to copyright. We also need to be mindful of the input data that we provide LLMs, avoiding the sharing of confidential or sensitive information. The risk of producing biased or hallucinated content can be mitigated by the use of **prompt engineering** techniques, that are usually employed to enhance prompt efficacy to get the LLM to do determined tasks. Effective prompts, once they have proven their usefulness and reliability, may be shared in a common library and reused in similar circumstances.

We should also keep in mind that AI is very high in **energy costs**,¹⁴¹ so, to be environmentally conscious, its routine use should be limited to those tasks where its added value is proven.



7.6. Conclusions

This chapter provides an overview of the potential use of automation tools in the lifecycle of a KCE project. The field of automation and AI is in full expansion, and some of the investigated tools were already obsolete by the end of writing this chapter, while others keep emerging.

This rapidly evolving landscape, together with the ethical and methodological questions that arise, warrant a constant **technological watch** on behalf of our research team. We suggest to **create a shared repository of efficient automated strategies that may be incorporated in our process** (including data analysis code snippets and prompts of proven efficacy). We suggest also the creation of a living document where we can collaborate and report our experiences and reflections on automation, within the institution and with the wider public.



■ BIBLIOGRAPHY

1. Garritty C, Nussbaumer-Streit B, Hamel C, Devane D, Cochrane Rapid Reviews Methods G. Rapid reviews methods series: assessing the appropriateness of conducting a rapid review. *BMJ Evid Based Med*. 2024.
2. Garritty C, Hamel C, Trivella M, Gartlehner G, Nussbaumer-Streit B, Devane D, et al. Updated recommendations for the Cochrane rapid review methods guidance for rapid reviews of effectiveness. *BMJ*. 2024;384:e076335.
3. Garritty C, Tricco AC, Smith M, Pollock D, Kamel C, King VJ, et al. Rapid Reviews Methods Series: Involving patient and public partners, healthcare providers and policymakers as knowledge users. *BMJ Evid Based Med*. 2023.
4. Moher D, Shamseer L, Clarke M, Gherzi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.
5. Klerings I, Robalino S, Booth A, Escobar-Liquitay CM, Sommer I, Gartlehner G, et al. Rapid reviews methods series: Guidance on literature search. *BMJ Evid Based Med*. 2023.
6. Nussbaumer-Streit B, Klerings I, Wagner G, Heise TL, Dobrescu AI, Armijo-Olivo S, et al. Abbreviated literature searches were viable alternatives to comprehensive searches: a meta-epidemiological study. *J Clin Epidemiol*. 2018;102:1-11.
7. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Med Res Methodol*. 2016;16(1):127.
8. Marshall IJ, Marshall R, Wallace BC, Brassey J, Thomas J. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. *J Clin Epidemiol*. 2019;109:30-41.
9. Affengruber L, Wagner G, Waffenschmidt S, Lhachimi SK, Nussbaumer-Streit B, Thaler K, et al. Combining abbreviated



- literature searches with single-reviewer screening: three case studies of rapid reviews. *Syst Rev*. 2020;9(1):162.
10. Klein-Fedyshin M, Ketchum AM. PubMed's core clinical journals filter: redesigned for contemporary clinical impact and utility. *J Med Libr Assoc*. 2023;111(3):665-76.
 11. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40-6.
 12. Gianola S, Jesus TS, Barger S, Castellini G. Characteristics of academic publications, preprints, and registered clinical trials on the COVID-19 pandemic. *PLoS One*. 2020;15(10):e0240123.
 13. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Med Res Methodol*. 2017;17(1):64.
 14. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet*. 2000;356(9237):1228-31.
 15. Regev-Yochay G, Lustig Y, Joseph G, Gilboa M, Barda N, Gens I, et al. Correlates of protection against COVID-19 infection and intensity of symptomatic disease in vaccinated individuals exposed to SARS-CoV-2 in households in Israel (ICoFS): a prospective cohort study. *Lancet Microbe*. 2023;4(5):e309-e18.
 16. Gilbert PB, Donis RO, Koup RA, Fong Y, Plotkin SA, Follmann D. A Covid-19 Milestone Attained - A Correlate of Protection for Vaccines. *N Engl J Med*. 2022;387(24):2203-6.
 17. Plotkin SA. Recent updates on correlates of vaccine-induced protection. *Front Immunol*. 2022;13:1081107.
 18. Andrews NJ, Waight PA, Burbidge P, Pearce E, Roalfe L, Zancolli M, et al. Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study. *Lancet Infect Dis*. 2014;14(9):839-46.
 19. Williamson DA, Bray GA, Ryan DH. Is 5% weight loss a satisfactory criterion to define clinically significant weight loss? *Obesity (Silver Spring)*. 2015;23(12):2319-20.
 20. Ciani O, Manyara AM, Davies P, Stewart D, Weir CJ, Young AE, et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine*. 2023;65:102283.
 21. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med*. 2007;357(21):2109-22.
 22. Investigators A-H. The role of niacin in raising high-density lipoprotein cholesterol to reduce cardiovascular events in patients with atherosclerotic cardiovascular disease and optimally treated low-density lipoprotein cholesterol: baseline characteristics of study participants. The Atherothrombosis Intervention in Metabolic syndrome with low HDL/high triglycerides: impact on Global Health outcomes (AIM-HIGH) trial. *Am Heart J*. 2011;161(3):538-43.
 23. la Cour JL, Brok J, Gotzsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. *BMJ*. 2010;341:c3653.
 24. Grey P, Grey A, Bolland MJ. Outcomes, Interventions and Funding in Randomised Research Published in High-Impact Journals. *Trials*. 2018;19(1):592.
 25. Villena R, Kriz P, Tin Tin Htar M, Burman C, Findlow J, Balmer P, et al. Real-world impact and effectiveness of MenACWY-TT. *Hum Vaccin Immunother*. 2023;19(2):2251825.
 26. Smith PG. Retrospective assessment of the effectiveness of BCG vaccination against tuberculosis using the case-control method. *Tubercle*. 1982;63(1):23-35.



27. FDA. Emergency Use Authorization (EUA) for EVUSHELD [Web page].2022 [cited 2 May]. Available from: <https://www.fda.gov/media/156674/download>
28. FDA. FDA announces Evusheld is not currently authorized for emergency use in the U.S. [Web page].2023 [cited 2 May]. Available from: <https://www.fda.gov/drugs/drug-safety-and-availability/fda-announces-evusheld-not-currently-authorized-emergency-use-us>
29. Fernandez-Felix BM, Lopez-Alcalde J, Roque M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. *BMC Med Res Methodol*. 2023;23(1):44.
30. Thompson J, McClure R, Scott N, Hellard M, Abeysuriya R, Vidanaarachchi R, et al. A framework for considering the utility of models when facing tough decisions in public health: a guideline for policy-makers. *Health Res Policy Syst*. 2022;20(1):107.
31. Ciolino JD, Kaizer AM, Bonner LB. Guidance on interim analysis methods in clinical trials. *J Clin Transl Sci*. 2023;7(1):e124.
32. Borg Debono V, Mbuagbaw L, Paul J, Buckley N, Thabane L. Sharing some interim data in trial monitoring can mislead or unmask trial investigators: A scenario-based survey of trial experts. *Contemp Clin Trials Commun*. 2017;7:81-5.
33. Counsell N, Biri D, Fraczek J, Hackshaw A. Publishing interim results of randomised clinical trials in peer-reviewed journals. *Clin Trials*. 2017;14(1):67-77.
34. Wayant C, Vassar M. A comparison of matched interim analysis publications and final analysis publications in oncology clinical trials. *Ann Oncol*. 2018;29(12):2384-90.
35. Jayk Bernal A, Gomes da Silva MM, Musungaie DB, Kovalchuk E, Gonzalez A, Delos Reyes V, et al. Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *N Engl J Med*. 2022;386(6):509-20.
36. Kampmann B, Madhi SA, Munjal I, Simoes EAF, Pahud BA, Llapur C, et al. Bivalent Prefusion F Vaccine in Pregnancy to Prevent RSV Illness in Infants. *N Engl J Med*. 2023;388(16):1451-64.
37. Gewandter JS, Kitt RA, Hunsinger MR, Poku J, Lozano J, Chaudari J, et al. Reporting of data monitoring boards in publications of randomized clinical trials is often deficient: ACTION systematic review. *J Clin Epidemiol*. 2017;83:101-7.
38. Van Norman GA. Data safety and monitoring boards. Should be required for both early-and late-phase clinical trials. *JACC: Basic to Translational Science*. 2021;6(11):9.
39. Garner P, Hopewell S, Chandler J, MacLehose H, Schunemann HJ, Akl EA, et al. When and how to update systematic reviews: consensus and checklist. *BMJ*. 2016;354:i3507.
40. Tugwell P, Welch VA, Karunanathan S, Maxwell LJ, Akl EA, Avey MT, et al. When to replicate systematic reviews of interventions: consensus checklist. *BMJ*. 2020;370:m2864.
41. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med*. 2007;147(4):224-33.
42. Cumpston MF, E. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors) *Cochrane Handbook for Systematic Reviews of Interventions version 6.4- Chapter IV: Updating a review*. 2023 2023. Available from: <https://training.cochrane.org/handbook/current/chapter-iv>
43. Kim JSM, Pollock M, Kaunelis D, Weeks L. Guidance on review type selection for health technology assessments: key factors and considerations for deciding when to conduct a de novo systematic review, an update of a systematic review, or an overview of systematic reviews. *Syst Rev*. 2022;11(1):206.
44. Robinson KA, Chou R, Berkman ND, Newberry SJ, Fu R, Hartling L, et al. Twelve recommendations for integrating existing systematic



- reviews into new reviews: EPC guidance. *J Clin Epidemiol*. 2016;70:38-44.
45. Roberfroid D, Fairon N, San Miguel L, Paulus D. Method – Rapid reviews. Brussels: Belgian Health Care Knowledge Centre (KCE); 2016. KCE Process Notes D/2017/10.273/01 Available from: https://kce.fgov.be/sites/default/files/2021-11/Rapid_Review_0.pdf
46. Chung M, Newberry SJ, Ansari MT, Yu WW, Wu H, Lee J, et al. Two methods provide similar signals for the need to update systematic reviews. *J Clin Epidemiol*. 2012;65(6):660-8.
47. Goossen K, Bieler D, Hess S, Becker M, Kalsen M, Flohe S, et al. An adapted 'Ottawa' method allowed assessing the need to update topic areas within clinical practice guidelines. *J Clin Epidemiol*. 2022;150:1-11.
48. Higgins JP, Lasserson T, Thomas J, Flemyng E, Churchill R. Methodological Expectations of Cochrane Intervention Reviews. . London: Cochrane; 2023.
49. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
50. Banzi R, Cinquini M, Gonzalez-Lorenzo M, Pecoraro V, Capobussi M, Minozzi S. Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. *J Clin Epidemiol*. 2018;99:24-32.
51. Gómez-García F, Ruano J, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, et al. Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool. *J Clin Epidemiol*. 2017;92:79-88.
52. Cochrane Collaboration. ROBINS-I tool [Web page].2017 [cited 2 May]. Available from: <https://methods.cochrane.org/robins-i>
53. Schunemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol*. 2019;111:105-14.
54. Rivoirard R, Langrand-Escure J, Oriol M, Tinquaut F, Chauvin F, Rancoule C, et al. Evaluation of the quality of the reporting of phase II clinical trials in oncology: A systematic review. *Crit Rev Oncol Hematol*. 2018;125:78-83.
55. Grellety T, Petit-Moneger A, Diallo A, Mathoulin-Pelissier S, Italiano A. Quality of reporting of phase II trials: a focus on highly ranked oncology journals. *Ann Oncol*. 2014;25(2):536-41.
56. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res*. 2020;7(1):7.
57. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol*. 2014;14:43.
58. Paiva Barbosa V, Bastos Silveira B, Amorim Dos Santos J, Monteiro MM, Coletta RD, De Luca Canto G, et al. Critical appraisal tools used in systematic reviews of in vitro cell culture studies: A methodological study. *Res Synth Methods*. 2023;14(6):776-93.
59. Gartlehner G, Nussbaumer-Streit B, Devane D, Kahwati L, Viswanathan M, King VJ, et al. Rapid reviews methods series: Guidance on assessing the certainty of evidence. *BMJ Evid Based Med*. 2023.
60. Vindrola-Padros C, Brage E, Johnson GA. Rapid, Responsive, and Relevant?: A Systematic Review of Rapid Evaluations in Health Care. *American Journal of Evaluation*. 2021;42(1):13-27.
61. Vindrola-Padros C. Doing Rapid Qualitative Research. First ed. London: SAGE Publications Ltd;; 2021.
62. Tan RKJ, Lim JM, Lo JJ, Teo AKJ, O'Hara CA, Ching AH, et al. Conducting rapid qualitative research to support sex workers' health



- and social needs in the face of COVID-19: capitalising on stakeholder networks from the HIV response in Singapore to drive policymaking (62). *Sex Transm Infect.* 2021;97(2):84.
63. Skillman M, Cross-Barnet C, Friedman Singer R, Rotondo C, Ruiz S, Moiduddin A. A Framework for Rigorous Qualitative Research as a Component of Mixed Method Rapid-Cycle Evaluation (92). *Qual Health Res.* 2019;29(2):279-89.
64. McNall M, Foster-Fishman PG. Methods of Rapid Evaluation, Assessment, and Appraisal. *American Journal of Evaluation.* 2007;28(2):151-68.
65. Bartsch D, Belgacem N. Real time evaluation of UNHCR's response to the emergency in Chad. Geneva, Switzerland: UNHCR; 2004. EPAU/2004/07 Available from: <https://www.unhcr.org/africa/sites/afr/files/legacy-pdf/4119d9c54.pdf>
66. Jamal A, Crisp J. Real-time Humanitarian evaluations: Some Frequently Asked Questions Geneva: United Nations High Commissioner for Refugees, Evaluation and Policy Unit; 2002. EPAU /2002/05 Available from: <http://www.unhcr.org/research/RESEARCH/3ce372204.pdf>
67. Sandison P. Desk review of real-time evaluation experience. New York: UNICEF; 2003. Available from: <https://www.livestock-emergency.net/userfiles/file/common-standards/Sandison-2003.pdf>
68. McNall MA, Welch VE, Ruh KL, Mildner CA, Soto T. The use of rapid-feedback evaluation methods to improve the retention rates of an HIV/AIDS healthcare intervention. *Evaluation and Program Planning.* 2004;27(3):287-94.
69. Sonnichsen R. High Impact Internal Evaluation: A Practitioner's Guide to Evaluating and Consulting Inside Organizations. London: Sage; 2000.
70. Wholey JS. Evaluation and effective public management. Boston: Little, Brown; 1983.
71. Beebe J. Rapid Assessment Process: An Introduction. AltaMira Press, U.S.; 2001.
72. Trotter R, II, Singer M. Rapid assessment strategies for public health: promise and problems. In: Trickett E, Pequegnat W, editors. *Community intervention and AIDS.* New York: Oxford University Press; 2005. p. 130-52.
73. Cornelis J, Dauvrin M, Desomer A, Gerkens S, Primus-de Jong C, Christiaens W. Organisation des soins médicaux en cas d'incidents CBRNE : Guidance. Health Services Research (HSR). Bruxelles: Centre Fédéral d'Expertise des Soins de Santé (KCE); 2023. KCE Reports (374BS)
74. Bentley ME, Pelto GH, Straus WL, Schumann DA, Adegbola C, de la Pena E, et al. Rapid ethnographic assessment: applications in a diarrhea management program. *Soc Sci Med.* 1988;27(1):107-16.
75. Guerrero ML, Morrow RC, Calva JJ, Ortega-Gallegos H, Weller SC, Ruiz-Palacios GM, et al. Rapid ethnographic assessment of breastfeeding practices in periurban Mexico City. *Bull World Health Organ.* 1999;77(4):323-30.
76. Kresno S, Harrison GC, Sutrisna B, Reingold A. Acute respiratory illnesses in children under five years in Indramayu, West Java, Indonesia: A rapid ethnographic assessment. *Medical Anthropology.* 1993;15(4):425-34.
77. Anker M, Guidotti RJ, Orzeszyna S, Sapirie SA, Thuriaux MC. Rapid evaluation methods (REM) of health services performance: methodological observations. *Bull World Health Organ.* 1993;71(1):15-21.
78. Johnson GA, Vindrola-Padros C. Rapid qualitative research methods during complex health emergencies: A systematic review of the literature (102). *Soc Sci Med.* 2017;189:63-75.
79. Krumkamp R, Mounier-Jack S, Ahmad A, Reintjes R, Coker R. Evaluating health systems' preparedness for emerging infectious



- diseases: A novel conceptual and analytic framework. *Health Policy*. 2010;98(2):91-7.
80. Lee-Kwan SH, DeLuca N, Adams M, Dalling M, Drevlow E, Gassama G, et al. Support services for survivors of ebola virus disease - Sierra Leone, 2014. *MMWR Morb Mortal Wkly Rep*. 2014;63(50):1205-6.
81. Carrión Martín AI, Derrough T, Honomou P, Kolie N, Diallo B, Koné M, et al. Social and cultural factors behind community resistance during an Ebola outbreak in a village of the Guinean Forest region, February 2015: a field experience. *Int Health*. 2016;8(3):227-9.
82. Kohn L, Christiaens W. The use of Qualitative Research Methods in KCE studies. *Methods*. Brussels: Belgian Health Care Knowledge Centre (KCE); 2012. KCE Reports (187C)
83. Challenor J, Georgaca E, Aloneftis R, Dlodlo N, Curran H. Listening for What Is Not Being Said: Using Discourse Analytic Approaches in Mental Health Research. In: Borcsa M, Willig C, editors. *Qualitative Research Methods in Mental Health: Innovative and Collaborative Approaches*. Cham: Springer International Publishing; 2021. p. 105-33. Available from: https://doi.org/10.1007/978-3-030-65331-6_6
84. Hanquet G, Benahmed N, Castanares-Zapatero D, Dauvrin M, Desomer A, Rondia K. Report on Post-intensive Care Syndrome (PICS) for general practitioners. COVID-19 KCE Contributions. Brussels: Belgian Health Care Knowledge Centre (KCE); 2020. COVID-19 KCE Contributions
85. Salomao A. Recherche primaire et recherche secondaire : Une analyse comparative [Web page]. *Mind the Graph*;2023 [updated 2023/01/09; cited 03/06/2024]. Available from: <https://mindthegraph.com/blog/fr/quest-ce-que-le-raisonnement-deductif-2/>
86. Adriaenssens J, Maertens de Noordhout C, Kohn L, Castanares-Zapatero D. Prévention du syndrome post-soins intensifs (PICS). Health Services Research (HSR). Bruxelles: Centre Fédéral d'Expertise des Soins de Santé (KCE); 2023. KCE Reports (364B)
87. Vindrola-Padros C, Chisnall G, Cooper S, Dowrick A, Djellouli N, Symmons SM, et al. Carrying Out Rapid Qualitative Research During a Pandemic: Emerging Lessons From COVID-19 (74). *Qual Health Res*. 2020;30(14):2192-204.
88. Higham R, Pini S, Quyn A, Kowal M, Helliwell J, Saman R, et al. Rapid qualitative analysis in a mixed-methods evaluation of an infection prevention intervention in a UK hospital setting during the COVID-19 pandemic: A discussion of the CLEAN study methodology (15). *Front Sociol*. 2022;7:958250.
89. Eguluz I, Sy A, Brage E, Gonzalez-Aguero M. Rapid qualitative health research from the Global South: Reflections and learnings from Argentina, Brazil, Chile, and Mexico during the COVID-19 pandemic (17). *Front Sociol*. 2022;7:983303.
90. Schopper D, Upshur R, Matthys F, Singh JA, Bandewar SS, Ahmad A, et al. Research ethics review in humanitarian contexts: the experience of the independent ethics review board of Médecins Sans Frontières. *PLoS Med*. 2009;6(7):e1000115.
91. Khayyat S, Walters P, Whittlesea C, Nazar H. Patient and public perception and experience of community pharmacy services post-discharge in the UK: a rapid review and qualitative study (57). *BMJ Open*. 2021;11(3):e043344.
92. Dos Santos Marques IC, Theiss LM, Johnson CY, McLin E, Ruf BA, Vickers SM, et al. Implementation of virtual focus groups for qualitative data collection in a global pandemic. *Am J Surg*. 2021;221(5):918-22.
93. Khan ME, Manderson L. Focus Groups in Tropical Diseases Research. *Health Policy and Planning - HEALTH POLICY PLANN*. 1992;7:56-66.
94. Graetz DE, Sniderman E, Villegas C, Ragab I, Laptsevich A, Maliti B, et al. Utilizing Multilingual Methods and Rapid Analysis for Global Qualitative Research During a Pandemic (22). *Glob Qual Nurs Res*. 2022;9:23333936221080969.



95. Scrimshaw SCM, Hurtado E. Rapid assessment procedures for nutrition and primary health care : anthropological approaches to improving programme effectiveness / Susan C.M. Scrimshaw and Elena Hurtado. 1989:ix, 70 p. :.
96. Green J, Thorogood N. Qualitative methods for health research. 4th ed. London: Sage Publications Ltd; 2018.
97. Singh H, Tang T, Thombs R, Armas A, Nie JX, Nelson MLA, et al. Methodological Insights From a Virtual, Team-Based Rapid Qualitative Method Applied to a Study of Providers' Perspectives of the COVID-19 Pandemic Impact on Hospital-To-Home Transitions (21). *Int J Qual Methods*. 2022;21:16094069221107144.
98. Vinck I, Benahmed N, Dauvrin M, Desomer A, Cornelis J, Jonckheer P, et al. Assessment and support of decisional capacity in persons with dementia or mental health problems. *KCE Reports*. 2022.
99. Ricour C, Desomer A, Dauvrin M, Devos C. How to improve Access to health care for people with intellectual disabilities? Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE); 2022. *KCE Reports* (361C)
100. Van den Heede K, Bouckaert N, Detollenaere J, Kohn L, Maertens de Noordhout C, Vanhooreweghe J, et al. Nurse staffing on Belgian intensive care units: the impact of two years of COVID-19 pandemic. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE); 2022. *KCE Reports* (353)
101. Benahmed N, Lefèvre M, Christiaens W, Devos C, Stordeur S. Towards integrated antenatal care for low-risk pregnancy. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE); 2019. *KCE Reports* (326)
102. Bruyère O, Kaux J-F, Locquet M. Traduction et validation d'un questionnaire. *KSI*. 2020:18-22.
103. Wanat M, Borek AJ, Pilbeam C, Anthierens S, Tonkin-Crine S. Conducting rapid qualitative interview research during the COVID-19 pandemic-Reflections on methodological choices (19). *Front Sociol*. 2022;7:953872.
104. Shenton A. Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*. 2004;22:63-75.
105. Roberti J, Jorro F, Rodriguez V, Belizan M, Arias P, Ratto ME, et al. Theory-driven, rapid formative research on quality improvement intervention for critical care of patients with COVID-19 in Argentina (46). *Glob Qual Nurs Res*. 2021;8:23333936211015660.
106. Horwood J, Pithara C, Lorenc A, Kesten JM, Murphy M, Turner A, et al. The experience of conducting collaborative and intensive pragmatic qualitative (CLIP-Q) research to support rapid public health and healthcare innovation (18). *Front Sociol*. 2022;7:970333.
107. Szilassy E, Barbosa EC, Dixon S, Feder G, Griffiths C, Johnson M, et al. PRimary care rEsponse to domestic violence and abuse in the COvid-19 panDEmic (PRECODE): protocol of a rapid mixed-methods study in the UK (52). *BMC Fam Pract*. 2021;22(1):91.
108. Keniston A, Patel V, McBeth L, Bowden K, Gallant A, Burden M. The impact of surge adaptations on hospitalist care teams during the COVID-19 pandemic utilizing a rapid qualitative analysis approach (23). *Arch Public Health*. 2022;80(1):57.
109. Nagawa CS, Lane IA, McKay CE, Kamberi A, Shenette LL, Kelly MM, et al. Use of a Rapid Qualitative Method to Inform the Development of a Text Messaging Intervention for People With Serious Mental Illness Who Smoke: Formative Research Study (16). *JMIR Form Res*. 2022;6(11):e40907.
110. Oliver DG, Serovich JM, Mason TL. Constraints and Opportunities with Interview Transcription: Towards Reflection in Qualitative Research. *Soc Forces*. 2005;84(2):1273-89.
111. Gilchrist K, Iqbal S, Vindrola-Padros C. The role of patient and public involvement in rapid qualitative studies: Can we carry out meaningful PPIE with time pressures? (14). *Res Involv Engagem*. 2022;8(1):67.



112. Pope C, Mays N. Qualitative methods in health research. In: Pope C, Mays N, editors. *Qualitative research in health care*. 3rd ed: Blackwell Publishing; BMJ Books; 2006. p. 1-11.
113. Palaganas EC, Sanchez MC, Molintas MP, Caricativo RD. Reflexivity in Qualitative Research: A Journey of Learning. *The Qualitative Report*. 2017;22(2):426-38.
114. Rankl F, Johnson GA, Vindrola-Padros C. Examining What We Know in Relation to How We Know It: A Team-Based Reflexivity Model for Rapid Qualitative Health Research (56). *Qual Health Res*. 2021;31(7):1358-70.
115. Bielinska AM, Archer S, Darzi A, Urch C. Co-designing an intervention to increase uptake of advance care planning in later life following emergency hospitalisation: a research protocol using accelerated experience-based co-design (AEBCD) and the behaviour change wheel (BCW) (32). *BMJ Open*. 2022;12(5):e055347.
116. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res*. 2017;19(11):e367.
117. Borcsa M, Willig C, Schröder-Werner S. Introduction: Qualitative Research in Mental Health—Innovation and Collaboration. In; 2021. p. 1-11.
118. Nunns H. Responding to the demand for quicker evaluation findings. *Social Policy Journal of New Zealand*. 2009(34):89-99.
119. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*. 2017;7(2):e012545.
120. Westgate MJ, Haddaway NR, Cheng SH, McIntosh EJ, Marshall C, Lindenmayer DB. Software support for environmental evidence synthesis. (2397-334X (Electronic)).
121. Gautamjeet Singh Mangat NP, Sugandh Sharma, Rito Bergemann. Do health technology assessment (HTA) bodies recommend the conduct and submission of artificial intelligence-based literature reviews (AILRs)? ISPOR (International Society of Pharmacoeconomics and Outcome Research); 2023.
122. Bubeck SaVCaREaJGaEHaEKaPLaYTLaYLaSLa. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv*. 2023.
123. Chen B, Zhang Z, Langren'e N, Zhu S. Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review. *ArXiv*. 2023;abs/2310.14735.
124. Kaddour JaJHaMMaHBaRRaRM. Challenges and Applications of Large Language Models. *arXiv*. 2023.
125. guide Pe. General Tips for Designing Prompts [Web page]. Available from: <https://www.promptingguide.ai/introduction/tips>
126. Putheti VK. What exactly is Prompt Engineering and how exactly it is useful while using Generative AI tools like ChatGPT [Web page]. Medium;2024. Available from: <https://medium.com/nerd-for-tech/what-exactly-is-prompt-engineering-and-how-exactly-it-is-useful-while-using-generative-ai-tools-18f207add75e>
127. Amatriain X. Prompt Design and Engineering: Introduction and Advanced Methods. *Arxiv*. 2024.
128. Kojima TaSSGaMRaYMaYI. Large Language Models are Zero-Shot Reasoners. *arXiv*. 2023.
129. Yao X, Kumar MV, Su E, Flores Miranda A, Saha A, Sussman J. Evaluating the efficacy of artificial intelligence tools for the automation of systematic reviews in cancer research: A systematic review. *Cancer Epidemiol*. 2024;88:102511.
130. Cowie K, Rahmatullah A, Hardy N, Holub K, Kallmes K. Web-Based Software Tools for Systematic Literature Review in Medicine: Systematic Search and Feature Analysis. *JMIR Med Inform*. 2022;10(5):e33219.



131. Kellermeyer Liz HB, Knight Shandra. Covidence and Rayyan. Journal of the Medical Library Association. 2018;106(4).
132. Mahmić-Kaknjó MA-O, Tomić V, Ellen ME, Nussbaumer-Streit B, Sfetcu R, Baladia E, et al. Delphi survey on the most promising areas and methods to improve systematic reviews' production and updating. (2046-4053 (Electronic)).
133. Haddaway NA-O, Grainger MA-O, Gray CT. Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. (1759-2887 (Electronic)).
134. Jin Q, Leaman R, Lu Z. PubMed and beyond: biomedical literature search in the age of artificial intelligence. (2352-3964 (Electronic)).
135. Uppalapati VK, Nag DS. A Comparative Analysis of AI Models in Complex Medical Decision-Making Scenarios: Evaluating ChatGPT, Claude AI, Bard, and Perplexity. (2168-8184 (Print)).
136. Sanii Ry Fau - Kasto JK, Kasto Jk Fau - Wines WB, Wines Wb Fau - Mahylis JM, Mahylis Jm Fau - Muh SJ, Muh SJ. Utility of Artificial Intelligence in Orthopedic Surgery Literature Review: A Comparative Pilot Study. (1938-2367 (Electronic)).
137. Affengruber L, Nussbaumer-Streit B, Hamel C, Van der Maten M, Thomas J, Mavergames C, et al. Rapid review methods series: Guidance on the use of supportive software. BMJ Evid Based Med. 2024.
138. Royer JaA, R and Parravano,S and Pathare,U and Kisielinska, M. Prospects for Automation of Systematic Literature Reviews (SLRs) with Artificial Intelligence and Natural Language Processing. ISPOR Europe; 2023.
139. Reason T, MSc, Langham, Julia, Ph.D, Gimblett, Andy, Ph.D, Malcolm, Bill, MSc, Klijn, Sven, MSc. Breaking Through Limitations: Enhanced Systematic Literature Reviews With Large Language Models. ISPOR Europe; 2023.
140. Birhane A, Kasirzadeh A, Leslie D, Wachter S. Science in the age of large language models. Nature Reviews Physics. 2023;5(5):277-80.
141. Bender EaG, Timnit and McMillan-Major, Angelina and Schitchell, Shmargaret. On the dangers of stochastic parrots: Can Language Models Be Too Big? : Association for Computing Machinery; 2021.
142. Yadav K, Lewis RJ. Immortal Time Bias in Observational Studies. JAMA. 2021;325(7):686-7.
143. Jardim PSJ, Rose CJ, Ames HM, Echavez JFM, Van de Velde S, Muller AE. Automating risk of bias assessment in systematic reviews: a real-time mixed methods comparison of human researchers to a machine learning system. (1471-2288 (Electronic)).



■ APPENDIX

APPENDIX 1. COMPARISON OF QUALITY APPRAISAL TOOLS

Below (Table 14) is a summary of the different items assessed by five different tools to assess bias in cohort studies: the Newcastle-Ottawa scale, the Robins-I tool (advised by Cochrane Collaboration), the current KCE process book assessment tool for cohort studies (available at <https://processbook.kce.be/74-critical-appraisal-observational-studies>), the Joanna Briggs Institute (JBI) appraisal tool for cohort studies and the Scottish Intercollegiate Guideline Network (SIGN) critical appraisal checklist for cohort studies.⁵⁶ The most commonly used tool is the NOS.⁵⁶ Other available assessment tools are the CASP cohort study checklist and the NIH quality assessment tool; these were not included in the comparison.

Table 15 compares the different tools according to their respective advantages and disadvantages. The Robins-I tool is arguably the most complete tool, but the wording is often very obtuse and it is generally difficult to follow. A more straightforward tool or checklist, focusing first on common types of bias before elaborating on very specific and complex types of bias, would be desirable.

None of the tools below mention or account for immortal time bias, which is increasingly recognized as an important type of bias in non-randomized intervention studies.¹⁴²



Table 15 – Comparison of the items included in the different assessment tools, by type of bias

	Newcastle-Ottawa scale	Robins-I tool	JB1 appraisal tool for cohort studies	SIGN critical appraisal checklist	KCE assessment tool
Selection bias	- Representativeness of the exposed cohort - Selection of the non-exposed cohort: from the same community as the exposed cohort? - Assessment of presence of outcome at the time of enrolment	- Selection into the study based on characteristics observed after the intervention? (selection related to an effect of the intervention/cause of the outcome and an effect of the outcome/cause of the outcome) - Do start of FU and intervention coincide? (is some FU time missed?) - o Were adjustment techniques used to adjust for this? - Were participants excluded due to missing data on intervention status (deviation from intended study sample) or excluded from analysis due to missing data on confounders? - o Proportion and reasons similar across cohorts? - o Results robust to presence of missing data?	- Were the two groups similar and recruited from the same population? - Assessment of presence of outcome at time of enrolment	- The two groups being studied are selected from the same source populations that are comparable in all respects other than the intervention - The study indicates how many people were asked to take part and did so (participation rate) - Likelihood that some subjects might have the outcome at the time of enrolment is assessed and taken into account in analysis - Are the results of this study directly applicable to the patient group in mind? (= question on representativeness?)	Assessed in this tool by considering: - Possibility of confounding by indication - Participation rate Elsewhere, there is also an assessment of presence of outcome at the time of enrolment (referred to as performance bias)
Confounding	- Comparability of the cohorts: does the study control for the most important factor (one star) and any additional factor (one star)	- Is there potential for confounding? - Assessment of time-varying confounding (when participants could switch between intervention groups); were switches related to factors prognostic	- Were confounding factors identified? - Were strategies to deal with confounding factors stated?	- The main potential confounders are identified and taken into account in design and analysis	Focuses on ways to decrease the impact of confounding: - Matching



		for the outcome? If so, was appropriate analysis method used to adjust for this? (inverse probability weighting) - Did the authors use appropriate analysis? - Measurement bias in confounders? - Controlling for post-intervention variables that could have been affected by the intervention? (collider bias)			- Statistical techniques (regression, stratification) - Confounding should be assessed separately for each outcome	
Detection bias / measurement bias	- Ascertainment of exposure: secure record or structured interview (one start) vs written self-report or no description - Assessment of outcome: one star for independent blind and record linkage; no star for self-report or no description - Follow-up long enough to measure all outcomes?	Measurement of intervention: - interventions well defined (dose, frequency, timing)? - Was the information used to define the intervention groups recorded at the start of the intervention; if no, could it have been affected by knowledge of the outcome? (differential misclassification) - Deviations from intended interventions due to expectations of a difference between intervention and comparator? - o Were these deviations unbalanced and likely to have affected outcome? - If you want to study effect of starting and adhering to intervention:	of	- Were exposures measured similarly across the cohort? - Was exposure measured in a valid and reliable way? - Were the outcomes measured in a valid and reliable way? - Follow-up long enough to measure all outcomes?	- Outcomes are clearly defined - Assessment of outcome is blind to exposure status; if this was not possible, this limitation is recognized in the study - The method for exposure assessment is reliable - The method of outcome assessment is reliable and valid (as demonstrated by evidence from other sources) - Exposure level or prognostic factor is assessed more than once	- Measurement of exposure - Measurement of outcomes - Assessment of outcome blind to exposure - Follow-up long enough to measure all outcomes?



		○ Co-interventions balanced? ○ Intervention implemented successfully? ○ Adherence? Measurement of outcomes: - influenced by knowledge of intervention? - Outcome assessors aware of the intervention? - Methods of assessment comparable across groups? - Differential misclassification of outcomes possible?			
Attrition bias	Adequacy of follow-up (FU): one star for complete follow-up or number lost less than 20% or description of those lost to follow-up (LTFU) not different from those not LTFU	Missing data: - Were outcome data available for all or a sufficient number of participants? - Was the missing data differential? - Was analysis adapted to this?	- Was FU complete, and if not, were the reasons described and explored? - Strategies to address incomplete FU utilized?	- What percentage of individuals or clusters dropped out before study completion? - Was a comparison made between full participants and those lost to follow-up, by exposure/ intervention?	- Assess attrition bias
Reporting bias /		Is the reported estimate likely to be selected based on the results, from... - Multiple outcome measurements? - Multiple analyses? - Different subgroups? Low risk of bias only if clear evidence that all reported results correspond to all intended outcomes, analyses	/	/	/



	and subgroups preregistration of the protocol	→	
Overall study / quality	/	- Was appropriate statistical analysis used?	- Does the study address an appropriate and clearly focused question? - Have confidence intervals been provided?

Table 16 – Comparison of advantages & disadvantages of the different appraisal tools

Tool	Advantages	Disadvantages
Newcastle-Ottawa scale	- Easy to follow - limited number of items, fits on one page - star system enables a quick visual assessment of risk of bias - questions and their interpretation are very specific, which makes it easy to use for inexperienced researchers	- Comparability: if the study controls for the two most important confounders, it is given maximum number of stars. However, it can be difficult to identify the most important confounders, and two may not be enough to minimize residual confounding - Biases/domains not addressed: reporting bias, participation rate differences (selection bias), immortal time bias, deviations from interventions, quality of statistical analysis - In general, the specificity of the questions has the downside of introducing errors when the specifics of the study are not included in the options. For example, in the case of outcome assessment: the only answer possibilities are “independent blind”, “record linkage” or “self-report”. There are however in reality also other possibilities (e.g. non-blinded assessment); for some outcomes, self-report may be the only way to evaluate (e.g. pain scales) - No cut-off defined to assess a study as “high” versus “low” risk of bias or “good” versus “poor” quality
ROBINS-I	- Thorough and relatively complete - Gives a nuanced score for each domain with the possibility of indicating the direction of the bias - Clear guidance on how to assess the overall bias	- Some signaling questions are difficult to understand or interpret correctly and could lead to erroneous assessment; more examples could enlighten some questions - Many items to consider, very lengthy; also because of the extensive explanation needed with many of the questions - Some of the assessed domains/biases may be rare/less important (e.g. deviations from interventions, time-varying confounding, pre-registered protocol, ...) - Biases/domains not assessed in this tool: - o Immortal time bias not addressed (there is a question on timing of intervention vs follow-up though, but it seems to pertain a different type of bias) - o Many signaling questions which recur in other tools for selection bias (e.g. same source population, participation rate, ...) are not addressed here; however, the tool states that imbalance between the groups as a result of selection bias can be addressed by controlling for confounding - o The tool does not specifically ask whether participants were free of the outcome at the start of the study (which is present in some other tools)



		○ Representativity of the study sample is not assessed (touched upon by the question on missing intervention status in which an intended study sample is mentioned) ○ General quality of study conduct (e.g. statistical analysis) not assessed in a signaling question; but the tool does regularly refer to appropriate statistical analysis in many of the domains and is therefore probably the most explicit in stating which analysis would indicate good quality
JB I appraisal checklist	- Limited number of questions (n=11); fits on one page - Easy to follow	- Overall appraisal at the end of the checklist is defined by the decision to include or exclude the paper, but it is not clear how this should be assessed - Confounding is assessed only superficially - Compared to the NOS, many signaling questions are kept vague (e.g. "were outcomes measured in a valid and reliable way" vs "one star for independent blind assessment and record linkage; no star for self-report or no description"); this is more subjective and thus less appropriate for inexperienced researchers; on the other hand, when used by experienced researchers, this could also be an advantage - Biases/domains not assessed: representativity, immortal time bias, reporting bias, deviations from interventions, participation rate
SIGN checklist	- Easy to follow - Guidance given to perform an overall appraisal - Assessment of confounding is concise, straightforward and efficient, compared to other tools	- Overall appraisal : "High quality (++) = majority of criteria met; little or no risk of bias. Acceptable (+): most criteria met; some flaws with an associated risk of bias." This is still vague; is there a difference between "majority" and "most"? How many identified biases count as "little to no risk"? - Quality of statistical analysis is solely assessed by the signaling question on confidence intervals - Biases/domains not assessed: immortal time bias, reporting bias, deviations from interventions, length of follow-up sufficient to detect outcome
KCE appraisal tool	- Concise (only 8 signaling question) but still with enough background information on how to assess in depth - Easy to follow	- No guidance on how to do overall appraisal - Biases/domains not assessed: immortal time bias, reporting bias, statistical analysis, deviations from interventions



APPENDIX 2. PROPOSAL FOR AN APPRAISAL TOOL OF *IN VITRO* STUDIES

Based on Barbosa et al, 2023 58

Study design	Is sample selection, collection and processing well described and uniform across the study?	
		☐ Yes: low risk of bias ☐ No: high risk of bias
	Are the procedures and experimental conditions well described?	☐ Yes: low risk of bias ☐ No: high risk of bias
	Were experimental conditions uniform across the different groups and entire duration of the study? If not, were attempts made to minimize the impact of the non-uniform conditions?	☐ Yes: low risk of bias ☐ No: high risk of bias
	Was an internal quality control used during the experiments?	☐ Yes: low risk of bias ☐ No: high risk of bias ☐ Not applicable
Randomization (when applicable)	Were outcome assessors blinded to the intervention groups?	☐ Yes: low risk of bias ☐ No: high risk of bias ☐ Not applicable
Data collection	Was evaluation of outcome done by independent observers?	☐ Yes: low risk of bias ☐ No: high risk of bias
	Were the methods for data collection identical between the different experimental groups?	☐ Yes: low risk of bias ☐ No: high risk of bias
Data analysis	Were outcomes defined beforehand and transparently presented?	☐ Yes: low risk of bias



☐ No: high risk of bias

Were methods of data analysis clearly explained and appropriate?

☐ Yes: low risk of bias

☐ No: high risk of bias



Appendix 2.1. Example of a RAP sheet

RAP SHEET CBRNE -

DEFINITION OF CBRNE EVENT

What is a CBRNe event?

ORGANISATION AT NATIONAL LEVEL

How is CBRNe care organised in Belgium?

ORGANISATION AT HOSPITAL LEVEL

How are the first receivers/responders organised to respond in case of a CBRNe event?

Who takes decisions? Operational decisions, but also hospital policy decisions.

What is the role of hospital emergency plans? How were they developed? How do we know that they work?

What does it cost for a hospital to be prepared for a CBRNe event? Who pays the bills? Is there a willingness to invest in something that stocked and unlikely to be used?

POSITIVE EVOLUTIONS – STRENGTHS

Regarding Staff? Enough deployable staff? Well-trained? Paid by whom?

Regarding Safety?

Regarding Structure?

Regarding supply? What is foreseen in the national stock? How can it be mobilised? How is it kept up to date? Who pays for it?

NEGATIVE EVOLUTIONS – WEAKNESSES

What are the main problems/needs for first receivers?

Regarding Staff? Enough deployable staff? How, when and by whom are staff trained for CBRNE incident/event? Paid by whom? What is the type of payment? Is there a staff reserve to mobilise in case of an incident?

Regarding Safety?

Regarding Structure?

Regarding supply? What is foreseen in the national stock? How can it be mobilised? How is it kept up to date? Who pays for it? Who finances the material (stocks of antidotes and other equipment) and human resources (training of



personnel, recruitment of additional personnel) related to CBRNE incidents?
And how?

SOLUTIONS – IMPROVEMENTS

What would be concrete solutions to improve the situation in Belgium?

AVAILABLE DATA

Is there a register of CBRNe victims?

Is it possible to identify these victims/expenses in the minimal hospital data (MZG/RHM)?

KEY PERSONS - ACTORS - STAKEHOLDER – EXPERTS

Who else should we talk to?

Who are the main stakeholders?

BEST PRACTICES TO LEARN FROM?

What are best practice countries/regions?

EXISTING INITIATIVES - NETWORKS - COLLABORATIONS

What are existing initiatives regarding the improvement of the organisation of CBRNe care?

How will the research findings of our research be able to contribute/ inform decisions?

ADDED VALUE OF THE KCE PROJECT

Source: KCE REPORTS 374C⁷³



APPENDIX 3. TIME/RESOURCE MATRIX OF NUNNS

Figure 8 – Time/resource Matrix¹¹⁸

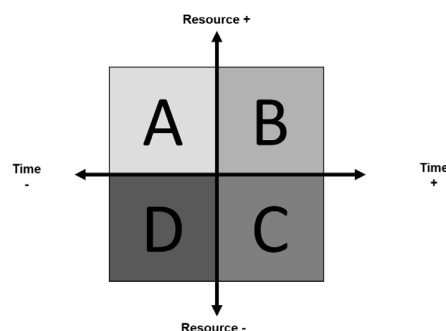


Figure adapted from Nunns, 2009 cited by Vindrola-Padros (2021)⁶¹

In KCE studies, RQR could then be useful in A and D situations, but sometimes only possible in quadrant A.

The resource-time matrix developed by Nunns identifies the following situations:¹¹⁸

“Higher level of resources and less time (quadrant A):

When to use: evaluations where findings were time-critical for stakeholders, but researchers had a team at their disposal and used rapid research and evaluation methods.

- Examples: evaluations with a formative purpose, such as increasing understanding of an evaluand and/or participants, diagnostic and/or problem-solving, learning and development, process involvement, or state-of-play assessment

Higher level of resources and more time (quadrant B):

When to use: evaluations that are important to stakeholders but are less constrained by time (or constrained in a different way, such as requiring regular feedback loops of findings) and more complex designs could use.

Examples: evaluations that measure change over time, evaluations where the level of attribution is important, outcome and impact evaluations, evaluations involving a summative judgement

Fewer resources and more time (quadrant C):

When to use: these studies are rather considered as ‘nice to do’, because the lack of resources might imply, they were not a priority for policy makers.

Examples: no examples mentioned

Fewer resources and less time (quadrant D):

When to use: these kinds of studies should only be carried out if the researchers had confidence in the quality of the available data. If new data needed to be collected or the quality of the existing data was questionable, these should be avoided.

Examples: no examples mentioned”



APPENDIX 4. AUTOMATION

Appendix 4.1. Prompt engineering with the complete set of prompt examples and their relative output.

To present a LLM with a general question without providing specific instructions or context is called **naïve prompting**. This is a simple technique, useful for quick interactions and for getting general information, but may result in generic or ambiguous responses¹²⁶.

To elicit a more precise response to execute a specific task, the structure of a prompt can include up to four essential elements (<https://www.promptingguide.ai/introduction/elements>). These elements—each critical in its own right—can be utilized in full or in selected combinations, depending on the specific requirements of the task:

Instructions that guide the task,

Context that provides background information or additional framing of the task, to steer the model in the desired direction

Input data that serves as the basis for the response

Output indicator that specifies the desired format or type of response.

For example, if the task is to explore what are the advantages for the patients of the available gestational diabetes screening strategies in the Belgian context, a naïve prompt could be as follows (*Prompt 1*) *What are the advantages of different gestational diabetes screening strategies?*

A prompt crafted to get a more precise response, and that includes all four elements could be as follows: (*Prompt 2*) *There are two types of gestational diabetes screening in Belgium, one is a one-step strategy, the other is a two-step strategy. Analyse advantages of both strategies and give the two main advantages for each strategy with three bullet points, from a patient's point of view*

ChatGPT3.5 responses to the example prompts

Prompt 1 What are the advantages of different gestational diabetes screening strategies?

Different gestational diabetes screening strategies offer various advantages depending on their specific approach and context. Some common screening strategies include:

One-Step Screening: This strategy involves administering a glucose challenge test (GCT) followed by a diagnostic oral glucose tolerance test (OGTT) if the GCT results are abnormal. The advantages of this approach include simplicity, as it requires only one testing step, and efficiency in identifying women at risk of gestational diabetes.

Two-Step Screening: In this approach, an initial screening is performed using a GCT, and if the results are above a certain threshold, a follow-up diagnostic OGTT is conducted. The advantages of two-step screening include lower cost and reduced burden on healthcare resources, as fewer women require the more time-consuming OGTT.

Risk Factor-Based Screening: Some healthcare providers may opt to screen only women with specific risk factors for gestational diabetes, such as obesity, family history of diabetes, or previous history of gestational diabetes. The advantage of this strategy is the potential for targeted screening, focusing resources on those at higher risk while minimizing unnecessary testing for low-risk individuals.

Universal Screening: This approach involves screening all pregnant women, regardless of individual risk factors. Universal screening ensures that no cases of gestational diabetes are missed, leading to early detection and intervention for all affected women. It may also help identify cases in women without apparent risk factors.

Point-of-Care Testing: Utilizing point-of-care testing methods, such as handheld glucometers, for screening can offer advantages such as rapid results, convenience, and reduced need for laboratory infrastructure. This approach may be particularly useful in resource-limited settings or for remote populations where access to traditional laboratory testing facilities is limited.



Home-Based Screening: With advancements in technology, home-based screening kits are being developed, allowing pregnant women to perform glucose testing in the comfort of their homes. This approach offers convenience and may improve compliance with screening recommendations, especially for women with logistical challenges accessing healthcare facilities.

Integrated Screening Programs: Implementing comprehensive, integrated screening programs that combine various screening methods with education, counseling, and support services can optimize the effectiveness of gestational diabetes screening. Such programs can address multiple aspects of care, including early detection, management, and long-term follow-up, leading to improved outcomes for both mother and baby.

Each screening strategy has its own set of advantages, and the choice of approach may depend on factors such as available resources, healthcare infrastructure, population characteristics, and healthcare provider preferences. Ultimately, the goal is to implement a screening strategy that maximizes detection of gestational diabetes while minimizing potential drawbacks such as cost, inconvenience, and unnecessary testing.

In the case of naive prompting, since the prompt did not give any input, the LLM cannot know that only universal one-step and two-step screening strategies are available in Belgium, so most part of the answer is not relevant, it spends a lot of text describing the screening strategies, but this was not a part of the task, there was no contextualisation on the fact that the task had to see the patient's point of view, so various different points of views are taken into consideration making the answer less relevant for the task.

Prompt 2 There are two types of gestational diabetes screening in Belgium, one is a one-step strategy, the other is a two-step strategy. Analyse advantages of both strategies and give the two main advantages for each strategy with three bullet points, from a patient's point of view:

Certainly, let's analyse the advantages of both gestational diabetes screening strategies in Belgium from a patient's perspective:

One-Step Strategy:

Convenience:

Single Visit: Patients appreciate the convenience of undergoing both the glucose challenge test (GCT) and the diagnostic oral glucose tolerance test (OGTT) in a single visit, saving time and minimizing disruptions to their daily routines.

Reduced Anxiety: Completing the screening process in one step may alleviate anxiety for pregnant women, as they receive immediate results and can proceed with necessary interventions or reassurance promptly.

Clarity of Diagnosis:

Clear Results: With the one-step approach, patients receive definitive results after the OGTT, providing clarity regarding their gestational diabetes status without the need for additional testing or uncertainty.

Early Intervention: Clear diagnosis allows for early initiation of treatment and lifestyle modifications if gestational diabetes is detected, potentially improving maternal and fetal outcomes and reducing the risk of complications.

Two-Step Strategy:

Minimized Testing:

Reduced Discomfort: Patients may appreciate undergoing only one initial screening test (GCT) instead of the more intensive OGTT, minimizing discomfort associated with prolonged fasting and blood draws.

Avoidance of Unnecessary Testing: Women who receive normal results on the initial GCT can avoid the need for additional testing with the OGTT, alleviating concerns about unnecessary medical procedures and potential false alarms.

Personalized Approach:

Tailored Evaluation: The two-step strategy allows for a tailored approach to screening, with further testing reserved for individuals who have an elevated



risk based on the initial GCT results or specific risk factors. This personalized evaluation may provide patients with a sense of individualized care.

Targeted Support: Patients who proceed to the second step of screening based on abnormal GCT results receive targeted support and interventions, focusing resources where they are most needed and potentially enhancing the effectiveness of management strategies.

From a patient's perspective, both one-step and two-step gestational diabetes screening strategies offer advantages in terms of convenience, clarity of diagnosis, minimized testing, and personalized care, ensuring that pregnant women receive appropriate evaluation and support tailored to their needs and preferences.

The answer to the structured prompt is more precise and to-the-point. There are only relevant information contained. The output follows a pre-definite length and structure

As we can see also from the examples, the right contextualisation is crucial to get a relevant response. A technique to steer the LLM in the desired direction is giving the LLM, in the prompt, a "role" to play (**role prompting**)¹²³. When the goal is to obtain a response within a specialized field, we may request the LLM to adopt the persona of an expert in the corresponding field. (**expert prompting**)¹²⁷. For instance, when the model is prompted to assume the role of a public health expert, it is inclined to offer a thorough and contextually precise response when questioned about matters pertaining to public health.

The different parts of the prompt can be demarcated by using **delimiters**, for instance you may use triple quotations to differentiate input data from the instructions.

A prompt example that incorporate all the aforementioned elements may be the following:

(Prompt 3) List two main advantages and two main disadvantages for the following gestational diabetes screening strategies available in Belgium

One-step screening strategy and two step screening strategy". Impersonate a concerned patient. Put the results in a table.

Screening Strategy	Advantages	Disadvantages
One-Step Screening	1. Convenience: Requires only one visit for testing, saving time and hassle.	1. Potential Overdiagnosis: May lead to unnecessary treatment for some patients.
	2. Immediate Clarity: Receive definitive results in a single step, reducing uncertainty.	2. Increased Stress: Immediate results may cause anxiety if gestational diabetes is detected.
Two-Step Screening	1. Minimized Testing: Only undergo one initial test, reducing discomfort and inconvenience.	1. Missed Cases: Some cases of gestational diabetes may be missed during the initial screening.
	2. Personalized Approach: Targeted evaluation based on initial results or risk factors.	2. Delayed Diagnosis: Additional testing required for those with abnormal initial results may cause delays.

Appendix 4.2. Zero-shot prompting and few-shots prompting

The examples of prompts from section 1.7.1 are examples of **zero-shot prompting**^{zz}: in this technique we rely on the training data of the LLM to perform the task, without giving any example or additional training. When zero-shot prompting fails to produce desired results, it is advisable to supplement the prompt with demonstrations or examples, transitioning to a **few-shots prompting** approach. Few-shot prompting involves providing a

^{zz} <https://www.promptingguide.ai/techniques/zeroshot>



language model with a small number of examples (shots) to illustrate the task at hand. Each example typically consists of an input followed by the desired output.

Zero-shot prompt: (Prompt 4) Find the acronyms in the following text concerning CBRNe preparedness in Belgium """"A pragmatic search was conducted in PubMed with keywords related to hospitals, preparedness, and emergencies, to retrieve review articles describing indicators related to the conceptualization and evaluation of hospital resilience. With this search it was aimed to understand which elements contribute to hospital resilience in case of CBRNe events and to integrate them in hospital emergency plans. Only articles written in English and published over the last 5 years were included. After screening on title, abstract and full-text, 14 studies 17-30 were included describing specific elements for implementation in order to enhance preparedness and consequently resilience of hospitals towards CBRNe emergencies. The search for key documents to provide a general description of 'what is CBRNe', 'health system resilience' and KCE Report 374 Organisation of medical care in CBRNe incidents – a guidance 'hospital resilience' and an illustration of (conceptual) framework, were conducted through desktop research and snowballing. We adopted the following strategy to select countries: countries with recent and well-developed CBRNe plans according to the Health System and Policy Monitor (HSPM) network of the European Observatory on Health Systems and Policies; countries with high threat score and better preparedness than Belgium based on the report of Booz Allen Hamilton (2017) ; countries that are selected for their 'best practices' according the Belgian stakeholders interviewed in chapter 4; neighbouring countries of Belgium, sharing common CBRNe risks (e.g. nuclear plants or SEVESO companys close to the frontier) and, eventually, cross-border cooperation agreements. Availability of the documents in Dutch, English, French, German or Spanish. Based on these 5 criteria, 5 countries were selected: France, Norway, The Netherlands, United Kingdom (UK) and United States of America (USA). This selection was validated by the Belgian stakeholder"""

Prompt 4 Output: CBRNe: Chemical, Biological, Radiological, Nuclear, and Explosive , KCE: Belgian Health Care Knowledge Center

The zero-shot prompt instructs the LLM to identify acronyms in the provided text without any additional training or specific examples. And in this case it is quite evident that the majority of acronyms have been missed. In order to enhance the performance we could provide **examples** of what exactly we would like the LLM to accomplish.

Few-shots prompt: (Prompt 5) Find the acronyms in the following text regarding CBRNe preparedness in Belgium """"Therefore the following requirements are relevant on top of the type II

CBRNe hospital requirements. An extensive hospital emergency plan (HEP), including decontamination SOP and agreements with for example the fire brigade and civil protection service could be aimed at to assure the processes and support needed to handle a large and rapid influx of victims. Elements to be considered as part of the decontamination SOP are for example designation of personnel responsibilities, upscaling and rotation of personnel, the available types and location of PPE, decontamination equipment, detectors, antidotes, and emergency clothing for the victims, and the designation of a separate entrance for regular patients"""" as it has been done in the following example:

Text: """"We adopted the following strategy to select countries: countries with recent and well-developed CBRNe plans according to the Health System and Policy Monitor (HSPM) network of the European Observatory on Health Systems and Policies; countries with high threat score and better preparedness than Belgium based on the report of Booz Allen Hamilton (2017) ; countries that are selected for their 'best practices' according the Belgian stakeholders interviewed in chapter 4; neighbouring countries of Belgium, sharing common CBRNe risks (e.g. nuclear plants or SEVESO companys close to the frontier) and, eventually, cross-border cooperation agreements. Availability of the documents in Dutch, English, French, German or Spanish. Based on these 5 criteria, 5 countries were selected: France, Norway, The Netherlands, United Kingdom (UK) and United States of America (USA). This selection was validated by the Belgian stakeholder"""



Acronyms

1 Acronym: CBRNe

Expansion: Chemical, Biological, Radiological, Nuclear, and Explosive

2. Acronym: KCE

Expansion: Belgian Health Care Knowledge Center

3. Acronym: HSPM

Expansion: Health System and Policy Monitor

4. Acronym: USA

Expansion: United States of America

5. Acronym: UK

Expansion: United Kingdom

6. Acronym: SEVESO

Expansion: Seveso Directive (European Union directive on control of major-accident hazards involving dangerous substances)

Put the results in a table

Prompt 5 Output: Here are the acronyms found in the provided text regarding CBRNe preparedness in Belgium, along with their expansions:

Acronym	Expansion
CBRNe	Chemical, Biological, Radiological, Nuclear, and Explosive
HEP	Hospital Emergency Plan
SOP	Standard Operating Procedure
PPE	Personal Protective Equipment

These are the acronyms identified in the text.

Appendix 4.3. Chain of thought prompting

In **chain of thought** prompting, we actively promote factual accuracy within the model by guiding it to adhere to a structured sequence of steps in its "reasoning" (Amatriain, 2024 #26, which is a step beyond the few-shots prompting method described in [section 1.7.2](#), since we do not limit ourselves only to providing examples of the desired output to a certain input, but we **describe the logical steps** that link input and output.

Chain of thought prompt (Prompt 6) Find the acronyms in the following text regarding CBRNe preparedness in Belgium. "Also the medical-advisors and FHI, described in our scenarios, should be able to access relevant databases. The databases should have features such as having a simple, easy-to-use and accessible system, using everyday language that is easy to understand and act upon, providing accurate and reliable information of high quality (preferably evidence based), providing integrated multi-device platforms available for PC, laptop, mobile data terminals, command and control systems, tablets, and mobile phones. Other examples of relevant databases are the medical management guidelines for acute chemical exposures from the Agency for Toxic Substances and Disease Registry (ATSDR), the Chemical Hazardous Emergency Medical Management (CHEMM – USA), or the European Chemicals Agency (ECHA) website, which the unique source of information for the chemicals manufactured and imported in Europe." , by making the following steps 1: find all words that are written mainly in capital letters (the acronyms) 2 expand the acronyms by using the expansion contained in the text 3 if the expansion is not contained in the text, expand the acronyms according to your own knowledge being mindful that the text is about CBRNe preparedness in Belgium Put the results in a table



Prompt 6 Output

Acronym	Expansion
FHI	Foreign Health Institutions
PC	Personal Computer
ATSDR	Agency for Toxic Substances and Disease Registry
CHEMM	Chemical Hazardous Emergency Medical Management
ECHA	European Chemicals Agency

Another technique that is used to elicit chain of thought “reasoning” from a LLM is **zero-shot chain of thought prompting**^{Kojima, 2023 #27} (that basically only involves adding a sentence akin to “Let’s think it step by step” to the original prompt.

Appendix 4.4. Prompt chaining

In the **prompt chaining** approach we need to break the task that we want to complete into subtasks. After identifying those subtasks, the LLM is prompted with one at a time, and its response becomes input for the next prompt in the sequence. This technique may be useful to complete complex tasks that a LLM could find challenging if faced with only one extremely detailed prompt^{aaa} (An example is if we wish to summarise a long or complex text we might want to first let the LLM to give us the main points of the text in bullet points, and after take the bullet points as an input and ask the LLM to make a fluent text.

^{aaa} https://www.promptingguide.ai/techniques/prompt_chaining

^{bbb} The data are fabricated for the purpose of the example

It is useful also for mathematical reasoning and data analysis and visualisation, as it is shown in the example below, where the output was produced using chatGTP 4

Prompt 7: Using the following table representing the Odds Ratios (OR) and Adjusted Odds Ratios (aOR) of the prevalence of a certain disease^{bbb} according to the month of the year, in the year 2020 compared to the year 2019: create another table which contain the natural logarithm of each OR and aOR and relative 95%CI.

OR (95%CI)	aOR (95%CI)*
1.11(1-1.23)	1.11(0.99-1.25)
1.11(1-1.24)	1.09(0.96-1.23)
1.12(1-1.26)	1.11(0.98-1.26)
0.98(0.89-1.1)	0.97(0.85-1.1)
1.07(0.96-1.18)	0.99(0.87-1.12)
1.15(1-1.27)	1.13(1-1.27)
1.37(1.25-1.5)	1.43(1.6-1.28)
1.15(1.05-1.26)	1.23(1.1-1.37)
1.15(1-1.26)	1.2(1.1-1.33)
1.17(1.1-1.28)	1.18(1.1-1.31)
1.32(1.2-1.45)	1.38(1.23-1.55)
1.25(1.14-1.38)	1.25(1.11-1.4)

Output Prompt 7:

Here is the table containing the natural logarithm (ln) of each OR and aOR along with their 95% confidence intervals for each month:



Month	ln(OR) (95%CI)	ln(aOR) (95%CI)*
January	(0.747, [0.693, 0.802])	(0.747, [0.688, 0.811])
February	(0.747, [0.693, 0.805])	(0.737, [0.673, 0.803])
March	(0.751, [0.693, 0.816])	(0.747, [0.683, 0.816])
April	(0.683, [0.637, 0.742])	(0.678, [0.615, 0.742])
May	(0.728, [0.673, 0.779])	(0.688, [0.626, 0.751])
June	(0.765, [0.693, 0.817])	(0.756, [0.693, 0.817])
July	(0.863, [0.811, 0.916])	(0.888, [0.956, 0.826])**
August	(0.765, [0.718, 0.814])	(0.802, [0.742, 0.862])
September	(0.765, [0.693, 0.814])	(0.788, [0.742, 0.845])
October	(0.775, [0.742, 0.824])	(0.779, [0.742, 0.834])
November	(0.842, [0.788, 0.893])	(0.867, [0.802, 0.936])
December	(0.811, [0.761, 0.867])	(0.811, [0.747, 0.875])

*** Please note that there is an inconsistency in the seventh row's aOR interval (2.43, [2.6, 2.28]) which has a lower bound greater than the upper bound after adding 1 to the original values. The logarithm of such an interval is not meaningful, as confidence intervals should have the lower bound less than the upper bound. This likely indicates a typographical error in the original data.*

Prompt 8: Switch the order of the confidence interval bounds in the adjusted OR of the seventh row so that it becomes consistent

Output Prompt 8: I have corrected the seventh row's aOR interval and recalculated the natural logarithms. Here's the updated table with the natural logarithm (ln) of each OR and aOR along with their corrected 95% confidence intervals for each month:

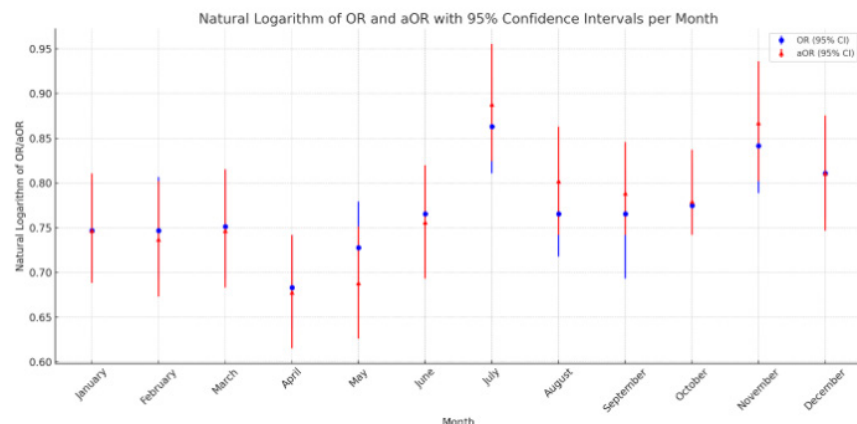
Month	ln(OR) (95%CI)	ln(aOR) (95%CI)*
January	(0.747, [0.693, 0.802])	(0.747, [0.688, 0.811])
February	(0.747, [0.693, 0.805])	(0.737, [0.673, 0.803])
March	(0.751, [0.693, 0.816])	(0.747, [0.683, 0.816])
April	(0.683, [0.637, 0.742])	(0.678, [0.615, 0.742])
May	(0.728, [0.673, 0.779])	(0.688, [0.626, 0.751])
June	(0.765, [0.693, 0.817])	(0.756, [0.693, 0.817])
July	(0.863, [0.811, 0.916])	(0.888, [0.824, 0.956])**
August	(0.765, [0.718, 0.814])	(0.802, [0.742, 0.862])
September	(0.765, [0.693, 0.814])	(0.788, [0.742, 0.845])
October	(0.775, [0.742, 0.824])	(0.779, [0.742, 0.834])
November	(0.842, [0.788, 0.893])	(0.867, [0.802, 0.936])
December	(0.811, [0.761, 0.867])	(0.811, [0.747, 0.875])

*** The corrected interval for July's aOR now has the bounds in the correct order, making it consistent*



Prompt 9: For the table you constructed, create a graph that visualises the OR and relative CI and aOR and relative CI for each month

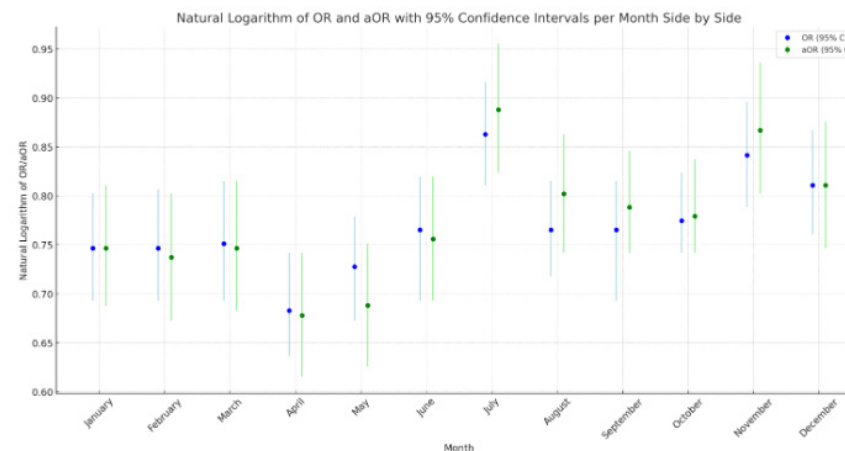
Output prompt 9:



Here's the visual representation of the natural logarithm of OR and aOR with their 95% Confidence Intervals for each month. The blue dots and lines represent the ORs, and the red triangles and lines represent the aORs. Each point indicates the mean of the natural logarithm, while the lines indicate the range of the confidence interval for that month.

Prompt 10: Instead of using dots and triangles, put the OR and relative confidence intervals and the aOR and relative confidence intervals, side by side in two different colours

Output Prompt 10:



The graph now displays the OR (in blue) and aOR (in green) with their respective 95% Confidence Intervals side by side for each month. This visualization helps to easily compare the OR and aOR for each month and see how they differ, including the spread of their confidence intervals.

In order to increase the explainability of the model, ChatGPT4 provides the python code relative to all of its data analysis.



Appendix 4.5. Tools and applications for specific research purposes

Tools and applications for retrieval and analysis of pertinent literature and scientific writing

Table 17 – Evaluated tools for retrieval and analysis of pertinent literature and scientific writing

Tool/ application	Web reference	Study phase	A	B	C	D	E	F	G	Candidate for Real life test
Ask your PDF	https://askyourpdf.com/	PDF data extraction	4	4	4	4	5	4	5	Yes
ChatPDF	https://www.chatpdf.com	PDF data extraction	4	3	4	4	4	3	4	Yes
Citation Chaser	https://estech.shinyapps.io/citationchaser/	snowballing	5	4	5	5	5	5	3	Yes
Consensus	https://consensus.app/	Semantic search engine	3	2	5	5	5	5	3	Yes
Covidence	https://www.covidence.org/	Literature management and streamlining	5	5	5	4	4	4	5	Yes
DistillerSR	https://www.distillersr.com/products/distillersr-systematic-review-software	Literature management and streamlining	5	5	5	4	5	2	5	Yes
Eppi reviewer	https://epi.ioe.ac.uk/cms/Default.aspx?tabid=2914	Literature management and streamlining	4	4	5	4	3	4	4	Yes
Jenni.ai	https://jenni.ai/	Writing assistance	4	3	4	5	5	4	5	Yes
Nested Knowledge	https://nested-knowledge.com/	Literature management and streamlining	4	4	5	4	4	2	5	Yes
Paper fetcher	https://paperfetcher-paperfetcher-web-app-paperfetcher-app-0w0vu2.streamlit.app snowballing	snowballing	4	4	4	5	5	5	3	No, all features present seem more functional in citation chaser
Perplexity	https://www.perplexity.ai/	Semantic search engine	3	3	5	5	5	5	4	Yes
Research rabbit	https://researchrabbitapp.com	snowballing	3	?	?	5	4	5	4	Risk related to serendipity
Robot Reviewer	https://www.robotreviewer.net/	RoB2 evaluation	4	4	5	na	na	5	5	Yes



Systematic review accelerator	https://sr-accelerator.com/	Suite for systematic literature review	3	5	5	3	3	5	4	Yes
--------------------------------------	---	--	---	---	---	---	---	---	---	-----

Tools and applications for qualitative data collection and analysis

Table 18 – Evaluated tools for qualitative data collection and analysis

Tool/ application	Web reference	Study phase	A	B	C	D	E	F	G	Candidate for Real life test
Amberscript	https://www.amberscript.com	Audio transcription	5	4	5	5	5	4	5	Yes
Happy scribe	https://www.happyscribe.com/	Audio transcription	5	4	5	5	5	4	5	Yes
Just: access	https://just-access.org/ai-transcription/	Audio transcription	5	5	1	5	5	4	5	No, servers based outside the EU
NoScribe	https://github.com/kaixxx/noScribe#readme	Audio transcription	3	3	5	3	5	5	5	No. Some advantages, but there seems to be issues with the tool
Otter.ai	https://otter.ai/transcription	Audio transcription			1					No, servers based outside the EU
Simon Says	https://www.simonsaysai.com/solutions	Audio transcription			1					No, servers based outside the EU
Transcribear	https://transcribear.com	Audio transcription			1					No, servers based outside the EU
Konch	https://www.konch.ai/product/features	Audio transcription	5	5	4	5	5	4	5	Yes
Trint	https://info.trint.com/knowledge/who-has-access-to-my-files	Audio transcription	5	5	4	5	5	4	5	Yes
Audiotype	https://www.audiotype.org/	Audio transcription			2					Not sure, use AI models from the UK and models from the US but servers based in the EU.
Rev	https://www.rev.com/services/auto-audio-transcription	Audio transcription			1					No, servers based outside the EU
SmartScribe	https://www.smartscribe.co/en/tcu.html	Audio transcription	5	4	5	5	5	3	5	Yes
Descript	https://www.descript.com/	Audio transcription			1					No, servers based outside the EU



Whisper	https://openai.com/research/whisper	Audio transcription		5	5	4	3	3	5	5	Yes
Atlas.ti	https://atlasti.com/	Qualitative analysis	content	5	5	5	5	5	4	5	Yes
Coloop	https://www.coloop.ai/	Qualitative analysis	content	5	5	5	5	na	na	4	No, treatment of the servers is based in the US
MAXQDA	https://www.maxqda.com/products/ai-assist	Audio transcription and qualitative analysis	content	5	5	5	5	5	na	4	Yes



APPENDIX 5. OBSOLETE TOOLS FOR AUTOMATION

currently unavailable so we were unable to get an idea on how it works. However, there is a possibility to get the tool directly on the machine.

Tools for PDF data extraction

ChatPDF is an application that can extract and interpret information from PDFs via APIs, enabling users to ask questions about the content of the document. Structured prompts designed for data extraction may be used in literature review, but also to generally have an overview on documents in PDF format. PDFs are accepted in any language and chats can be conducted in any language. Answers contain references in the original PDF. The website mentions that files are stored in a secure cloud storage and will never be shared. ChatPDF allows to use 2 PDF of up to 120 pages for free. The technology used is ChatGPT 3.5 (Chat Generative Pre-trained Transformer) and cannot chat with multiple PDFs at the same time. Paying subscription is 140 Euros/year, and it includes access to Jenny.ai.

Ask your PDF is a third-party application that can be integrated with the GPT language model to enable users to interact with external APIs and access information from PDF documents. They mentions that the documents uploaded in their servers are completely erased when we deleted them. Ask your PDF does not help to know what the content is, but maybe useful to resume or make a synthesis of the content, by asking question to the program. Access is free for 100 pages per document and 15MB uploaded/month, and 50 questions. After that you have to pay 13.70 Euros/month.

Tools for automated quality appraisal

RobotReviewer is an AI assisted tool for the automation in the quality appraisal of RCTs using RoB2 tool. The feasibility of adopting RobotReviewer was assessed within a national public health institute through a randomised, real-time, user-centered study involving 26 RCTs and six reviewers. The study found that reviewers were as likely to accept the judgment made by RobotReviewer as they were to accept each other's judgment during the consensus process¹⁴³. Unfortunately the web demo is