REIGN Framework

Use of **R**esearch **E**vidence to **I**nform **G**uidance regarding **N**ormative-ethical Topics

Dr. hum. biol. Corinna Klingler, M.Sc.
Faculty of Health Sciences Brandenburg, University of Potsdam

Dr. phil. Marcel Mertz, M.A.
Institute of Ethics, History & Philosophy of Medicine, Hannover Medical School

Contents

1. INTRODUCTION

a. Problem Description

With respect to the management of health care, modern innovative (health) technology; national and global health disparities; and political, social, and demographic changes pose questions that have often not yet been answered by established legal and nonlegal governance frameworks. As challenges surrounding the management of health care systems, clinical decision-making and medical research increase, the calls from decision-makers for guidance informed by academic research and, to a lesser degree, scholarly expertise increase accordingly.

In recent decades, such guidance informing health policy and clinical decision-making began to be based on evidence and expert consensus [e.g., Bosch-Capblanch 2012]. Evidence-based medicine (EBM) [e.g., Sacket et al 2000; Kunz et al 2007] or health care (EBHC) [e.g., Muir Grey 2001] provides an epistemological and methodological framework for fine-tuning development processes for guidance documents, such as clinical practice guidelines (CPGs), health technology assessment (HTA) reports and other forms of regulations and recommendations. Within the EBM paradigm, recommendations must be built on comprehensive surveys and appraisals of existing research evidence, especially by using systematic reviews and meta-analyses of data on efficacy, effectiveness [e.g., Sehon/Stanley 2003], adverse effects, and the cost of health technologies. Organizations, such as Cochrane, have grown as a consequence, thus constantly furthering the advancement of methods for searching, collecting, analysing and synthesizing research evidence.

Apart from effectiveness, harm and economic aspects that have to be considered in decision-making, most challenges in health governance are necessarily intertwined with ethical issues – or are even composed mainly of them. Evaluating the benefits and risks of a health technology, for example, is not solely a question of biomedical or clinical research evidence; such evaluations (implicitly) refer to value judgements that are ethical in nature and that also play a role in generating evidence and assessing facts [e.g., Molewijk et al 2008; Strech 2008a]. Additionally, values or principles other than (medical) benefit and harm, such as the autonomy of the patient, justice [e.g., Beauchamp/Childress 2009], equity, and human dignity, are crucial for decision-making in health care or health policy and, when conflicting, have to be balanced or prioritized transparently and reasonably, i.e., in an argumentatively justified way. It may therefore be problematic that CPGs and HTA reports rarely give (explicit) guidance related to these issues, which nevertheless must be addressed by decision-makers [Droste/Gerhardus/Kollek 2003; Mertz/Strech 2014].

While most organizations charged with developing guidelines, or guidance for practice in health care in general, have clearly delineated their processes for arriving at recommendations in domains such as effectiveness or cost-effectiveness, it is remarkable that ethical issues are seldom addressed in these manuals for guideline development (see chapter 2). Where "ethics" is addressed in HTA

reports, research has shown that it is addressed unsystematically [e.g., DeJean et al 2009, Polus et al 2019]. Literature reviews are sometimes conducted, but the methods for searching, identifying and analysing the literature are often unclear and underdeveloped. Furthermore, the findings tend to be superficial [DeJean et al 2009; Polus et al 2019]. In addition, even institutions engaged in developing ethics guidelines/guidance have not yet established clear methodological standards for developing ethics recommendations.[1] In many National Ethics Committees (NECs), recommendations were – and are – often developed via public and expert hearings and internal discussions among members of the NEC [e.g., Deutscher Ethikrat 2013].

However, in most institutions that either integrate ethical issues in CPGs or HTA reports or are developing ethics guidelines, no control of quality or a more rigorous mechanism (or at least a more transparent process) was, and mostly is, established – although several organizations or authors have begun discussing and proposing possible frameworks and methods [e.g., Burls et al 2011; Reiter-Theil et al 2011; Mertz/Strech 2014; EUnetHTA 2016; Lysdahl et al 2016a; IQWiG 2017; SBU 2017]. Especially the question of integrating research evidence – now at most unsystematically considered – is not at all or only insufficiently addressed thus far. It is yet unclear how the call for rooting recommendations firmly in evidence, as emphasized by the EBM movement, can be translated into the development of ethics guidelines.

b.   WHO's Position and Strategy

The World Health Organization (WHO), as the directing and coordinating authority for international (public) health within the United Nations system, develops guidance for various actors on how to improve health and combat diseases. One of the WHO's core functions is "articulating ethical and evidence-based policy options" [WHO 2014a, p. 20]. While guidelines issued by the WHO are not legally binding, they are highly influential in the global governance structure, as they represent important orientation points for decision-makers at the national and local levels. Approximately ten years ago, the recommendations from the WHO were often criticized as being based only on expert consultation and seldom on evidence that was systematically gathered and synthesized [Oxman/Lavis/Fretheim 2007]. To counter this tendency and to ensure high-quality issued recommendations in the future, the WHO has developed methodological standards for guideline development as described in *The WHO Handbook for Guideline Development* [WHO 2014b]. Empirical evidence is comprehensively identified, assessed, synthesized and then issued to underpin recommendations on technical questions, such as intervention effectiveness and harms. The codified standards build on the EBM paradigm, an established systematic review methodology and explicit evidence-to-decision framework (e.g., GRADE/DECIDE framework [Guyatt et al 2008]). At least since establishing

---

[1] Terminologically, the report uses "ethics guidance", "ethics guidelines" or "ethics recommendation" in favour of "ethical guidance", "ethical guidelines" and "ethical recommendation" because the latter could also be understood as guidance, guidelines or recommendations that fulfil ethical criteria but are not intended to give guidance about ethical topics.

more methodological guidance and clear procedures, WHO can "in its normative and standard set-ting work" be understood as a "a science- and evidence-based organization with a focus on public health" [WHO Evaluation Office 2017, p. i].

In addition to technical guidelines, the WHO also provides guidance on ethical issues and their ade-quate normative-ethical handling in guidance documents. The WHO recently published the first in-ternational guidelines on ethical issues in public health surveillance [WHO 2017] and guidelines on managing ethical issues in infectious disease outbreaks [WHO 2016]. In the *WHO Handbook for Guideline Development*, however, ethical issues are not addressed, and the processes for developing these types of guidance documents are not well established – especially not from the viewpoint of an organization that in its own understanding is, or strives to be, "science- and evidence-based" (see above). It is thus particularly unclear what role *evidence* should play in normative-ethical delibera-tions that aim to arrive at respective ethics recommendations, although recent guidelines have at-tempted to consider evidence more explicitly.

Given the lack of established standards for developing ethics guidelines, the *Global Health Ethics Unit* (Health Systems and Innovation Cluster) and the *Guidelines Review Committee* (Department of Information, Evidence & Research) of the WHO initiated a project – REIGN (*Use of Research Evidence to Inform Guidance regarding Normative-ethical Topics*) – to review current approaches and develop a framework (the REIGN framework, see chapter 4) for evidence integration in ethics guidelines. With this, the WHO hopes to take a first step towards strengthening its own processes and to prompt further discussions regarding methodology in understanding and using evidence for devel-oping normative-ethically justified recommendations.

c. Aims and Scope

REIGN was initiated to support the establishment of more transparent and systematic processes and methods for the use of research evidence when developing guidelines focusing on normative-ethical topics or, more exactly, when developing *ethics guidelines* (for further elaboration on differ-ent kinds of guidelines and policies, see chapter 3c). This discussion paper presents the REIGN frame-work developed by the two commissioned authors and further necessary conceptual groundwork. The REIGN framework can be seen only as a starting point for further discussions. This framework should **not** be viewed as a guideline for how to integrate evidence in ethics guideline development. Developing such methodological guidelines would require an even more comprehensive engage-ment with the relevant literature and possibly also bringing together the (or a purposive sample of) experts in the field in a consensus-seeking process. The framework therefore provides only the nec-essary conceptual groundwork for enabling further processes that might eventually result in the development of a guideline. Initiating such a process would be highly desirable. The REIGN frame-work will, however, support guideline developers by providing terminology and concepts that can help in structuring choices regarding the collection and use of evidence in guideline development.

The REIGN framework addresses what, from where and which type of evidence can be used for informing the development of ethics guidance in general and, more precisely, for ethics guidelines as an instrument of ethics guidance. In doing so, the framework also clarifies central concepts – such as "research evidence" in the context of ethics guideline development itself – to provide a useful terminology for future work in this area. The conceptual groundwork also addresses, regarding the relation of evidence and ethics guidance, general methodological issues that are not sufficiently settled in ethics, guideline development, or public health scientific communities.

The project did not directly address the (explicit) integration of ethical issues or recommendations in CPGs, public health guidelines, or HTA reports. However, ethical issues and recommendations addressed in these contexts can also be informed by research evidence, and the development of such guidelines and reports generally follows an evidence-based approach, thereby indicating some connection between these topics and the topic of ethics guideline development. Some aspects of tackling ethical issues in these areas are context-specific, and the REIGN framework will not help in addressing these issues. However, some overarching aspects might resonate with the deliberations in REIGN. In fact, the discussions about the integration of ethical issues and recommendations in CPGs, HTA reports and public health guidelines were deemed relevant as a possible source of information for developing the REIGN framework (see chapter 2 for publications consulted). Therefore, many conceptual discussions and parts of the REIGN framework (see chapter 4) can possibly be used or adapted for these purposes.

Completely out of the scope of REIGN, however, is the development of policies with (quasi-) legal force or of codes of ethics for specific institutions/organizations (e.g., a local hospital). Additionally, the proposed framework does not contribute to ethics as a theoretical discipline; e.g., the framework does not discuss how ethical norms or values can be justified or which ethical theories are better (more justified) than other theories. Furthermore, REIGN does not address applied ethics interests, such as which ethical theories are more appropriate for analysing and assessing ethical issues in the health care or public health context.

d.   Target Audience

REIGN is thus primarily of interest to *guideline developers* who *develop ethics guidelines* and researchers who are involved in such processes (e.g., consulting ethics experts or literature review groups)**.** Because of the proximity of the topics, REIGN can also in part be of interest to developers of CPGs, public health guidelines, and HTA reports that aim to explicitly address ethical issues alongside clinical or economic issues. REIGN will be of particular interest to WHO staff involved in developing ethics guidance documents. Furthermore, REIGN can be of scholarly interest to those devising methods for developing ethics guidance or for integrating ethics in CPGs, HTA reports and similar forms of regulation or (non-governmental) governance.

e.   Approach to framework development

The REIGN framework was commissioned by the WHO, which determined the basic direction and conditions of the work conducted. The authors were given four months (September to December 2017) to conduct the work on the framework. As part of this work, the authors were asked to screen institutional and academic publications for relevant content for framework development. Due to limited time, the authors could conduct only a scoping review of the academic and grey literature; this review provided the basis for developing the framework. Summaries of the literature are provided as part of this discussion paper. However, the set goal was not only to qualitatively synthesize and describe the relevant literature but also to move beyond this and develop a framework based on further conceptual analysis. The literature accordingly provided only the baseline for framework development. As part of their approach to framework development, the authors employed classical philosophical tools, such as critical reading, developing and structuring arguments, checking arguments for validity and soundness, examining the consistency of terminology and ideas proposed, thinking, discussing with each other and additional experts in the field, etc. The framework presents the outcomes of this work.

To further improve the quality of the work, a discussion paper presenting the framework was sent out for peer review in 2018 (April to November 2018). Four experts in the field agreed to review the framework and provided critical comments and questions (see Acknowledgements). Based on the feedback, the framework was revised by the authors.

f.   Structure

As a first step, **(a)** the authors undertook an explorative screening of possibly relevant organizations and institutions (e.g., clinical guideline developers, HTA organizations, and public health institutions) in all six WHO regions to capture the state of the art of using research evidence for informing guidance related to normative-ethical topics (see chapter 2). Additionally, the academic literature on this topic was analysed as relevant background information for developing the framework (see also chapter 2).

Additionally, **(b)** theoretical background assumptions regarding ethics, (ethics) guideline development, and the understanding of "evidence" were explicated, and related terminological decisions were made (see chapter 3).

Based on these results, **(c)** the REIGN framework of how research evidence can be used to inform ethics guidance was devised (see chapter 4).

Afterwards, **(d)** the framework was applied to two case studies. Specifically, the framework was applied to two WHO ethics guidelines to reflect on their use of evidence in guideline development (see chapter 5). New insights gained by applying the framework led to refinements of the framework itself.

Finally, **(e)** the results of REIGN are summarized, open questions are raised, and implications for practice are addressed (see chapter 6). Additionally, an Appendix (see chapter 8) contains further material – including the *REIGN toolkit,* providing an orientating summary of the framework for ethics guideline developers at the WHO.

2. STATE OF CURRENT PRACTICE

a.   Organizations/Institutions

As an important information base for the framework, the websites of several relevant organizations were analysed to see whether guidance for considering evidence in developing ethics guidelines had already been developed and published. To avoid too much ("Western") bias because of the nationalities and prior experiences of the REIGN project members and to take the responsibility of the WHO as a global organization seriously, it was highly important to review several countries and their relevant organizations. The following types of institutions were considered potentially relevant: **(a)** public health agencies, **(b)** organizations producing HTA reports, **(c)** clinical and public health guideline development agencies/groups, and **(d)** national bioethics committees across all WHO regions (see figure 1). Assessments were focused primarily on national institutions.

Due to language barriers (see below) and time constraints, the scope of this review for identifying the state of the current practice was limited. Given the constraints, it was possible to review only some selected countries. Therefore, this screening has to be understood as an *explorative* and a *selective*, not a comprehensive, screening.

Regions and Countries

Three countries in each WHO region (see figure 1) were selected for screening, thus allowing for a diverse but manageable sample size. The selection of countries was primarily based on prior knowledge of potentially relevant documents, income and development level, size, and language (it was feasible to screen documents written only in German and English).
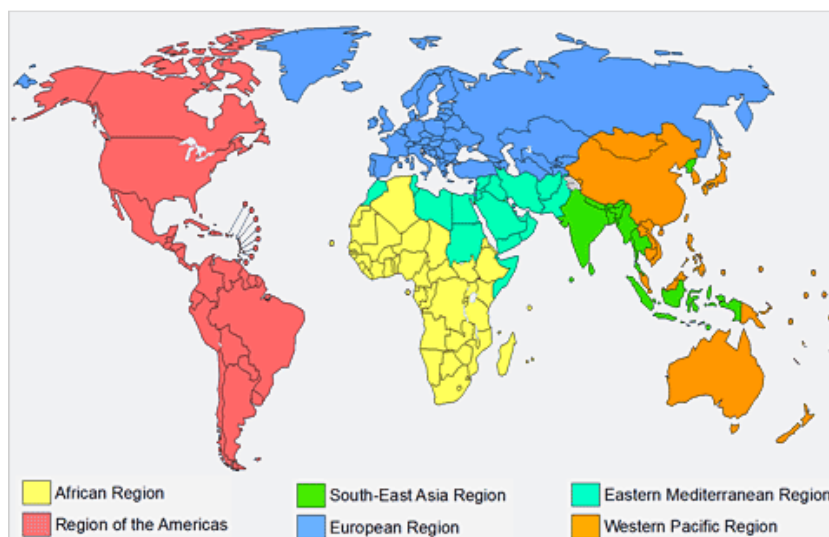
Figure 1: Overview WHO regions
(from http://www.who.int/about/regions/en/)

In total, 18 countries (see table 1 below, which includes the rationales for selecting the countries) were selected and their relevant organizations and institutions reviewed for information about the role of research evidence when developing ethics guidance. Selected organizations were also contacted via email to inquire whether internal documents guiding the development of ethics recommendations exist and whether these

documents could be shared with the REIGN project. In the appendix, an overview of screened insti-
tutions within those 18 countries is provided (see Appendix A).

WHO Region	Countries selected
African Region	
Countries were chosen based on language (one of the official languages is English) and income level (upper-middle-income country or higher) be-cause it was assumed that countries of higher income levels were most likely to invest limited resources in the topic of interest. Population size was an additional argument to narrow down the list of potential candidates for screening.	Republic of **Botswana**
	Republic of **Namibia**
	Republic of **South Africa**
Region of the **Americas**	
Countries were chosen based on their official language (English) and income level (high income). As an additional criterion, population size was introduced to prioritize among the available candidates.	Republic of **Trinidad and Tobago**
	United States of **America**
	Canada
South-East Asia Region	
Countries chosen based on their official language (English: only India). For those that are not English-speaking, it was decided to prioritize those of higher income (upper-middle income or higher) as most likely to invest limited resources in the topic of interest and publish findings in English.	Republic of **India**
	Kingdom of **Thailand**
	Republic of **Maldives**
European Region	
Countries were chosen based on language (German- or English-speaking), pre-screening knowledge about esp. HTA agencies that are already addressing ethics in their processes/documents, income level and size.	Federal Republic of **Germany**
	United Kingdom of Great Britain and Northern Ireland
	Kingdom of **Sweden**
Eastern Mediterranean Region	
Countries were chosen based on language (English speaking), income level (upper-middle-income country or higher) and population because it was assumed that these countries were most likely publishing guidelines that could be accessed (in terms of language) and have resources to invest in the topic of interest.	Kingdom of **Saudi Arabia**
	United Arab Emirates
	Islamic Republic of **Iran**
Western Pacific Region	
Countries were chosen based on language (English-speaking) and pre-screening knowledge about esp. HTA agencies/activities in the area of interest.	Commonwealth of **Australia**
	Republic of **Korea (South Korea)**
	Japan

Table 1: Countries selected for screening according to WHO regions

In addition, relevant international organizations that develop CPGs, HTAs or other types of guide-
lines were screened for published documents addressing this question (see table 2).

Relevant International Organizations	
The International Network of Agencies for Health Technology Assessment **(INAHTA)**	INAHTA is a network of 52 HTA agencies that support health system decision-making. INAHTA provides a forum for exchange and pursuit of interest of the relevant national agencies.
Health Technology Assessment international **(HTAi)**	HTAi is the global scientific and professional society for all those who produce, use, or encounter HTAs. HTAi has members from over 65 countries and embraces all stakeholders, including researchers, agencies, policy makers, industry, academia, health service providers, and patients/consumers.
European Network for Health Technology Assessment **(EUnetHTA)**	EUnetHTA was established to create an effective and sustainable network for HTA across Europe. EUnetHTA is dedicated to helping develop reliable, timely, transparent and transferable information to contribute to HTAs in European countries.
The Network to Strengthen Collaboration among HTA Agencies in Asia **(HTAsiaLink)**	HTAsiaLink is a network to support collaboration between Asian HTA agencies. HTAsiaLink focuses on facilitating HTA research by accelerating information and resource sharing and developing an efficient methodology for HTA in the region.
The International Bioethics Committee **(IBC)** of the United Nations Educational, Scientific and Cultural Organization **(UNESCO)**	The International Bioethics Committee (IBC), which is a body of 36 independent experts, follows progress in the life sciences and their applications to ensure respect for human dignity and freedom. The IBC was involved in the development of various declarations regarding bioethical topics.

Table 2: International organizations selected for screening

Summary of Results

By searching institutional websites, potentially relevant documents were identified for six national and three international organizations. Three additional national institutions supplied us with relevant publications after being contacted directly via email. In total, 17 publications of potential interest were identified (see table 3 for an overview). In this first step of identifying potentially relevant publications, the authors were particularly inclusive to ensure that they missed nothing relevant. Accordingly, publications with minimal chance of containing relevant information were also included. Subsequently, the identified documents were read and summarized, and their relevance to the research question was assessed. In the following, an overview of the findings focusing on the arguments most relevant to REIGN's underlying research question is provided. A detailed overview of publications, including the summaries and evaluations of relevance, is supplied in Appendix A.

Name of Institution	Name of Document
National Institutions	
National Collaborating Centre for Healthy Public Policy (NCCHPP), Canada	Blog article by Michael Keeling and Olivier Bellefleur: Finding Traction in Public Health Ethics: Reflections and Practical Resources
Canadian Agency for Drugs and Technology in Health (CADTH), Canada	Guidelines for the Economic Evaluation of Health Technologies: Canada (Sections on "Equity")
National Institute for Health and Clinical Excellence (NICE), UK	Social Value Judgements – Principles for the development of NICE guidance
Nuffield Council on Bioethics, UK	Website: How does the party gather evidence?
National Health and Medical Research Council (NHMRC), Australia	Different sources: • Ethical Guidelines for Organ Transplantation from Deceased Donors (Section titled "Process report") • Ethical guidelines on the use of assisted reproductive technology in clinical practice and research (Section titled "Process report") • Personal communication, via email, used as an additional information source
Institute for Quality & Efficiency in Health Care (IQWiG), Germany	Different sources: • Allgemeine Methoden Version 5.0 (Section 6.5.3., titled "Ethik") • In addition, IQWiG granted us access to two internal documents (templates) used to structure analysis of ethical aspects
Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU), Sweden	Assessment of methods in healthcare – a handbook ("Chapter 12: Ethical and social aspects")
Presidential Commission for the Study of Bioethical Issues, USA	Bioethics for every generation: Deliberation and education in health, science, and technology ("Chapter 2: Democratic Deliberation in Bioethics")
International Organizations	
European Network for Health Technology Assessment (EUnetHTA)	HTA Core Model Version 3.0 (Section titled "Ethical Analysis")
International Network of Agencies for Health Technology Assessment (INAHTA)	INAHTA's Working Group on Handling Ethical Issues – Final Report
Health Technology Assessment International (HTAi)	Different sources from conferences/meetings: Slides • Bond, Ken: Appraising the primary ethics literature • Bond, Ken: Introduction to Ethics in Health Technology Assessment • Scott, Anna Mae & Sacchini, Dario: Reporting on ethics in HTA: Methods, the results and interpretation Workshop report/summary • Stoklosa, Anna & Bond, Ken: Workshop on Methodology in Ethics for Health Technology Assessment: Assessing the Need for and Quality of Ethics Analyses in HTA

Table 3: Overview of the results of screening organizations/institutions

Publications of No Relevance

Among these 17 publications, three were found to be irrelevant to the research question. One publication from the NCCHPP (Canada) discussed how ethics literacy can be built among public health practitioners. The other two publications published by the CADTH (Canada) and NICE (UK) discussed the ethics of guideline development (see below for definition). More precisely, these publications addressed which normative considerations should underlie *empirical* guideline development (e.g., consideration of equity issues in economic evaluations).

Publications of Limited Relevance

Three publications were found to be of limited relevance, but they still influenced the thought process of developing the framework (see chapter 4). The Nuffield Council (UK) detailed how they try to ensure that different voices are heard during guideline development (e.g., by using stakeholder surveys or consultations); the council consider these voices *evidence*. However, the council did not specify how *research* evidence in particular should be considered in the process. The NHMRC (Australia) emphasized the importance of research evidence for guideline development. They introduce evidence to the process by inviting experts from the scientific community to the groups developing guidance and by demanding that all evidence supplied during public consultation periods has to be considered. However, the specific role of evidence and how it is to be used in guideline development is not fully explained. The IQWiG's (Germany) publicly accessible documentation of methods specifies various approaches to address ethical issues in health technology assessments (e.g., principlism [Beauchamp/Childress 2009] or the Socratic approach [Hofmann 2005]); however, these approaches do not specify which role evidence is to play in these assessments.

Publications of Relevance

The internal documentation of the IQWiG, however, is relevant to REIGN's research questions. In delineating more clearly the processes of ethical analysis[2], the internal documents emphasize the importance of evidence. The IQWiG specifies three ways in which information (evidence) on the ethical aspects of a health technology can be generated. Which combination of methods will be chosen depends on the ethical contentiousness of the technology. Ethical aspects/arguments can be identified by (a) an exploratory search of various publications (including but not limited to scientific publications) – this will always be part of the process, (b) application of Hofmann's [2005] list of questions to the intervention of interest, and (c) stakeholder discussions based on the same list. Whichever combination of methods is chosen to identify ethical implications, effectiveness or cost-effectiveness studies included in the HTA should be analysed as well for relevant information on ethical aspects.

---

[2] The discussion is, however, applicable only to the products pertaining to the rubric "ThemenCheck Medizin". In personal communications, IQWiG has further emphasized that it is still in the early phases with "ThemenCheck Medizin" and is in the process of expanding and developing its methods.

The Presidential Commission for the Study of Bioethical Issues (USA) also describes its approach to guideline development in one document of moderate relevance. Its working mode to arrive at recommendations is democratic deliberation, where consensus regarding bioethical issues will have to be built in a cooperative process that is inclusive regarding diverse viewpoints. The commission proposes that democratic deliberation should be structured in accordance with five steps: (1) begin with an open question and consider distinct points of view; (2) time deliberation for maximum impact; (3) invite input from experts and the public; (4) foster open discussion and debate; and (5) develop detailed, actionable recommendations. As part of the third step, evidence generation and consideration are discussed. It is explicitly stated that the quality of the guideline depends on the guideline being based on trustworthy evidence. Evidence is understood mainly as empirical evidence (see chapter 3 for differentiation between empirical and normative evidence) and – although the report is not particularly clear on this – should be provided mainly by expert hearings. Establishing the evidence base is no one-time effort but should be a continuous effort as new information might become available. Additionally, guideline developers might not know from the beginning what information will be needed. Ethical analyses might also be used as supportive material but more as "best-practice examples" for good ethical reasoning. These analyses, however, are not understood as forming part of the evidence base.

Three publications – published by the Swedish SBU, INAHTA and EUnetHTA – describe the approaches of the respective institutions to integrate ethics into HTAs. These approaches are also highly relevant. Although the approaches differ, they are quite similar in how they explain in detail the role of evidence in addressing ethical aspects.[3] Therefore, these three organizations will be described together. All three emphasize that analysing ethical aspects should be seen as a two- or three-step process: The first step is the identification of ethical issues for the technology in question. For the identification of ethical issues, lists of questions should be used (as, for example, developed by Hofmann [2005]). As a second step, the ethical issues identified need to be analysed. Different methods for analysing ethical issues are mentioned; the listed methods are most often quite similar to the one IQWiG provides in its public documentation (e.g., principlism, see above). All three institutions emphasize that the analysis of ethical issues should be informed by evidence. Evidence seems mostly (although not always clearly) to imply *empirical* evidence. All three institutions point out that literature reviews can be a useful tool, but at least two point out that relevant information can also be collected via additional (or alternative?) ways: gathering views of affected parties, looking at (other) guidelines, or having an ethicist conduct an ethical analysis. As a last and often not explicitly mentioned step, a summary of the relevant points of the analysis and – where it is within the mandate of the respective institution – the final recommendation is to be provided.

---

[3] This is not surprising as often the same organizations are members of INAHTA and EUnetHTA. EUnetHTA also explicitly admits that it built on work conducted earlier by INAHTA. SBU, on the other hand, explicitly refers to the publications of international organizations as having influenced its approach.

Four further relevant publications were found on the HTAi website. While various aspects are addressed, these publications focus on critical (quality) appraisal of ethical analyses.[4] All publications argue that to address ethical aspects, both *empirical* and *normative literature* will have to be considered. As quality criteria for empirical literature are established, the discussion centres on normative literature. One of the publications (a workshop summary) argued specifically that different quality criteria might have to be considered for ethical arguments, the ethical literature (all arguments identified as an information base for ethical analysis), and/or the whole ethical analysis (including arriving at a final recommendation). For the ethical arguments/the whole ethical literature, two quality criteria are proposed: (a) the validity/soundness of the argument according to formal and informal logic and (b) the thoroughness/comprehensiveness of the arguments identified (a similar differentiation is proposed by our framework, see chapter 4). As part of one publication, a checklist for establishing the soundness (the first criterion) of arguments is proposed (but mainly as a starting point for further discussions). Another publication argues that certain arguments (e.g., majority opinion) should, in accordance with McCullough et al [2004], always be considered bad. To ensure comprehensiveness (the second criterion), it is proposed to conduct systematic reviews (by using the methodology of Strech/Sofaer [2011] as a blueprint); to use checklists based, for example, on principlism [Beauchamp/Childress 2009]; or to implement public consultations. It might also be important to combine strategies because systematic reviews might be biased if they, for example, include only English literature. It is further pointed out that checking for the validity/soundness of an argument should not be conflated with checking for "relevance and force" of the argument; the latter is presumably not the same as the former (this point is also stressed by the REIGN framework). Where the quality of the whole ethical analysis is considered (including arriving at a final recommendation), the relevance and force of the arguments will have to be considered as well, not just their soundness.

b. Academic discourse

As with material published by organizations and institutions, the academic discourse shows that the topic of interest is covered by only a few publications.

Four different academic discourses were identified as potentially relevant for REIGN: **(a)** general reflections on "evidence" and empirical data in (medical) ethics ("evidence-based ethics" and "empirical ethics"); **(b)** methods/procedures for developing ethics guidelines or guidelines for normative-ethical topics; **(c)** integration/consideration of ethical issues, aspects, or arguments in HTA reports and clinical or public health guidelines; **(d)** methods/concepts of (systematic) literature reviews in ethics as a way of synthesizing evidence for ethics guidance. For all four discourses, scoping literature reviews were conducted; the scoping reviews were complemented by already known

---

[4] One of the identified publications was a summary of two workshops. It was sometimes difficult to bring the various strands of discussion together and understand what exactly was debated. This summary should therefore be read as a somewhat subjective reconstruction.

literature the authors of this discussion paper were aware of. The search strategy for the scoping reviews and a list of examples of the literature in these four discourses is provided in Appendix B.

Summary Results

Evidence-based Ethics/Evidence and Ethics

The discourse about the role of empirical data and (empirical) evidence in ethical analysis and decision-making takes place especially in what is called "empirical (bio)ethics", "empirically informed (bio)ethics", or "evidence-based (bio)ethics" [e.g., Ashcroft 2003; Tyson et al 2003; Molewijk et al 2004; Borry et al 2005; Goldenberg 2005; McMillan/Hope 2008; Strech 2008b; Düwell 2009; Kon 2009; Ives/Dunn/Cribb 2017; and many more]. However, this discourse is directed to more general questions about the relationship between "the empirical" and "the normative", whether in the context of ethical decision-making or especially in the context of developing and refining ethical theory. This discourse thereby also touches on classical meta-ethical topics (e.g., the fact-value distinction, the is-ought gap, and moral epistemology) [e.g., DeVries/Gordijn 2009]. This discourse is often also about methods for interdisciplinary bioethics, medical ethics or public health ethics, e.g., about how quantitative and qualitative socio-empirical methods can be properly combined with normative-ethical reasoning while upholding quality standards that might vary between the disciplines [e.g., Ives/Draper 2009; Dunn et al 2012; Salloch et al 2012; Mertz et al 2014]. Research regarding "empirical ethics" is therefore not specifically directed to the context of guideline development, HTA reports, or other governance-related guidance documents, although such research can be considered as providing theoretical and methodological background information for the latter topics (see chapter 3).

The literature discussing "evidence-based ethics" is not distinguishable in its topics from that discussing "empirical ethics" [e.g., Jansen 1997; Kim 2004; Halpern 2005; Bonneux 2007; Sieber 2009], is directed to clinical-ethical decision-making "at the bedside" [e.g., Major-Kincade/Tyson/Kennedy 2001; Frize/Walker/Ennett 2003; Tyson/Stoll 2003], or analyses – mostly criticizes – the use of empirical evidence as a basis for ethical decision-making due to possible misuse of evidence [Goldenberg 2005; Strech 2008b]. The authors of these studies point out the danger of misrepresenting ethical issues as mere empirical issues [Goldenberg 2005]. These authors also emphasize that empirical evidence can justify only instrumental norms ("if x should be achieved, y is the best means for achieving it") and not categorical norms ("x should be done!"), for which ethical analysis would presumably be needed [Salloch 2012]. Furthermore, the legitimacy of referring to "evidence" in medical ethics is questioned because of unclear epistemic standards, especially regarding social science research [Strech 2008b, 2008c].[5]

---

[5] Strech, however, argues not *against* evidence-based approaches but for an "information-critical" approach that, although not directly using the concept of "evidence", follows the methodological focal point of evidence-based approaches, namely, to systematically assess availability, quality, and relevance of (empirical) information used.

The only academic publications discussing the role of evidence – both *normative* and *empirical* evidence – in developing ethics guidelines are co-authored by one of the authors of this discussion paper [Mertz 2011; Reiter-Theil et al 2011b].[6] These papers argue, inter alia, that on the basis of a systematic search and synthesis of normative-ethical (and legal) literature, (external) evidence informing ethics guidelines can be generated. These papers also stress the role of empirical evidence in a wide array of ethically relevant aspects (e.g., also regarding psychological biases in ethical decision-making). The theoretical explication of the respective understanding of "evidence" can be found mainly in Mertz [2011] and has significantly impacted the development of the REIGN framework.

Guideline Development in Ethics/for Ethical Topics

A search for the topic of guideline development on ethical issues produced the fewest publications. Most of this literature does not address (systematically) the role of evidence in such procedures and is therefore not relevant to the topic of this discussion paper. Although some of the literature is interested in maintaining the quality of and providing methodological reflection on the process of guideline development, the focal point often is – in addition to following a structured process – ensuring the involvement of ("all") relevant stakeholders (especially the prospective "users" of the guideline) and/or establishing a multidisciplinary composition of the Guideline Development Group (GDG) [e.g., Reiter-Theil et al 2011a; Cho 2014; Jox 2014; Neitzke et al 2015; Kangasniemi et al 2017; Riedel 2017]. Where the role of empirical evidence or "additional research" in guideline development is addressed, this role is problematized mainly as a means to make final decisions about primarily normative recommendations. Additionally, empirical research is regarded as important background information for ethical deliberation [Cho 2014]. Furthermore, why (institutional) policies or ethics guidelines are actually needed, what goals and value they have [Jox 2014; Riedel 2017], and implementation issues are addressed [Reiter-Theil et al 2011a; Jox 2014]. Importantly, however, these publications often discuss ethics guidelines that are *locally* developed for a specific hospital or another health institution [Jox 2014; Neitzke et al 2015; Riedel 2017]. These studies do not claim that their conclusions are necessarily transferable to other institutions or settings (e.g., those operating at the national or even international levels). This is also partly true for the approach of Reiter-Theil et al [2011a, 2011b] and Mertz [2011], who, as already indicated above, propose that the development of ethics guidelines should be oriented towards procedures for developing clinical practice guidelines – thus implying an evidence-based approach. Therefore, conducting (systematic) literature searches for normative and empirical information relevant to the topic of the guideline is proposed (a simplified model of the advocated interaction of empirical information and normative-

---

[6] In their review of methods for integrating ethical issues in HTA reports, Assasi et al [2014] do actually refer to "ethical evidence" that could be combined "with other types of evidence" in some of the frameworks they reviewed (p. 211). However, the authors do not explicate what "ethical evidence" means; it seems to be the same as either "ethical considerations" (or the respective information) or empirical evidence (e.g., benefit, cost-effectiveness) that is (highly) relevant for determining if a health technology is ethical.

ethical information in the context of guideline content and usage is given in Albisser Schleger et al [2012, p. 68], although there, the term "evidence" is used only for empirical evidence).

Integrating Ethical Issues in Guidelines or HTA Reports

Another academic discourse that is relevant to the topic of interest is literature focusing on how ethical issues, aspects, or arguments can be integrated in HTA reports as well as clinical or public health guidelines. Most of the identified literature is situated within the HTA context. Although more literature is to be found on this topic than for ethics guideline development, only a few publications specifically address the role of evidence. The focus is often on different theoretical approaches (e.g., principlism, casuistry, participatory approaches, and "eclectic" approaches) or "question lists" that can be used by HTA professionals to conduct an ethical analysis and/or evaluation of a particular health technology [e.g., Autti-Rämö/Mäkelä 2007; Saarni et al 2008; Sacchini et al 2009; Burls et al 2011; Heintz et al 2015; Lysdahl et al 2016b; for an overview of different approaches, see Assasi et al 2014]. Interestingly, all of these approaches are seldom actually applied [Hofmann et al 2015; see furthermore Droste/Gerhardus/Kollek 2003]. Some publications mention additional approaches, such as stakeholder involvement techniques (questionnaires/interviews, discussion rounds, etc.) or consultations of "ethics experts" [e.g., Lehoux/Williams-Jones 2007; Bombard et al 2011; Mittelstadt/Stahl/Fairweather 2013; Lysdahl et al 2016b] as providing important information for HTA reports. Some authors mention literature searches (literature reviews) as having a role to play in identifying ethical issues, arguments or other ethically relevant information (e.g., empirical data). However, these authors often do not further explicate how systematic searches should be implemented, exactly what literature should be sought, or the relevance of the results for arriving at recommendations, etc. [e.g., Saarni et al 2008; Mittelstadt/Stahl/Fairweather 2013]. Nevertheless, some publications argue in more detail that *systematic* literature reviews should be striven for and provide hints regarding their quality and process [Scott et al 2017; Lehoux/Williams-Jones 2007]. Although the respective authors do not directly label the results of such reviews as "evidence" (or these reviews as part of evidence-gathering strategies), they seem to imply that the results play such a function. Mittelstadt/Stahl/Fairweather [2013] talk about "evidence" in this context, but it is unclear whether only empirical evidence is implied. For integrating ethical issues and recommendations in CPGs, Mertz/Strech [2014] describe systematic reviews as part of providing empirical and "argument-based" (i.e., normative) evidence about ethical issues; however, the authors also mention further methods (e.g., expert surveys) as part of a "multi-method approach" for identifying a possible "full spectrum" of ethical issues for the topic currently being addressed. Furthermore, the authors propose a tabular form for systematically representing each ethics recommendation, divided as (a) (an empirical) presupposition of the recommendation; (b) the recommendation itself; (c) its justification in terms of, e.g., ethical principles; (d) any elucidations or comments; and (e) references (academic studies) used to arrive at the recommendation.

Methods/Concepts for (Systematic) Reviews in Ethics

If (systematic) literature reviews are seen as a way of synthesizing (research) evidence for normative-ethical issues, the academic discourse about how such reviews can be conducted and which concepts might be useful also becomes relevant. Although many such reviews have already been published in academic journals [cf. a systematic review of such reviews in Mertz/Kahrass/Strech 2016 and Mertz/Strech/Kahrass 2017], published methodological reflections or conceptual discussions are rare [e.g., McCullough/Coverdale/Chervenak 2007; Sofaer/Strech 2012; Strech/Sofaer 2012; McDougall 2014; McDougall 2015]. The authors of these reflections or discussions discuss mainly how to search and synthesize normative evidence. Strech/Synofzik/Marckmann [2008] further discuss methods for reviews of *empirical ethics*, i.e., social science research about, e.g., stakeholder perceptions or preferences that can be ethically relevant. Some authors address more specific questions, for example, the ways ethics literature can be searched [e.g., Droste 2008; Droste/Dintsio/Gerber 2010] or how a quality appraisal of the normative literature or information is possible [McCullough/Coverdale/Chervenak 2004; Scott et al 2016; Mertz 2017; for empirical studies in ethics, see Strech 2010]. Although not focusing on the methods of systematic reviews *per se*, Mertz/Strech [2014] also partially describe the methods for analysis and synthesis of information on ethical issues.

c.   Discussion and Implications

Neither the institutions or organizations examined nor the academic discourses provide a framework that systematically explains and elaborates the possible role of (research) evidence in developing ethics guidelines. Many questions relevant to developing such a framework are, however, at least touched upon in the reviewed documents (e.g., normative vs. empirical evidence, relation of empirical data and normative reasoning, the structure of development processes, the role of additional primary research methods, systematic literature review techniques or quality criteria for normative evidence). The topic of interest is thus treated rather fragmentarily. Additionally, the definition of "evidence" is seldom explicit, nor do authors elucidate whether the concept of "evidence" can – or should – also be applied to results from normative-ethical research.

Importantly, however, some documents stress the need for (more) systematic literature reviews or for additional primary research (e.g., application of ethical theories to the question of interest) as an important part of developing ethics guidance. Therefore, the idea of an "eminence-based" approach where only consulted "ethics experts" or the members of the GDG themselves provide information (evidence) is increasingly abandoned. The notion that addressing ethical issues is somehow a "sleight of hand" that does not have to justify its methods or make its processes transparent is concurrently relinquished. However, some authors also associate certain dangers with applying concepts and procedures of evidence-based approaches to ethics guidance documents (e.g., one such danger is the tendency to wrongly present ethical issues as mere empirical issues, thereby

misusing empirical evidence to promote ethically unjustified courses of action, etc.) [e.g., Goldenberg 2005; Strech 2008b; Salloch 2012].

The relevant documents found and analysed in these two scoping reviews therefore provided the basis for designing a comprehensive framework for the role of evidence in developing ethics guidance. Some of the documents have profoundly influenced the conceptual work behind the REIGN framework, others less so. Where the framework draws heavily on ideas set out in some of these documents, this is of course indicated. As, however, insufficient definitions and conceptual ambiguities were often encountered, the following chapter aims to explicate and define several crucial concepts as the necessary groundwork for the framework.

3. DEFINITIONS AND CONCEPTUAL CLARIFICATIONS

This chapter aims to clarify central concepts related to the development of ethics guidelines (or ethics guidance in general), especially when considering research evidence as a part of such development processes. It also aims to establish a consistent terminology, being transparent about important background assumptions and clarifying some contested methodological issues relevant to the REIGN framework. This chapter thereby proposes certain ways of understanding these issues, e.g., the role of evidence in arriving at ethics recommendations.

Although not all of the following sections have to be read and understood in full to use the REIGN framework, they provide the necessary theoretical groundwork for devising the framework and are indispensable, as they provide the justificatory backbone of the framework.

a.   The Meaning of "Normative"

The meaning of "normative" has to be clarified upfront, as the WHO tends to use the term in a more product- and process-oriented way than is common in philosophy, ethics, and most of the social sciences. For REIGN, both meanings have to be considered, although the meaning of the term within ethics will be more central.

The work of the WHO is, in part, described as *normative* by the WHO itself: the "WHO was established as an intergovernmental organization with the authority to adopt and approve normative instruments" [WHO Evaluation Office 2017, p. 10]. The term "normative instrument" covers both (a) normative *products* (e.g., conventions, regulations, regulatory recommendations, Secretariat guidelines, and health trend assessments) and (b) normative *functions* (activities in normative processes or in policymaking) [WHO Evaluation Office 2017]. Normative products can encompass pieces of World Health Assembly (WHA)-based constitutional "soft laws" (such as the Global Code of Practice on the International Recruitment of Health Personnel) or other strong binding standards (such as the International Nonproprietary Names or the food quality norms and standards of the Codex Alimentarius) [WHO Evaluation Office 2017, p. 44]. In addition, WHO issues non-binding, WHO Secretariat-based scientific and technical normative products, such as "technical guidelines and standards on, e.g., immunization, safe motherhood, financing, malaria, etc." [WHO Evaluation Office 2017, table 2, p. 13], which are quantitatively larger than the more binding normative products. They are also normative products that are of interest to the REIGN project.

In this understanding often implied in WHO documents, "normative" refers mainly to (devising) *"technical" norms* or *standards* (also *standardization* processes) for medical diagnosis, therapy or prevention. Therefore, the WHO establishes, for example, a norm or "best practice" standard of how to treat certain diseases, thereby giving orientation to national guidelines or directly to practitioners by determining how one (ideally) *should* act or decide (to achieve a certain outcome, i.e., treating a disease effectively and safely, etc.).

This latter aspect of WHO's understanding of "normative" is a bridge to the understanding stipulated in philosophy, ethics and most of the social sciences. In these disciplines, "normative" does not denote specific products or processes but is a fundamental *logical* and *semantic* category that is contrasted with terms such as "descriptive", "factual", "empirical", "constative", or "explanatory" [e.g., Kambartel 2004].[7] In summary, (mere) descriptive or empirical statements just describe "what there is" (e.g., they describe reality), while normative statements entail "what should be" or often more specifically "what should be *done*" (e.g., pre- or proscribing certain behaviour, at least recommending certain courses of action as *better* than others, or defining what would be considered an ideal state of reality). Normative statements, thus, in general refer to norms, principles, or values.[8] Statements (e.g., certain norms) are considered normative irrespective of the kind of "product" where they are or could be contained or the process that brings them about. This is partially because such norms often exist only informally and are not written down in any document. The underlying processes leading up to normative judgements are also often not formal but embedded in social and cultural practices with historical dimensions; therefore, the norms generally invoked are *social* or *cultural norms* or *values*, not "technical" norms or standards.

In ethics, more specifically *ethical* norms, principles and values are referred to. These *can* be codified or at least written down (e.g., in professional codes, political position papers, religious texts, and philosophical treaties) but do not necessarily have to be to regulate behaviour. Often, they will be part of the informal or "implicit" knowledge of a person socialized in a specific society. The label "normative-ethical" refers to *how one should act or decide ethically* based on ethical norms, principles and values, or a respective ethical decision-making process.

In this discussion paper, "normative" usually means "normative-ethical" as described above. It will be explicitly stated if the term is meant to denote something else, such as "normative-epistemically" (e.g., about what should be done to gain knowledge about something).

b.   Bioethics, Normative-ethical Theory, Meta-Ethics –Positioning

(The following subsections can be bypassed by readers uninterested in or familiar with some standard topics of ethics as a theoretical discipline. These subsections are relevant, however, for addressing possible critical queries by the ethics community, e.g., about which theoretical positions undergird the REIGN framework or from which theoretical perspective in ethics REIGN is approached in the first place. The following elaborations are thus part of the conceptual clarification and an attempt to fulfil scientific justification requirements).

---

[7] Sometimes, the differentiation used is *ontologically/metaphysically*, i.e., that "facts" are (objective) constituents of reality, while "values" are not – or at least not in the same way (see, e.g., the "fact-value debate" in philosophy).

[8] However, when, e.g., norms are just described, the statements are *descriptive*, not normative. For more detail, see chapter 3d, "Differentiating Empirical Evidence from Normative Evidence".

Stance towards Bioethics, Public Health Ethics and "Empirical Ethics"

Bioethics and public health ethics, which are the main academic orientation points for the kind of ethics guidance documents this discussion paper is about, are traditionally understood as so-called applied ethics disciplines (and, thus, part of moral philosophy or ethics conceptualized as sub-fields of philosophy). However, these disciplines differ from other applied ethics disciplines in (a) the considerable extent of interdisciplinarity in these fields on personal, institutional and methodological levels, (b) the often strong practice orientation, and (c) the resulting relevance of contributions of descriptive ethics[9] for identifying and understanding ethical issues and for providing solutions for ethical challenges. Consequently, bioethics and public health ethics are regarded as mostly independent, "self-contained" fields by the authors. This implies that expertise and relevant academic backgrounds or training are not considered to be solely situated within philosophy (or theology) but are located in a broad range of disciplines, such as medicine and public health itself; the social, political and economic sciences; health policy; biology; and information technology.

Because of the resulting interdisciplinarity and the relevance of descriptive ethics for understanding an existing (ethical) practice, a distinctive methodological position or "program" called "empirical (bio)ethics" (see chapter 2b) originated in bioethics approximately two decades ago. Although there are several conceptualizations of what "empirical ethics" or "empirically informed ethics" is and how respective research should be conducted, a general tendency to actively combine normative-ethical considerations with empirical research can be observed, e.g., "EE research […] is normatively oriented bioethical or medical ethical research that directly integrates empirical research. Key elements […] are […] that it encompasses (a) empirical research as well as (b) normative argument or analysis, and (c) attempts to integrate them in such a way that knowledge is produced which would not have been possible without combining them" [Mertz et al 2014, p. 2; see comparable conceptualizations in Molewijk et al 2004; McMillan/Hope 2008; Salloch/Schildmann/Vollmann 2012; Ives et al 2018; and others]. At least, a penchant for "empirical ethics" can be observed in bioethics and public health ethics that allows for empirical studies alongside conceptual/philosophical analysis as an accepted part of the overall research effort in these fields. Even though no specific conception or methodological approach of "empirical ethics" is embraced in this discussion paper, general accordance with such a position is seen as a conceptual prerequisite for developing a framework regarding the use of research evidence in ethics guidelines.

Stance towards Normative-Ethical Theories

As often in bioethics and public health ethics, a pluralistic, "nonpartisan" approach towards theories of normative ethics (such as Kantian deontology, utilitarianism, contractualism, and virtue ethics)

---

[9] "Descriptive ethics" tries to describe and explain, e.g., moral behaviour and attitudes, or explores how morality is learned and is therefore traditionally part of disciplines such as moral and evolutionary psychology, sociology, anthropology and other social sciences and humanities. "Normative ethics", in contrast, tries to answer questions about what is – reasonably, justifiably – the morally good or right action and tries to provide coherent general values, principles, norms or criteria for determining the morality of a (proposed, actual) action.

was favoured for the development of the REIGN framework. Choosing one theory as the "right" or "only" theory would introduce a theoretical bias regarding the guideline development of ethics guidelines (and other guidance). A decision towards a specific theory – or "ethical world view" – cannot be justified in the context of the framework sought (see chapter 4). Nonetheless, a general tendency towards more procedural-oriented approaches (such as discourse ethics) frames certain background assumptions (see chapters 3d and 3e).

Stance towards Meta-Ethical Topics

Similar to the stance towards normative-ethical theories, the discussion paper and the framework, respectively, do not favour a specific meta-ethical[10] theory or position. Thus, some sort of "neutrality thesis" for meta-ethical backgrounds is proposed; i.e., whatever meta-ethical stance is taken, it does not preclude the use of the framework. There is, however, one crucial exception to this:

The authors would argue that without at least *some* acceptance of the possibility of meaningful rational argumentation in ethics and the basic assumption that there can be "good" and "bad" guidance about normative-ethical topics, there is no epistemological or methodological rationale for any kind of structured ethics guideline development or for ethics guidance whatsoever.[11] Therefore, at least *strong* versions of so-called noncognitivism (these versions assume that moral sentences, such as norms, principles or judgements, are not propositions/statements and therefore cannot be true or false) are incompatible with the framework that is described in the following chapters. Consequently, positions such as emotivism (which understands moral norms or judgements as mere expressions of one's feelings) or decisionism (which understands moral norms or judgements as mere products of decisions made by, e.g., political or legal bodies) are not easily compatible with REIGN. However, the framework does not presuppose (strong) cognitivism (the view that moral sentences can be true or false) or moral objectivism (the view that what is morally right or wrong is independent of the beliefs or feelings of persons). Instead, the framework rests on the conviction that there can be some intersubjective and rational agreement regarding moral norms and judgements and that a rational discourse that implies exchanging arguments and critiques also regarding ethical issues at least makes sense.[12]

---

[10] *Meta-ethics* functions as a kind of "philosophy of science" for (normative) ethics and examines the logical, linguistic, ontological, epistemological and other presuppositions and aspects of (normative-)ethical theories [cf. Marckmann et al 2012].

[11] Apart from mere instrumental or pragmatic rationales, such as that it is better to have *some* guidance to avoid complete arbitrariness and to reduce feelings of disorientation of agents even if the guidance used is in no way morally better than any alternative guidance (because it is rejected that there can be morally "better" and "worse" ways of action, etc.).

[12] This stance is quite compatible with discourse ethics [e.g., Habermas 1991], although outright discourse ethics would go further; this stance can also be easily reconciled with some strands of American pragmatism [e.g., Kitcher 2011] (for reasons of transparency, one of the authors, MM, is also oriented towards these two positions).

c. "Ethics Guidelines", "Ethics in Guidelines" and "Ethics of Guidelines"

After having explained in detail REIGN's positioning regarding bioethics, normative-ethical theories and meta-ethical stances, this section conceptually clarifies what is understood by the term "ethics guidelines". In doing so, this section will also delimit the concept from other similar concepts.

First, it is assumed that "ethics guidelines" formulate recommendations with regard to *actions, rules of action or (social) practices*.[13] The problem with the term "ethics guidelines", understood this way, is that ultimately, most guidelines have an ethical dimension because they are based on normative assumptions. A guideline that recommends treatment A over treatment B on the basis of effectiveness studies has to assume that effectiveness (or the underlying goal of medical treatment, which is health) has a normative value. Only this underlying assumption allows arriving from research findings – *evidence* – at a recommendation for treatment A. Some normative assumptions might be as uncontested as the value of health, but others might be more contentious, for example, whether health gains in the future should be discounted and thereby a reduced value assumed in cost-effectiveness assessments or not [Broome 1994]. Furthermore, some institutional manuals for guideline development require consideration of further normative values that are not already inherent in the measure of interest (e.g., effectiveness). An example would be the *WHO Handbook for Guideline Development* [WHO 2014b], which requires incorporating values such as equity, human rights, and gender equality in all steps of guideline development. Sometimes, these ethical aspects are addressed explicitly by measuring, e.g., effects on equity of various public health measures. Often, these aspects are addressed more implicitly, e.g., by inviting specific people with expertise in these areas to serve in the GDG. Either way, the implicit normativity of seemingly solely empirically based guidelines could be seen as turning all guidelines into "ethics guidelines".

However, a pragmatic definition that allows differentiating among various types of guidelines can and needs to be proposed at this point; the following differentiations among "ethics guidelines", "ethics in guidelines" and "ethics of guidelines" and among different kinds of (ethics) guidelines, would need further empirical work to be better suited for especially practical uses. Nevertheless, theoretically, it is useful for at least some differentiations to be made.

The definition is grounded in the assumption that guidelines can be sorted into those that focus *primarily* on empirical questions and those that focus *primarily* on normative-ethical questions (of course, much depends in practice on how "primarily" is interpreted). The term "ethics guidelines" is therefore used to refer to guidelines that address primarily ethical issues and formulate recommendations with regard to ethical issues and where formulating ethical recommendations is probably also the intention of developing the guideline. The WHO has published several reports that would qualify as *ethics guidelines* according to this characterization; examples of such reports are

---

[13] This implies that the goals and potential consequences of actions will be relevant. However, the focus on actions also means that virtues or intentions will not be the object of interest (or will at least be the object of less interest) for guideline developers. While a pluralistic stance to (bio)ethical theories is embraced (see chapter 2a), this means that certain theories will be less relevant for guideline development in ethics.

"Guidance on ethics of tuberculosis prevention, care and control" [WHO 2010], "Guidance for managing ethical issues in infectious disease outbreaks" [WHO 2016], and "Guidelines on ethical issues in public health surveillance" [WHO 2017]. These guidelines then address primarily issues such as under what circumstances name-based reporting of infectious diseases is ethically acceptable or whether it is acceptable to isolate tuberculosis patients to protect their relatives and others close by. Guidelines that address primarily empirical questions (e.g., effectiveness, safety or cost-effectiveness of various actions), even if clearly exhibiting an ethical dimension as discussed above, will not be considered ethics guidelines. Neither will guidelines that address ethical issues merely in passing be considered "ethics guidelines". Many clinical practice guidelines fall in this category. Clinical practice guidelines that, for example, discuss primarily empirical questions of diagnosis or of which treatment is most effective in slowing disease progression but that also touch on the normative question of how to handle the loss of decision-making capacity [Knüppel et al 2013] would not be considered *ethics guidelines*; rather, the normative questions addressed in such guidelines should be referred to as "ethics in guidelines". However, explicitly integrating ethical issues in such "non-ethics" guidelines can also be based on evidence [see Mertz/Strech 2014]; the same is true of course for HTA reports that incorporate an ethics domain.

For the sake of precision, it should be emphasized that implicit normativity arises not only from value judgements internal to the measure of interest (e.g., effectiveness) but also, as described above, from those externally introduced. Certain institutions, such as the WHO and UK's National Institute for Health and Clinical Excellence (NICE) [NICE 2008], have identified normative considerations that need to be considered alongside the main measure of interest.

By introducing these additional considerations, the abovementioned institutions want to ensure that the guideline development process as such is ethically acceptable (e.g., by ensuring that an implementation of a recommendation does not unintentionally disadvantage any group unfairly). What is at stake here is how guidelines should be developed from an ethical point of view, or the "ethics of guidelines".[14] This debate also includes reflections on the procedural requirements that ethically acceptable guideline development processes have to fulfil and the positive and possibly negative consequences of guideline (or policy) development as such [Winkler 2005]. However, as the point here was only to clearly delineate these distinct, although related, concepts, there will be no further engagement with this debate. While this discussion paper focuses on ethics guidelines, the more implicit normative judgements in primarily empirical guidelines also warrant further academic reflection.

---

[14] Quite obviously, the question what the main measure of interest should be (e.g., effectiveness or costs) is also a question regarding the ethics of guidelines development. This shows that the discourses on the "ethics in guidelines" and the "ethics of guidelines" overlap. (This differentiation between "ethics in guidelines" and "ethics of guidelines" and the possible overlapping is comparable to the differentiation between "ethics in HTA" and "ethics of HTA" [see, e.g., Braunack-Mayer 2006; Hofmann 2014]).

Furthermore, the ethics guidelines that are of interest mainly in this context are not those locally developed to be used only by specific institutions, hospitals or even wards to manage recurring ethical questions or issues [cf. Jox 2014; Neitzke et al 2015; Riedel 2017]. In such guidelines, searching and using evidence has arguably often lower priority than understanding the perspectives of those involved in handling issues and setting ethical "minimal standards" for the respective institutions. The ethics guidelines of interest are those that claim to have a more general applicability (on a national or even an international level) regarding the topic they are addressing (see also table 4). Such guidelines can be conceived as a "normative product" [WHO Evaluation Office 2017] in WHO terminology (see chapter 3a). Finally, ethics guidelines should be differentiated from mere *Codes of Ethics* or *Codes of Conduct* that describe, e.g., the (general) values or virtues of a specific organization or profession. Neither should ethics guidelines be confused with policies that have – more or less – legal force (i.e., that make some courses of action legally binding) and have undergone some democratic legitimization process.

Type of guideline/policy	Applicability of the REIGN framework
Policy with legal force	Not given
Code of Ethics/Code of Conduct	Not given
(Empirical) Guideline with (implicit) ethical dimensions (e.g., clinical practice guideline, public health guideline, professional guideline)	Not given
Clinical Practice Guideline or HTA report with explicitly integrated ethical issues	Given in part
Ethics Guideline, local legitimacy (e.g., specific institution, hospital, ward)	Given in part
Ethics Guideline, general legitimacy (e.g., national, international)	**Given**

Table 4: Overview of different types of guidelines/policies and applicability of the REIGN framework

As the guiding question of this discussion paper is how such ethics guidelines as characterized above can use evidence in their development, it is furthermore necessary to define "evidence" for this context.

d.   Research Evidence

General Understanding of "Evidence"

It is not the goal of the following elaborations to define "evidence" as a general term of philosophy of science or the like but solely as a concept for *evidence-based* approaches in health care, more precisely in the context of ethics guideline development in health care. Therefore, it is assumed that the concept of "evidence" is always in some way related to informing *decision-making processes* (this is not necessarily the case in, e.g., basic science research).

Understanding of Evidence in the EBM Tradition and the Limits of such Understanding

In the EBM tradition, evidence is generally understood in a rather narrow fashion as systematically aggregated empirical (mainly quantitative) data that determine stochastically if an effect or a phenomenon exists and which are assessed for quality and relevance [in part Strech 2008c]. Often, the effect to be shown by evidence in this sense is the *effectiveness* of an intervention.

No word, however, has a "natural meaning" – all the more if epistemological "elevator words", such as "evidence", are concerned[15] –, and there are no epistemologically compelling reasons, not even from the philosophy of science, to understand "evidence" only so narrowly [Sehon/Stanley 2003; Kulkarni 2005]. This is especially true when *research questions other* than those directed to the effectiveness of therapeutic interventions, prevalence of symptoms, diagnostic testing accuracy and other biomedical interests are concerned. As there are movements to expand the idea of evidence-based approaches beyond the realm of medical (cost-)effectiveness [Young et al 2002], insisting on such a narrow definition would prove an unnecessary barrier to, for example, evidence-based policy making. Rycroft-Malone et al, as just one example, propose to define "evidence" as "knowledge derived from a variety of sources that has been subjected to testing and has [*sic*] found to be credible" [Rycroft-Malone et al 2004, p. 83] to make the concept of "evidence" more fruitful for a more general evidence-based practice.

Broadening the Understanding of Evidence

What is therefore needed is a *broader* or, perhaps more precisely, a more *abstract* definition of evidence that can accommodate the specificities of evidence-based guideline development in ethics but also incorporate the more traditional understanding of evidence [for a comparable goal in HTA and for a summary of other understandings of "evidence" in the HTA context, see Stoklosa 2013]. To arrive at such a definition, it is necessary to more closely examine the different understandings of evidence. A main tenet of EBM (and EBHC) is that it is possible to prioritize or "grade" different kinds of evidence. Traditionally, the systematic review ranks above an individual randomized controlled trial (RCT), the individual RCT above cohort studies, etc., thereby defining "levels of evidence" [see, e.g., CEbM 2009]. Therefore, what differentiates the understanding of "evidence" in EBM and alternative approaches in other fields is often essentially the question of *which evidence* is regarded as (most) valuable [Kulkarni 2005] or as the "gold standard" for researchers or guideline developers. Unsurprisingly, the "gold standard" for one area of research does not necessarily have to be the "gold standard" for another: RCTs and subsequent meta-analyses are relevant when a research question is about *causal relationships* that can be tested by an experimental design [cf. Backmann 2017] but are completely meaningless in regard to, for example, merely descriptive/phenomenological questions, e.g., how to accurately describe deeply held attitudes or preferences of patients

---

[15] Following Hacking [1999]: "[…] In philosophical discussions, these words are often made to work at a different level than words for ideas or words for objects, so I call them elevator words. Facts, truth, reality, and even knowledge are not objects in the world […]. The words are used to say something about the world or about what we say or think about the world. They are at a higher level" (pp. 22-23).

or citizens, and even more so when conceptual and normative questions have to be explored, such as which ethical issues have to be considered in a certain context.

Consequently, any conceptualization of evidence that wants to include more traditional understandings of evidence has to support the central idea of being able to determine what the "best" evidence could (ideally) be as in some way epistemically privileged information (implying the information-critical aspect of handling "evidence"). Additionally, a conceptualization of evidence will have to push this idea to a more abstract or general level (e.g., not referring solely to RCTs or possibly systematic reviews) to be able to accommodate different traditions and research questions.

Furthermore, *implicit normativity,* which is part of an evidence-based approach [e.g., Goldenberg 2005; Molewijk et al 2008; Strech 2008c], has to be considered, although this normativity can be part of a normative-ethical and a more epistemic approach (about what we want to know, about how we should organize knowledge production processes, about what constitutes "good research" in this regard, etc.). Therefore, additionally, the conceptualization has to accommodate that mere information without any ascribed function of informing decision-making (about health care-related actions) may not be considered "evidence" at all. Thus, information becomes "evidence" because of a specific interest that is related to the decision-making processes: Never will evidence be generated or synthesized without having at least an epistemic – but most often also practical or even *ethical* – interest in it. In evidence-based approaches, one does not seek empirical evidence about the effectiveness of a clinical intervention just out of pure (scientific) curiosity. Empirical evidence is sought because, in the end, there is an interest in answering the question of whether the intervention should be recommended. This question is of practical, even ethical, interest.

More broadly, evidence should answer a research question or verify or falsify a hypothesis, for example, "Intervention X is effective (under conditions Y for population Z, etc.)". The truth or falsity of the statement (or at least the plausibility of accepting or rejecting it) is important because it provides an argument for or against a decision, namely, either to employ this intervention or not. This decision is directly linked to practical or even ethical interests (i.e., it would be – ethically – problematic to employ interventions that have no benefit).

Or framed in logical terms: evidence will support a (central) *premise* or will directly provide a premise in an *argument* where the *conclusion* is a statement about what should be done (or not be done).[16] In the example given above, the conclusion could be "Intervention X should be promoted

---

[16] Concerning terminology, the following decisions are made: *data/information* is a particular statement or comprises several statements that are contentwise related to each other and are generated by a distinct epistemic process (e.g., scientific method). An *argument* is a combination of statements (e.g., information) where one statement (the *conclusion*) is intended to be justified (or supported) by the other statements (the *premises*); premises can be conclusions of other arguments, i.e., can also be in need of justification/support, and arguments can also be just *enthymemes* (arguments with at least one unstated/suppressed premise – mostly because of practicality – where the unstated/suppressed premise(s) is (are) often obvious for the immediate context and/or supplemented by the audience).

instead of existing intervention X' for population Z", and the central premise for justifying this conclusion (statement) is "Intervention X is effective (under conditions Y for population Z, etc.)".

Stoklosa also defines "evidence" for the HTA context as the "available data relevant to the issue being addressed, question asked, decision made, etc." [Stoklosa 2013, p. 95]; this definition is compatible with our understanding. Stoklosa argues, furthermore, that the "issue being addressed …" has to be interpreted broadly to accommodate the "variety of reasons and purposes" of undertaking an HTA [Stoklosa 2013, p. 96]; this can also be said in the context of guideline development to also accommodate the understanding of "evidence" for *ethics* guideline development.

This all means to accept that evidence is always *normatively laden* – as it has to contribute to giving direction to courses of action, i.e., evidence has to help answer the question of what action should be done – while this is not necessarily the case regarding mere information. However, this does not mean that evidence *per se* is already "directed" to a specific course of action; evidence only *contributes* to giving such direction by providing arguments for accepting or rejecting a respective statement (conclusion) about what should be done.

Defining "Evidence"

Based on the points raised above, the following definition is proposed as groundwork for the REIGN framework [based on a proposal in Mertz 2011]:

> *evidence (general)* = a piece or body of information that is, by varying degrees, qualified within an existing knowledge system to either provide or support a central premise of a possible rational argument for holding a statement (conclusion) true, plausible or right (or false, implausible, or wrong) in a context of decision-making or directing actions.

This definition generally presupposes a conceptual approach oriented by informal (and formal) logic, where even *empirical* evidence "just" provides an argument or supports a central premise as part of an argument [see also Upshur/Colak 2003]. This definition is also corroborated through the orientation to discourse-theoretical approaches (i.e., at the end of the day, nearly everything is part of a discourse between agents exchanging arguments for or against particular statements or actions). (For further explanation and justification, see [Mertz 2011]).

As demanded, the definition *by itself* does not stipulate *particular* information or data as (more or less) qualified or epistemically privileged (this is, for example, perfectly reconcilable with alternative proposals, such as that of Stoklosa mentioned above [Stoklosa 2013]). At first, it is irrelevant whether this information – and related arguments – consists of empirical data points, hermeneutical or historical clues, philosophical arguments, or interpretations of juridical norms. The question of what counts as qualified information becomes relevant only when the statement (that one is seeking to substantiate with evidence, i.e., the conclusion of the argument) and the respective knowledge system are specified. The quality and relevance criteria deeming certain information (and the related arguments) as "qualified" (to a certain degree) have then to be determined against

the backdrop of (a) the subject area that the statement is concerned with; (b) the already established knowledge base in the subject area; and (c) the epistemological/methodological possibilities of exploring whether the statement is true, plausible or right (or false, implausible, or wrong).[17] Therefore, if the statement (conclusion) is "Intervention X is more effective than intervention Y", certain criteria will have to be met by information or data points to be considered more or less qualified to work as a central premise for substantiating the statement (conclusion) or to consider the statement (conclusion) true (e.g., the results of a meta-analysis are more qualified than results of a cohort study). Furthermore, different criteria will be employed for evaluating data used to substantiate statements such as "For action X, it has to be safeguarded that ethical principle Y is not violated" or "Person A in situation S is obliged to do Z".

As already suggested above, quality and relevance criteria for substantiating information correspond to or are equal to the criteria for checking validity claims of statements; these criteria are both also part of the knowledge system. If the statements' validity has been accepted, the statement can become a new part of the knowledge system (see figure 2).

The definition, furthermore, incorporates the epistemological function evidence has, i.e., its function in justifying – or *not* justifying – statements that are claimed to be, e.g., true, plausible, or right in the context of decision-making (= conclusions). However, this definition reflects a *logical* reconstruction and *does not* stipulate any causal or procedural *sequence* between seeking information and arriving at conclusions, nor does this definition say something about the procedures of evidence gathering or synthesizing. Therefore, the definition does *not* entail that one first sets up a conclusion (that has to be upheld under all circumstances), e.g., "Intervention X is effective", and then, afterwards, one just seeks evidence that supports this very conclusion and disregards all other evidence. This would, indeed in most circumstances, render the piece or body of information as *not qualified* for its use as an argument for decision-making.

Finally, the definition does not suggest that all persons are always aware that they are actually setting up, discussing or contesting arguments (their logic or the truth or plausibility of their premises, etc.) when they refer to evidence. Nevertheless, "evidence" can be logically reconstructed in this way even when persons do not reflect upon the function of evidence as providing premises for a specific conclusion (e.g., decision), as they probably seldom do in praxis.

Although not part of the definition itself but of the process of ethics guideline development and similar processes (such as HTA), there is a last point to be addressed regarding the proposed understanding of "evidence": Does the proposed understanding violate the differentiation between "assessment" and "appraisal" (this differentiation is common in the HTA context

---

[17] Saying a statement is, e.g., true, plausible, or right means making a validity claim. Not every statement makes the same validity claim. For example, the statements "'Berlin is the capital of Germany' is true", "'Dark matter exists' is plausible", and "'It is wrong to murder' is right" involve different validity claims, and each type of validity implicates arguments that are necessary for justifying, falsifying or just contesting them (albeit also different ones).

[Garrido/Zentner/Busse 2008, p. 61f; see also Sandman/Heintz 2014] but can also be applied to ethics guideline development)? In HTA, "assessment" has to do with "evaluating relevant aspects of the technology to form a basis for decision", while "appraisal implies some form of recommendation about the implementation of the technology, based on this assessment" [Sandman/Heintz 2014]. "Appraisal" in this meaning is often explicitly not the task of scientific experts (in WHO terminology: the review group), who should only conduct the "assessment" (e.g., collect and synthesize all relevant information). Rather, "appraisal" is the task of other, often more politically situated, persons or groups (in WHO terminology: the guideline development group (GDG)).[18] The definition proposed might be read to imply that it is already the task of scientific experts (review groups) to provide a recommendation (which equals the conclusion in the rational argument that should be justified via the evidence gathered). However, as initially said, who is allowed to finally draw the conclusions and formulate recommendations is not part of the mere definition of "evidence" but of the guideline development processes (or HTA processes) used. Additionally, "evidence" is *not* defined as being already the conclusion (e.g., decision) but only as the available premise(s) for (not) justifying a possible conclusion for the matter being addressed. Therefore, it is unproblematic, even given this definition of "evidence", to conceptualize the task of scientific experts (review group) to be that of **assessment only**, i.e., of working out which evidence is available (and how good it is); this also means: which premises or already whole arguments could justify (or not justify) a possible conclusion (decision), which in the end has to be drawn by the GDG (or other persons, groups or institutions different from the scientific experts/review group). (See also chapters 3e and 4a for further remarks about the process and the tasks of review groups and GDGs).
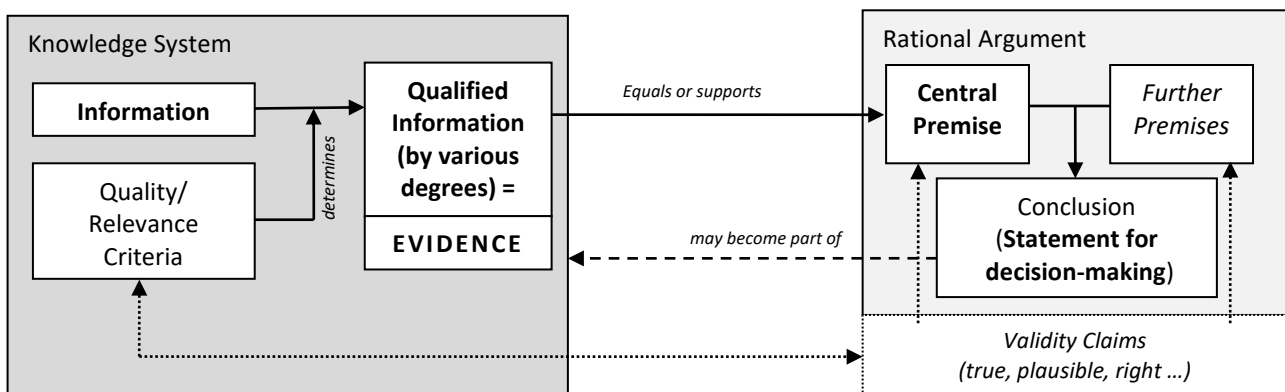
Figure 2: Simplified depiction of general understanding of "evidence" in the context of decision-making

---

[18] Although regarding ethical aspects of a technology, the differentiation can be more difficult to uphold due to the nature of ethical analyses that normally include explicit normativity regarding whether a particular technology is ethically uncritical or not [Sandman/Heintz 2014]. However, as is argued above, a certain amount of normativity is always part of evidence, also regarding empirical evidence, in the context of EBM or EBHC approaches.

Understanding of "Research Evidence"

As this discussion paper aims to specifically fathom the role of "*research* evidence", this term will also have to be defined more clearly. The main difference in this *specification* (*viz* subtype) of the general definition of "evidence" is that the more or less qualified information (that provides a central premise in an argument or supports such a premise) – and the knowledge system and the corresponding criteria for assessing the quality of information and the validity of the corresponding statement – is bound to the state of the art of a scientific discipline or an inter-/transdisciplinary field. Alternatively, the quality and relevance criteria (for considering information more or less qualified) will be defined mainly against a *scientific* knowledge system (e.g., clinical medicine, a particular social science, or academic ethics). The knowledge system is characterized by current knowledge (mostly in the form of scientific literature) as well as methods and methodological considerations (e.g., quality criteria of good research). This also means that the evidence has to be produced by *research actions* and is generally "external" to the persons seeking and using the evidence. While the research is external to its users, the knowledge system and accordingly the research actions are oriented towards particular decision-making contexts (e.g., medical research is producing knowledge to inform medical decision-making). The research is therefore generally designed (particularly in applied research) to inform a certain group of people in particular decision-making contexts (and thereby becomes evidence):

> **research evidence** = a piece or body of information that is, by varying degrees, qualified within the accessible knowledge system (i.e., scientific literature and scientific community) of a scientific discipline or of an inter-/transdisciplinary field to either provide or support a central premise of a possible rational argument for holding a statement (conclusion) true, plausible or right (or false, implausible, or wrong) in a context of decision-making or directing action; additionally, research evidence is information that accounts for the scientific methods used for its generation/justification and subsequently allows for assessing its quality on this basis.

As said previously, research information is produced to fulfil the quality criteria stipulated in the specific knowledge system which again is geared towards particular decision-making contexts. Research information might, however, also be used to inform statements it was not originally intended to inform (e.g., where ethics publications are not used to inform ethical decision-making but, e.g., funding policies for ethics departments). Research evidence used for non-intended decision-making contexts can no longer be judged against established discipline-specific quality criteria, but the specific decision-making context will dictate the criteria to be used.

Empirical and Normative Evidence

After having clarified the understanding of "evidence" and "research evidence", the focus will be on differentiating between two types of (research) evidence that are relevant in the development of ethics guidelines or other ethics guidance: *empirical (research) evidence* and *normative (research) evidence* [see also Reiter-Theil et al 2011; Mertz 2011; comparable: Scott et al 2016, who

differentiate between descriptive and normative ethics analyses in HTA]. It will be important to differentiate between these two, as they will – to some degree – have to be handled differently [Stoklosa/Bond 2013; Scott et al 2016]. The focal point of the following discussion will be *normative* evidence.

Differentiating Empirical Evidence from Normative Evidence

Whether evidence should be considered normative or empirical (the differentiation criterion) will be determined primarily by the *kind of information* sought or which kind of argument and/or central premise should be provided in an argument. Thus, *normative evidence* as "evidence on normative-ethical aspects" is information about values, ethical principles, norms, rules, criteria, or arguments with normative-ethical conclusions –, e.g., arguments prescribing an action (ethical obligation), permitting an action (ethical permission), forbidding an action (ethical prohibition), or allowing or forbidding an action only when certain conditions are met ("safeguards" and "cut-offs", respectively). Generally, normative information and normative evidence, then, is about what *should, or should not, be done ethically* (or perhaps legally) or what *is valuable ethically*.[19] Furthermore, to be considered normative, information has to be intended to be *used normatively* (e.g., "One should do X") and not just be *mentioned descriptively* (e.g., "Person A or all people working in discipline B think(s) 'One should do X'"). In the latter cases, the information is empirical information and thus constitutes *empirical* evidence.

Empirical evidence is accordingly constituted by empirical information, such as whether a phenomenon or effect exists, in which way it exists, how it is perceived by those affected, how it is related to other phenomena or effects (e.g., causal or statistical/functional relationships), or why the phenomena or effect came into existence in the first place (e.g., causes). This information can be quantitative or qualitative.

As mentioned above in the subchapter on the definition of "evidence", empirical evidence entails implicit normativity. Thus, in distinguishing empirical evidence from normative evidence, it is not meant that empirical evidence (even "pure" empirical evidence – see below) does not comprise implicit normativity whatsoever; this would contradict the statements regarding the general characterization of "evidence".

However, the *kind* and the *level* of normativity involved is different: implicit normativity has to do with either epistemic normativity (decisions about how something should be researched, what are relevant quality criteria in research, etc.) or a practical or an ethical *interest* in gathering evidence (e.g., the empirical evidence is to be used to argue for or against normative-ethical conclusions). The first is not the same kind of normativity as in normative evidence (where normative-ethical, not epistemic normativity, is implied), and the latter is not on the same level as normative evidence

---

[19] It would be more correct to refer to *evaluative* information in the latter case; for the sake of simplicity, both are subsumed under "normative information" and, hence, "normative evidence".

because normative evidence explicitly wants to answer questions such as what should be done eth-ically *in a specific situation* or in light of a *certain ethical question* (e.g., is it ethically defendable to use a health technology under the given circumstances, etc.?). Empirical evidence, in contrast, wants to answer questions such as "what are the possible consequences of doing X in a specific situation" without already evaluating (from a normative perspective) the action or condition. How-ever, this empirical information might be used in a normative-ethical context as part of an argument for or against a certain behaviour (as the conclusion of the whole argument). The ethical interest that can constitute the implicit normativity in empirical and normative evidence itself is not about "what should be done ethically in situation X" but about, e.g., "why do we want to know the effects of action X" or "why do we want to know what should be done in situation X".

Therefore, it is important to remember when considering the following that when empirical and normative evidence is discussed, the focal point is whether this evidence contains normative state-ments (such as prohibitions) about a situation or question that the ethics guideline is interested in. Normative evidence does contain such statements, while empirical evidence does not (at least not without additional, possibly only implicitly given normative-ethical premises; see the discussion be-low regarding a "mixed" status of "practical" evidence).

"Practical" Normative Evidence

In practice, normative evidence is often (already) "mixed" with empirical information[20]: The (quali-fied) information that is sought and/or synthesized (e.g., in a systematic review) from the research literature (see below) is, in most cases, *contextualized*. Depending on the concrete "piece" of nor-mative information sought, this can mean different things. For *values*, contextualized information may mean that a specific empirical state of affairs is considered valuable (e.g., "curing AIDS" on the basis of the values of health and wellbeing). For criteria, certain empirical characteristics may have to be fulfilled (e.g., "To understand information in an informed consent process, a patient must be able to recapitulate the given information accurately in its own words" on the basis of a general criterion of understanding information). For *principles,* contextualized information may mean that that they are specified (e.g., "Respect patient autonomy in cases of reduced capacity of judgement by empowering patients through adequate information processing" on the basis of the principle of respecting patient autonomy). For *norms* or *rules,* contextualized information may mean that they are "applied" to specific contexts (e.g., "Every obesity prevention program should make sure that children of low-income families have the same opportunities to engage in sportive activities" on the basis of a norm of equal opportunities or social justice). Finally, for *arguments,* contextualized infor-mation may mean that they include empirical research premises (e.g., "If there is a high probability

---

[20] Although this is acknowledged for information found and used in (social) practice, rejected is a meta-ethical position that supposes that "the empirical" and "the normative" cannot be separated *analytically logically* and are thus (onto-logically?) intertwined in such a way that differentiating them is fruitless from the start (for such a position held in empirical ethics, see e.g., Molewijk et al [2004]). The meta-ethical position is closer to the conception of so-called "mixed judgements", thus indicating that concrete moral judgements are based on partly normative/evaluative and empirical *judgements* and that the overall validity of concrete moral judgements depends on both parts [Düwell 2009].

of adverse risk effects and a low probability of benefit for participants, then, the research should not be undertaken. In research X, there is a high risk of severe adverse effects and only limited benefit for participants. Therefore, research X should not be undertaken").

Empirical and Normative Evidence: A Continuum

The "mixed" status of evidence is not limited to "practical" normative evidence. In the case that information on the probability of an effect taking place is of interest, the information (and subsequently, evidence) is admittedly empirical because, as said previously, the main differentiation criterion for both types of evidence rests on which information (on how things are or should be) is sought. However, as the interest in gaining this information is regularly tied to (implicit) normative considerations – we are interested in knowing the probability of an effect taking place because this is relevant for our decision-making in health care (see above and chapter 3c) – it is quite obvious that empirical evidence is also often in one way or another "mixed" with normative information.

However, even if this "mixing" of normative and empirical information is often the case in actual practice, this mixing does not preclude differentiating between normative and empirical *evidence* – although this mixing requires thinking of these two types of evidence rather as being on a *continuum* than as always having clear-cut distinctions (see figure 3). Nonetheless, there are "pure" forms of these types at either end of the continuum. For example, there *are* values (such as "health", "well-being", "justice", and "the good life"), principles or norms (such as "Do respect patient autonomy", "Do no harm", etc.) without any (or at least without much) contextualization or application.[21] Such "pure" forms of principles and norms can be understood as being hypothetical in the way of "If situation A *were to be* realized [context], then principle/norm B *would* apply". This can be further analysed as follows: "If situation A *were to be* realized, then action B *should* be done [content of the norm]". For values, it is quite the same: "If situation A *were to be* realized [context], then B *would be* valuable". Evidence of this kind will often be found in more philosophically inclined articles that, for example, argue for the special moral value of health or try to justify the basic principle of "respecting autonomy" without further explicating what this would mean in (different) practice(s). Often, such information by itself might not be useful for practitioners, as such abstract values will have to be operationalized for specific contexts. How the wellbeing of people can be furthered might be very different depending on whether we are discussing the context of palliative care or obesity prevention.

---

[21] Philosophically, also such "pure" values and norms rest upon fundamental anthropological or action-theoretical empirical assumptions or facts. For example, "Respect for patient autonomy" rests upon the assumption that persons can be free to make decisions (or have "free will"), and "Do not kill" rests upon the fact that there are empirically possible actions that can (intentionally) end the life of a person. These kinds of fundamental empirical assumptions or facts, however, are normally regarded as "given" in the context of ethics guideline development and, thus, are of no further interest for differentiating normative from empirical information or evidence. Furthermore, the question how values and norms *themselves* are justified – often the main topic of philosophical ethics – and what role empirical data may play in these justifications, is of no interest here.

Nevertheless, such "pure" information can exactly be what is sought after when normative evidence is synthesized: Knowing which empirical states of affairs would be valuable, which principles or norms apply or which actions should be done or not *if* an empirical state of affairs is realized – even if it is not (yet) settled whether this empirical state of affairs actually exists (e.g., potential or hypothetical ethical issues when implementing a specific health technology).[22] Such information can be and often is the result of research in medical ethics or public health ethics. The role of empirical evidence in formulating normative recommendations, then, is especially to settle whether the empirical state of affairs referred to in hypothetical normative statements actually *is* realized. Therefore, ethics guideline development (at least in applied contexts as the one the WHO is working in) will never rely only on pure normative evidence, as there will always be a need for contextualization. Ethics guideline development will not be able to rely only on empirical evidence, as a normative conclusion (recommendation) can never simply follow from descriptive (empirical) statements.

"Pure" Normative Evidence	"Practical" Normative Evidence	"Pure" Empirical Evidence
Non-contextualized or hypothetical values, principles, norms, rules, criteria, and arguments with normative-ethical conclusions *("If A is the case, then B should be done")*	**Contextualized** values, principles, norms, rules, criteria, and arguments with normative-ethical conclusions with empirical premises	Statements about the following: existence of phenomena or effects, characteristics of phenomena/effects, how phenomena/effects are perceived by those affected, relations to other phenomena/effects, causes of phenomena/effects, etc.

Continuum of Normative and Empirical Evidence

Figure 3: Continuum of normative and empirical evidence

Empirical Literature and Normative Literature

Following from this, it also needs to be fully explained where this (research) evidence can be mainly found. Regarding the *sources* of empirical and normative research evidence, a parallel differentiation between *empirical (academic) literature* and *normative (academic) literature* has to be introduced. The latter is sometimes also dubbed "argument-based" or "reason-based" literature [e.g., McCullough/Coverdale/Chervenak 2007; Strech/Sofaer 2012; McCarthy/Gastmans 2015], although this distinction is not convincing given that empirical literature also contains arguments.

---

[22] Strictly speaking, in normative evidence, one can always analytically differentiate "pure" normative information (e.g., hypothetical norms) from "pure" empirical information and, for example, examine the corresponding empirical evidence separately. This, actually, is partly reflected later in the REIGN framework by referring to different complexes of questions or interests when seeking evidence (see chapter 4). Most often, however, due to practical limitations, this analysis cannot be done excessively.

Differentiating Empirical Literature from Normative Literature

These two kinds of literature are primarily differentiated by the methods used to produce the respective (academic) literature. A published empirical study is empirical literature because a respective scientific method was employed (e.g., a survey, RCT or interview method). In normative literature, other methods, such as philosophical-argumentative, hermeneutical, theoretical or other scholarly methods, are used. Additionally, however, goals play a role in determining whether a publication should be considered empirical or normative. Normative literature aims to morally (or legally) evaluate judgements, decisions, actions, institutions, etc., or aims to determine/prescribe which decision or course of action is morally (or legally) obligatory, forbidden, or permissible.[23] Empirical literature, in contrast, aims (only) to describe, explain or predict phenomena (including possibly actions) or effects (of, e.g., actions), whether as a result of actual empirical research or by providing a theory based on synthesizing results of several empirical research studies over time.

Typical examples of empirical literature are clinical trials, economic studies and socio-empirical studies (including qualitative studies about perceptions or opinions). Typical examples of normative literature would be philosophical (specialist) articles about ethical questions, legal literature, governance-related literature, and sometimes (more theoretically inclined) social science articles. While in many cases, these criteria will allow clear classification of papers, there will be some borderline cases where it might be difficult to decide to which kind of literature the specific article belongs. This will prove particularly true for empirical-ethical studies – studies following the idea of "empirical ethics" (see chapter 3b) – which combine empirical methods and aims with normative-ethical methods and aims. However, in many cases, these should probably be subsumed under "normative literature" because of their overarching normative aim.

Correspondence of Normative and Empirical Evidence and Normative and Empirical Literature

Having differentiated empirical evidence from normative evidence and empirical literature from normative literature, there is no one-to-one correspondence. Empirical evidence will build mainly on empirical literature but may also be based on normative literature if the normative information provided is "bracketed" (see below). Normative evidence, however, will in many cases be generated from normative and empirical literature because in many cases, empirical literature can help answer a normative question. If one, such as Knüppel et al [2013], is interested in identifying, for example, (potential) ethical issues in dementia care, such information can also be found in articles considered empirical literature, although the information will be introduced (originally) as empirical evidence. One example would be studies describing prevalence of late diagnosis of dementia; although this prevalence is presented as an empirical fact, this can also be understood as an ethical issue or used

---

[23] More in detail: "Normative literature (i) aims to morally or legally evaluate judgements, decisions, actions, (social) practices, technologies, institutions, organizations or generally states of affairs, and/or to determine/prescribe which decision or course of action is morally or legally obligatory, forbidden, or permissible, or should be so; or (ii) aims to develop, interpret or criticize evaluative or prescriptive concepts that are required for the former aim" [Mertz 2017, own translation]. Point (ii) allows for including theoretical literature that does not, in itself, aim for evaluation or prescription but, e.g., tries to descriptively clarify ethically relevant concepts.

as an argument for increasing information campaigns – provided certain normative principles or values that render late diagnosis ethically problematic (e.g., because of consequences or by hindering patient autonomy) are taken for granted.

However, there is an argument why normative literature is especially valuable for finding normative evidence: this kind of literature is *meant* for providing normative information; is generated accordingly with suitable methods, procedures and discourses; and is controlled for quality in this regard. The results of empirical literature, in contrast, were not meant to be used and read as normative evidence and are thus somehow alienated when used as such and are, generally, without a similar support regarding methods and discourses.

Lastly, one more point already alluded to above has to be emphasized. Primarily normative information – because in the context of the literature or paper at least used to answer a normative question – can be used to answer an empirical question and thereby turn into empirical evidence, and *vice versa*. This, however, is only possibly under particular circumstances: To turn normative studies into empirical evidence, the contained information will *always* have to be "bracketed" or interpreted as an expression of an attitude, conviction or belief of someone. Accordingly, normative information will have to be turned into the form "people of community X believe Y" to be considered empirical evidence. One example would be a study or review describing beliefs about right or wrong actions in the community of professional ethicists according to their publications. Empirical studies, however, can contribute to normative evidence only if the empirical evidence is presupposed and interpreted through *normative assumptions*, e.g., by means of a normative framework.[24] Put differently: Empirical information or evidence can be used to answer an ethical question (e.g., whether an information campaign is ethically acceptable) because some normative claims are (implicitly) already presupposed. For example, an increase in stigmatization after an information campaign can be seen as an ethical issue of or argument against the campaign only where some form of a principle of beneficence is accepted.

Normative assumptions might of course also guide the information collection from normative literature (e.g., what constitutes an ethical issue can be defined by recourse to various ethical theories). However, this does not necessarily have to be the case, as the normative assumptions of the publication itself (possibly varying across articles) can simply be accepted by the evidence collectors (e.g., the first paper defines ethical issues by using a consequentialist framework, and the second paper by using a deontological framework). This is generally unproblematic for evidence collection because in describing what ethical issues are discussed, one just follows the framework of the author. It can be a problem, however, in regard to evidence *synthesis*, for example, by means of systematic reviews (see also Appendix C).

Table 5 provides a short summary of the difference between normative and empirical literature.

---

[24] Which is just a consequence of the is-ought gap and Hume's Law ("One cannot deduce an 'ought' from an 'is'!").

Normative (academic) literature	Empirical (academic) literature
Differentiation by the methods used to produce the respective (academic) literature	
• e.g., philosophical-argumentative, hermeneutical or theoretical methods	• e.g., experimental (RCT), survey, interview, and document analysis methods
Differentiation by the goal of the literature	
• aims to morally (or legally) evaluate judgements, e.g., decisions, actions, or institutions; or aims to determine/prescribe which decision or course of action is morally (or legally) obligatory, forbidden, or permissible	• aims (only) to describe, explain or predict phenomena (including possibly actions) or effects (of, e.g., actions)
Typical examples	
• philosophical (specialist) articles about ethical questions, legal literature, governance-related literature, perhaps (more theoretically inclined) social science articles	• clinical trials, economic studies or socio-empirical studies (including qualitative studies about perceptions or opinions)
Relation to normative evidence	
• is "meant" for providing normative information • provides pure or practical normative evidence	• is somewhat "alienated" from its original purpose when used for providing normative information • provides under certain conditions (if interpreted through normative assumptions) primarily practical normative evidence
Relation to empirical evidence	
• can provide empirical information if "bracketed" (e.g., what people think is right or wrong) • can under certain conditions (see above) provide pure empirical evidence	• is "meant" for providing empirical information • provides pure empirical evidence

Table 5: Characteristics of normative and empirical literature

e.   What Type of Evidence can contribute to Ethics Guideline Development

Before a framework for considering evidence in ethics guidelines can be presented (see chapter 4), one final conceptual question has yet to be answered: It will be fundamental to consider the proper role evidence or evidence-based procedures can and should play in the *process* of (ethics) guideline development. It might sometimes be held that evidence in the context of ethics guideline development should determine whether one should – ethically, perhaps also legally – perform or not perform a certain action or what has to be considered – from an ethical or legal and, thus, normative point of view – when performing a certain action. This section will therefore more closely examine

whether evidence *alone* can determine a recommended course of action. This section will thereby specify the proper role of evidence in ethics guideline development.[25]

Moral and Ethical Expertise and their Implications for Evidence

While empirical evidence is clearly needed to develop recommendations for ethics guidelines, the justificatory burden will instead rest with normative evidence. This section will therefore focus on normative evidence and its ability to determine by itself (or in combination with empirical evidence) the recommendations to be produced by guideline developers. While this question has thus far not been discussed in the academic literature, this question resembles one that has been more thoroughly addressed already: the possibility and role of moral and ethical expertise or moral and ethical experts, especially in drafting recommendations. Differently put, the question is whether an expert in ethics is justified – or legitimized in the context of a formal process, such as guideline development – to decide which recommendation is the "right" one (simply because she/he has moral or ethical expertise).

Quite a few papers have been written on this subject [e.g., Weinstein 1994; Yoder 1998; Grunwald 2004; Cowley 2005; Gesang 2010; Birnbacher 2012]. These articles converge largely on the point that experts in ethics ("ethicists") cannot decidedly answer what should be done; i.e., ethicists do not have *moral expertise* [Grunwald 2004; Cowley 2005; Birnbacher 2012]. However, ethicists add an important perspective to committees or boards tasked with developing ethics guidance because of their specialized skills (e.g., constructing logically valid argumentations) and knowledge (e.g., of various ethical theories and principles). This set of skills, which is called *ethical expertise*, will facilitate the analysis of ethical issues, elucidate normative principles that various viewpoints are based on and allow mediation between various positions. By using these skills, ethicists *improve* final recommendations while *not determining* them.

While this position is not without critics, especially because it often rests on a particular meta-ethical stance that rejects the possibility of objective moral knowledge [see, e.g., Gesang 2010], the authors argue that it is the most plausible and pragmatic position in the context of established processes of guideline development. Therefore, it is analogously assumed that while normative evidence brings important information (e.g., an overview of all raised arguments for and against a certain policy) to the GDG and can thereby improve resulting recommendations, such evidence cannot give an ultimate answer to the normative question at issue.

---

[25] In fact, evidence "itself" never determines anything, as assessments and decisions – which are inevitably value laden – are always made by *persons* who generate/synthesize the evidence or are informed by the evidence [e.g., Rycroft-Malone et al 2004].

Proper Role of Evidence in Ethics Guideline Development

It is thus assumed that evidence as such cannot determine a recommendation due to the *normativity* of the questions posed in ethics guidelines. Evidence can only inform deliberations by providing "good" arguments for what should be recommended; evidence cannot conclusively answer questions about "what should be done". The burden of justifying the recommendations given ("conclusions", see chapter 3d) rests with the *GDG*. The task of the group is to enter into a *consensus process* in which, for example, certain ethical issues will be prioritized or ethical principles (such as those regarding risks and benefits) will be balanced to reach a recommendation ("appraisal" step vs. "assessment" step, see also chapters 3d and 4c). Evidence in normative-ethical contexts therefore cannot and should not replace deliberations and consensus-seeking processes, as it cannot obviate deliberations about what should be done (this resembles the approaches of at least some National Ethics Commissions [Presidential Commission for the Study of Bioethical Issues 2016]).[26]

In this regard, how ethics guidelines are developed, however, does not differ from how empirical guidelines are developed, even though the importance of arguments might be slightly higher when developing ethics guidelines. As discussed above (see chapter 3c), most guidelines exhibit at least an implicit normative dimension. This implicit normativity makes it necessary to enter into deliberations to critically reflect upon the importance of various values or ethical principles at stake and balance and weigh them either way. In practice, evidence may seem to make deliberations unnecessary: Where the literature review conclusively shows that intervention X is more effective than intervention Y, the recommendation will in most circumstances be to administer intervention X. Therefore, one might be tempted to believe that evidence can indeed obviate discussions and determine recommendations.

However, such an attitude towards evidence has to be explicitly warned against – particularly in the context of ethics guidelines. Evidence *per se* provides *arguments* for those informed by it to make a decision regarding a recommendation (see also the used understanding of evidence, chapter 3d). An argument might be so conclusive that there is not much of a debate, but this does not imply that there is no deliberation at work at all. In the example given, the (ethical) value – benefit for patients – and its valuation – benefit is a highly important outcome – is *implicitly* assumed, straightforward, and barely disputed; therefore, it seems as if the evidence "alone" closes further discussion.

However, as soon as further evidence conclusively shows that the same intervention X, although effective, bears considerable risks for a patient or restricts his or her freedom in important ways, the evidence alone does not "close" but rather opens an *ethical* discussion about balancing the respective ethical values or principles. This shows that deliberation and consensus processes are also inevitable in developing empirical guidelines, especially in developing ethics guidelines. This also

---

[26] This is also highly compatible with – or would even follow from – a discourse ethics approach (see chapter 3b).

demonstrates how important – apart from a solid evidence base – the composition and decision-making processes of the GDG are for high-quality ethics guidelines.[27]

Informational and Bias Reduction Functions of Evidence

If evidence cannot determine the final recommendation, what role can evidence play in ethics guideline development? First, evidence fulfils an *informational function*: Evidence adds relevant theoretical background information, brings into the discussion arguments the GDG had not considered yet, clarifies points raised and generally broadens the perspectives of those involved in guideline development. Considering the evidence enables consideration of the complete (or a more complete) set of relevant ethical issues, arguments, and ethically relevant facts. This is important because formulating recommendations depends on weighing and balancing *all* relevant issues and arguments, thus making it possible to choose the policy supported by the "best arguments", e.g., the policy most ethically well justified given, i.a., the stated norms, principles or values [comparable: Koplin/Selgelid 2015]. Guideline processes that systematically consider evidence will accordingly be primarily better informed and may therefore reach better, more considered recommendations.

Second, evidence fulfils a *bias reduction function*. Active engagement with evidence will guard against potential capture of the process by powerful individuals with agendas of their own. People might have biases (or intellectual conflicts of interest [Bion 2009]) towards certain ethical issues or certain arguments for all kinds of reasons and might therefore prefer to focus guidelines on one particular topic or to provide one specific recommendation (e.g., because such people are personally concerned, have themselves witnessed certain forms of mistreatment, or have specific political/ideological interests). Or, as the ethicist Dan Brock, serving on the President's Commission for the Study of Ethical Problems in Medicine, USA, stated in 1987: "[W]e can become wedded to particular views or general theories so that we fail to recognize or acknowledge the difficulties facing them" [Brock 1987]. If members of GDGs have to justify their choices against the backdrop of evidence – whether an overview of ethical issues arising in a certain context or arguments to act in a certain way – it will be more difficult to push the process in one particular direction to satisfy purely personal preferences. Explicit consideration of evidence might accordingly be an important tool for counteracting the influence of biases existing within GDGs on the decision-making process. This does not mean that engagement with evidence *per se* will reduce the biases people hold. It often takes more to change peoples' deeply held beliefs. However, introducing a process of evidence consideration might reduce the influence of biases on the outcomes because such a process requires one to explicitly defend one's own views.

---

[27] It would also be highly important to address what the process of such deliberations should look like (for example, should it be based on Delphi methods [e.g., Linstone/Turoff 2002] or a nominal group technique [e.g., Van de Ven/Delbecq 1974]), who should be part of such a deliberation group process (should participants include various experts, patient representatives, and other stakeholders), and what role may ethical theories and approaches play in regard to weighing and balancing different ethical values or principles, as the quality of the resulting guidelines will (possibly largely) depend on these factors. However, this discussion paper focuses exclusively on those factors related to an evidence-based approach to developing guidelines and respective recommendations.

The guiding question for the framework was accordingly *what information* is needed to best inform guideline development but not how evidence can provide an ultimate answer to the normative question at issue or completely obviate deliberations. Having these considerations in mind, it is now the time to closely examine how empirical and normative research evidence can be used to develop ethics guidance.

4. REIGN FRAMEWORK

The REIGN framework is based on three meta-questions that a guideline developer has to answer:

- "**For what** (is evidence needed)?" (consequences of actions, arguments for actions, etc.)

- "**From where** (is evidence gained)?" (sources, materials, methods, etc.)

- "**Which type** (of evidence should be used)?" (regarding quality, "level of evidence", etc.)[28]

In the following, the framework will be presented roughly following these three meta-questions.

a. Evidential Support Components (ESC)

Ethics guidelines can contain recommendations for or against actions (*direct recommendations*, which are mostly substantial/material) or recommendations to avoid ethically problematic behaviour when a specific action is planned (*indirect or "safeguard" recommendations*, which are mostly procedural). Sometimes, recommendations about how an institution or social system should be "designed" (e.g., regarding incentives, internal rules, and hierarchical structures) can be part of ethics guidelines (*design recommendations*). All these types of recommendations are (ideally) based on empirical and normative research evidence.

In conceptualizing how evidence can inform guideline development, various normative questions were identified that will generally (explicitly or implicitly) have to be answered by guideline developers to arrive at recommendations. These overarching normative questions will have to be addressed through deliberations in the GDG. However, these questions can be split into various sub-questions, of which some will be answerable by evidence. Such complexes of questions can be seen as *components* of the justificatory system of statements underlying ethics recommendations. Answering these questions can, and possibly should, be supported or informed by empirical and/or normative research evidence. These complexes are called *evidential support components* (ESCs) in the REIGN framework.

The framework differentiates five such ESCs: **(a)** *value base* or *ethical corridor* (this ESC identifies which basic normative principles are assumed to be action guiding in the area the guideline targets), **(b)** *conceptual disambiguation* (this ESC clarifies the various meanings of central concepts or terms), **(c)** *need for action* (this ESC identifies ethical problems and substantiates the need for ethics guidance), **(d)** *strategies for addressing needs* (the strategies composing this ESC identify solutions for the ethical problems), and **(e)** *(hypothetical) arguments for actions* (the arguments composing this ESC are evaluations of different solutions and accordingly actions, including consideration of the probable "outcome" of solutions). (See figure 4 for an overview of the decisions for which there can

---

[28] "For whom (is evidence gained)?" can be considered, too, as it is relevant for stakeholder orientation. However, as the evidence should inform a GDG, this question is already sufficiently answered in the context at issue and, thus, will not be further elaborated.

be evidence from each ESC and how different ESCs relate to each other. See also table 5 and figure 5 at the end of this section for further information).

Evidential Support Components (ESC)   *Operational Steps*

```
┌─────────────────────────────────────────┐
│           ESC 1: Value Base               │
└─────────────────────────────────────────┘
                        ┌ · · · · · · · · · · · · · · · · · · · ┐
                        :   Definition of basic principles      :
                        └ · · · · · · · · · · · · · · · · · · · ┘
┌─────────────────────────────────────────┐
│      ESC 2: Conceptual Disambiguation      │
└─────────────────────────────────────────┘
                        ┌ · · · · · · · · · · · · · · · · · · · ┐
                        :   Definition/use of basic concepts     :
                        └ · · · · · · · · · · · · · · · · · · · ┘
┌─────────────────────────────────────────┐
│          ESC 3: Need for Action for...     │
├──────────────┬──────────────┬─────────────┤
│ ...Concept #1│ ...Concept #2│ ...Concept #n│
└──────────────┴──────────────┴─────────────┘
                        ┌ · · · · · · · · · · · · · · · · · · · ┐
                        :   Definition of relevant problems      :
                        └ · · · · · · · · · · · · · · · · · · · ┘
┌─────────────────────────────────────────┐
│       ESC 4: Strategies for addressing...  │
├──────────────┬──────────────┬─────────────┤
│ ...Problem #1│ ...Problem #2│ ...Problem #n│
└──────────────┴──────────────┴─────────────┘
                        ┌ · · · · · · · · · · · · · · · · · · · ┐
                        :   Definition of relevant solutions      :
                        └ · · · · · · · · · · · · · · · · · · · ┘
┌─────────────────────────────────────────┐
│   ESC 5: (Hypothetical) Arguments for...   │
├──────────────┬──────────────┬─────────────┤
│ ...Solution #1│ ...Solution #2│ ...Solution #n│
└──────────────┴──────────────┴─────────────┘
                        ┌ · · · · · · · · · · · · · · · · · · · ┐
                        :   Definition of relevant arguments      :
                        └ · · · · · · · · · · · · · · · · · · · ┘
```
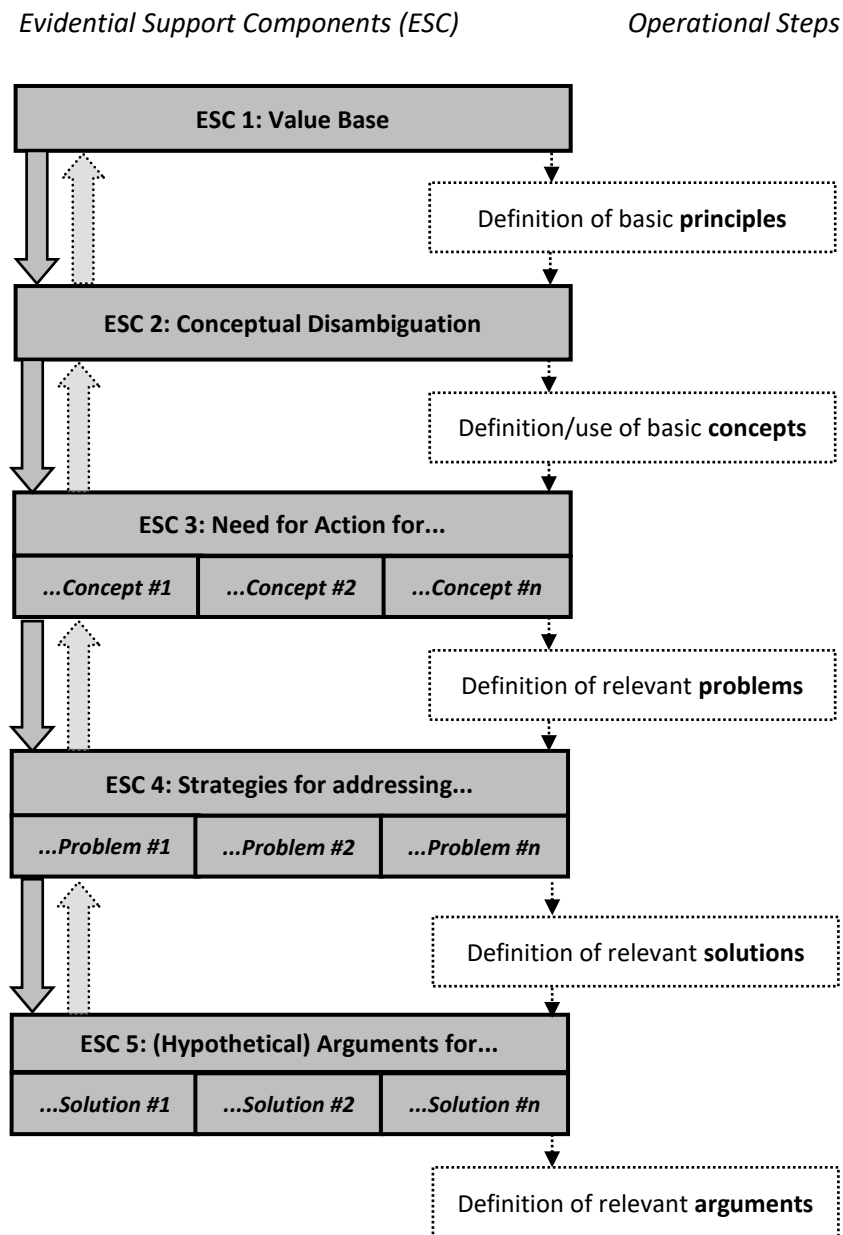
Figure 4: Overview of the REIGN Framework

For each of these five ESCs, (research) evidence is possible and – at least ideally – also required. In practice, the importance of supporting decisions associated with the various ESCs with evidence will vary. It will be the task of the GDG to judge how important further evidence will be against the backdrop of already known or available evidence or the current state of the art of an ethical debate. For example, it can be undisputed in the scientific literature, by experts and/or the public that something is an ethical problem and that there is a clear need for action. Thus, in such a situation, it will be more relevant to direct resources to find evidence regarding strategies for addressing this need

or for identifying arguments for (ethically obliged or at least permitted) or against (ethically forbidden) implementation of the identified strategies – or arguments for specific conditions that have to be guaranteed for an action to be ethically unproblematic (ethically permissible only if conditions X, Y … are met). It will make less sense in this context to invest in collecting evidence to substantiate the need for ethics guidelines. Additionally, which ethical issues should be addressed by the institutions commissioning the guideline might be externally fixed, thus rendering further collection of evidence in the "need for action" ESC unnecessary.

Accordingly, further evidence will not always improve – or will sometimes only marginally improve – the decision-making process in terms of stability of the underlying justificatory system. ESCs should therefore be seen as mere theoretical possibilities where (further) research evidence might be needed, particularly because it is seldom possible to cover all ESCs due to costs, time consumption, workload, and available expertise. It will be the responsibility of the GDG to take a stance – grounded in arguments and transparently made value judgements – which ESCs require further (research) evidence collection to arrive at justified ethics recommendations. Sometimes balancing the need to limit costs and the need for further substantiation of the justificatory system will become necessary.

When deciding whether to engage in more evidence collection, the GDG might want to consider three relevant dimensions. (a) *Relevancy*: Whether the ESC is already answered for external reasons should be considered. If the WHO, for example, has fixed the value base for the GDG or specifically commissioned the group's guideline developers to address specific ethical issues, further evidence consideration will not change any decisions on these dimensions and is therefore futile or irrelevant. (b) *Knowledge base*: Whether the knowledge base the GDG has access to is already sufficient should be considered; for example, in cases where all the relevant experts who have written all the relevant papers are already represented on the GDG, conducting a systematic review for relevant papers might not be very useful. Further evidence collection is unnecessary also because systematic reviews on relevant aspects have already been conducted and can be considered by the GDG. There already exists, for example, several systematic reviews collecting ethical issues, values and norms and arguments for various contexts [for an overview, see Mertz/Kahrass/Strech 2016; Mertz/Strech/Kahrass 2017]. (c) *Proportionality*: Whether the financial, time and other costs associated with evidence collection are justified given the expected benefits of evidence collection should be considered. For example, if guidelines are urgently needed in a crisis situation and not having them has significant costs in terms of delaying disaster response, engaging in a year-long thorough evidence collection process might be less justifiable. When considering proportionality, it should also be considered that there are various methods of collecting further evidence apart from systematic reviews (see chapter 4b on sources of evidence) and that these methods require different levels of resource investment. Additionally, each method can be implemented more or less stringently and comprehensively. Clearly, these methodological choices will impact the quality of the evidence base (see chapter 4c on the quality of the evidence body), but deciding against or for a less

reliable method for evidence collection might be justifiable in light of scarce resources. These choices should be made transparent by the GDG.

Additionally, although, in figure 4 above, the ESCs are depicted mainly linearly to avoid information overload in the figure, in practice, there can be feedback loops between all ESCs, not only those directly following each other (e.g., it is possible to return to ESC 1 while working on ESC 3; see also "ESC Order and Logic", the concluding section of subchapter 4a).

After these preliminary remarks, each ESC will now be described in more detail:

ESC 1: Value Base (or Ethical Corridor)

In guideline development, it might be necessary to decide on the basic (ethical) values, norms, principles or even rights (in a more legal sense) that are considered to be "valid" and "binding" in the context of the guideline. All ethical recommendations that, in the end, will be formulated during the guideline development will have to conform to these values, norms, principles or rights, which can be relevant for identifying and especially *resolving* ethical issues, even though the latter cannot be resolved by evidence alone but will have to be resolved through deliberation among the GDG (see ESC 3/2 below and chapter 3e).

As presupposed earlier (see chapter 3b), it is not the task of a methodological framework such as the REIGN framework to define which ethical theories and approaches or, more specifically, which ethical values should underlie ethics guideline development. However, GDGs may decide on a specific theory or a "set" of theories or principles to guide the development of recommendations for the groups' respective guidelines. Such a choice can be considered comparable to deciding upon a specific ethical approach as a method for integrating/evaluating ethical issues in HTA [e.g., Saarni et al 2011; Lysdahl et al 2016a]. Deciding on a "value base" is also comparable to employing a "normative corridor" or "ethical corridor" [Reiter-Theil et al 2011]. The idea of an "ethical corridor" is that certain ethical positions can be excluded in the analysis or rather synthesis of the (normative) evidence because they are regarded as untenable by the institution that develops the guideline. Additionally, the final recommendations must lie within what this "ethical corridor" allows; that is, they have to be compatible with the "value base" (which, however, is something that goes beyond what the REIGN framework can depict, as the "value base" is part of the deliberation process of guideline development).

For deciding on a "value base" and/or "ethical corridor", there can be normative evidence, namely, evidence about which values, norms, principles or rights are seen as central for a certain context (e.g., for public health, nursing, or surgery) or for certain topics of interest (e.g., for handling "big data" applications or biobanks) in guideline development. The normative evidence will be taken mainly from normative literature. An example would be a systematic review of public health ethics frameworks that attempt to provide an overview of the various principles considered relevant in

the context of public health [Lee 2012].[29] However, the "value base" might in some circumstances already be settled because it is predefined by a code of ethics, professional standards, or even legal obligations of the institution or organization developing the guideline.

ESC 2/3: Conceptual Disambiguation

Ethics guideline developers face the challenge that ethical discourses are sometimes conceptually convoluted; therefore, it might be difficult to discuss certain ethical issues. One reason is that the same topic can be discussed under different labels (e.g., abortion, termination of pregnancy or foeticide) and/or are discussed within different academic disciplines with their established and preferred nomenclature. Although this is true for other scientific issues as well, ethical debates might more often lack standardization of central terms possibly because ethical issues themselves (such as abortion) and their underlying central concepts or terms (e.g., human dignity or welfare) might be politically, ideologically or scientifically contested because such concepts or terms (e.g., "post-trial access of drugs" or "community-based participatory research") can themselves be (implicitly) ethical.[30] Even if these concepts or terms (e.g., "assistive technology", "big data" or "biobank") are mainly descriptive, they may have normative implications by influencing the framing of ethical issues (e.g., more as a problem of autonomy or more as a problem of justice) or by affecting which issues are seen as (especially) relevant. Guideline developers need to decide which terms the planned guideline should use. Guideline developers should therefore have an overview of terms used in the public and academic discourse with regard to the issue of interest *and* the normative implications of the various terms. Otherwise, the impact of terminological choices on guidelines cannot be fully reflected and might likely complicate reception of the guidelines by potential users.

Furthermore, even if terms are well established, they may be used in different ways or, in other words, have different meanings or definitions. For example, in a systematic review conducted on definitions of individualized/personalized medicine, Schleidgen et al [2013] found 1459 ends (e.g., to further the development of new treatment measures) and 1025 different means (e.g., by using genetic information) used to specify the meaning of the terms. Using specific ways of understanding a concept or term can thus lead to identifying different ethical issues and respective solutions and may even lead to quite different ethics guidelines in the end. An example would be "incidental findings" that can be understood as encompassing any health-relevant diagnostic finding that was not intended by the diagnostic means used or as (also) including negligent and false positive findings [e.g., Schmücker 2016]. Therefore, it might be necessary to first clarify the variety of understandings of important terms to determine one understanding that will guide further development; otherwise,

---

[29] Further examples of relevant publications (some reviews, some frameworks based on conceptual work, etc.) for different contexts/topics are ten Have et al 2010, Petrini 2010, Abbasi et al 2017 (public health), Shahriari et al 2013 (nursing), Emanuel et al 2008 (biomedical research), and McDougall/Notini 2013 (overriding parents' wishes).

[30] Concepts/terms that are implicitly normative are sometimes also called "crypto-normative", as they seem to be descriptive only on the "outside" but are normatively laden on the "inside"; "post-trial access of drugs" is of course normatively laden (*justice*), as is "community-based participatory research" (*self-determination*, *autonomy*).

conceptual ambiguity will possibly complicate communication within the group and influence the outcome of the guideline development without being noticed by the group and/or hamper dissemination of the guidelines.

Accordingly, this ESC is oriented to answering questions about what terms are used in the academic, societal or political debate on the topic of the guideline and how these terms are understood or defined. An overview of terms or definitions in usage should be considered normative or empirical evidence (depending on the concept in question); however, overviews of arguments for or ethical implications of choosing one particular definition/term should be considered normative evidence. Empirical and normative literature will probably have to be consulted. While various methods of evidence generation will be discussed at a later point, successful systematic reviews of normatively relevant terms (definitions, concepts, etc.) have already been conducted [see, e.g., Schleidgen et al 2013; Sørenson et al 2012].

Either way, in guideline development, the review group cannot decide which term to use or how a term is to be understood in the context of the guideline. The task of the group is to provide only an overview of the different terms, their meanings, and their normative implications (i.e., their general normative implications or implications for further guideline development). While the review group can be asked to propose a specific term or definition on the basis of the found evidence, the final decision to use a specific term or definition always lies with the GDG as a whole.

Depending on the concrete guideline to be developed, this ESC (conceptual disambiguation) and the following ESC (*need for action*) might have to swap places. Generally, the broader or more general the topic of the guideline is, the more likely that it will be necessary to first clarify terms before determining the need for action, as (given the conceptual ambiguity) it is not clear enough what issues may be at stake at all. It might furthermore become necessary to engage (again) in conceptual clarification after having collected evidence for ESC 3 if the new evidence has uncovered further ambiguous concepts.

ESC 3/2: Need for Action

Clearly, one important question is whether ethical guidance is actually needed, or better, for *what* ethical issues guidance is needed. The idea of ethics guidelines rests upon the presuppositions that there actually is (a) an ethical problem, challenge or conflict that calls for action and is (b) widespread, recurrent or prevalent in a current social practice or has critical consequences (although the ethical problem, challenge or conflict might not be widespread) and where (c) there is insufficient ethical orientation and/or established (professional, institutional, legal, etc.) regulation [cf. Neitzke et al 2015]. The justification for developing an ethics guideline – and the usefulness of the respective ethics recommendations – is weak when the guideline is supposed to address a perceived ethical problem that might not sufficiently be seen as such in current ethical debates and maybe by agents working on the ground. It might also be unjustified to develop ethics guidelines when the problem

in question is rarely witnessed by relevant institutions/organizations or generally in the world or when the negative effects of possible unethical behaviour are negligible.

It could be argued that this question should not be addressed by the GDG, and in part, this argument is true. When the WHO decides to commission guidelines, it should have established beforehand whether there is a need to develop these guidelines. The authors of REIGN generally consider this responsibility to lie with the WHO. In addition, where the WHO has already clearly answered this question, there is no need for further evidence collection for the GDG (see also subchapter 4a on the criterion of the relevancy of further evidence collection). However, the WHO often commissions guidelines that are not supposed to address one specific ethical issue but a broader context of healthcare provision or field of research (e.g., public health surveillance or HIV/AIDS). Those contexts or fields are often plagued with many ethical issues that cannot all be discussed within the guideline. Accordingly, at least in these cases, it will be the responsibility of the GDG to decide which ethical issues to address in the guidelines, or differently put, which ethical issues to prioritize.

As part of the need for action ESC, different kinds of evidence might therefore have to be collected. In guideline development, it will be necessary to delineate more clearly what various ethical issues arise in the given context. This delineation has, for example, been done in the context of public health surveillance [Klingler et al 2017], care for patients with amyotrophic lateral sclerosis [Seitzer et al 2016] or big data in biomedical contexts [Mittelstadt/Floridi 2015], where the authors conducted systematic reviews to provide overviews of ethical issues arising in the respective contexts. Such information indicates whether ethical orientation *per se* is needed in this context and will be necessary background information to help determine what particular issues to focus on (although this again cannot be determined by evidence alone but has to be decided by the GDG). With respect to ethical issues, an overview that is provided as part of guideline development might also help to ensure all relevant issues are considered and nothing of importance is missed. Information collected on ethical issues should be considered normative evidence and should rely on normative and empirical literature.

It will also be important to further investigate the relevance of specific ethical problems or challenges that the GDG plans to focus on and the respective urgency of dealing with these challenges. This investigation includes examining the empirical prevalence of (known) ethical problems or challenges (whether their occurrence is widespread or rather marginal). Furthermore, it might be necessary to probe stakeholders' or affected people's perceptions and opinions regarding these ethical problems or challenges (or the social practice itself) or to investigate preferences of the relevant population. If an ethical issue is not perceived as important, it might be less justified to invest resources in providing orientation with regard to this particular question. The matter of relevance and urgency may be answered mainly by empirical evidence; as "need for action", however, is an evaluative concept, this ESC might also be considered normative evidence, as it can be seen as collecting normative arguments as to why a specific ethical problem is very important and should be

addressed. This evidence, however, relies primarily on empirical literature. Additionally, with respect to urgency, arguments for prioritizing certain ethical issues can be subsumed and collected under the need for action ESC. These arguments are considered normative evidence and found predominantly in the corresponding normative literature.

Lastly, it might also be necessary to collect evidence on which guidance has already been published on the topic of interest. This evidence will be needed to point out the gaps in published guidance and help to focus the guideline on the points where orientation is most severely needed. This type of empirical or normative evidence (the type depends on the perspective) will most likely be found in normative (governance) publications.

To summarize the argument, the need for action ESC is important to address the overall question of why there is a need for ethics guidance in the first place and to decide which ethical issues do need intensive or urgent consideration and should accordingly be addressed by the guideline.

Where different terms or concepts are used (as identified in ESC 2), it might become necessary to differentiate which ethical problems arise for which term or concept. However, in most cases, guideline developers should decide on clear terminology and definitions before engaging in evidence collection for ESC 3.

ESC 4: Strategies for Addressing Need

When ethical problems or challenges have been identified and the need for action (i.e., providing ethical orientation, improving social practice, etc.) is documented (ideally) by evidence (see ESC 3), the following questions will inevitably centre around what can be done to prevent, mitigate or resolve identified problems or challenges that were deemed relevant.

ESC 4 is therefore dedicated to questions related to finding and describing different strategies (course of actions, changes to social practices, etc.) for solving identified problems so that the GDG can decide which strategies should be explicitly considered in developing recommendations. Sometimes, however, it might be common sense what options or solutions are conceivable or even actually available, so there might be no need to seek evidence for different strategies. In contrast, the more complex the social practice or ethical problem is, the more complex possible strategies can be. When discussing adequate public health responses to, for example, a high prevalence of obesity, it might be worthwhile to know the variety of obesity prevention programs developed and implemented as relevant background information for sensible recommendations. When identifying approaches, however, not just singular programs but also more systemic approaches (including those often discussed in terms of "nudging" [Bowie 2009]) should possibly be considered. Empirical evidence generated for ESC 4 will be taken mainly from empirical studies.

When several ethical problems or challenges have been identified – maybe even on the basis of different understandings of central concepts or terms (see ESC 2) – it will be necessary to identify

solutions for each ethical problem or challenge, while additional evidence generation will not necessarily always be demanded.

ESC 5: (Hypothetical) Arguments for Actions

Normative-ethical questions are about what should be done. Understanding the problem and knowing possible solutions to it are just one prerequisite for answering such questions. Essential to arriving at justified recommendations are arguments for or against specific courses of actions (or arguments for doing something only when a certain condition, a "safeguard", is given) and, thus, arguments for or against possible solutions to an identified problem.

ESC 5 is therefore dedicated to the question of which course of action addressing a certain ethical problem or challenge should be preferred to other options, i.e., why, from an ethical point of view, a solution is better than other proposed solutions. This question will have to be answered by the GDG to arrive at final recommendations. Evidence that can support the GDG in arriving at a justified decision provides an overview of relevant arguments. One example of where such an overview of relevant arguments is attempted is a systematic review conducted concerning whether (or not) research participants should have access to trial drugs after the end of the clinical trial [Sofaer/Strech 2011]. The information collected here is normative-ethical in kind but will again likely rely on empirical and normative literature. Most often, however, such information will rely on normative literature. Accordingly, the arguments sought as part of ESC 5 are probably often hypothetical because their empirical underpinnings – if applicable – are not or insufficiently explored (e.g., the argument "Obesity program A can increase stigma" might be raised purely hypothetically to argue against said program, but whether it is probable or even likely that the program will actually increase stigma is not shown or even discussed). A hypothetical argument can be relevant if plausible, e.g., if the argument may plausibly warn against a possible harm, albeit it is, of course, even more relevant if the argument is also empirically validated.

Arguments are not limited to ethical arguments in any kind of narrow sense.[31] In ESC 5, all arguments that give direction towards a specific solution (or leads away from it) can be considered[32] because they are, ultimately, bound to values, norms, principles, and so on. While many arguments might be of the kind "Action A should (not) be done because principle A (e.g., respect for autonomy) is not adequately considered", relevant arguments include, for example, economic arguments (e.g., those concerning costs or insurance coverages), organizational arguments (e.g., those concerning

---

[31] Identifying arguments as "ethical" always inevitably depends on a normative framework (e.g., ethical theory) that determines what counts and does not count as "ethical" (e.g., even costs can be conceptualized as ethical reasons because of considerations of justice in a system with limited resources). No attempt will be made here to differentiate "genuine ethical" arguments from "non-ethical" ones. However, what is important is that in the end, it becomes clear within ESC 5 why the arguments have some ethical bearing for the discussed solutions. This underlines again the necessity of having a normative framework when developing ethics guidelines (see chapter 6, "Open Questions").

[32] See wide reflective equilibrium approaches and generally coherentist justification programs in ethics [Arras 2007; Daniels 2017].

hindrances to implementation), social arguments (e.g., those concerning compatibility with the established moral norms or views or acceptability/preferences of relevant stakeholders), or legal arguments (e.g., those concerning whether the proposed solutions are currently illegal or may have legally problematic consequences).[33] Therefore, generally speaking, ESC 5 includes arguments regarding the feasibility of the proposed solution with respect to implementation barriers and other practical hindrances of a strategy identified in ESC 4. The latter could also include – or be based on – descriptions of (problematic) behavioural patterns of social actors. Understanding why some actors behave in a given system in a certain way (e.g., because of an incentive structure, bureaucratic provisions, or pertinent ideological beliefs) can provide an argument for preferring a specific strategy that has better chances of success than a rival strategy does. Therefore (as has been hinted at above), recommendations to be formulated do not necessarily describe only specific courses of action for individual actors but how institutions or societies may have to change (e.g., what should be altered to improve the feasibility of ethically preferable courses of actions).

For a broad ethical evaluation of different courses of action, however, providing an overview of hypothetical arguments is insufficient. Assessing the (expected) consequences of these actions or the probability of realizing certain consequences when implementing an action is of the utmost importance – even if one is not embracing a consequentialist ethics. Having a plausible idea about the (probable) "impact" or "outcome" of actions according to the identified and considered solutions can support or empirically substantiate merely hypothetical arguments and allows for consideration of the intended – and maybe *un*intended – effects of actions. Accordingly, as part of ESC 5, evidence can be collected to answer which outcomes are actually realized (or with a certain probability realized) when certain actions are put in place.

Consequences of interest might be, for example, actual harm; the effectiveness of an intervention; adverse effects; cost-effectiveness of a health technology; and especially social and psychological impacts, such as an increase of stigmatization, problematic incentives, discrimination of certain groups and whatever other consequences have been identified as normatively (hypothetically) relevant (either through evidence or by the GDG). Against this backdrop, it is unsurprising that to answer this sub-question, mainly empirical evidence from empirical studies is collected and analysed (normative literature may hint at further consequences that might be important, but it will generally not help substantiate whether proclaimed consequences actually materialize or with what probability).

---

[33] To obtain an idea of the types of arguments (e.g., arguments referring to (a) health impact, (b) feasibility or (c) costs) that could become relevant in ethics guideline development, frameworks identifying decision criteria (or types of arguments) for the health policy context might be insightful. However, as these guidelines are generally developed for a specific decision context (e.g., resource allocation or priority setting), additional types of arguments might become relevant, and guideline developers should not restrict themselves to the dimensions proposed. The following publications might be particularly interesting to those trying to further structure the search for arguments that can become relevant in ethics guidelines: Guido et al 2012; Tromp/Bultussen 2012; Goetghebeur et al 2008; and Rehfuess et al 2019.

However, it might not always be possible to find salient empirical evidence for the relevant potential consequences (which have been identified as hypothetical arguments for or against an action). Sometimes, empirical research that can produce information about these consequences is not even conceivable or would be unethical (e.g., testing whether putting obese children under severe psychological pressure will harm them emotionally, even if they may reduce their weight under the given pressure). In such cases, it is important to be transparent about the lack of (sufficient and available) empirical evidence and to consider this when developing ethics recommendations; e.g., it is important to openly acknowledge that some empirical assumptions are indeed only *assumptions*, maybe plausible ones, but still unsupported by (good) evidence.

(Relevant arguments may have already been identified by evidence gathering in ESC 3 (*need for action*). Furthermore, exploring arguments in ESC 5 can lead to new ethical issues not already identified in ESC 3).

ESC Order and Logic

To emphasize again: Although, for reasons of clarity, the five ESCs have been presented as separate complexes of questions and sub-questions, the ESCs are all, of course, often intertwined: arguments for prioritizing certain ethical issues might be recyclable as arguments for choosing particular strategies. Furthermore, upon considering various strategies for action, one might realize that certain new ethical issues arise when choosing one particular strategy.

Moreover, the five ESCs are not arranged in a *causal* order but in an order reflecting the *logic of action*. In practice, one will go back and forth between the five ESCs because, as mentioned above, the ESCs have substantial overlaps.

Table 6 provides an overview of the different parts of each ESC and thereby summarizes the preceding discussions. For each ESC, the main question that has to be addressed by the GDG – and which cannot be answered by evidence because of the normativity of the question – is identified. The table further identifies the evidence that can and possibly should support the GDG in arriving at answers for the questions associated with each ESC.

ESC 1	**Question for GDG:** What basic normative principles should guide action and serve as orientation points for the topic of the guideline?
	Evidence to support GDG decision-making: ✓ Overview of normative principles commonly used in the context [NE]
ESC 2	**Question for GDG:** What terms (e.g., abortion or foeticide) should be used for the main topics discussed in the guideline, and how should these terms be defined?
	Evidence to support GDG decision-making: ✓ Overview of the terms used for the main topics discussed [NE/EE] ✓ Overview of the definitions provided for the main topics [NE/EE] ✓ Overview of the ethical implications of/arguments for choosing particular terms/definitions [NE]

ESC 3	**Question for GDG:** What ethical issues should be addressed by the guideline?
	Evidence to support GDG decision-making ✓ Overview of the ethical issues associated with the topic of the guideline [NE] ✓ Overview of the data on the urgency of ethical issues (prevalence, consequences, etc.) [EE] ✓ Overview of the arguments for prioritizing ethical issues [NE] ✓ Overview of the regulatory documents addressing (certain) ethical issues to determine whether additional guidance is needed [EE/NE]
ESC 4	**Question for GDG:** Which strategies for addressing ("solving") identified ethical issues should be considered by the guideline?
	Evidence to support GDG decision-making: ✓ Overview of the strategies for addressing prioritized ethical issues [EE]
ESC 5	**Question for GDG:** Why should specific strategies be recommended by the guideline, and what has to be considered when follow-ing this strategy?
	Evidence to support GDG decision-making: ✓ Overview of the (hypothetical) arguments given to choose either of the identified strategies [NE] ✓ Overview of the data on the (probable) consequences of choosing either of the identified strate-gies (to substantiate merely hypothetical (consequentialist) arguments) [EE] ✓ Overview of the possible implementation barriers and other practical hindrances of the identi-fied strategies [EE]

GDG = Guideline Development Group; NE = normative evidence; EE = empirical evidence

Table 6: Overview of the different parts of the ESCs

Finally, Figure 5 below depicts the function of evidence in the different ESCs in ethics guideline de-velopment as a whole (see also chapter 3e).
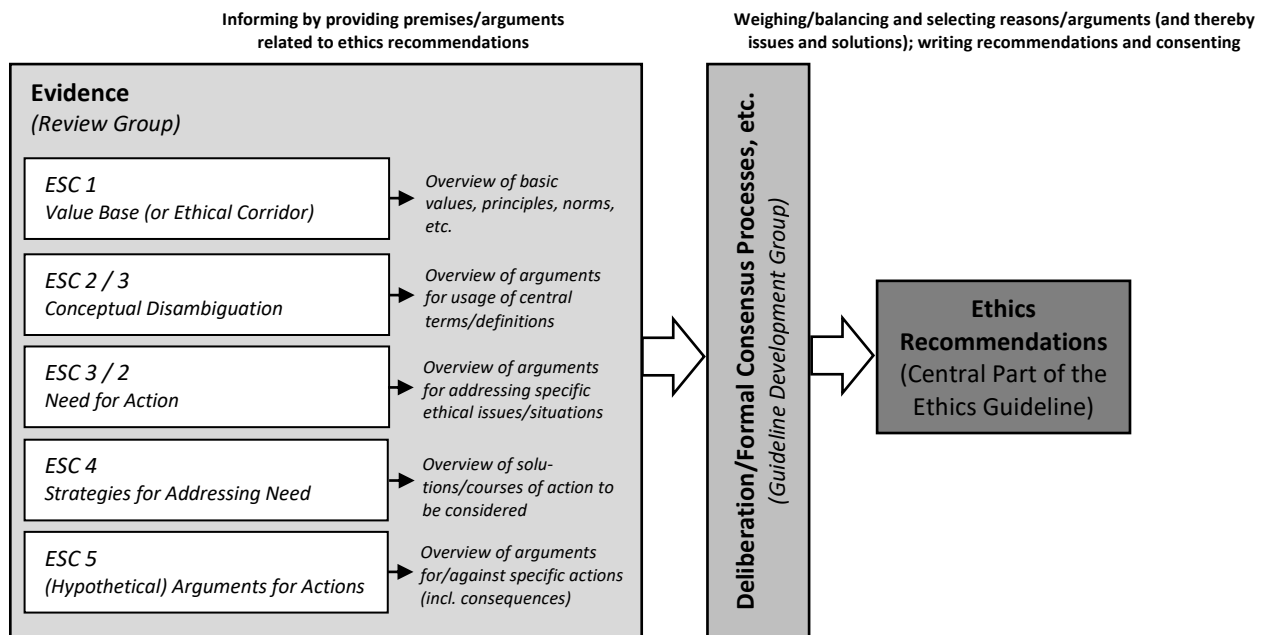

Figure 5: Function of evidence in the process of ethics guideline development

To guide decisions on further evidence collection and analysis, a "toolkit" based on the five ESCs was developed. It is a preliminary toolkit that would greatly benefit from further expert input and user feedback. However, it might be helpful to more clearly structure the process of evidence consideration in ethics guideline development (see Appendix D).

b.   Sources of Evidence for ESCs

As discussed above, to answer the questions that arise in the various ESCs, different kinds of evidence are needed. Some ESCs will rely more heavily on empirical evidence (e.g., especially when looking at policy options or consequences of certain actions), and some more on normative evidence (e.g., when trying to provide an overview of ethical issues that can arise in certain contexts or when trying to provide arguments for justifying various options for action). Next, this discussion paper turns to the question of where to collect evidence (and how to choose and access various evidence sources) for ethics guidelines.

Sources of Empirical Evidence

Much has been written on searching, collecting and synthesizing empirical evidence, and tools (such as the GRADE approach [Guyatt et al 2008]) have been developed to, for example, assess the quality of evidence. As shown, many questions relevant to ethics guideline development can be answered by empirical evidence. Processes already established at the WHO for empirical evidence collection and analysis will be useful in these cases and can probably be translated into ethics guidelines with only minor adjustments. For example, studies on the impact of various obesity prevention programs on stigma will have to fulfil the same quality criteria as other studies examining the adverse effects in nonnormative contexts, and evidence can be collected and analysed by established methods. The *WHO Handbook for Guideline Development* [WHO 2014b] should be consulted in these cases to provide methodological orientation.[34] However, how to approach normative evidence collection and analysis has thus far received only scarce academic attention. Therefore, although collecting (socio-)empirical evidence in the bioethics context might have specific considerations that have to be considered and might pose specific challenges that have to be overcome [see also Strech/Synofzik/Marckmann 2008], the following elaborations will be focused on normative evidence.

Sources of Normative Evidence

When considering sources of normative evidence, one might be inclined to think first and foremost of academic discourse, especially in (interdisciplinary) bioethics or public health ethics but also in philosophy and related disciplines. However, it is important to realize that academics engaged in

---

[34] Additional guidance is constantly being developed for specific areas of the WHO's work. The project "Retrieval, Synthesis and Assessment of Evidence on Complex Health Interventions", for example, tries to provide more specific guidance for complex health interventions. Updates on the project can be found here: https://www.who.int/maternal_child_adolescent/guidelines/development/complex-health-interventions/en/ (24.10.2019).

academic discussions are only *one* group of people who can possibly provide relevant information – although in most cases only the information provided by academics can be considered *research* evidence. In addition, academics will be – generally speaking – most transparent about the origins of the information provided and about the interests that originators have. While normative evidence might also be provided by other groups knowing, working in, having an interest in or being affected by the context/topic of interest, possible conflicts of interest or cognitive biases due to special interests and particular life experiences must be considered when handling evidence stemming from such sources. In the healthcare context, groups that might often have an important perspective are, for example, patients, public health experts, policy-makers or medical practitioners. These groups will be referred to as *stakeholders* in the following. By using various methods (e.g., interviewing or focus groups) or integrating these groups directly in the GDG as representatives, normative evidence from these groups can be collected by asking them to share their viewpoints. Sometimes, researchers might engage with stakeholders and represent their views in academic publications, and thereby, their views become accessible via academic discourse. However, this will not always be the case. Furthermore, these publications present only a mediated view of stakeholders. The voices of various individuals, communities and organizations should therefore also be considered alongside "classical" research evidence as relevant sources of normative evidence (in this, the REIGN approach resembles the approach of the Nuffield Council, as described in chapter 1).

It might be particularly important to engage additional (nonacademic) stakeholders if certain voices are not represented by the academic literature (see also chapter 4c on the quality of the body of evidence). The views presented in the academic literature might, for example, predominantly represent the perspectives of those living in high-income, urban contexts. In these situations, given the context dependency of viewpoints, it might be important to also engage actors from low-income and/or rural contexts. People living in low-income contexts might identify, for example, ethical issues that those living in other contexts are unaware of. As elaborated below, there are several methods for engaging nonacademic stakeholders. In addition, both discourses (academic and nonacademic) can be accessed via different sources: direct interaction with the respective stakeholder/researcher or written documentation. For academic discourse, relevant information can be accessed via the academic literature or the researchers themselves. As most guideline developers will resort to written sources, Table 7 below provides a summary and illustration of different written sources of normative and empirical evidence for each ESC. The list does not attempt to be comprehensive regarding written sources. The aim of the table is to give only some ideas about what, e.g., studies can be gathered or (research) fields approached for collecting and analysing evidence for the respective ESCs. Obviously, which sources are relevant has to be operationalized for the concrete context of the guideline to be developed.

ESC	Written Sources of Normative Evidence	Written Sources of Empirical Evidence
ESC 1 *Value Base (or Ethical Corridor)* *As part of this of interest:* ➢ *Overview of values, principles, norms, etc., relevant in the given context [NE]*	*Normative academic literature, e.g.,* ▪ Philosophical papers/books/book chapters ▪ Theoretical papers/books/book chapters on medical ethics (or, e.g., bioethics or public health ethics) *Empirical academic literature, e.g.,* ▪ Interview studies or surveys about which values, norms, principles are found relevant by participants, etc. *Further publications or written sources, e.g.,* ▪ Reports (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Reports/statements by ethics councils or ethics committees ▪ Laws or regulations (national/international)	– Not applicable for this ESC –
ESC 2/3 *Conceptual Disambiguation* *As part of this of interest:* ➢ *Overview of relevant terms/definitions [NE/EE]* ➢ *Overview of ethical implications of using specific terms/definitions [NE]*	*Normative academic literature, e.g.,* ▪ Philosophical papers/books/book chapters ▪ Theoretical papers/books/book chapters on medical ethics (or, e.g., bioethics or public health ethics) *Empirical academic literature, e.g.,* ▪ Studies with stakeholders about how concepts are understood or used ▪ Possibly theoretical social science papers explicating/discussing certain concepts ▪ Other studies explicating certain terms as parts of larger studies *Further publications or written sources, e.g.,* ▪ Reports or other documentation (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Reports/statements by ethics councils or ethics committees ▪ Laws or regulations (national/international)	*Normative academic literature, e.g.,* ▪ Philosophical papers/books/book chapters ▪ Theoretical papers/books/book chapters on medical ethics (or, e.g., bioethics or public health ethics) *Empirical academic literature, e.g.,* ▪ Studies with stakeholders about how concepts are understood or used ▪ Possibly theoretical social science papers explicating/discussing certain concepts ▪ Other studies (as parts of larger studies) explicating certain terms *Further publications or written sources, e.g.,* ▪ Reports or other documentation (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Laws or regulations (national/international)

ESC 3/2 *Need for Action* *As part of this of interest:* ➢ *Overview of ethical issues [NE]* ➢ *Overview of data regarding urgency of the issues identified [EE]* ➢ *Overview of arguments for prioritizing certain issues* [NE] ➢ *Overview of gaps in guidance documents [EE/NE]*	*Normative academic literature, e.g.,* ▪ Philosophical papers/books/book chapters ▪ Theoretical papers/books/book chapters on medical ethics (or, e.g., bioethics or public health ethics) ▪ "Empirical-ethics" studies *Empirical academic literature, e.g.,* ▪ Document analyses of already existing guidance ▪ Interview studies or surveys on perceived ethical issues arising in a certain context ▪ Possibly theoretical social science papers discussing challenges arising in the context of interest *Further publications or written sources, e.g.,* ▪ Reports (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Reports/statements by ethics councils or ethics committees ▪ National laws or further regulation ▪ Media (newspapers) ▪ Court cases	*Empirical academic literature, e.g.,* ▪ Studies involving those affected by/knowing the context about how prevalent ethical issues are, their probability or severity, or the impact of unethical behaviour ▪ Sociological or psychological research about how those affected experience a specific situation ▪ Medical, psychological or economic outcome research ▪ Science-and-technology (STS) studies ▪ Epidemiological studies regarding the prevalence of certain (health-related) issues ▪ Document analyses of media reports (reports describing how those affected experience the situation, how implementation of a certain intervention has unfolded, etc.) *Further publications or written sources, e.g.,* ▪ Reports (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Reports/statements by, e.g., ethics councils or ethics committees ▪ National laws or further regulation
ESC 4 *Strategies for Addressing Need* *As part of this of interest:* ➢ *Overview of strategies for addressing prioritized ethical issues [EE]*	– Not applicable for this ESC –	*Normative academic literature, e.g.,* ▪ Philosophical papers/books/book chapters ▪ Theoretical papers/books/book chapters on medical ethics (or, e.g., bioethics or public health ethics) ▪ "Empirical-ethics" studies *Empirical academic literature, e.g.,* ▪ Studies involving experts or other stakeholders about possible strategies/solutions ▪ Document analyses of already existing guidance ▪ Document analyses of, e.g., public health program descriptions ▪ Possibly theoretical social science papers *Further publications or written sources, e.g.,* ▪ Reports (on of public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Reports/statements by, e.g., ethics councils or ethics committees

		▪ Program descriptions (e.g., public health programs)
ESC 5 *(Hypothetical) Arguments for Actions* *As part of this of interest:* ➢ *Overview of normative arguments for or against certain activities [NE])* ➢ *Data on real-world consequences of certain activities* ➢ *Overview of practical hindrances to certain activities*	*Normative academic literature, e.g.,* ▪ Philosophical papers/books/book chapters ▪ Theoretical papers/books/book chapters on medical ethics (or, e.g., bioethics or public health ethics) ▪ "Empirical-ethics" studies ▪ Possibly theoretical social science papers ▪ Legal studies *Empirical academic literature, e.g.,* ▪ Studies with experts or stakeholders about their reasons for preferring/rejecting certain actions ▪ Document analyses of already existing guidance ▪ Document analyses of media reports (e.g., reports describing reasons of stakeholders) ▪ Health economics studies ▪ Implementation research (on, e.g., barriers to new practices in a health care system) ▪ Psychological research (e.g., about motivation or explaining behaviour) ▪ Health services research *Further publications or written sources, e.g.,* ▪ Reports (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Reports/statements by, e.g., ethics councils or ethics committees ▪ Laws/regulations (national/international) ▪ Media (newspapers)	*Normative academic literature, e.g.,* ▪ "Empirical-ethics" studies *Empirical academic literature, e.g.,* ▪ (Clinical) effectiveness studies ▪ (Clinical) safety studies ▪ Health economics studies ▪ Psychological or sociological research (on, e.g., practical effects on people) ▪ Health services research ▪ Implementation research ▪ Generally (health) technology assessment studies *Further publications or written sources, e.g.,* ▪ Reports (on public health, HTA, etc.) by national or international institutions or organizations (governmental or non-governmental) ▪ Evaluations of (public health) programs

NE = normative evidence; EE = empirical evidence

Table 7: Possible written sources for collecting/synthesizing evidence for ESCs

Strategies for Collecting Evidence from Various Sources

It has been shown that there are different sources of normative evidence (academic vs. nonacademic and written vs. personal). Additionally, there are different methods or strategies for gaining access to these sources. Academic discourse, for example, can be accessed not only indirectly via academic publications by using systematic or unsystematic reviews [McDougall 2015] but also via direct consultations by conducting presentations, discussions, or consensus groups with academics. Researchers can also be asked to conduct further primary research, e.g., in the form of *theory application,* if no research on the topic of interest has been conducted. This means that one particular theoretical standpoint or various theoretical standpoints (for example, utilitarianism, deontological approaches or principlism [e.g., EUnetHTA and INAHTA; see chapter 2a and Appendix A]) are used to identify and analyse ethical issues in the context of interest or to provide arguments to act in a certain way. The policy discourse will often be accessible via written documentations of some kind (e.g., reports) that can be (systematically) reviewed, but policy makers can also be involved in more direct ways (e.g., via expert consultations). There are also various methods for capturing the perspectives of additional stakeholders (e.g., those affected): open consultation where all are invited to contribute their opinions on (a set of questions concerning) the topic of interest in, e.g., writing, face-to-face workshops/discussions, interviews, or surveys [see, e.g., Nuffield Council; chapter 2a and Appendix A].[35] Table 8 below provides an overview, including short descriptions, of selected strategies for accessing various sources of evidence.

---

[35] Obviously, the realm of research and additional stakeholders might not always be that clearly differentiable. Researchers might be asked to conduct a stakeholder survey, thereby blurring the lines between the academic and non-academic perspectives.

Sources	Evidence collection strategy	Explanation
*Written Sources: Academic Literature**	**Systematic Review**	A literature review that methodically follows ex ante defined steps to identify, synthesize and present relevant research (see Appendix C).
	Unsystematic or Narrative Literature Review	A literature review that identifies, synthesizes and presents relevant research without following an explicated process.
	Several Single Papers	A convenience sample of papers supplies the evidence base.
	Single Paper (n=1)	A single paper supplies the evidence base.
Persons as Sources: Academic Experts	**Consensus Process**	Consensus among academic experts is built regarding the topic of interest by using, for example, Delphi methods.
	Workshop	A face-to-face meeting allowing various experts to present their research and discuss findings among themselves (and with the GDG).
	Commissioned Theory Application	A researcher is asked to analyse the question of interest (e.g., ethical issues in a given context) by using specific theoretical lenses (principlism, consequentialism, etc.).
	Consultation (written or verbal)	Academic experts are asked to present their position on a specific topic or question in writing or verbally during a meeting.
Persons as Sources: Additional Stakeholders	**Interviews/Focus Groups**	Stakeholders share their views in interviews or group discussions.
	Opinion Survey	Stakeholders are asked to share their views in a (postal or online) survey. Compared with interviews/focus groups, this type of survey allows more people to be approached; however, no deeper engagement with stakeholder positions will be possible.
	Consensus Process	Consensus among stakeholder representatives is built regarding the topic of interest by using, for example, Delphi methods.
	Workshop	A face-to-face meeting allowing various stakeholder representatives to present their positions and discuss findings among themselves (and with the GDG).
	Consultation (written or verbal)	Stakeholder representatives or the public are asked to present their positions on a specific topic or question in writing or verbally during a meeting.
*Persons as Sources: GDG members***	**View of a single (or various) member(s) in the GDG**	A single GDG member or a group of GDG members presents the group's views (possibly based on research or field experience) on the question of interest as part of the development process.
	Consensus of all members of the GDG	Consensus among GDG members is built regarding the topic of interest by using, for example, Delphi methods.

* Similar strategies are available for accessing further written documents representing nonacademic discourses. As academic discourse will be the most relevant for the reasons named above, it is the focus here.

** Clearly, the opinions and experiences of a member or various members of the GDG can also be considered evidence. They are always considered part of the process of guideline development and are therefore of less interest here but should still be listed to be as comprehensive as possible. The evidence base for guidelines that rely only on this kind of evidence is particularly weak in these cases.

Table 8: Strategies for accessing sources of normative evidence

Systematic Reviews of Normative Evidence

One method of collecting and synthesizing evidence from research publications has become the "gold standard" and a requirement for developing high-quality guidelines [see also chapters 2a and 2b and examples in Appendices A and B]: systematic reviews (SRs), or at least more systematic literature searches. As already indicated in the chapters above, SRs are also conducted in medical ethics and public health ethics. SRs are used for synthesizing socio-empirical findings (e.g.,

preferences, attitudes or experiences of stakeholders towards or with ethical issues or regulation) and normative-ethical discussions [Strech/Synofzik/Marckmann 2008; Mertz/Krahrass/Strech 2016; Mertz/Strech/Kahrass 2017]. However, the numbers of these SRs are currently in no way comparable to the number of SRs in clinical medicine or health economics.[36]

One reason for the limited use of SRs in medical ethics might be that especially regarding SRs addressing normative-ethical questions, there are still many open questions concerning, inter alia, the aims; types of suitable information; methods for searching, analysing and synthesizing information; and possible reporting standards [Mertz/Strech/Kahrass 2017]. Another reason might be that SRs are not the standard approach in more normatively oriented disciplines. As a method, one could say they were imported into the normative disciplines. Additionally, empirically, it is unclear whether compared to classical more conceptually inclined approaches, systematic reviews can ensure a more thorough overview of issues, arguments or principles.

As SRs involve searching and analysing mainly normative literature, they are sometimes labelled "SRs of argument-based literature" [e.g., McCullough/Coverdale/Chervenak 2007] or directly "SRs of normative literature" [e.g., Mertz/Krahrass/Strech 2016; Mertz 2017]. Within the REIGN framework, SRs are called "systematic reviews for normative evidence" (SRNE), as it was established that normative evidence – which such SRs should ultimately provide – can be generated by relying on normative and empirical literature, albeit with a clear tendency towards normative literature. For the information SRs try to synthesize in the context of ethics guideline development, various subtypes can be differentiated (see table 9).

Type of SRNE	Explanation and Example
SRNEs of **normative arguments**	In focus: ethically relevant arguments for a specific topic (e.g., whether post-trial access to trial drugs is morally proscribed [Sofaer/Strech 2011])
SRNEs of **ethical issues**	In focus: ethical issues (e.g., conflicts) that have to be considered for a specific topic (e.g., assistive technologies for the elderly [Zwijsen/Niemeijer/Hertogh 2011])
SRNEs of **normative concepts**	In focus: ethically relevant definitions or concepts and approaches (e.g., "moral distress" [McCarthy/Gastmans 2015])
SRNEs of **ethical values, norms or principles**	In focus: values, norms or principles relevant to specific courses of actions or (clinical) fields of action (e.g., plastic surgery [Chung/Pushman/Bellfi 2009])

Table 9: Types of systematic reviews for normative evidence (SRNEs)
(originally from Mertz [2017], slightly adapted and shortened; own translation)

These subtypes can be used in different ESCs as a means of evidence collection and synthesis; not all SRNEs are useful for all ESCs (see also Table 7). With respect to normative literature, other types of SRs (e.g., SRs of all-things-considered ethical conclusions as proposed by McCullough et al [2007])

---

[36] The meta-review of Mertz/Kahrass/Strech [2016] found 183 SRs in ethics published between 1997 and 2015; of these, 84 were classified as being SRs of normative literature.

have been conducted that are, however, less relevant to the REIGN framework and are therefore not further elaborated here.

As the WHO has expressed heightened interest in whether and how SRs can be used in the context of normative evidence, a short overview of some methodological issues related to such reviews will be provided in the Appendix (see Appendix C). No similar discussion can be provided for other methods of evidence collection and analysis because of lack of resources. This, however, does not imply that these methods are of no or less value for ethics guideline development.

c.   The Quality of the Evidence Body for Particular ESCs

After having discussed the sources of empirical and especially normative evidence and various strategies to access these sources, it will also be important to reflect on the quality of the (normative) evidence collected by the various strategies presented. This is often called "quality appraisal" (or "critical appraisal") in the context of systematic review methodology [Higgins/Green 2008] and will here be generally used as a term for evaluating evidence regarding its quality. It is important, however, not to mistake "quality appraisal" for the appraisal step in ethics guideline development or HTA processes ("assessment" vs. "appraisal", see chapter 3d). Assessing the *quality* of evidence ("how good is the evidence?") will be the responsibility of those collecting and analysing evidence (in WHO terminology: the review group) and is thus part of the "assessment" step.

For quality appraisal regarding particular ESCs, it is necessary to first reflect upon what constitutes quality in this context. First, it must be pointed out that various pieces of information can be assessed regarding their quality or strength. To reflect on the concept of quality in the context of normative evidence, "arguments" (e.g., arguments for implementing action A instead of action B) are exemplarily used as the information unit of interest. However, the same logic would apply if the information unit were ethical issues, principles or concepts. Assuming for the moment that one is interested in hypothetical arguments for action, then one can assess the following for quality: **(a)** the individual argument, **(b)** all the arguments collected via various strategies that constitute the body of evidence, or **(c)** the ethical analysis and resultant recommendation based on the collected evidence and undertaken by the GDG (similar differentiations are introduced by Scott et al [2016, 2017] and Stoklosa/Bond [2013]). The quality criteria will be different for each of these pieces of information. As the goal of REIGN is not to discuss the decision-making process of the GDG (the "appraisal" as used in the distinction related to HTA), only the first two aspects of quality are elaborated upon in the following.

Quality of Individual Information Units

Obviously, the quality of the body of evidence (e.g., all arguments) will depend on the quality of each individual information unit (e.g., each individual argument or each ethical issue). As part of the evidence collection, analysis and synthesis, each individual piece of evidence (e.g., arguments) should be appraised regarding quality (see footnote 17, chapter 3d for the terminology used).

"Quality", as a term, subsumes all (fulfilled) criteria that are seen as relevant to the individual information unit.

Although there are quality criteria and appraisal methods from philosophy in general and more specific ones from argumentation theory and informal and formal logic [e.g., Fogelin/Sinnott-Armstrong 2005; Copi 1998; Thomson 1999; Føllesdal/Walløe/Elster 2010; Tetens 2010; and many others] regarding the quality of arguments, it is still often unclear how these criteria and methods can be operationalized for quality appraisal in the context of especially systematic reviews of normative literature [see more in detail in Mertz 2017, esp. p. 17]. Or differently put: There are no tools (e.g., checklists or frameworks) that would allow appraisal of the quality of relevant information units by someone not fully competent in argumentation theory and (in)formal logic. Being competent in using methods (or concepts) of, say, informal logic (e.g., having an understanding of what basic kinds of arguments there are and how they are assessed, how good definitions can be formulated, or what common formal and informal fallacies have to be avoided) presupposes having studied informal logic first – and even then, it is not directly obvious which criteria or methods (concepts) are actually useful (in which way) for the quality appraisal of individual units of normative evidence.

Furthermore, the most extensive discussions can be found on quality criteria for individual arguments as individual normative information units; authors working on the topic affirm that, as quality criteria, arguments should be (logically) valid and sound [e.g., Scott et al 2016; Droste et al 2011]. It is less clear, however, how quality can be understood when the object of interest consists of neither arguments nor their premises but ethical issues, principles or concepts. Most likely, one can use considerations of definition theory (which is part of informal and formal logic) and generally criteria for evaluating concepts [as, e.g., proposed in Thomson 1999], such as *coherence,* with other (accepted) concepts and ethical principles, etc.; however, it would still be necessary to investigate in more detail how exactly this could be worked out (e.g., clearly defined and practically checked) to be truly useful for researchers conducting quality appraisals.

However, even when only arguments are considered, how exactly a quality appraisal can be conducted is unclear from the onset. Of course, *consistency* (the idea that statements that are part of an argument should not contradict each other) is an important criterion for arguments but does not go far. However, defining a quality appraisal of arguments as a check for *logical validity* (*validity* for short), for example, would also be insufficient, as this criterion can be attributed only to *deductive* arguments (where the conclusions should follow necessarily from the premises due to the formal-logical structure) and not to *inductive* arguments (where premises support only the conclusion, but the conclusion is not necessarily true even when the premises are true; i.e., the conclusion is true by only a certain probability). Neither is logical validity applicable to *abductive* arguments ("arguments to the best explanation"; these arguments are sometimes considered to be a subtype of inductive arguments, as the conclusion for them is also not necessarily true but supported by the premises only to a certain degree) (see also table 10 below for an overview and short descriptions).

In addition, a valid deductive argument does not have to be *sound* – i.e., even when the argument allows for deducing the conclusion from the premises given, the premises may be wrong (and the conclusion accordingly not justified; such an argument would not be invalid, but unsound). Soundness refers directly to the question of how the truth – or at least plausibility – of the premises can be assessed. For inductive and abductive arguments, neither validity nor soundness is an applicable criterion, but *strength* is. Strength is not a dichotomous criterion – as validity or soundness is – but a gradual one (to be assessed in a continuum between "very weak" and "very strong"); strength is also associated with assessing the truth of the premises (or of checking if there are additional relevant premises that have to be considered and may reduce the strength). The same is true for abductive arguments, although for such arguments, *explanatory power* is normally used as the main criterion.[37]

In sum, while there exists a sophisticated theoretical debate on the quality of arguments (and to a lesser extent for normative concepts, issues and norms/values), there is a lack of practical, contextualized tools to help appraise the quality of individual pieces of normative evidence in the context of ethics guideline development – although checklists to assess the *overall* ethical analyses have been developed [Scott et al 2016, 2017]. Against this backdrop, the REIGN framework cannot offer a ready-made solution to how to assess the quality of arguments or other objects of interest. In Table 10 below, however, some applicable or at least possible (or especially in the case of norms/principle/values and ethical issues, exemplary) criteria are depicted, although it cannot, also due to length considerations, be explained in detail how one has to conduct the quality appraisal according to such criteria.

---

[37] Both inductive and abductive arguments can illustrate the problem of quality appraisal very well. While it can be relatively easy to learn how to assess the validity of a deductive argument (though that does not mean it is always easy, especially when being confronted with complex arguments), learning to assess the strength of an inductive or abductive argument can be much more difficult – not least because there are many different types of inductive arguments, e.g., inductive generalizations, statistical syllogisms, conclusions by analogy, and causal reasoning, which all may have additional/specific criteria for their assessment. It may not be a coincidence that, for example, the chapters discussing the assessment of inductive and abductive arguments in the informal logic book by Fogelin/Sinnott-Armstrong [2005] are considerably lengthier than those concerned with the assessment of deductive arguments.

Individual Information Unit	Short Explanation	(Possible/Exemplary) Criteria or Method
Argument	Multiple statements where one statement (conclusion) is justified/supported by the other statement(s) (premises) due to a logical (or argumentative) structure (at least this is claimed)	• Consistency (statements do not contradict)
Deductive Argument	An argument where the conclusion follows necessarily from the premise(s) due to the formal-logical structure. For example: Premise 1: The capital of Germany is either Berlin or Bonn. Premise 2: In 2019, the capital is not Bonn. Conclusion: Berlin is the capital of Germany.	• Consistency (statements do not contradict) • Validity (logical structure is correct) • Soundness (premises are true or at least plausible)
Inductive Argument	An argument where the conclusion is supported by the premise(s) but is not necessarily true (with only a certain probability). For example: Premise: Most nights are dark. Conclusion: It will be dark tonight.	• Consistency (statements do not contradict) • Strength (how well the premises can support the conclusion/increase its probability)
Abductive Argument	An argument where the conclusion is proposed as being the "best" hypothesis that explains a given phenomenon or "set" of observations (which are, i.a., expressed in the premises); the argument type can also be used outside of the explanations, e.g., in arguing for the "best theory", "best decision" or "best solution". For example: Premise 1 (observation to be explained): The lock on my apartment door is broken. Premise 2: There have been many break-ins in the area lately. Premise 3: There is no notice on the door from the police or from fire workers that they had to forcibly enter the apartment due to an emergency. Conclusion (hypothesis): My apartment was broken into.	• Consistency (statements do not contradict) • Explanatory power (how good the explanation is (e.g., does it explain all facts and is it coherent with other existing knowledge)) • (Strength (how well do the premises support the conclusion/increase its probability))
Premise	A statement used to justify a statement, i.e., a premise in an argument.	• Intelligibility (linguistically/semantically) • Truth of the statement (weaker: plausibility; even much weaker: trustworthiness, e.g., when statement is only assured by testimony) • Independence from other premises in an argument (can the premise stand alone, or must it be combined with other premises to be useful argumentatively?)

Concept	A definition or a set of interconnected statements ("theory") that characterize a pertinent (ethical) phenomenon or typical case that is debated, researched, etc. (e.g., "moral distress", "post-trial-access", "consent", or "euthanasia"); a concept can (in parts) be a premise of an argument	• Intelligibility (linguistically/semantically) • Criteria for good definitions (e.g., not too broad, not too narrow, not only negative, i.e., defining what is *not* meant with the concept; cf. definition theory) • Coherence (do other related concepts and ethical principles, etc., support this understanding of the phenomenon; and does the concept has implications that contradict other concepts, ethical norms or judgements that are considered correct, etc.?)
Norm/Principle/Value	A normative or evaluative statement that a) obliges, permits or forbids certain actions (norms and, in part, principles) or gives b) orientation regarding the goal or the desired "outcome" of actions (in part principles, esp. values); a norm, principle, or value can be either a premise or a conclusion of an argument	• Intelligibility (linguistically/semantically) • Coherence (with other related norms, principles and values, or concepts) • Existing counterexamples (are there examples/cases that the norm/principle covers but which are ethically judged differently from what the norm would entail, with the result that the norm/principle, not the judgement of the particular case, is considered incorrect?)
Ethical Issue	A loose category of single or more complex statements that should be considered for a specific topic or action in practice and are often operationalized as, e.g., ethical conflicts or dilemmas, (recurrent) ethical shortcomings or risks of ethical wrongs	• Intelligibility (linguistically/semantically) • Theoretical consistency/coherence (e.g., are issues defined by referring to ethical principles, theory, concepts, etc.?) • Empirical adequateness (is the issue correctly described, does the issue exist in practice, or does the issue have a certain probability of occurring to avoid discussing mere hypothetical issues?)

Table 10: (Possible/exemplary) criteria/methods for quality appraisal of individual information units

Further reflection on quality appraisal is provided as part of the discussion on quality appraisal as one step in SRNEs (see Appendix C). However, this discussion can also be insightful for quality appraisals using other methods, as quality appraisal of single information units will be important irrespective of the chosen strategy for collecting information.

Either way, as already noted, assessing the quality of, e.g., an argument cannot mean the same as appraising the argument – or all found arguments together (see also the following subchapter) – in view of recommending a certain course of action, e.g., regarding implementation of a health technology. It is the task of the GDG to formulate, by building a consensus, recommendations that will have to consider more than just the quality (e.g., the validity and soundness of a deductive argument or the coherence of a concept used) of the individual information units; these recommendations will also have to balance and weigh individual information (especially arguments) or prioritize them (especially issues) against the backdrop of an existing health care system, political and legal constraints, cultural sensitivities and other considerations that might not even appear on the level of the individual information units extracted from the literature. (However, REIGN was not commissioned to address this task of the GDG, and this paper will thus not further discuss how this kind of appraisal could or should be conducted.)

Quality of the Body of Evidence

Furthermore, the quality of the body of evidence (again, for example, all arguments) will depend on whether *all the relevant*[38] arguments, principles or issues, etc. *(that are of sufficient quality)* have been identified, i.e., whether the sample is skewed. A skewed sample would likely and unjustifiably impact the final decisions of the GDG.

Argumentative/Thematic Saturation as the Main Criterion

The REIGN framework thereby stipulates the following quality criterion for the body of normative evidence: *argumentative or thematic saturation*.[39] This means that the body of evidence is considered of high quality when all relevant information units (of sufficient quality) are included and no relevant arguments, principles or issues are missed. This criterion somewhat incorporates criteria stipulated by frameworks addressing quality assessment of the whole ethical analysis (option c above), namely, completeness and bias [Scott et al 2016, 2017]. The academic (or public) discourse might be biased in various ways, and therefore certain perspectives (and accordingly relevant arguments, issues, principles or concepts) might be missing. It might, however, also be incomplete for other reasons.

If the goal of evidence collection and synthesis is the capture of all relevant information (of sufficient quality) for the question at issue, it is plausible to assume that not all strategies for evidence collection will be equally well equipped to reach this aim. Three aspects impact how far this aim can be reached: **(a)** the attributes of the strategy chosen, **(b)** the context, and **(c)** how the strategy is implemented.

Attributes of the Strategies for Evidence Collection

Above, an overview of strategies for evidence collection was provided (see table 8). Certain attributes of some strategies increase the probability that relevant information will or will not be missed. Compared to a systematic review that devotes resources to identifying a comprehensive sample of papers dedicated to the topic of interest, examining a convenience sample of studies will increase the risk that by chance (or choice) certain information will be missed. Similarly, a workshop with a convenience sample of researchers may be more likely than a systematic review to miss certain

---

[38] (Ethical) relevance can probably also be considered a quality criterion applied more on the level of single arguments or issues than the whole body of evidence [see also Mertz 2017]. However, it might be more accurate to conceptualize relevance as an inclusion criterion. When, for example, systematic reviews are used to identify arguments or issues, the definition of clear inclusion criteria should ensure that only relevant ethical arguments or issues are included in the analysis in the first place.

[39] Apart from completeness (saturation), there might also be better and worse ways to synthesize and present the findings, or differently put: collecting, analysing, synthesizing and presenting issues or arguments is no straightforward task and involves (to some extent) subjective interpretation. As outcomes might therefore differ between researchers, it will be important to reflect on the best way to implement this task. However, this will rather be a question of the quality of the methods of synthesis and will necessarily vary depending on what method (qualitative or quantitative) for collating and presenting normative evidence is chosen. Therefore, no overarching quality criterion can be proposed, as it will depend on the method chosen.

relevant aspects, as only the voices of those partaking in the workshop can be heard, while the review will include all those who have voiced their opinions through academic writing. While some strategies – especially those that rely just on (members of) the GDG – are clearly less likely to reach argumentative or thematic saturation, it is not always clear what strategy should be preferred from the perspective of saturation. It is an open question how, for example, a stakeholder survey compares to a systematic review. In many cases, how they compare will depend on additional factors, of which the most important are discussed below.

However, combining various strategies will definitely increase the chances of generating a truly comprehensive overview, especially if this mixed strategy allows the inclusion of various sources of evidence (e.g., systematic reviews of the academic literature *and* stakeholder surveys). While certain strategies (especially mixed strategies containing systematic reviews to collate research evidence) might be more able than others to achieve comprehensiveness, the REIGN framework does not propose a hierarchy among identified strategies.[40] As long as research on this subject is so scarce, one should be careful not to formulate premature conclusions regarding the ability of various strategies to generate comprehensive or saturated evidence bodies. Such research would therefore be particularly valuable.

Contextual Factors

Furthermore, the context that is investigated will heavily impact the usefulness of various strategies. Some contextual factors that might impact the effectiveness of various strategies to attain argumentative/thematic saturation are, for example, whether the topic or technology of interest is new or not well researched. If guidelines are to be formulated for a new technology or a newly emerged public health threat, it is possible that no paper has yet been written on the topic of interest. Additionally, if no technology or threat already known and written about is sufficiently similar – because the existence of literature on a similar technology or threat might make it possible to search the literature on this technology or health threat instead and to draw analogies between the two cases[41] – every strategy that relies on searching the research literature will be almost useless.

In such cases, it will be more valuable to commission someone to conduct primary normative research in the form of applying ethical theories to the question at issue or to use other strategies

---

[40] Although an overarching hierarchy cannot be proposed, the various strategies belonging to "Written Sources: Academic Literature" (see table 8) can be sorted with regard to quality. A systematic review will without question be better prepared than a non-systematic review to provide a comprehensive overview of normative issues (i.e., reach argumentative/thematic saturation), a non-systematic review will be superior to a convenience sample of papers, and examining such a sample will be better than examining just a single paper. One should be more hesitant to propose similar hierarchies for other strategy types, as it is less clear how well these strategy types achieve argumentative/thematic saturation.

[41] A recent and good example of this are ethical issues related to genome editing, where many of the issues discussed (biological risks, modification of the germline, therapy vs. enhancement, etc.) are not entirely new but are already discussed in (older, more established) gene therapy interventions [see, e.g., Nuffield Council on Bioethics 2016].

involving direct interaction with researchers. It might be equally valuable to engage one way or another with further stakeholders.

Furthermore, in the literature, certain disciplines, theoretical approaches or views (e.g., of affected communities) might not be represented because of power imbalances or certain trends (which could be called "perspective bias") within the academic discourse. This was, for example, the case in the academic discourse on ethical issues in public health surveillance; in this discourse, the issues particular to a developing country context were severely underrepresented [Klingler et al 2017]. In these cases, it will probably be most helpful to implement a mix of strategies (e.g., combining a systematic review, a stakeholder survey of neglected groups, and possibly a commissioned theory application). However, whether the discourse is in any way skewed might not be known from the beginning but rather become apparent through engaging with relevant sources of evidence. To ensure that knowledge gaps become visible, it would be helpful to introduce an additional step of explicit reflection on this question to all evidence collection and synthesis strategies. It might be particularly fruitful before and after conducting, for example, a systematic review to reflect which information/viewpoints are expected and then compare these with what was indeed found. Therefore, the need for further engagement with additional evidence sources might become apparent.

Implementation of Strategies

A third aspect impacting the saturation of the evidence body is how the strategy chosen is implemented, as each of these strategies can come in many shapes and colours. The following example of a systematic review is used again for illustration: The quality of the evidence body generated by this strategy will, for example, depend on the number and disciplinary diversity of databases searched, whether the topic of interest itself or an analogous topic (for lack of relevant publications) was searched, and whether publications from normative researchers were actually included in the sample. Analogously, it will make a difference whether a stakeholder discussion group consists of a diverse set of stakeholders presenting various perspectives or a homogenous group or whether the group follows formal procedures that ensure everybody is heard or more informal procedures where speaking time is potentially allotted according to more personal characteristics (such as command of a common language or shyness). For each strategy, Table 11 below presents examples of up- and downgrading implementation factors that allow more explicit reflection on the manner of implementation and how it might have affected the quality of the evidence body in terms of thematic/argumentative saturation. This overview is meant to enable more structured discussions and reflections on the effects certain implementation choices will have on the resulting body of evidence. This overview cannot, however, define clear-cut guidelines or thresholds for identifying (un)acceptable strategies, as much more research on the effects of the various up-/downgrading factors is needed. The overview should therefore be seen as a starting point (not an end result) for further discussions that complement and specify the list provided.

In conclusion, the strategy chosen for evidence collection (and synthesis)[42] impacts the quality of the evidence body, while contextual factors also play a role. The responsibility for choosing such strategies lies (or should lie) with the GDG. The group should accordingly remember that while the goal of argumentative/thematic saturation should to some extent guide the choice of strategies for evidence collection and synthesis, this goal should not be overemphasized. Each strategy will consume different amounts of resources, and systematic reviews – which will often be the best method from the perspective of argumentative/thematic saturation – are particularly intensive in terms of time and human resource investments. Given that guidelines are sometimes urgently needed and that (financial) resources are generally scarce, considerations of saturation need to be balanced with resource considerations. In some cases, only limited insights will be expected to be gained from engagement with different types of evidence because, for example, the field is already well understood or the most pressing issues are already clear. In these cases, it might be justifiable to employ methods less reliable than systematic reviews and rely, for example, solely on internal evidence generation through members of the GDG. This is particularly important to remember because there will often be an extensive list of relevant questions that will warrant further investigations, but full-blown systematic reviews of all questions will not be possible. Nevertheless, it would be important to transparently report in the guideline that because of such restrictions, an evidence collection and synthesis strategy has been utilized that has more limits than some other strategies.

Reporting Quality

While the preceding discussions have focused on content-related quality criteria, reporting quality should not be forgotten. The methods of collecting, analysing and synthesizing normative evidence should be transparently documented, and possible limitations explicitly reflected. This will allow the GDG and users of the resulting guideline to make their own judgements about the evidence base (and the resulting guideline). For the HTA context, Scott et al [2017] argued that when systematic reviews are employed to identify ethical issues, the research question, literature search strategy, inclusion and exclusion criteria, the perspective and the ethics framework chosen should be transparently stated (the authors also suggest how to make choices with regard to these five aspects) [regarding reporting aspects of systematic reviews, cf. Mertz/Strech/Kahrass 2017]. Reporting guidelines developed for systematic reviews of empirical data (particularly the PRISMA guidelines [Moher et al 2009]) might also be helpful as orientation points, as long as no guidelines specifically targeting systematic reviews of normative information are available.[43] When methods other than

---

[42] The elaborations here do not differentiate between methods of evidence collection and synthesis. However, a further differentiation between these two is possible and warranted because most of the evidence collection strategies described can be combined with various methods of synthesizing evidence (qualitative or quantitative and the variations among both). However, such further differentiation would exceed the scope of this discussion paper and would probably not further the discussion about quality. It should be considered the responsibility of the review group – or whatever term is chosen for those collecting and synthesizing the evidence – to choose the adequate synthesis methods for the question at issue. See also Appendix C for a more detailed discussion.

[43] However, a PRISMA adaptation for – as REIGN would call them – systematic reviews of normative evidence is currently being developed and is registered at EQUATOR: "PRISMA-Ethics: an extension to PRISMA for SRs on ethics literature";

systematic reviews are chosen, whether reporting guidelines are available should be investigated. For this, the EQUATOR network has developed a database containing reporting guidelines and further information regarding reporting. The database can be accessed via the following link: http://www.equator-network.org/.

Sources		Evidence collection strategy	Up-/downgrading implementation factors (from the perspective of thematic/argumentative saturation)
External Sources	Academic Literature	Systematic Review	• Number of databases searched • Diversity of databases in terms of traditions/disciplines included • Number of traditions/disciplines included in the paper sample • Normative expertise represented/not represented in the sample • The topic of interest searched directly/an analogous topic searched **Additionally, for a single paper:** • Paper explicitly/implicitly discusses the normative issue of interest
		Unsystematic or Narrative Literature Review	
		Several Single Papers	
		Single Paper (n=1)	
	Academic Experts	Consensus Process	• Expert variation (particularly regarding disciplinary background) • Formal/informal procedures • Open/closed for unexpected contributions
		Workshop	
		Commissioned Theory Application	• Multiple or only one theory/approach applied
		Consultation (written or verbal)	• Expert variation (diversity of disciplinary backgrounds) • Strategies for dissemination ensure all relevant/diverse groups can participate in the process
	Additional Stakeholders	Interviews/Focus Groups	• Stakeholder variation (diversity of perspectives) • Formal/informal procedures • Openness to unexpected contributions
		Opinion Survey	
		Consensus Process	
		Workshop	
		Consultation (written or verbal)	• Stakeholder variation (diversity of perspectives) • Strategies for dissemination ensure all relevant/diverse groups can participate in the process
Internal Source	GDG members	View of a single (or various) member(s) of the GDG	• Researcher is (not) an expert on the topic of interest
		Consensus of all members of the GDG	• Stakeholder variation (diversity of perspectives) • Formal/informal procedures • Open/closed to unexpected contributions

Table 11: Evidence collection strategies including up- and downgrading implementation factors

---

see http://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-systematic-reviews/#4102018 (24.10.2019)

5. CASE STUDIES

The following are two case studies that build on published WHO ethics guidelines: **(a)** the older *Guidance on Ethics of Tuberculosis Prevention, Care and Control* [WHO 2010] and **(b)** the newer *Guidelines on Ethical Issues in Public Health Surveillance* [WHO 2017]. The guidelines were selected as case studies after consultation with the WHO departments responsible for the REIGN project. The central idea was to have a (newer) guideline that explicitly tried to rely more on (normative) evidence and an (older) guideline where no explicit (more) evidence-based approach was attempted.

The analysis of the guidelines follows the REIGN framework. The five ESCs are thus used as orientation points for discussing whether some sort of evidence gathering or synthesizing took place – and if yes, how – or how the evidence base for the guideline could have been increased. This is, to avoid excess detail, not always done exhaustively, but often only exemplarily. Furthermore, any critical appraisal of the two guidelines is always to be understood against the backdrop of an evidence-based approach and the REIGN framework, not as a general critique of the guidelines; additionally, any critical appraisal of these guidelines is always directed at *methodological* aspects, never at the content itself.

a.   Case Study 1: "Guidance on Ethics of Tuberculosis Prevention, Care and Control"

The guideline reacts to the "growing breadth and complexity of TB efforts [*sic*]", thus, according to the guideline document, implying "a greater range of concerns associated with the ethics of action, inaction and specific approaches to clinical, public health and research interventions" [WHO 2010, p. 1].

Characteristics of the Guideline Document

Structure

The guideline document consists of 38 pages. It is structured into four explanatory sections and eight content sections, which are numbered from one to eight (see Table 12 below).

Explanatory sections	Content sections
Acknowledgements Introduction Background on TB	
	1. Overarching goals and ethical values
	2. The obligation to provide access to TB services

	3. Information, counselling and the role of consent
	4. Supporting adherence to TB treatment
	5. The gap between the availability of drug susceptibility testing and access to M/XDR-TB treatment
	6. Health care workers' rights and obligations
	7. Involuntary isolation and detention as last-resort measures
	8. Research on TB care and control
References	

Table 12: Structure of the guideline document ("TB ethics guideline", case study 1)

Development

As indicated above, the guideline document does not feature a separate section on the method used or about the development process of the guideline. However, information related to the development process can be found in the *Acknowledgements* [WHO 2010, p. v-vi] and in the *Introduction* [WHO 2010, p. 1-2]. These sections explain that the guideline was drafted mainly by the "WHO Task Force on Addressing Ethical Issues in TB Care and Control Programmes" (an expert group consisting of 21 persons); this task force was informed by four discussion/background papers (about "access to diagnosis and treatment", "obligations and rights of health-care workers and patients", "public health measures", and "research"), which were commissioned beforehand [WHO 2010, p. v and p. 1]. These papers were discussed at the first meeting of the task force (December 2008), where it was decided which "main points" raised by the papers should be included in the guidance document. The outline of the guidance was drafted in the second meeting (August 2009) [WHO 2010, p. 1]. A consultation "with additional representatives of civil society and national TB programmes enabled further valuable input" (October 2009). Shortly afterwards (November 2009), a refined outline "was endorsed by WHO's Strategic and Technical Advisory Group for Tuberculosis (STAG-TB)" [WHO 2010, p. 2]. Further feedback (or "contribution") was provided by various groups (e.g., the aforementioned STAG-TB), participants from conferences or other meetings (e.g., the 8th Global Summit of National Bioethics Advisory Bodies, 2010), and various institutions (e.g., the Open Society Institute and the World Medical Association) [see WHO 2010, pp. v-vi]. The work was coordinated and guided by three persons from two WHO departments (WHO *Stop TB Department* and *WHO Department of Ethics, Equity, Trade and Human Rights*) [WHO 2010, p. v]. Although "major efforts were undertaken to reach consensus", the guideline's recommendations do not "necessarily reflect the agreement of all members of the Task Force" [WHO 2010, p. v].

Presentation of Ethical Issues/Courses of Actions/Recommendations

All ethical issues or possible courses of actions – and other content in the main sections – are presented as *questions*, for example, "Does the obligation to provide free care include diagnosis and other services?" [WHO 2010, p. 10], "How should programmes respond to patients who do not adhere to treatment despite repeated efforts?" [WHO 2010, p. 17], or "What general ethical principles should govern TB research?" [WHO 2010, p. 24] and "What is the relationship between ethical

values and human rights principles?" [WHO 2010, p. 5]. Sometimes, the questions refer to a "bundle" of ethical issues (e.g., "How does the ethical concept of 'informed consent' apply to TB testing and treatment?" [WHO 2010, p. 14]) and thus function as a kind of 'superscription'". At other times, the questions directly address specific issues (e.g., "What kind of information should individuals be given about TB tests and treatments?" [WHO 2010, p. 13]) or address concrete courses of actions (e.g., "Is it ever appropriate to compel treatment of TB patients over their objection?" [WHO 2010, p. 23]).

The "answers" to the questions vary accordingly. In some cases, an answer consists of (longer) elucidations about relevant principles, obligations and rights or about criteria or "safeguards" that should be adhered to; these elucidations are sometimes illustrated by (short) examples. In other cases, where for example, concrete courses of action are addressed in the question, the answer section will start with a "yes" or "no", and this answer is followed by an explanation or justification. Recommendations (in the form of "should" sentences or similar prescriptive formulations) can be found in the guideline (an example of a recommendation is "Patients should be informed, at the initiation of treatment, that they will be contacted if they do not show up for their appointments […]" [WHO 2010, p. 17]); however, recommendations are included in the elucidations, explanations, or justificatory texts accompanying each question and are not, for example, necessarily at the beginning or end of an answer.

Analysis and Appraisal

ESC 1: Value Base (or Ethical Corridor)

In the guidelines, ethical values that are seen as "particularly important to TB care and control" [WHO 2010, p. 6] are explicitly defined. Additionally, the relationship of these values to human rights principles is described. Mentioned and explained briefly are "social justice/equity", "solidarity", "common good", "autonomy", "reciprocity", "effectiveness", "subsidiarity", "participation", and "transparency and accountability" [WHO 2010, p. 6-7]. These "key values" are not referenced. How this list of values came into being and the rationale for including values as guiding principles in the context of the guideline are not described.

Nevertheless, these values and the described relationship to human rights can be interpreted as a "value base" in the terminology stipulated by the REIGN framework (see chapter 4). As a value base, these values underlie the guideline and/or provide an "ethical corridor" (especially by referring to human rights). The role of these values in the discussion of ethical issues or in justifying recommendations in the later sections is not made clear, although later sections sometimes refer to these values (e.g., "In addition, enablers empower patients to take an active role in their care, thereby promoting the ethical value of individual autonomy" [WHO 2010, p. 16]). Regarding the specific "evidence source" that informed the choice of these values, it can only be hypothesized. Most likely,

the evidence source was either the "theory application" (albeit only implicit, as no ethical theory or approach is mentioned) or the "consensus process" of experts or of the GDG itself.

ESC 2: Conceptual Disambiguation

The guideline does not dedicate a section to conceptually clarifying the central terms of the guideline (e.g., "tuberculosis prevention" or "tuberculosis control"). However, in section 1, the "overall goal of TB care and control programmes" is briefly summarized by referring to the already existing WHO's Stop TB strategy [WHO 2010, p. 5]. This can be understood as some low-level conceptual clarification, although no possible normative implications of this usage of the terms are discussed. Neither is the understanding contrasted with other possible definitions of the terms. In addition to referencing the WHO strategy, no further references or referral to evidence is provided.

However, it might not have been necessary to further elucidate the understanding of the central terms, as they can be regarded as "given" in the context of WHO guidelines due to the established WHO approach for stopping TB. Further evidence about the usage of terms may also be unnecessary because they are often sufficiently straightforward (e.g., "reduce human suffering and socioeconomic burden associated with TB", "protect poor and vulnerable populations from TB, TB/HIV, and MDR-TB" [WHO 2010, p. 5]). Nevertheless, "prevention", "care", and "control" (e.g., "protect and promote human rights in TB prevention, care and control" [WHO 2010, p. 5]) could be differently understood – what does "prevention" exactly entail in contrast to "care"? – and hence could have different normative implications.

ESC 3: Need for Action

The guideline has a background section (*Background on TB* [WHO 2010, p. 3-4]), but this section does not address the need for action for *ethics* guidance in general or for specific issues. The main argument of this section is that there is a need for action *against TB*: the document shows, also by referring to empirical data, why TB is a major public health problem that has to be addressed. Only two sentences that can be understood as a statement regarding the need for action (in terms of providing ethics guidance) can be found in the *Introduction*: "With the growing breadth and complexity of TB efforts [*sic*] today comes a greater range of concerns associated with the ethics of action, inaction and specific approaches to clinical, public health and research interventions" and "In 2006, the documented emergence of extensively drug-resistant tuberculosis (XDR-TB), including a dramatic and lethal outbreak in South Africa, brought forward urgent issues of public health ethics given the imposition in some programmes of involuntary detention of persons suspected and/or confirmed of being ill with drug-resistant TB under the justification of public safety" [WHO 2010, p. 1]. The second sentence also references an academic publication.

The aforementioned second sentence, however, shows only why the issue of "involuntary detention of persons" in the context of TB treating and control programmes needs ethics guidance (which is only one section – section 7 – in the guideline document). The first sentence, which is more general, is substantiated by neither normative evidence (e.g., why some actions or inactions are ethically

problematic) nor empirical evidence (e.g., how stakeholders and those affected negatively experience certain actions and inactions, where they see problems, and where they need help/guidance).

Nevertheless, ethical issues were identified, prioritized and selected for analysis: "The aim was to undertake an analysis of selected priority ethical issues in TB [*sic*] […]" [WHO 2010, p. 1]. Obviously, at least implicitly, a need for ethics guidance was recognized for these issues. However, nothing is said about how the ethical issues were identified (i.e., whether they were identified by referring to academic publications, by consensus rounds, by stakeholder survey, etc.), why and how they were subsumed under (broader) topical sections (e.g., "Information, counselling and the role of consent" or "Research on TB care and control"), and how they were selected in the first place from a possibly even broader "list" of issues. While the latter goes beyond issues of consideration of evidence – as issues has to be prioritized by the GDG – it is unclear what the evidence base was for **(a)** the ethical issues in total (e.g., was the "list" of issues comprehensible, and where did normative information about these issues come from?) and **(b)** the prioritizing task (e.g., were arguments collected about why specific issues should be prioritized, or was empirical information available about which stakeholders deem a specific issue as especially important?). Again, in the terminology of the REIGN framework, the need for action was probably substantiated only by evidence attained via strategies, such as "workshops" or "consensus processes" with experts or additional stakeholders.

ESC 4: Strategies for Addressing Need

The questions used to structure the guideline not only point to certain ethical issues but also often hint at possible solutions to the identified issues (e.g., "Is the use of 'enablers' an ethically justifiable strategy for promoting adherence to treatment?" [WHO 2010, p. 16]). However, solutions can also be found in the answers to the questions mostly already integrated into the recommendations (e.g., "Programmes should work with peer advocates and community leaders to design mechanisms for providing information that will be appropriate for individuals from diverse linguistic, educational and cultural backgrounds" [WHO 2010, p. 13]). In the guideline document, therefore, there is no separate listing or discussion of strategies.

Although often the strategies to be addressed are possibly straightforward enough to accept that there is no explicit evidence for them mentioned – especially if the question posed allows only "yes" or "no" for an answer – from the perspective of an evidence-based approach, one can still question how strategies were identified and selected. It seems particularly questionable whether all relevant strategies have been identified. For example, in a text passage accompanying the following question, "How does the ethical concept of 'informed consent' apply to TB testing and treatment?" [WHO 2010, p. 14-15], various strategies for ensuring informed consent can possibly be mentioned (and many actually are). As there are few references given to academic publications or (at that time) already established guidance documents, a reader cannot fathom if the strategies mentioned are well grounded in, e.g., the academic discourse or accepted, widely shared standards.

ESC 5: (Hypothetical) Arguments for Actions

As with strategies, arguments are – reasonably – given in the answers to the guiding questions. In a strict sense, however, there actually are a few instances where no arguments are provided in the elucidations. For example, in the text answering the question "How should health-care providers make decisions about the care of individual patients when governments do not fulfil their obligation to ensure the availability of quality-assured drugs?" [WHO 2010, p. 11], courses of action are mentioned (e.g., "In some cases, they may reasonably conclude that it would be ethically preferable to give a patient drugs of unknown quality rather than forego treatment entirely" or "There is an additional duty to notify the national government about this particular problem, and advocate for an urgent rectification"). However, no arguments are provided to support that the proposed courses of action are ethically acceptable or even demanded. In some cases, arguments are mentioned more descriptively as possible arguments (e.g., "[…] an argument based on a humanitarian principle (beneficence, solidarity, etc.) might appeal to the fact that fellow human beings require relatively cheap interventions that could easily and dramatically improve their lives" [WHO 2010, p. 11]). In other cases, arguments are provided to actively justify ethical obligations (e.g., "There are several reasons to ensure that individuals undergoing TB testing and treatment receive complete and accurate information about the risks, benefits, and alternatives available to them. First, at the most basic level, people have a right to know what is being done to their bodies and why it is being done. […]" [WHO 2010, p. 13]). The recommendations themselves (the "should" sentences) are sometimes justified by mentioning arguments (or underlying premises), sometimes they are not.

Arguments in the guideline are sometimes linked to the values mentioned in the first section of the guideline document (see above). This could be understood as some kind of "theory application". In rare cases, recommendations are implicitly justified by referring to publications, for example, an already existing guidance (e.g., "TB programmes should provide assistance and support to patients who undertake to notify their contacts (27)" [WHO 2010, p. 13], which refers to *Opening up the HIV/AIDS epidemic – Guidance on encouraging beneficial disclosure, ethical partner counselling & appropriate use of HIV case-reporting* of the WHO). As such, the guideline referred to should be considered as part of the normative evidence base that is provided here (although it is unclear whether other relevant documents and arguments were also considered).

In one notable case, an assertion of existing empirical evidence is made but not substantiated with any reference to primary studies or secondary research synthesizing evidence ("Is it ethically acceptable to refuse to initiate treatment when it appears that a particular patient is unlikely to adhere to the prescribed regimen? – No. There is no evidence that anyone can accurately predict whether an individual will adhere to treatment." [WHO 2010, p. 18]). Although one may assume that the statement is true, looking from an evidence-based framework, it is remarkable that the existence of evidence is mentioned, but the evidence is not presented. A similar example is the following: "There is an urgent need to develop an enhanced evidence base for TB prevention and treatment and to improve the standard of care" [WHO 2010, p. 24]. Again, no arguments – e.g., based on empirical

or normative evidence – are given to justify this statement, which, again, is probably a consensus statement of the experts involved.

Consequences of actions (also providing justifying arguments for actions) are also seldom directly addressed in the guideline document. Of course, it is often implicitly assumed that certain courses of action lead to ethically positive outcomes. For example, it seems to be implied that the safeguards mentioned for cases where involuntary isolation or detention *is* implemented (e.g., "based on a legitimate objective", "the least restrictive and intrusive means available") [WHO 2010 p. 23] can promote more ethical "outcomes" than *not* implementing these safeguards. These claims are also not substantiated by evidence.

However, it must be considered that it is generally not easy to provide (empirical) evidence for such cases. Possible candidates for evidence are negative stakeholder experiences when actions were implemented without such safeguards in place, or perhaps general preferences regarding such situations. Such studies may imply that without the safeguards mentioned, the ethical "outcome" would be worse. Future research might be necessary to elucidate how ethical "outcomes" can be assessed empirically – particularly where consequentialist/utilitarian ethics approaches are chosen.

In addition to these more general remarks, there are actually some examples in the guideline document where empirical evidence could have been gathered and would have been helpful in substantiating recommendations or other statements related to them. Most often, statements about the *effectiveness/efficacy* of specific courses of actions could have benefited from further substantiation by evidence. For example, it is stated that "[…] [d]irectly observed therapy is an effective way to ensure adherence to treatment" [WHO 2010, p. 16]. This is an empirical statement that, in principle, can be shown by evidence to be true (or at least plausible) or false. The same is true for the following example: "Is it ethically acceptable to give TB patients financial or other incentives in exchange for completing treatment? – […] Whether to give patients incentives to complete treatment should be based on judgements about both the expected efficacy of such practices and sensitivity to local norms" [WHO 2010, p. 17]. Again, it would have been possible to look for empirical evidence regarding the efficacy of such practices or how such practices are experienced by various stakeholders. Additionally, when discussing health risks for health-care workers looking after TB patients, medical evidence might have been available to substantiate claims ("Are the risks associated with looking after TB patients sufficiently great to absolve health-care workers of a duty to care? – In general, no. With reasonable training, supplies, equipment, infrastructure, support, and access to proven methods of care and treatment, HCWs can legitimately be expected to look after patients with TB." [WHO 2010, p. 20]). At least one reference is given to support assertions about effectiveness, for example: "In addition, community-based care reduces burdens on health-care facilities and is more cost effective than facility-based treatment (16), thereby enabling governments with limited resources to serve the greatest proportion of those in need [WHO 2010, p. 11-12]. The reference, however, is to another WHO guideline (*Guidelines for the programmatic management of drug-*

resistant tuberculosis. Emergency update 2008); therefore, it is difficult to directly assess how well the underlying evidence base substantiates the statement. In one case, it was also admitted that research about effectiveness for some courses of action is needed ("Further research should be conducted to determine the most effective methods to promote adherence." [WHO 2010, p. 17]). Nevertheless, providing more transparency about the available evidence or the evidence base used by the guideline developers would generally have been feasible (although, arguably, it might not have been reasonable to conduct full blown systematic reviews, it would have been at least possible to hear expert opinions and assess the evidence available to them).

Overall, there is no systematic approach to finding, categorizing/synthesizing and appraising the various relevant arguments for (normatively) justifying courses of actions or "safeguards" for actions. How recommendations are justified (with arguments) appears to be inconsistent throughout the document. However, of course, it has to be considered that the number of ethical issues presented and discussed – considering the various topical sections containing several ethical issues, actions and recommendations – would have made it difficult to thoroughly assess the empirical and normative evidence for all issues, particularly considering the time constraints and workload. Notwithstanding, it would have been possible to be more explicit about the argumentative basis for the recommendations given while also substantiating the arguments themselves by evidence – at least for some selected (prioritized) issues.

Concluding Remarks

As a preliminary note, it has to be remarked that applying the REIGN framework to analyse and appraise the TB ethics guideline can be difficult because ethics guidelines are (currently) not structured along the ESC categories (or comparably), thus making it difficult to identify what parts belong to, e.g., "conceptual disambiguation" or where "arguments for action" can be found. Nonetheless, the case study showed that, in principle, the framework *can* be applied and used to assess which kind of evidence was used – or not used.

This said, the guideline analysed is only marginally substantiated by evidence, as seen from the REIGN framework's perspective, and even more often the involvement of evidence remains unclear – even in cases where both empirical and normative evidence obviously would have been, generally speaking, available. Evidence might have been more profoundly gathered in the discussion/background papers (e.g., academic publications reviewed) and synthesized; however, this possibility cannot be assessed on the basis of the guideline document, and nothing is written within the document concerning the methodology of the discussion papers. Neither is it pointed out whether and where the discussion papers can be (publicly) accessed.

Where academic publications were referenced (as a possible way of supporting statements with evidence), the referencing was unsystematic. For example, in answering the question "What general ethical principles should govern TB research?", a list of eight considerations is provided; however,

just one of these eight is referenced [WHO 2010, p. 24]. As would be expected, the literature in the reference section seems to be mainly normative literature, but only a few publications can be subsumed under *research* evidence in a stricter sense, as mostly other guidelines are referenced. Additionally, it is unclear how the body of considered literature/the evidence base came about. Apparently, there were no systematic attempts to identify not only relevant normative but also empirical literature; this might have mitigated risks of bias in using literature and its content.

Especially regarding ESCs 1, 2 and 3, systematic reviews would probably have been an asset. In ESCs 1 and 2, SRNEs of normative concepts and/or of ethical values, norms or principles were possible; at least, more references to ethical "background" theories (which would fall in the category of "theory application" in the terminology of the REIGN framework) and/or more transparency on how and why the documented list of values were created would have been advantageous from the viewpoint of an evidence-based approach. In ESC 3, an SRNE of ethical issues would have ensured comprehensiveness and increased methodological accountability regarding the various ethical issues addressed. SRs of arguments in ESC 5 would also have been an option, although, as already remarked, presumably solely for selected, high-priority issues because of the associated workload. More reliance on empirical evidence to assess the probable consequences of actions/outcomes or to substantiate crucial empirical claims that were used in arguments for justifying recommendations might sometimes also have been reasonable.

Generally, thus, it is difficult to assess – without engaging in further evidence collection – for example, if all "important" issues are addressed in the guideline. This is at least true for someone not (already) an expert in ethical issues in the context of TB care, prevention and control. Evaluating the guidelines is particularly difficult because barely any information about the method or development process is given, and no evidence base is provided that one could access, appraise, and perhaps "reproduce". This also renders it difficult to assess the quality of the process and of the content.[44] This is aggravated by insufficient transparency regarding additional procedural aspects, such as what was the exact contribution or kind of feedback from the various other stakeholders (institutions, participants at conferences). Did the stakeholders provide additional ethical issues, additional arguments or explanations regarding existing issues, or was the feedback limited to wording or to remarks about how to structure and format the guideline document?

Finally, the presented ethical issues and courses of action vary greatly regarding the operationalization of questions and respective answers; no "standard form" that would have made it easier to discern recommendations from explanations/elucidations, specifications and justifications, and thus, any evidence used was utilized. However, it might be more "user friendly" to present recommendations in this manner, even if doing so is more inexpedient from a methodological point of view. As a side note, further evidence is also needed regarding the following issues: which

---

[44] This implies *not* that the quality *is* actually low but just that it is difficult for someone not having partaken in developing the guideline to appraise its quality or the presumable quality of the content.

structuring, formatting, and operationalizations of ethical issues or recommendations and the whole guideline document are especially helpful for the target audience of ethics guidelines. This will, however, necessitate respective meta-research that is currently seldom conducted in the ethics community.

b.   Case Study 2: "Guidelines on Ethical Issues in Public Health Surveillance"

The WHO's *Guidelines on Ethical Issues in Public Health Surveillance* are claimed to be the first over-arching (meaning not disease-specific) ethical framework for conducting public health surveillance. The goal of the guideline development project was to "identify key ethical considerations to guide resolution of controversies that may arise in surveillance" [WHO 2017, p. 13]. This was the first WHO ethics guideline that tried to more explicitly consider evidence in its development and is therefore particularly interesting as a case study.

Characteristics of the Guideline Document

Structure

The guideline document consists of 55 pages. The document is structured mainly around 17 guidelines (formulated as *normative demands*) that are couched within seven explanatory sections that contextualize the guidelines (see table 13 below):

Explanatory sections	Guidelines
Foreword Acknowledgements Background Framing the ethics of surveillance	
	Guideline 1: Countries have an obligation to develop appropriate, feasible, sustainable public health surveillance systems. Surveillance systems should have a clear purpose and a plan for data collection, analysis, use and dissemination based on relevant public health priorities
	Guideline 2: Countries have an obligation to develop appropriate, effective mechanisms to ensure ethical surveillance
	Guideline 3: Surveillance data should be collected only for a legitimate public health purpose
	Guideline 4: Countries have an obligation to ensure that data collected are of sufficient quality, including being timely, reliable and valid, to achieve public health goals
	Guideline 5: Planning for public health surveillance should be guided by transparent governmental priority-setting
	Guideline 6: The global community has an obligation to support countries that lack adequate resources to undertake surveillance

	Guideline 7: The values and concerns of communities should be taken into account in planning, implementing and using data from surveillance
	Guideline 8: Those responsible for surveillance should identify, evaluate, minimize and disclose risks for harm before surveillance is conducted. Monitoring for harm should be continuous, and, when any is identified, appropriate action should be taken to mitigate it
	Guideline 9. Surveillance of individuals or groups who are particularly susceptible to disease, harm or injustice is critical and demands careful scrutiny to avoid the imposition of unnecessary additional burdens
	Guideline 10: Governments and others who hold surveillance data must ensure that identifiable data are appropriately secured
	Guideline 11: Under certain circumstances, the collection of names or identifiable data is justified
	Guideline 12: Individuals have an obligation to contribute to surveillance when reliable, valid, complete data sets are required and relevant protection is in place. Under these circumstances, informed consent is not ethically required
	Guideline 13: Results of surveillance must be effectively communicated to relevant target audiences
	Guideline 14: With appropriate safeguards and justification, those responsible for public health surveillance have an obligation to share data with other national and international public health agencies
	Guideline 15: During a public health emergency, it is imperative that all parties involved in surveillance share data in a timely fashion
	Guideline 16: With appropriate justification and safeguards, public health agencies may use or share surveillance data for research purposes
	Guideline 17: Personally identifiable surveillance data should not be shared with agencies that are likely to use them to take action against individuals or for uses unrelated to public health
The shifting boundaries of surveillance Reference	

Table 13: Structure of the guideline document ("PHS ethics guideline", case study 2)

Development

One author (CK) was involved in developing the guidelines as the lead of the literature review group and is therefore aware of the development process (see the section titled "How Evidence was Integrated in the Development Process"). However, the document itself does not elaborate (much) on the methods and processes underlying the development of the guidelines. There is no section devoted to the development process and methods; however, some information is provided in the acknowledgement section [WHO 2017, p. 7-8]. The document specifies who within WHO was responsible for preparing the document (the WHO *Global Health Ethics Team* was mainly responsible). The document also specifies who was part of the GDG and provides information on the chair and cochair of the group. The GDG consisted of 25 members. It is also specified who contributed to developing the guidelines as observers (five experts), as part of the literature review group (five

researchers), as part of the internal steering group (17 experts) and as external reviewers (nine experts). Additional WHO colleagues and consultants (seven experts) are also acknowledged for their input. As part of the acknowledgement section, who provided which text for the guideline was further described, with the chair of the GDG being identified as the lead writer and editor. Funding sources are also named. However, while persons involved in guideline development are listed and identified as writers of certain parts, no information whatsoever is provided on the underlying processes (e.g., how or where the GDG met, what information was searched by the literature review group, how information was fed back to the group, and how discussions among the group were structured) – at least not as part of the acknowledgement section.

Throughout the guideline, however, are paragraphs that give some insights into the methods and processes involved, albeit to a very limited degree. For example, as part of the introduction, the reasons for selecting particular people as members of the GDG are described: "[The guidelines] were prepared by an international group of experts in surveillance, epidemiological research, bioethics, public health ethics and human rights. The authors of these guidelines represent leading research institutions and nongovernmental organizations (NGOS) either involved in surveillance or representing groups or populations with a vital interest in both the benefits and burdens of surveillance. The authors also represent countries of both the south and north, with different political systems, social values and priorities" [WHO 2017, p. 12]. The work of the literature review group is concretized in the following way: "The guidelines are based on a systematic literature review of relevant research and grey literature in accordance with the WHO *Handbook for Guideline Development*" [WHO 2017, p. 13].

Presentation of Ethical Issues/Courses of Actions/Recommendations

The document is organized around 17 guidelines that are formulated as *normative demands* (see table 13). Preceding the guidelines are two sections – "Background" and "Framing the ethics of surveillance" – that (a) provide a working definition of public health surveillance [WHO 2017, pp. 14-16]; (b) situate the topic within a legal, ethical and historical discourse [WHO 2017, pp. 16-18]; (c) establish the need for guidelines against the backdrop of existing guidance [WHO 2017, primarily pp. 19-20]; and (d) finally, provide four overarching principles as the normative backbone of the guideline [WHO 2017, pp. 21-23]. The overarching principle "considered central to making decisions in the specific context of public health surveillance" [WHO 2017, p. 22] consists of *the common good*, *equity*, *respect for persons*, and *good governance*.

The 17 guidelines are accompanied by text that elaborates on their meaning. Mostly, these accompanying passages provide arguments that justify the recommendations given; for example, with regard to guideline 5, it is said: "Transparency is important because it fosters trust and creates conditions for citizens to advance the common good individually and collectively" [WHO 2017, p. 31]. The arguments provided often refer back to the overarching principles presented in the preceding explanatory sections. In addition to arguments, the accompanying paragraphs contain specifications,

meaning the accompanying passages more clearly delineate the implications of the stipulated guide-line. These passages address how certain aspects of the guidelines are to be understood; for exam-ple, guideline 10 specifies: "Security in this context consists of operational and technological safe-guards to protect personal data from unauthorized access or disclosure" [WHO 2017, p. 37]. This often means that these sections formulate more specific normative demands that follow from the overarching guidelines; for example, as part of guideline 6, it is formulated: "An obligation to sup-port does not give the global community license to ignore the priorities of countries that require support or resources" [WHO 2017, p. 32]. Sometimes, instead of providing explicit normative de-mands, these passages also point out normative criteria that will have to be weighed case by case; for example, for guideline 13, it is said: "The communication of knowledge is a double-edged sword: on the one hand, knowledge may clearly empower; on the other, it may lead to injury, stigmatization or discrimination" [WHO 2017, p. 41].

The individual guidelines refer to each other and somewhat build on each other. The elaborations are presented as free text, and the more specific normative demands formulated there are not clearly marked as such (neither are arguments provided). For some guidelines, the arguments for demanding certain actions are not clearly discernible from the elaborating text provided (particu-larly in guideline 11).

Analysis and Appraisal

Reporting

Most importantly, the reporting of the guideline document with regard to evidence integration and other (particularly consensus building) processes can be improved immensely. In the context of eth-ics guidelines – which can rely only on limited methodological guidance – it might be particularly important to be transparent with regard to the processes that led up to the final recommendations given. With regard to the evidence base, it should be pointed out, e.g., what questions were re-searched; what methods were employed for searching, identifying, analysing and synthesizing data; what sources were accessed; and how the data were fed back to the GDG. With regard to further processes, it would be important to know how often the group met, who attended the meetings, how discussions were structured, who moderated the discussions, how consensus was built and reached (or whether no consensus with regard to recommendations was reached after all). Alt-hough it would also be illuminating to know who wrote the final guidelines (information that is pro-vided in the guideline document), the information on the consensus building process might be even more important because – presumably – the authors formulated only what had before been dis-cussed and agreed upon within the GDG.

Addressing ESCs and Considering the Evidence

Because of the limited reporting, there is only so much one can say about the evidence base of the guideline (based on reviewing the document alone). The document clearly addresses many of the

aspects specified in the ESCs. The document dedicates a whole section to clarifying central terms and the "value base" (or, in the terminology proposed in the REIGN framework, ESCs 2 and 1, respectively). The document clearly explains why there is a need for action in terms of limited guidance and lists several (presumably unresolved) ethical issues that arise in this context (ESC 3). The ethical issues, solutions (ESC 4) and arguments for action (ESC 5) are addressed jointly as part of the 17 specific recommendations. Claims that are substantiated by arguments are most often formulated as hypothetical arguments but sometimes also formulated as empirical claims (e.g., regarding harm realized by conducting public health surveillance). The guideline does refer to various sources of literature (including, to a limited extent, empirical sources, e.g., Graeme et al [2015]), but because of the limited reporting, it is unclear whether the literature was considered and used more than selectively.

Given the extended elaborations provided as part of case study 1 and the marginal additional insights to be expectantly gained from applying the REIGN framework in detail to this guideline document, no further appraisal will be provided. Instead, further sections will be devoted to reconstructing the integration or consideration of evidence in the guideline development process, as witnessed by one of the authors (CK). Additionally, how the REIGN framework could have improved the process will be considered.

Evidence-integration in Developing the Guideline

Primary Question Formulation

During the primary meeting of the GDG at the *Brocher Foundation* in Geneva on May 26-27, 2014, a primary list of questions to be researched by a literature review group was devised. At this point, the members of the review group were not yet identified and, therefore, not present. No experts on information retrieval or systematic review methodology were present during this workshop either. Therefore, the questions devised were not particularly well suited to be informed by evidence – at least not by using a systematic review methodology.

To provide just one example of a question posed by the GDG to the review group: "How should one balance the imperative to collect identifiable surveillance data with the need to protect privacy and confidentiality?" Balancing these issues (or providing recommendations) is the task of the GDG, and therefore, the question cannot be answered by evidence. Evidence might, however, have informed this task (e.g., a systematic review might have collected arguments for and against name-based reporting), but the question would have to be reformulated (and the expectations of the GDG adapted). This would probably not have happened had the review group been identified before the guideline development had started. Early involvement of the reviewers (or researchers, to use a more inclusive terminology) with the GDG will be helpful in understanding (and possibly adapting) each other's expectations and formulating questions that are (a) informative to the GDG but (b) also answerable by evidence. Having the REIGN framework as the orientation point for the GDG might

have also helped because the framework would have specified what kinds of questions can be answered by evidence (and what kinds of questions are left to be answered by the GDG itself).

Furthermore, there seemed to be a gap between the expectations of the WHO team leading the development process and the GDG. Most of the questions formulated by the GDG were geared more towards philosophical or conceptual analysis (in the terminology of the REIGN framework: *theory application* as one viable strategy for evidence collection). The WHO, however, was planning to conduct systematic reviews that need questions of a certain format to be workable. Again, the REIGN framework might have helped bridge this gap by identifying various sources of evidence that the GDG could have built on and pointing out various strategies for evidence collection. Therefore, the framework might have helped clarify that the WHO and the GDG were thinking of employing different strategies for evidence collection (although some experience with SRNE methodology would still have been needed to realize the specific requirements of question formulation in this context).

Secondary question formulation

After the literature review group was established, it reformulated the questions to allow them to be searched and analysed by using a systematic review methodology. The list of questions was fed back to the GDG, which provided comments to ensure that the questions were relevant to them. This resulted in a list of ten questions:

- Question 1: How is public health surveillance defined/what are the defining elements of public health surveillance?

- Question 2: In what cases has public health surveillance contributed significantly to improvements in health?

- Question 3: What harms with regard to stigma and discrimination have been documented as consequences of public health surveillance?

- Question 4: What privacy breaches have been documented after sharing public health surveillance data?

- Question 5: What reasons are brought forward concerning the level of care that should be provided to communities/individuals identified by public health surveillance as sick/at risk of becoming sick?

- Question 6: Is informed consent relevant in the context of public health surveillance, and for what reasons (it is not)? If authors find informed consent to not be relevant in the context of public health surveillance, what additional conditions have to be fulfilled to allow foregoing informed consent?

- Question 7: What are the ethical issues of using digital technology (e.g., electronic medical records) in public health surveillance?

- Question 8: What are the ethical issues raised by public health surveillance for untreatable conditions?

- Question 9: What kinds of oversight mechanisms/models of governance have been proposed (and possibly implemented) in the literature?

- Question 10: What ethical issues arise in the context of public health surveillance?

Several aspects about this list of questions can be seen as worthy of improvement (and could have been improved with the REIGN framework):

First, the list was too extensive given the limited resources available for developing the guidelines. The GDG had simply put out all questions that they considered potentially interesting in the context without fully realizing the amount of resources needed to conduct a systematic review. It might have been necessary to further prioritize or to decide to examine all questions but with less-resource-intensive methods. This decision, however, was – more or less – left to the review group (see below). The REIGN framework explicitly acknowledges the need to balance the need for additional information and resource constraints. Prioritization decisions, however, should be made by the GDG and not the review group.

Second, some of the questions proved irrelevant at later stages of the development. After the GDG had decided upon a definition of public health surveillance, the question about the level of care (Question 5) seemed to fall outside the remit of public health surveillance and was accordingly no longer relevant. Searching evidence for irrelevant questions might have been prevented if the stepwise process proposed by the REIGN framework had been followed. The framework explicitly points out how the ESCs relate to and build on each other (although it also acknowledges the interconnections between the different steps): It makes no sense to formulate questions with regard to arguments (ESC 5) if the central terms (ESC 2) to use or the issues (ESC 3) to discuss are not yet determined.

Third, the GDG might have missed questions for which further evidence would have been valuable. For example, the guideline incorporates four ethical principles that are stipulated as central orientation points for decision-making in public health surveillance. It might have been helpful when deciding on these guiding principles to have an overview of principles that are generally invoked by not only academics but also practitioners in this context. The REIGN framework might have been helpful in pointing out other areas (or ESCs) where evidence can be helpful in terms of improving decision-making. However, whether the GDG did not realize that further evidence might have been of benefit or whether the group – justifiably so – chose to not collect evidence to address these further questions cannot conclusively be shown because CK did not participate in the first meeting where the set of questions was developed (and presumably their relevance and the relevance of further questions was discussed). Either way, having a checklist or something similar based on the developed ESCs to guide the choice of questions would ensure that no important questions are – unintentionally – missed.

Collecting and Analysing Evidence

CK was involved only in collecting and analysing six of the ten questions (Questions 1, 5-8, 10). Accordingly, the following elaborations pertain only to these questions. The others were addressed by another member of the review group (Michael Vaughn, Columbia University School of Public Health, USA). Due to resource constraints, questions 5-8 and 10 were searched together. However, except

questions 7, 8 and 10 (as question 10 already contained questions 5 and 6), these questions were addressed separately in the analysis. Instead of prioritizing questions, the review group therefore decided to "cut down" on the quality of the review by searching most of the questions jointly with a focus on question 10. Accordingly, two (systematic) literature reviews (or better, searches) were conducted: one of normative concepts (public health surveillance) and one of ethical issues (that arise in public health surveillance); these reviews included searches of solutions and safeguards. The first review concentrated on academic discourse and international policy documents, and the second concentrated only on academic discourse. The authors limited the latter search to two databases (PubMed and Google Books). One of the reviews is published in an academic journal [Klingler et al 2017].

Again, at this stage, the REIGN framework might have improved the process. As part of the review on ethical issues, it was realized that the perspective of developing countries was heavily underrepresented in the academic literature (as already described above). Therefore, there was a risk that ethical issues particular to this context could not be identified (and might not have been addressed) as part of the guideline. The REIGN framework stipulates thematic/argumentative saturation as a quality criterion for the body of evidence and accordingly invites consideration of whether there are gaps in the evidence. Had the REIGN framework been applied, accessing further evidence sources (e.g., through further involvement with stakeholders from developing countries) might have been explicitly considered in this case. This might have added an important perspective (although, again, this cannot conclusively be shown, as further research would be needed to prove that issues were truly missed). Furthermore, the REIGN framework can raise awareness that SRNEs are not the only and even not always the best option for collecting and synthesizing (at least normative) evidence. This might be particularly important, as some WHO officials involved in the process – particularly those not socialized in ethics or philosophy – assumed that SRNEs always produce the best or most reliable results.

Using the Evidence

The findings from the first review (on understandings of public health surveillance) were presented to the GDG in the form of a report. The report was 21 pages long and contained a list of the definitions found and a more detailed analysis of the different components contained in definitions (and their varying specifications). The findings from the second review (on ethical issues in surveillance) were also presented in a report (of 51 pages including an overview of ethical issues with example quotes from various publications). In addition, the findings were presented at two meetings of the GDG. The preliminary findings were presented in Prato, Italy, on June 9-10, 2015. The final results were presented on Kish Islands, Iran, during the last meeting on December 15-16, 2015.

The REIGN framework does not explicitly address how to present results – particularly because no standards have been developed or consensus reached in this regard and because how the results are presented might also depend on the particular questions asked and sources accessed to collect

evidence. However, in describing possible ways to conduct SRNEs, the report clearly points out the importance of stakeholder orientation in presenting results. In the review conducted, the authors chose to present the ethical issues in a "phase model". This means that the different phases of surveillance (background; system design and implementation; data collection, analysis, and storage; and data reporting, sharing, and using) were used to sort and structure the different ethical issues found. This structure was chosen because it was found to be the most helpful for practitioners implementing or working in surveillance systems. However, the report was not prepared for this group of stakeholders, and this structure might have been less helpful for the GDG. It might have been more helpful for the group (the GDG) primarily using the report to use a normative framework (e.g., the four principles specified in the report) to sort the various ethical issues found. The REIGN framework might provide additional orientation in this regard. However, many questions are left unanswered. Therefore, it will be particularly interesting to implement empirical studies to see what ways of synthesizing and presenting evidence are most helpful for ethics guideline developers. These ways might differ across the identified ESCs.

Concluding Remarks

It has been shown in both case studies that the REIGN framework might have been helpful in strengthening guideline development. However, the process used in guideline development has been important in making the REIGN framework possible in the first place. This process was the first attempt to explicitly consider evidence in ethics guideline development, and much was learned during this process. The process, for example, raised awareness of the importance of clarifying the meaning of basic terms upfront (because public health surveillance is a particularly contested term). This process has therefore heavily impacted the development of the REIGN framework. It will thus be particularly important to use the REIGN framework to learn about existing blind spots and to develop the framework further according to future experiences.

6. CONCLUSIONS

The preceding chapters provide a framework specifying the potential roles of evidence in the development of ethics guidelines. First, the strategies used for identifying institutional approaches and academic publications addressing this question were described. The identified approaches (or approaches to subquestions) were also presented as a conceptual basis that informed the development of the REIGN framework. Second, several conceptual clarifications were provided. The ethical assumptions grounding this framework were made explicit, as was the underlying understanding of ethics guidelines, explicating the scope of the REIGN framework. It was furthermore important to clarify the understanding of evidence – both normative and empirical – and its relation to normative and empirical literature. Based on these clarifications, the REIGN framework was introduced; the framework specified **(a)** the (complex of) *questions* or evidential support components (ESCs) possibly arising in ethics guideline development and that could benefit from being informed by evidence; **(b)** the *sources* from which particularly normative evidence can be drawn; and **(c)** how *quality* should be understood in the context of normative evidence and how quality should impact the choice of strategies for collecting evidence. Finally, the clarified concepts and the REIGN framework were used to analyse and critically appraise two existing ethics guidelines of the WHO as case studies of how the REIGN framework might have supported guideline development.

a. Limitations

This framework was developed under time and resource constraints. The authors were given only four months (September to December 2017) to conduct the work on the framework; this time constraint was relaxed only for developing the toolbox (which was developed at the beginning of 2018). The authors are convinced that they have provided solid and sound analyses despite the short timeframe, but the literature review, particularly, was not as comprehensive as would have been desirable.

State of Current Practice: Organizations/Institutions

Only a few countries were screened for approaches to integrating evidence in ethics guideline development. In addition, only a limited number of institutions (public health institutions, (clinical) guideline development agencies, organizations involved in health technology assessment, and national ethics committees or commissions) within their respective countries could be screened. As the authors were additionally unfamiliar with most of the health care systems and due to language barriers, it cannot be guaranteed that the institutions of interest were correctly identified and accordingly searched in each national context. However, the authors were also overinclusive with regard to the institutions screened to ensure nothing relevant was missed. Not all institutions, therefore, perfectly align with the types of organizations identified above. Nevertheless, excluded countries or institutions might have developed approaches that were highly relevant. Furthermore,

relevant approaches of even the included institutions might have been inaccessible due to language restrictions.

The risk that relevant publications were not accessed because they were not publicly available was mediated by additionally approaching relevant institutions via email. However, it was not always possible to identify relevant contact persons, and requests for support had been sent to general emails seemingly without being forwarded to the relevant contact person.

State of Current Practice: Academic Discourse

It was not possible to conduct a full systematic review of approaches to evidence integration in ethics guideline development: only a scoping review of a limited number of databases and a focus on four prespecified discourses was conducted. The resulting sample of papers was likely not comprehensive, but both lead authors of this discussion paper are intimately involved with the literature through their own academic work and are aware of recent developments. Accordingly, the risk that very relevant publications were missed is considered small.

While full summaries of all institutional approaches are provided within the discussion paper (see Appendix A), due to the number of papers, only short summaries of the most important points raised in the four discourses were provided. However, the papers identified were read and considered in developing the REIGN framework.

REIGN Framework

The REIGN framework is built on certain assumptions (normative, meta-ethical and definitional) that can be contested. For example, the definition of "evidence" is strongly influenced by examining "evidence" from the perspective of informal logic/argumentation theory while also assuming epistemologically that in the end, at least in lived praxis, all epistemic and practical issues can be properly described as taking place in a (more or less) open – and often, in a modern and complex society, "never-ending" – argumentative discourses where beliefs are exchanged, defended and criticized (orientation to a discourse theoretical approach). This could be contrasted with approaches that are oriented to more direct evidentialism, i.e., where sense data can directly justify a statement without having to make a detour via beliefs and subsequent arguments (which can all be contested in a discourse). Additionally, the differentiation made between empirical and normative evidence can be contested when assuming a theoretical stance where a strict division between the "is" and the "ought" is denied. Generally, the orientation towards empirical ethics approaches, or at least a more empirically inclined bioethics (or, e.g., public health ethics), can be contrasted with approaches that conceptualize ethics explicitly as having nothing to do with empirical work, as the latter has to be conducted solely by social scientists. Furthermore, a possible objection to the assumption of having a "value base" or an "ethical corridor" is that providing such a "corridor" gives too much leeway for ethics guideline development, as it, in principle, could rule out important normative-ethical or legal

requirements (e.g., human rights) from the very start. Finally, stressing the relevance of systematic reviews for ethics is based on methodological assumptions informed generally by philosophies of evidence-based medicine and health care and of a certain way of (also) "doing ethics". This approach can be contested by (general) critics of evidence-based approaches and some (philosophical) ethicists who do not agree that such a method is a proper way of working on ethical topics.

The assumptions made are, however, well suited for the context given in this discussion paper. Every framework starts from some premises; thus, this should not be considered problematic, although readers may reject some of the underlying central assumptions (e.g., the orientation to evidence-based approaches *per se*). While the REIGN framework is built on the insights shared in academic and institutional publications, it is still the conceptual work of only two researchers with particular academic biographies, (inter-)disciplinary backgrounds, preferred theoretical stances, and research interests. Therefore, other researchers may disagree with some of the understandings and conceptualizations. Particularly as the authors have, as part of the development exercise, made suggestions for handling or understanding contested issues thus far (and not just summarized already developed and generally approved solutions), this framework should be seen only as a starting point for further discussions. As already stated above, it will be particularly important in this context to bring together various experts from guideline development, evidence synthesis, ethics, health care policy, and other fields to discuss proposals made in this discussion paper and to develop more concrete guidance in the future.

b.   Open Questions

Because of the limitations mentioned above and for further reasons, many questions cannot be answered in this discussion paper.

Those developing ethics guidelines will particularly miss practical or more concrete guidance on "when to do what how". However, a manual for developing ethics guidelines cannot be provided yet; instead, this paper can provide only a framework for structuring deliberations (on, e.g., which source should be considered under what circumstances for collecting or synthesizing normative evidence) and some preliminary practical advice (see Appendix D) because many conceptual questions are still underdeveloped and highly contested in the scientific communities. (Examples of such conceptual questions include the understanding of evidence in the context of ethics guidance; the relevance of empirical evidence for arriving at ethics recommendations; the role of ethical theory when faced with evidence-based approaches; and the quality appraisal of normative information, such as individual arguments or ethical issues.) It is even contested, especially in the (medical) ethics community, whether an evidence-based approach is appropriate at all for devising *ethics* guidance.[45] The proposed alternative to such an evidence-based approach in ethics is to rely on ethical theories, the careful consideration of ethical issues, and then arguments for specific courses of action. While

---

[45] This, at least, is the recurring experience of the authors.

such an approach also relies on studying the academic literature, this approach does so less systematically. This approach, however, mirrors the "traditional" (research) processes in (philosophical) ethics and the content of many publications in the discipline. Whether such "traditional" research processes should be implemented by dedicated ethics experts (e.g., philosophers) or by guideline developers also (sufficiently) trained in ethics (see also below) cannot be answered by the authors. Either way, in terms of an evidence-based approach and REIGN's own terminology, such an approach would imply relying solely on expert consultations or internal discussions of the GDG as an evidence source.

At the moment, however, there is no empirical research regarding which approach achieves better or optimal results. Nevertheless, while the above-sketched alternative approach (i.e., relying on ethics theories, etc.) to developing ethics guidance is less complex and needs fewer resources, this approach is also less transparent and possibly results in less accountability – and less valid results. Recommendations developed by this alternative approach are therefore harder to criticize from an academic perspective. Consequently, there is a danger that ethics guidelines developed with this alternative approach are considered ideological products that are easily dismissed as opinion based. Additionally, proponents of this alternative approach might underestimate the risk of bias in ethical analysis and argumentation; this risk can be mitigated to at least some extent by an approach that is more evidence based.

Before more definite methodological and practical guidance can be developed – such a development should definitely be the goal for the future – conceptual issues raised in this paper have to be addressed first, and at least some partial consensus in the relevant communities reached. To achieve such a consensus, it might be particularly important to refine the methodology for conducting systematic reviews to obtain normative evidence. An overview of approaches to the different steps of a systematic review is provided in the Appendix (see Appendix C) of this discussion paper. While indicating the contentious points, this overview acknowledges that no "best practice standard" has yet been established. As was done for empirical evidence, it might be important to bring experts together in a consensus-seeking process (e.g., a Delphi process) to address contested issues and develop guidance documents comparable to AGREE-II [Brouwers et al 2010], PICO [Higgins/Green 2008] or GRADE [Guyatt et al 2008]. For systematic reviews in ethics, a reporting guideline comparable to PRISMA [Moher et al 2009] would also be helpful [Mertz/Strech/Kahrass 2017].[46] As a stop-gap measure, however, it might be helpful to have questions or noticeable points provided as a sort of "checklist" that guides evidence gathering. As part of this framework, a toolkit (including a checklist based on the identified ESCs) is therefore provided (see Appendix D).

---

[46] However, the process for developing such reporting guidelines is already underway. See: "PRISMA-Ethics: an extension to PRISMA for SRs on ethics literature"; see http://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-systematic-reviews/#4102018 (24.10.2019)

Apart from further conceptual work to develop more concrete guidance, it might also be helpful to test the usability of the framework and toolkit, assuming they might be used in the meantime to strengthen processes. It might be instrumental to engage with stakeholders (meaning those potentially using the developed tools) to refine the framework and toolkit based on their feedback. In addition, any application of the framework and toolkit in a WHO guideline development process should be accompanied by an evaluation to see whether their use is feasible and useful. It might be particularly interesting to compare experiences with frameworks and toolkits across a variety of development processes of varying complexity.

Additionally, the WHO might consider stipulating a fixed value base (or ethical corridor in accordance with Reiter-Theil et al [2011b]) instead of shifting the responsibility for defining ethical baseline principles to the GDG. Such an ethical corridor denotes the boundaries of acceptable recommendations the GDG might decide to issue and therefore limits the possible recommendation the GDG can reach. Therefore, ethical corridors have to be broad enough to leave room for pluralistic values, but ethical corridors also have to introduce clear boundaries. In national guideline development, the law of the country is sometimes used to provide crucial normative content for defining this corridor [Reiter-Theil et al 2011b]. The WHO is part of the United Nations system and, as such, guided by the Universal Declaration of Human Rights. Thus, it might make sense to use human rights approaches as normative-ethical and normative legal side constraints for devising ethics recommendations [see, e.g., WHO 2010]; however, this cannot be determined in this discussion paper but will need further discussion within the WHO itself or possibly the *World Health Assembly*.

Even if the WHO decides not to introduce "ethical corridors" to the decision-making process, certain ethical assumptions will likely have to be made before engaging in collecting and synthesizing normative evidence. As previously noted, if literature – normative and empirical – is searched for normative information, certain normative assumptions will have to be made: for example, what renders an issue an "ethical issue" (e.g., because it conflicts with certain predefined ethical principles) or a piece of information an "ethical argument". The WHO will likely have to take a stance on these questions in accordance with its statute. Former systematic reviews that have informed WHO ethics guidelines have used principlism to define ethical issues [Klingler et al 2017]; principlism is also one of the approaches often advocated in the HTA context but might not best fit WHO's international mandate.

Furthermore, following the basic idea of evidence-based approaches, it will be necessary to empirically evaluate the *effectiveness* of ethics guidelines, as is the case with other "normative products", such as the more "technical" guidelines of the WHO [WHO Evaluation Office 2017, p. 45]. This means that there have to be evaluation studies about whether ethics guidelines (or their content) are sufficiently disseminated, whether the target audiences comply with these guidelines, and – more challenging – whether these guidelines actually improve the practice *ethically* as intended. Similarly – as some sort of an "ethics of doing ethics" – it has to be evaluated whether the implementation of

the ethics guidelines leads to (unintended) ethically problematic side effects. Such an assessment can also be conducted before actual implementation of the guidelines in the form of an "ethics consequence assessment" ("Ethikfolgenabschätzung" [Strech 2008c]) by, e.g., examining especially socio-empirical insights from the implementation of comparable ethics instruments or "tools". These assessments can be understood as a kind of "extended" ESC 5 (focusing on the *consequences of actions*), even though the empirical information gathered is arguably more directed towards the ethics guideline as a whole than towards specific actions that are recommended in the ethics guideline.

Apart from these aspects related to an evidence-based approach to ethics guideline development, other aspects of the process also deserve further attention. As has been pointed out several times before, the final recommendations are decided and formulated by the GDG, and the GDG will also decide what kind of evidence is collected and considered. Accordingly, the GDG has considerable power over the final guidelines, and the quality of the guidelines will depend significantly on the GDG – particularly on its composition, decision-making processes, and potentially available resources. It will therefore be particularly important to better understand the influences of these factors and formulate guidance with regard to *who* participates in the decision-making and *how* this process is structured. The *WHO Handbook for Guideline Development* specifies what groups have to be represented in the GDG for more technical questions [WHO 2014b]: technical experts, end-users of the measure in question, affected communities, and method experts. It will have to be discussed whether adjustments have to be made when the goal of the guideline is explicitly ethical in nature. Ethics expertise should be represented among the group, but ethics experts might just be one specification of "technical" experts. With regard to consensus building, it will be particularly important to prevent domination of the discussion by individuals or subgroups to ensure that rational *arguments* – at best resulting from a solid evidence base, including empirical and especially normative evidence – not power, tips the scale. In developing these processes, consulting frameworks for good decision-making in ethically charged situations might be indispensable [see, e.g., Bennett/Gibson 2006; Jiwani 2015].[47]

In this context, what qualifications or skills are necessary for searching and synthesizing research evidence for ethics guideline development must also be addressed. Even when a methodological framework (such as the REIGN framework) or even more practical guidance in the future is provided, guideline developers unfamiliar with ethical analyses and evaluations might be overburdened with detecting and appreciating relevant research evidence – especially in regard to *normative* evidence. Although it will likely *not* be necessary to have a dedicated ethics expert involved in evidence identification and synthesis, at least some competence regarding ethics, ethical analyses and decision-making in normative contexts might be a prerequisite.

---

[47] For further information on the Good Decisions frameworks, see also Jiwani's website: http://incorporatingethics.ca/

c.   Implications for Practice

For developing ethics guidelines – or ethics guidance in general – the following points resulting from the discussions in this discussion paper should be considered:

- The five ESCs should be used to structure guideline development from an early stage. The ESCs provide an overview of points or questions that have to be addressed during the decision-making process, which may benefit from (further) evidence. The GDG should use the ESCs to take an explicit stance towards engaging (or not) in (further) evidence collection and synthesis. Using the ESCs as orientation points enables the GDG to outline how decisions that have to be taken relate to each other and will prevent the GDG from formulating questions that become irrelevant at later points. A toolkit to support decision-making with regard to further evidence collection is provided in Appendix D, although the toolkit will have to be further assessed for understandability (by users) and comprehensiveness (by experts).

- The development process should be transparently reported in resulting guidelines (or accompanying documentation) with regard to methods used to allow readers to assess the quality of the development process (and possibly resulting recommendations and conclusions). Such reporting can be structured in accordance with the ESCs provided by the REIGN framework. Guideline developers should list for each ESC which (normative or empirical) information or sources of information were used, how these sources were accessed – and assessed – and what limitations are associated with the evidence base available. Reporting can even be extended to each individual recommendation (e.g., in a standardized form [Mertz/Strech 2014]) so that for each recommendation, the evidence base is apparent to the reader – whether the evidence base consists of various academic publications (which should be referenced), an expert consensus, the results of a stakeholder survey, etc.

- For a reliable evidence base, conducting systematic reviews should be actively considered and discussed during development; reviews for normative evidence should especially be conducted, but reviews for (socio-)empirical evidence should also be conducted when needed. Systematic reviews can be crucial when there is much discussion about the "right" course of action, as they help mitigate possible risks of bias when only a few sides are "heard" during development or when a topic is discussed by various (sub)disciplines. As discussed in more detail in chapter 4c, systematic reviews will be less valuable in certain contexts, e.g., where the topic is comparably new and only a few academic publications on the topic of interest are to be expected. However, conducting systematic reviews is always also a question of available time and human resources. Therefore, how important or valuable a review could be for the final recommendation must be discussed (e.g., for each ESC). There are also other possibilities (stakeholder surveys, discussion rounds, theory application, etc.) for gathering or directly generating evidence during guideline development; therefore, the pros and cons should be discussed beforehand, and it should be determined whether other strategies, which may be less reliable than systematic reviews, also

suffice. Either way, the people tasked with collecting and analysing evidence should be involved in these decision-making processes. Because such people (should) bring specific methodological expertise to the table, they will be able to ensure that the formulated questions fulfil relevant criteria, for example, that the questions are searchable by systematic reviews and relevant to guideline developers.

- Persons charged with collecting and synthesizing evidence in ethics guideline development should have sufficient proficiency regarding "ethics" to be able to identify important sources and to assess the relevance of identified information for planned recommendations. Although as discussed before, no clear skill requirements can be offered as part of this framework, it will be important to define such requirements in the future. Possible criteria include a certain degree of familiarity with ethics (topics, analysis, argumentation, etc.), understood as a normative enterprise (i.e., not seen solely from a more sociological, descriptive perspective). In addition, knowledge about different ways of gathering and synthesizing normative information (including at least basic knowledge about systematic reviews for normative evidence) and some understanding of the structure of guideline development processes might be important.

- Generally, it has to be acknowledged that in reality, not everything can be done to improve the evidence base of a guideline, as it would simply be too costly. However, using a framework such as REIGN, which formulates how evidence "ideally" should be considered in guideline development in ethics will still facilitate the reporting of methods. Guideline developers should make when they depart from the ideal approach and for what (pragmatic) reasons transparent. For example, there may be no need to clarify central terms or identify the strategies for addressing needs in many situations. What is essential, however, is to transparently state where one departed from an ideal and provide the rationale for doing so (see also the second point raised above). "Departing from the ideal" might also be justified in regard to choosing people to serve on the GDG or the review group (e.g., persons with the ideal experience and skills set might not exist, but there may still be persons capable of doing the work).

- Finally, those involved in guideline development should be aware of the unanswered questions and contested conceptual issues surrounding the role of evidence in ethics guidelines. Guideline developers should not assume that there is widespread methodological consensus when, in fact, there is not. However, REIGN will hopefully raise awareness of the importance of clearly explaining and justifying how and why an ethics guideline was developed. This not only increases the knowledge base for those developing guidelines but also provides possibilities for evaluating specific ways of developing ethics guidelines. REIGN further allows us to compare different processes and results and to start conversations with others involved in ethics guideline development. These benefits provided by REIGN will be crucial prerequisites for eventually arriving at a consensus on how to optimally develop guidance documents on normative-ethical topics.

Acknowledgements

7. References

Abbasi M, Majdzadeh R, Zali A, Karimi A, Akrami F (2018) The evolution of public health ethics frameworks: systematic review of moral values and norms in public health policy. *Medicine, Health care and Philosophy* 21(3):387-402

Albisser Schleger H, Mertz M, Meyer-Zehnder B, Reiter-Theil S (2012) *Klinische Ethik – METAP. Leitlinie für Entscheidungen am Krankenbett*. Springer, Heidelberg

Arras JD (2007) "The way we reason now: reflective equilibrium in bioethics". In: Steinbock B (ed) *The Oxford Handbook of Bioethics*. Oxford University Press, Oxford/New York; p. 46-71

Ashcroft RE (2003) Constructing empirical bioethics: foucauldian reflections on the empirical turn in bioethics research. *Health Care Analysis* 11(1):3-13

Assasi N, Schwartz L, Tarride J-E, Campbell K, Goeree R (2014) Methodological guidance documents for evaluation of ethical considerations in health technology assessment: a systematic review. *Expert Review of Pharmacoeconomics & Outcomes Research* 14(2):203-220

Autti-Rämö I, Mäkelä M (2007) Ethical evaluation in health technology assessment reports: An eclectic approach. *International Journal of Technology Assessment in Health Care* 23:1-8

Backmann M (2017) What's in a gold standard? In defence of randomised controlled trials. *Medicine, Healthcare and Philosophy* 20(4):513-523

Beauchamp TL, Childress JF (2009) *Principles of Biomedical Ethics*. 6th edition. Oxford University Press, New York/Oxford

Bennett MD, Gibson JM (2006) *A field guide to good decisions: values in action*. Praeger, Westport, Connecticut/London

Birnbacher D (2012) Can there be such a thing as ethical expertise? *Analyse & Kritik* 34(2):237-249

Bion J (2009) Financial and intellectual conflicts of interest: confusion and clarity. *Current Opinion on Critical Care* 15(6):583-590

Bombard Y, Abelson J, Simeonov D, Gauvin FP (2011) Eliciting ethical and social values in health technology assessment: A participatory approach. *Social Science & Medicine* 73:135-144

Bonneux L (2007) From evidence based bioethics to evidence based social policies. *European Journal of Epidemiology* 22:483-485

Borry P, Schotsmans P, Dierickx K (2005) The birth of the empirical turn in bioethics. *Bioethics* 19(1):49-71

Bosch-Capblanch X, Lavis JN, Lewin S, Atun R, Røttingen J-A, Dröschel D, Beck L, Abalos E, El-Jardali F, Gilson L, Oliver S, Wyss K, Tugwell P, Kulier R, Pang T, Haines A (2012) Guidance for evidence-informed policies about health systems: rationale for and challenges of guidance development. *PLoS Medicine* 9(3):e1001185

Braunack-Mayer AJ (2006) Ethics and health technology assessment: Handmaiden and/or critic? *International Journal of Technology Assessment in Health Care* 22(3):307-312

Bowie NE (2009) How empirical research in human cognition does and does not affect philosophical ethics. *Journal of Business Ethics* 88:635-643

Brock DW (1987) Truth or consequences: The role of philosophers in policy-making. *Ethics* 97(4):786-791

Broome J (1994) Discounting the future. *Philosophy & Public Affairs.* 23(2):128-156

Brouwers M, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna S, Littlejohns P, Makarski J, Zitzelsberger L for the AGREE Next Steps Consortium (2010) AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *Canadian Medical Association Journal.* doi:10.1503/cmaj.090449

Burls A, Caron L, Cleret de Langavant G, Dondorp W, Harstall C, Pathak-Sen E, Hofmann B (2011) Tackling ethical issues in health technology assessment: a proposed framework. *International Journal of Technology Assessment in Health Care* 27:230-237

Burns PB, Rohrich RJ, Chung KC (2011) The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery* 128(1):305-310

Copi IM (1998) *Introduction to Logic*. Prentice Hall College Div, Upper Saddle River (NJ)

CEbM (Centre for Evidence-Based Medicine) (2009) *Levels of Evidence*. http://www.cebm.net/blog/2009/06/11/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ (24.10.2019)

Cho MK (2014) Ethics and empiricism in the formation of professional guidelines. Editorial. *American Journal of Bioethics* 14:1-2

Chung KC, Pushman AG, Bellfi LT (2009) A systematic review of ethical principles in the plastic surgery literature. *Plastic and Reconstructive Surgery* 124(5):1711-1718

Cowley C (2005) A new rejection of moral expertise. *Medicine, Healthcare and Philosophy* 8:273-279

Daniels N (2011) "Reflective Equilibrium". In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*. (Fall 2008 Edition). http://plato.stanford.edu/archives/spr2011/entries/reflective-equilibrium/ (24.10.2019)

DeJean D, Giacomini M, Schwartz L, Miller FA (2009) Ethics in Canadian health technology assessment: A descriptive review. *International Journal of Technology Assessment in Health Care* 25(4): 463-469

Deutscher Ethikrat (2019) *Geschäftsordnung.* https://www.ethikrat.org/der-ethikrat/#m-tab-0-geschftsordnung(24.10.2019)

DeVries R, Gordijn B (2009) Empirical ethics and its alleged meta-ethical fallacies. *Bioethics* 23(4):193-201

Droste S, Gerhardus A, Kollek R (2003) *Methoden zur Erfassung ethischer Aspekte und gesellschaftlicher Wertvorstellungen in Kurz-HTA-Berichten – eine internationale Bestandsaufnahme*. Deutsche Agentur für Health Technology Assessment des Deutschen Instituts für Medizinische Dokumentation und Information (DAHTA@DIMDI), Cologne

Droste S (2008) Systematische Gewinnung von Informationen zu ethischen Aspekten in HTA-Berichten zu medizinischen Technologien bzw. Interventionen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 102:329-343

Droste S, Dintsios CM, Gerber A (2010) Information on ethical issues in health technology assessment. Ho wand where to find them. *International Journal of Technology Assessment in Health Care* 26:441-449

Droste S, Herrmann-Frank A, Scheibler F, Krones T (2011) Ethical issues in autologous stem cell transplantation (ASCT) in advanced breast cancer: a systematic literature review. *BMC Medical Ethics* 12:6

Dunn M, Sheehan M, Hope T, Parker M (2012) Toward methodological innovation in empirical ethics research. *Cambridge Quarterly of Healthcare Ethics* 21:466-480

Düwell M (2009) Wofür braucht Medizinethik empirische Methoden? Eine normativ-ethische Untersuchung. *Ethik in der Medizin* 21(3):201-211

Emanuel EJ, Grady C, Wendler D (2008) "An ethical framework for biomedical research". In: Emanuel EJ, Grady C, Crouch R, Lie RK, Miller FG, Wendler D (eds) *The Oxford Textbook of Clinical Research Ethics*. Oxford University Press, New York; p. 123-135

EUnetHTA Joint Action 2, Work Package 8 (2016) *HTA Core Model ® Version 3.0*. https://eunethta.eu/hta-core-model/(24.10.2019)

Frize M, Walker RC, Ennett CM (2003) Development of an Evidence-Based Ethical Decision-Making Tool for Neonatal Intensive Care Medicine. *Proceedings of the 25th Annual International Conference of the IEEE EMBS Cancun, Mexico, September 17-21, 2003*

Fogelin RJ, Sinnott-Armstrong W (2005) *Understanding Arguments. An Introduction to Informal Logic*. Wadsworth/Thomson Wadsworth, Belmont (CA)

Føllesdal D, Walløe L, Elster J (2010) *Rationale Argumentation. Ein Grundkurs in Argumentations- und Wissenschaftstheorie*. Walter de Gruyter & Co., Berlin/New York

Garrido MV, Zentner A, Busse R (2008) "Health systems, health policy and health technology assessment". In: Garrido MV, Kristensen FB, Nielsen CP, Busse R (eds) *Health Technology Assessment and Health Policy-Making in Europe. Current Status, Challenges and Potential*. World Health Organization, on behalf of the European Observatory on Health Systems and Policies; p. 53-78

Gesang B (2010) Are moral philosophers moral experts? *Bioethics* 24(4):153-159

Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson L, Rindress D (2008) Evidence and value: Impact on DEcisionMaking – the EVIDEM framework and potential applications. *BMC health services research* 8:270

Goldenberg MJ (2005) Evidence-based ethics? On evidence-based practice and the 'empirical turn' from normative bioethics. *BMC Medical Ethics* 6:11

Graeme L, Stevens L, Jones KH, Dobbs C (2015) *A review of evidence relating to harm resulting from uses of health and biomedical data*. Nuffield Council on Bioethics, Oxford

Grunwald A (2004) The normative basis of (health) technology assessment and the role of ethical expertise. *Poiesis & Praxis* 2:175-193

Guindo LA, Wagner M, Baltussen R, Rindress D, van Til J, Kind P, Goetghebeur MM (2012) From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking. *BMC Cost Effectiveness and Resource Allocation* 10(1):9

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ (for the GRADE Working Group) (2008) GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal* 336:924

Habermas J (1991) *Erläuterungen zur Diskursethik*. Suhrkamp, Frankfurt a.M.

Hacking I (1999) *The Social Construction of What?* Harvard University Press, Cambridge/London

Halpern SD (2005) Towards evidence based bioethics. *British Medical Journal* 331:901-903

Hannes K & Lockwood C (2012) Synthesizing qualitative research: Choosing the right approach. Wiley-Blackwell & BMJ Books, Oxford, Chichester, Hoboken.

Heintz E, Lintamo L, Hultcrantz M, Jacobson S, Levi R, Munthe C, Tranæus S, Östlund P, Sandman L (2015) Framework for systematic identification of ethical aspects of healthcare technologies: the SBU approach. *International Journal of Technology Assessment in Health Care* 31(3):124-130

Higgins JPT, Green S (2008) Cochrane Handbook for Systematic Reviews of Interventions. Wiley-Blackwell, Chichester

Hofmann B (2005) Towards a procedure for integrating moral issues in health technology assessment. *International Journal of Technology Assessment in Health Care* 21:312-318

Hofmann B (2014) Why not integrate ethics in HTA: identification and assessment of the reasons. *GMS Health Technology Assessment* 10:Doc04

Hofmann B, Lysdahl KB, Droste S (2015) Evaluation of ethical aspects in health technology assessment: more methods than applications? *Expert Review of Pharmacoeconomics & Outcomes* Research 15:5-7

IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) (2017) *Allgemeine Methoden.* Version 5.0, 10.07.2017.
https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf (20.12.2017)

Ives J, Draper H (2009) Appropriate methodologies for empirical bioethics: It's all relative. *Bioethics* 23(4):249-258

Ives J, Dunn M, Cribb A (2017, eds) *Empirical Bioethics. Theoretical and Practical Perspectives*. Cambridge University Press, Cambridge

Ives J, Dunn M, Molewijk B, Schildmann J, Baeroe K, Frith L, Huxtable R, Landerweer E, Mertz M, Provoost V, Rid A, Salloch S, Sheehan M, Strech D, deVries M, Widdershoven G (2018) Standards of practice in empirical bioethics research: towards a consensus. *BMC Medical Ethics* 19:68

Jansen R (1997) Evidence-based ethics and the regulation of reproduction. *Human Reproduction* 12:2068-2075

Jiwani B (2015) Ethically justified decisions. *Healthcare Management Forum* 28(2):86-89

Jox R (2014) Entwicklung einer ethisch-rechtlichen Klinik-Policy. *Hessisches Ärzteblatt* 5:268-281

Kahrass H, Strech D, Mertz M (2017) Ethical issues in obesity prevention for school children: a systematic qualitative review. *International Journal of Public Health* 62(9):981-988

Kambartel F (2004) "Normativ". In: Mittelstrass J (ed) *Enzyklopädie Philosophie und Wissenschaftstheorie. Band 2.* J.B. Metzler, Stuttgart/Weimar; p. 1034-1035

Kangasniemi M, Arala K, Becker E, Suutarla A, Haapa T, Korhonen A (2017) The development of ethical guidelines for nurses' collegiality using the Delphi method. *Nursing Ethics* 24:538-555

Klingler C, Silva DS, Schuermann C, Reis A, Saxena A, Strech D (2017) Ethical issues in public health surveillance: a systematic qualitative review. *BMC Public Health* 17(1):295

Knüppel H, Mertz M, Schmidhuber M, Neitzke G, Strech D (2013) Inclusion of ethical issues in dementia guidelines: a thematic text analysis. *PLoS Medicine* 10(8):e1001498

Kon AA (2009) The role of empirical research in bioethics. *American Journal of Bioethics* 9(6-7):59-65

Kim SY (2004) Evidence-based ethics for neurology and psychiatry research. *NeuroRx* 1:372-377

Kitcher P (2011) *The Ethical Project*. Harvard University Press, Cambridge (MA)

Koplin JJ, Selgelid MJ (2015) Burden of proof in bioethics. *Bioethics* 29(9):597-603

Kulkarni AV (2005) The challenges of evidence-based medicine: a philosophical perspective. *Medicine, Healthcare and Philosophy* 8:255-260

Kunz R, Ollenschläger G, Raspe H, Jonitz G, Donner-Banzhoff N (2007, eds) *Lehrbuch Evidenz-basierte Medizin in Klinik und Praxis*. Deutscher Ärzte-Verlag, Cologne

Lee L (2012) Public health ethics theory: review and path to convergence. *Journal of Law, Medicine and Ethics* 40(1):85-98

Lehoux P, Williams-Jones B (2007) Mapping the integration of social and ethical issues in health technology assessment. *International Journal of Technology Assessment in Health Care* 23:9-16

Linstone HA, Turoff M (2002, eds) *The Delphi Method: Techniques and Applications.* https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf (20.12.2017)

Lysdahl KB, Mozygemba K, Burns J, Chilcott JB, Brönneke JB, Hofmann B (2016a, eds) *Guidance for assessing effectiveness, economic aspects, ethical aspects, socio-cultural aspects and legal aspects in complex technologies*.
http://www.integrate-hta.eu/downloads/ (20.12.2017)

Lysdahl KB, Oortwijn W, van der Wilt GJ, Refolo P, Sacchini D, Mozygemba K, Gerhardus A, Brereton L, Hofmann B (2016b) Ethical analysis in HTA of complex health interventions. *BMC Medical Ethics* 17:16

Major-Kincade TL, Tyson JE, Kennedy KA (2001) Training pediatric house staff in evidence-based ethics: An exploratory controlled trial. *Journal of Perinatology* 21:161-166

McCarthy J, Gastmans C (2015) Moral distress: a review of the argument-based nursing ethics literature. *Nursing Ethics* 22(1):131-152

McCullough LB, Coverdale JH, Chervenak FA (2004) Argument-based medical ethics: A formal tool for critically appraising the normative medical ethics literature. *American Journal of Obstetrics and Gynecology* 191:1097-1102

McCullough LB, Coverdale JH, Chervenak FA (2007) Constructing a systematic review for argument-based clinical ethics literature: the example of concealed medications. *Journal of Medicine and Philosophy* 32:65-76

McDougall R (2014) Systematic reviews in bioethics: types, challenges, and value. *Journal of Medicine and Philosophy* 9:89-97

McDougall R (2015) Reviewing literature in bioethics research: increasing rigour in non-systematic reviews. *Bioethics* 29(7):523-528

McDougall R, Notini L (2014) Overriding parents' medical decisions for their children: a systematic review of normative literature. *Journal of Medical Ethics* 40(7):448-452

McMillan J, Hope T (2008) "The possibility of empirical psychiatric ethics". In: Widdershoven G, McMillan J, Hope T, Van der Scheer L (eds) *Empirical ethics in psychiatry*. Oxford University Press, New York; p. 9-22

Mertz M (2011) *Evidenzbasierte Klinische Ethik. Philosophische Untersuchungen zur Verwendung von Empirie und Evidenz in der (Medizin-)Ethik*. GRIN Verlag, München/Ravensburg

Mertz M (2017) Qualitätsbewertung in systematischen Übersichtsarbeiten normativer Literatur. Eine Problemanalyse. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 127-128:11-20

Mertz M, Strech D (2014) Systematic and transparent inclusion of ethical issues and recommendations in clinical practice guidelines: a six-step approach. *Implementation Science* 9:184

Mertz M, Inthorn J, Renz G, Rothenberger GL, Salloch S, Schildmann J, Wöhlke S, Schicktanz S (2014) Research across the disciplines: a road map for quality criteria in empirical ethics research. *BMC Medical Ethics* 15:17

Mertz M, Kahrass H, Strech D (2016) Current state of ethics literature synthesis: a systematic review of reviews. *BMC Medicine* 14:152

Mertz M, Strech D, Kahrass H (2017) What methods do reviews of normative ethics literature use for search, selection, analysis and synthesis? In-depth results from a systematic review of reviews. *Systematic Reviews* 6:261

Mittelstadt B, Floridi L (2015) The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics* 22(2):303-341

Mittelstadt B, Stahl B, Fairweather B (2013) PHM-Ethics and ETICA: complementary approaches to ethical assessment. *Studies in Health Technology and Informatics* 187:117-135

Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine* 151:264-269

Molewijk B, Stiggelbout AM, Otten W, Dupuis HM, Kievit J (2004) Empirical data and moral theory. A plea for integrated empirical ethics. *Medicine, Healthcare and Philosophy* 7:55-69

Molewijk B, Stiggelbout AM, Otten W, Dupuis HM, Kievit J (2008) First the facts, then the values? Implicit normativity in evidence-based decision aids for shared decision-making. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 102:415-420

Muir Gray JA (2001) *Evidence-based healthcare: how to make health policy and management decisions*. Churchill Livingston, Edinburgh/New York

Neitzke G, Riedel A, Brombacher L, Heinemann W, Herrmann B (2015) Empfehlungen zur Erstellung von Ethik-Leitlinien in Einrichtungen des Gesundheitswesens. *Ethik in der Medizin* 27(3):241-248

NICE (National Institute for Health and Clinical Excellence) (2008) *Social Value Judgements. Principles for the Development of NICE Guidance.* 2nd Edition. https://www.nice.org.uk/Media/Default/About/what-we-do/Research-and-development/Social-Value-Judgements-principles-for-the-development-of-NICE-guidance.pdf (24.10.2019)

Nuffield Council on Bioethics (2016) *Genome Editing. An Ethical Review.* http://nuffieldbioethics.org/wp-content/uploads/Genome-editing-an-ethical-review.pdf (24.10.2019)

Oxman AD, Lavis JN, Fretheim A (2007) Use of evidence in WHO recommendation. *The Lancet* 369:1883-1889

Petrini C (2010) Theoretical models and operational frameworks in public health ethics. *International Journal of Environmental Research and Public Health* 7(1):189-202

Polus S, Mathes T, Klingler C, Messer M, Gerhardus A, Stegbauer C, Willms G, Ehrenreich H, Marckmann G, Pieper D (2019) Health technology assessment of public health interventions published 2012 to 2016: An analysis of characteristics and comparison of methods. *International Journal of Technology Assessment in Health Care* 35(4):280-290

Presidential Commission for the Study of Bioethical Issues (2016) *Bioethics for every generation: deliberation and education in health, science, and technology.* Washington, DC. https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/PCSBI_Bioethics-Deliberation_0.pdf (24.10.2019)

Rehfuess E, Stratil JM, Scheel IB, Portela A, Norris SL, Baltussen R (2019) The WHO-INTEGRATE Evidence to Decision Framework Version 1.0: Integrating WHO norms and values and a complexity perspective. *BMJ Global Health* 4:e000844

Reiter-Theil S, Mertz M, Albisser Schleger H, Meyer-Zehnder B, Kressig RW, Pargger H (2011a) Klinische Ethik als Partnerschaft – oder wie eine Leitlinie für den patientengerechten Einsatz von Ressourcen entwickelt und implementiert werden kann. *Ethik in der Medizin* 23(2):93-105

Reiter-Theil S, Mertz M, Schürmann J, Stingelin N, Meyer-Zehnder B (2011b) Evidence – competence – discourse: the theoretical framework of the multi-centre clinical ethics support project METAP. *Bioethics* 25(7):403-412

Riedel A (2017) Ethik-Leitlinien als Verfahren der Ethikberatung – Stellenwert und Beitrag zur ethischen Reflexion und Entscheidungsfindung in der Behindertenhilfe. *EthikJournal* 4:1 http://www.ethikjournal.de/fileadmin/user_upload/ethikjournal/Texte_Ausgabe_1_06-2017/Riedel_Ethik-Leitlinien_als_Verfahren_der_Ethikberatung_EthikJournal_4_2017_1.pdf (24.10.2019)

Saarni SI, Hofmann B, Lampe K, Lühmann D, Mäkelä M, Velasco-Garridod M, Autti-Rämöa I (2008) Ethical analysis to improve decision-making on health technologies. *Bulletin of the World Health Organization* 86:617-623

Saarni SI, Braunak-Mayer A, Hofmann B, Jan G (2011) Different methods for ethical analysis in health technology assessment: an empirical study. *International Journal of Technology Assessment in Health Care* 27:305-312

Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB (2000) *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd Edition. Churchill Livingstone, Edinburgh

Sacchini D, Virdis A, Refolo P, Pennacchini M, de Paula IC (2009) Health technology assessment (HTA): ethical aspects. *Medicine, Health Care and Philosophy* 12:453-457

Salloch S (2012) "Evidenzbasierte Ethik"? Über hypothetische und kategorische Handlungsnormen in der Medizin. *Ethik in der Medizin* 24:5-17

Salloch S, Schildmann J, Vollmann J (2012) Empirical research in medical ethics: how conceptual accounts on normative-empirical collaboration may improve research practice. *BMC Medical Ethics* 13(1):5

Sandman L, Heintz E (2014) Assessment vs. appraisal of ethical aspects of health technology assessment: can the distinction be upheld? *GMS Health Technology Assessment* 10:Doc05 (doi: 10.3205/hta000121)

Sehon SR, Stanley DE (2003) A philosophical analysis of the evidence-based medicine debate. *BMC Health Services Research* 3(1):14

Seitzer F, Kahrass H, Neitzke G, Strech D (2016) The full spectrum of ethical issues in the care of patients with ALS: A systematic qualitative review. *Journal of Neurology* 263(2):201-209

Schleidgen S, Klingler C, Bertram T, Rogowski WH, Marckmann G (2013) What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Medical Ethics* 14:55

Schmücker R (2016) "Incidental findings: definition of the concept". In: Weckbach S (eds) *Incidental Radiological Findings*. Medical Radiology Series. Springer, Cham

Scott AM, Bond K, G Gutiérrez-Ibarluzea I, Hofmann B, Sandman L (2016) Quality assessment of ethics analyses for health technology assessment. *International Journal of Technology Assessment in Health Care* 32(5):362-369.

Scott AM, Hofmann B, Gutiérrez-Ibarluzea I, Lysdahl KB, Sandman L, Bombard Y (2017) Q-SEA – a tool for quality assessment of ethics analyses conducted as part of health technology assessments. *GMS Health Technology Assessment* 13:Doc02

Shahriari M, Mohammadi E, Abbaszadeh A, Bahrami M (2013) Nursing ethical values and definitions: A literature review. *Iranian Journal of Nursing and Midwifery Research* 18(1):1-8

Sieber JE (2009) Evidence-based ethical problem solving (EBEPS). *Perspectives on Psychological Science* 4: 26-27

Sofaer N, Strech D (2011) Reasons why post-trial access to trial drugs should, or need not be ensured to research participants: a systematic review. *Public Health Ethics* 4(2):160-184

Sofaer N, Strech D (2012) The need for systematic reviews of reasons. *Bioethics* 26(6):315-328

Sørenson K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H & (HLS-EU) Consortium Health Literacy Project European (2012) Health Literacy & Public Health: A systematic review and integration of definitions and models. *BMC Public Health* 12:80

Stoklosa A (2013) *The Concept of Evidence in Health Technology Assessment (HTA)*. Thesis. Institute for the History and Philosophy of Science and Technology, University of Toronto. https://tspace.library.utoronto.ca/bitstream/1807/69007/1/Stoklosa_Anna_201306_PhD_thesis.pdf.pdf (24.10.2019)

Stoklosa A, Bond K (2013) *Workshop on Methodology in Ethics for Health Technology Assessment: Assessing the Need For and Quality of Ethics Analyses in HTA*. https://htai.org/wp-content/uploads/2018/02/Ethics_Checklist_Workshops_-Edmonton_and_Cologne_-_FINAL_1_Apr_2014.pdf (24.10.2019)

Strech D (2008a) *Evidenz und Ethik. Kritische Analysen zur Evidenz-basierten Medizin und empirischen Ethik*. Lit Verlag, Berlin

Strech D (2008b) Evidence-based ethics – what it should be and what it shouldn't. *BMC Medical Ethics* 9:16

Strech D (2008c) Evidenz-basierte Ethik. Zwischen impliziter Normativität und unzureichender Praktikabilität. *Ethik in der Medizin* 20:274-286

Strech D, Synofzik M, Marckmann G (2008) Systematic reviews of empirical bioethics. *Journal of Medical Ethics* 34(6):472-477

Strech D (2010) How factual do we want the facts? Criteria for a critical appraisal of empirical research for use in ethics. *Journal of Medical Ethics* 36(4):222-225

Strech D, Sofaer N (2012) How to write a systematic review of reasons. *Journal of Medical Ethics* 38(2):121-126

SBU (Swedish Agency for Health Technology Assessment and Assessment of Social Services) (2017) *Assessment of Methods in Health Care. A Handbook.* Preliminary Version, August 2017 http://www.sbu.se/globalassets/eng_metodboken.pdf (20.12.2017)

Tantivess S (2014) Social and ethical analysis in health technology assessment. *Journal of the Medical Association of Thailand* 97(Suppl.5):S81-S86

Ten Have M, de Beaufort ID, Mackenbach JP, van der Heide A (2010) An overview of ethical frameworks in public health: can they be supportive in the evaluation of programs to prevent overweight? *BMC Public Health* 10:638

Tetens H (2010) *Philosophisches Argumentieren. Eine Einführung*. C.H. Beck, München

Thomson A (1999) *Critical Reasoning in Ethics. A Practical Introduction*. Routledge, London/New York

Tromp N, Baltussen R (2012) Mapping of multiple criteria for priority setting of health interventions: an aid for decision makers. *BMC health services research* 12:454.

Tyson JE, Stoll BJ (2003) Evidence-based ethics and the care and outcome of extremely premature infants. *Clinics in Perinatology* 30:363-387

Upshur REG, Colak E (2003) Argumentation and evidence. *Theoretical Medicine* 24:283-299

Van De Ven AH, Delbecq AL (1974) The effectiveness of nominal, Delphi, and interacting group decision making processes. *Academy of Management Journal* 17(4):1-17

Weinstein BD (1994) The possibility of ethical expertise. *Theoretical Medicine* 15:61-75

Winkler E (2005) The ethics of policy writing: how should hospitals deal with moral disagreement about controversial practices? *Journal of Medical Ethics* 31:559-566

World Health Organization (WHO) (2010) *Guidance on Ethics of Tuberculosis Prevention, Care and Control*. Geneva.
http://apps.who.int/iris/bitstream/10665/44452/1/9789241500531_eng.pdf (24.10.2019)

World Health Organization (WHO) (2014a) *Twelfth General Programme of Work 2014-2019. Not merely the absence of disease.* Geneva.
http://apps.who.int/iris/bitstream/10665/112792/1/GPW_2014-2019_eng.pdf?ua=1 (24.10.2019)

World Health Organization (WHO) (2014b) *Handbook for Guideline Development*. 2nd edition. Geneva.
http://apps.who.int/medicinedocs/en/d/Js22083en/ (24.10.2019)

World Health Organization (WHO) (2016) *Guidance for Managing Ethical Issues in Infectious Disease Outbreaks.* Geneva.
http://apps.who.int/iris/bitstream/10665/250580/1/9789241549837-eng.pdf?ua=1 (24.10.2019)

World Health Organization (WHO) (2017) *WHO Guidelines on Ethical Issues in Public Health Surveillance.* Geneva.
http://apps.who.int/iris/bitstream/10665/255721/1/9789241512657-eng.pdf?ua=1 (24.10.2019)

World Health Organization (WHO) Evaluation Office (2017) *Evaluation of WHO's Normative Function. Volume 1: Evaluation Report*. Geneva.
http://www.who.int/about/evaluation/who_normative_function_report_july2017.pdf (24.10.2019)

Yoder SD (1998) The nature of ethical expertise. *The Hastings Center Report* 28(6):11-19

Young K, Ashby D, Boaz A, Grayson L (2002) Social science and the evidence-based policy movement. *Social Policy and Society* 1(3):215-224.

Zwijsen SA, Niemeijer AR, Hertogh CM (2011) Ethics of using assistive technology in the care for community-dwelling elderly people: an overview of the literature. *Aging & Mental Health* 15(4):419-427

8.   A PPENDIX

Appendix A: Overview of Organizations/Institutions Screened

Overview of Organizations/Institutions per Country

The following institutions were screened in each country. The websites of some of the institutions that were identified could not be accessed due to language barriers or could not be found online despite being mentioned in several documents. Those documents are marked as not accessible (NA) in the list below.

African Region

Botswana
- Ministry of Health
- Ministry of Health Research and Development Committee
- University of Botswana
- Botswana Health Professions Council

Namibia
- Ministry of Health
- Health Professions Council of Namibia
- Medical Association of Namibia
- Namibia Medicines Regulatory Council

South Africa
- Department of Health
- Health Professional Council of South Africa
- South African Health Technology Assessment Society (the society has no official function within the South African healthcare system)
- National Health Research Ethics Council
- Ethics Committee of the South African Medical Research Council

Region of the Americas

Canada
- Health Canada
- Public Health Agency of Canada
- National Collaborating Centres for Health Public Policy
- Canadian Medical Association (the association publishes the Clinical Practice Guidelines Infobase)
- Canadian Task Force on Preventive Health Care
- Canadian Agency for Drugs and Technology in Health
- Patented Medicine Prices Review Board
- Commission de l'éthique de la science et de la technologie
- Canadian Institutes of Health Research (particularly, the Ethics Office, Standing Committee on Ethics)

Trinidad and Tobago
- Ministry of Health
- Caribbean Health Research Council
- Caribbean Public Health Agency
- Medical Board of Trinidad and Tobago

United States of America
- U.S. Department of Health and Human Services
- Centers for Disease Control and Prevention (particularly, the Public Health Ethics Section in the Office of the Associate Director for Science)
- Substance Abuse and Mental Health Services Administration
- Office of Human Research Protection
- Agency for Health Care Research & Quality (particularly, Effective Health Care Program, Evidence-based Practice Centers)
- Technology Assessment Program (at the Agency for Health Care Research & Quality/Centers for Medicare & Medicaid Services)
- U.S. Preventive Service Task Force
- The Presidential Commission for the Study of Bioethical Issues

South-East Asia Region

India
- Ministry of Health and Family Welfare
- National Health Mission
- National Health Systems Resource Centre
- Public Health Foundation of India
- Indian Council of Medical Research
- National Bioethics Committee at the Department of Biotechnology at the Ministry of Science and Technology (partially NA)

Maldives
- Ministry of Health
- Maldives Medical and Dental Council

Thailand
- Ministry of Public Health (NA)
- National Health Security Office
- Health Systems Research Institute (partially NA)
- Health Intervention and Technology Assessment Program
- Forum for Ethical Review Committees in Thailand (NA)

European Region

Germany
- Federal Ministry of Health
- Robert Koch Institute
- German Association of Scientific Medical Professional Societies
- German Institute for Medical Documentation and Information (including, the German Agency for Health Technology Assessment)
- Institute for Quality and Efficiency in Health Care
- German Ethics Council

Sweden
- National Board of Health and Welfare
- The Public Health Agency of Sweden
- Medical Products Agency Sweden
- Swedish Agency for Health Technology Assessment and Assessment of Social Services
- The Swedish National Council on Medical Ethics

United Kingdom
- Department of Health
- Medicines & Healthcare Products Regulatory Agency
- Health Technology Assessment Program (at the Evaluation, Trials and Studies Coordinating Centre, National Institute for Health Research)
- Policy Research Programme (at the Evaluation, Trials and Studies Coordinating Centre, National Institute for Health Research)
- National Institute for Health and Clinical Excellence
- Centre for Reviews and Dissemination
- Public Health England
- National Health Service England
- All Wales Medicines Strategy Group
- Scottish Intercollegiate Guidelines Network
- Scottish Medicines Consortium
- The Nuffield Council on Bioethics

Eastern Mediterranean Region

Islamic Republic of Iran
- The Ministry of Health and Medical Education (NA)
- Department of Health Technology Assessment at Yazd University of Medical Sciences (NA)
- Health Technology Assessment in the Standardization and Tariffs Office, Ministry of Health (NA)
- Iranian Committee on Bioethics (NA)
- National Committee for Ethics in Science and Technology (NA)

Kingdom of Saudi Arabia
- The Saudi Center for Evidence Based Healthcare at the Ministry of Health
- Deputy Minister for Public Health at the Ministry of Health
- National Committee of Bioethics, King Abdulazizi City for Science & Technology

United Arab Emirates
- Ministry of Health (partially NA)
- Health Regulation Department, Dubai Health Authority
- Department of Health (Abhu Dhabi)
- Medical Research Section at the Department of Health (Abhu Dhabi)

Western Pacific Region

Commonwealth of Australia
- National Health and Medical Research Council, Australian Government
- Australian Health Ethics Committee, National Health and Medical Research Council, Australian Government
- Health Technology Assessment, Department of Health, Australian Government
- Medical Services Advisory Committee, Department of Health, Australian Government
- Adelaide Health Technology Assessment, University of Adelaide
- Centre for Applied Health Economics, Griffith University
- Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University

Japan
- Department of Health Policy and Technology Assessment, National Institute of Public Health
- Medical Information Network Distribution Service, Japan Council for Quality Health Care
- Bioethics and Biosafety Commission, Council for Science and Technology, Ministry of Education, Culture, Sports, Science and Technology (NA)

Republic of Korea (South Korea)
- Ministry of Health and Welfare
- Health Insurance Review and Assessment Service, Ministry of Health and Welfare
- Korean Association of Health Technology Assessment (NA)
- National Evidence-based Healthcare Collaborating Agency
- National Bioethics Committee

Summary of Institutional Documents

Institution	Document	Summary	Classification
National Institutions			
National Collaborating Centre for Healthy Public Policy **(NCCHPP)**	Blog Article by Michael Keeling and Olivier Bellefleur: Finding Traction in Public Health Ethics: Reflections and Practical Resources	This article reflects on the work of the NCCHPP in public health ethics. Particularly, the article reflects on the importance of and difficulties associated with developing ethical literacy – given the variety of tasks and actors in public health. Particularly, the article points to public health ethics frameworks, deliberations based on those frameworks, and case studies (of ethically challenging situations) as important ways to develop ethical literacy.	**Irrelevant**: The article discusses building ethical literacy among public health professionals (not developing ethics guidelines).
Canadian Agency for Drugs and Technology in Health **(CADTH)**	Guidelines for the Economic Evaluation of Health Technologies: Canada (Section on Equity)	The guideline discusses how to handle equity issues in conducting economic evaluations. Specifically, it recommends (a) making equity assumptions underlying the evaluation transparent (e.g., using Quality Adjusted Life Years can favour the young, as each unit of measurement is considered equally), (b) identifying the primary beneficiaries and disadvantaged groups from the intervention in question and (c) providing information on the impact (e.g., benefits, harms, and costs) of the measure on various groups and the cost-effectiveness for various groups.	**Irrelevant**: The guideline discusses the ethics of guidelines: Additional normative considerations that should be considered in economic evaluations are discussed.
National Institute for Health and Clinical Excellence **(NICE)**	Social Value Judgements – Principles for the Development of NICE Guidance	The document describes the (normative) principles that NICE should follow in designing the processes it uses to develop its guidance (recommendations) and in developing individual pieces of guidance. NICE develops only empirical guidelines in the REIGN terminology (focusing on effectiveness and cost-effectiveness data). The document discloses which (further) normative considerations will have to be considered in developing recommendations. The document delineates the fundamental ethical principles underlying NICE work (e.g., the four principles: respect for autonomy, non-maleficence, beneficence, and justice) and specifies how these principles are to be interpreted in practice (e.g., through active stakeholder engagement or particular care when recommending interventions to a specific group to avoid discrimination).	**Irrelevant**: The document discusses the ethics of guidelines: It makes transparent the normative foundation of – empirical – guideline development at NICE.
Nuffield Council on Bioethics	How does the party gather evidence? (Website)	The online text describes various ways to collect mainly non-research evidence (e.g., surveys with particular groups of people) to ensure voices of	**Limitedly relevant**: This online text discusses other types of evidence.

	Available at: http://nuffieldbioeth-ics.org/about/how-council-works/working-party-gather-evi-dence	relevant stakeholders are heard. No explicit stance is taken towards research evidence.	
National Health and Medical Research Council (**NHMRC**)	Ethical Guidelines for Organ Transplantation from Deceased Donors (Section: process report), Ethical Guidelines on the Use of Assisted Reproductive Technology in Clinical Practice and Research (Section: process report) and personal communication via email	The process for developing the ethics guidelines for NHMRC is set out in the NHMRC ACT of 1992 and further legislation. In accordance with this legislation, guideline development involves the following steps: (1) The establishment of formal working committees (including external experts and consumer representation, whose conflicts of interest should be identified and adequately managed) (2) The development of draft materials (3) The conduct of public and/or targeted consultations (4) The review of all feedback from these consultations, including any evidence that those making submissions to the consultation wish to provide (5) The revision, as necessary, of the draft materials (6) Review and approval by the Australian Health Ethics Committee and the NHMRC Council (7) Issuing of the guideline by the CEO of the NHMRC (8) Tabling of the approved guidelines in the Australian Parliament (if necessary) (9) Publication of the guidelines on the NHMRC website	**Limitedly relevant:** While research evidence is considered a necessary part of guideline development and integrated by involving external experts from the research community, the role of evidence is not explicitly explained and reflected.
Institute for Quality and Efficiency in Health Care (**IQWiG**)	Allgemeine Methoden Version 5.0 (Section "6.5.3. Ethik")	The very short segment on ethics names various methods for analysing ethical aspects without providing any detail: • Principlism • Socratic approach [Hofmann 2005] • Social shaping of technology • Virtue ethics • Triangular approaches IQWiG favours Hofmann's Socratic approach [2005] but allows for variance in methodological approaches.	**Limitedly relevant**: This section identifies different approaches to addressing ethical aspects in health technology assessments without specifying the role of research evidence.
	Internal documents (templates) for analysis of ethical aspects in products pertaining to the rubric	The IQWiG has specified several ways to generate information on ethical aspects. Which (combination of) methods will be chosen depends on the ethical relevance (or contentiousness) of the question/technology at issue, on	**Relevant:** This document addresses what evidence is to be considered in addressing ethical aspects as part of HTA reports.

	"ThemenCheck Medizin" (not publicly available)	which the IQWiG decides internally before commissioning an HTA report. The methods are • Exploratory search of various publications (including but not limited to scientific publications) • Application of Hofmann's [2005] list of questions to the intervention at issue • Stakeholder discussion based on Hofmann's [2005] list of questions How exactly an exploratory search should be conducted is not further specified in the documents (however, the publication "Allgemeine Methoden" [IQWiG 2017] discusses in more detail how IQWiG conceptualizes an exploratory search – although not specifically for ethics); in addition, the document provides further thoughts on certain aspects, e.g., databases. Whichever combination of methods is chosen to identify ethical implications, effectiveness or cost-effectiveness studies included in the HTA should be analysed as well for additional relevant information on ethical aspects. An exploratory search is also always required. Identified ethical aspects and arguments can be presented in different ways but should be presented in the form of a table.	
Swedish Agency for Health Technology Assessment and Assessment of Social Services **(SBU)**	Assessment of methods in healthcare – a handbook (Chapter 12: Ethical and social aspects)	The SBU has developed an elaborate approach to addressing ethical and social issues (which are jointly addressed). Addressing the ethical aspects of an intervention involves three steps: (1) Identifying relevant ethical and social aspects (2) Analysing and discussing ethical and social aspects (3) Providing a summary Ethical issues are identified by the project group guided by a set of questions based on Hofmann's work [2005] and EUnetHTA and INAHTA publications (see below). The questions relate to (1) impacts on health, (2) compatibility with ethical values, (3) systematic factors, and (4) long-term ethical consequences. The analysis and discussion can (but does not have to) be supported by either or all of the below: (1) Literature search and review of studies (It is pointed out that it might sometimes prove useful to conduct a systematic literature review of ethical analyses and empirical data. If possible, the	**Relevant**: The report addresses literature reviews in ethics and the role of such reviews in HTA development. While the usefulness of literature reviews (and thereby in a certain reading the importance of research evidence) is emphasized, many questions remain unanswered (e.g., when should a literature review be implemented, how should it be implemented and how should the information be used).

		literature search should be conducted in accordance with Droste et al [2010]) (2) Gathering experiences from affected parties (3) Ethical analysis supported by an ethicist Finally, the pros and cons of implementing a certain intervention are summarized, and it is assessed whether it is possible to modify the measure/its implementation to avoid ethical problems.	
Presidential Commission for the Study of Bioethical Issues	Bioethics for every generation: Deliberation and education in health, science, and technology (Chapter 2: Democratic Deliberation in Bioethics)	The report sets out to describe the working mode of the Presidential Commission for the Study of Bioethics Issue (the Commission). It is argued that policy-making on bioethical topics in general, due to their complexity and contentiousness, should be guided by democratic deliberation. Democratic deliberation implies that the process for arriving at a recommendation is inclusive with regard to various viewpoints and the recommendation is reached in collaboratively (consensus building). Deliberation is not simply discussion, as deliberation aims to arrive at a shared policy (not just at understanding an issue). It is argued that introducing democratic deliberation into policy-making processes has several benefits: (1) it improves decisions by basing them on relevant facts and reasoned judgements, (2) it fosters mutual respect between people of differing viewpoints and increases the legitimacy of decisions because of stakeholder involvement, and (3) it might help build consensus when other methods have failed (although it itself is obviously not foolproof). It is furthermore pointed out that democratic deliberation might make sense on various levels: international, national, and institutional (e.g., in hospital ethics committees). Central to democratic deliberation is mutual respect and reciprocal reasoning because the quality of the deliberation legitimizes the resulting consensus. Reciprocal reasoning is characterized by accessibility (i.e., the reasons used appeal not only to members of one particular group and are understandable to all), morality (the reasons can be and are applied to all people in the same circumstances without making exceptions), respect (other positions are taken seriously and accommodation of all viewpoints is attempted), revisable (conclusions will have to be revised after some time). It will, however, also be important who participates in deliberations. Bioethics bodies should therefore be constituted in a balanced way with regard to diverse backgrounds and disciplines. Furthermore, the Commission deliberates publicly so everybody with a stake can participate and comment.	**Relevant**: The report addresses the role of empirical (and normative) evidence in the Commission's specific framework for developing recommendations. As the framework of democratic deliberations is also almost close to how the decision-making process was conceptualized in the report, this publication is particularly insightful. Nevertheless, many questions (e.g., the role of systematic reviews) are left open.

The Commission has furthermore formulated five steps that should guide their deliberations.

(1) <u>Begin with an open question and consider distinct points of view</u>: This point particularly addresses what kinds of questions should be addressed by democratic deliberations. It is argued that these questions are complex and no clear consensus is yet reached, they involve deep disagreement, have broad public impact and require concrete action.

(2) <u>Timing of deliberation for maximum impact</u>: Deliberations should be conducted when relevant facts are established that allow reasonable deliberation and a window of opportunity for policy change has opened up as otherwise deliberations will not lead anywhere. It needs to be ensured that there is enough time to conduct deliberations. If the topic of interest is emergency management, deliberations should best be started before the emergency occurs.

(3) <u>Invite input from experts and the public</u>: **Considering the empirical evidence will be highly important.** Only recommendations based on reliable and validated evidence will be trustworthy. Establishing relevant facts is no one-time effort but will be an iterative process not only because the information base can change but also because during deliberations, participants might realize that additional information is necessary to arrive at conclusions. It might be helpful to hear a diverse set of experts to hear all relevant (also differing) reasons and come to a balanced conclusion. Normative evidence is mentioned as being relevant in deliberations, but only as best practice examples of good reasoning: "Materials also might include a reasoned argument on different sides of a question, to model for participants how to use established facts and ethics principles to form opinions with fully formed justifications" (p. 39). In choosing who should participate in the deliberations, it will be necessary to strike a good balance among including differing perspectives, maximizing efficiency, and minimizing costs.

(4) <u>Foster open discussion and debate</u>: The deliberations should be characterized by mutual respect and reason giving. It is argued that this will be important from both an instrumental and ethical perspective. Facilitators should be trained to create an atmosphere where those values are nurtured.

		(5) Develop detailed, actionable recommendations: This will be important to ensure the recommendations are actually fed back into and considered in the policy-making process	
		The report ends with three recommendations for public policy making in circumstances of complex ethical concern.	
		(1) Guide bioethics policy decisions with democratic deliberation	
		(2) Conduct deliberative activities in ways conducive to mutual respect and reason-giving among participants in accordance with best practice	
		(3) To further the practical contribution of deliberation in bioethics, conduct additional research on the effectiveness of deliberative methods	

International Organizations and Research Projects			
European Network for Health Technology Assessment (**EUnetHTA**)	HTA Core Model Version 3.0 (Section "Ethical Analysis")	The HTA Core Model describes how ethical aspects (of the technology itself and of conducting the HTA) should be addressed in an HTA. In doing so, the Core Model stipulates that two interconnected steps are necessary for addressing ethical aspects: (1) identify ethical/moral issues for which a list of questions/issues and corresponding assessment element cards are provided and (2) conduct an ethical analysis to answer the relevant issues identified.	**Relevant**: The report discusses how to arrive at ethics recommendations in the context of an HTA and what role (research) evidence should play in this. However, many questions are unanswered (e.g., when to conduct a systematic review and how this information is to be considered in the different methods of ethical analysis). Complicating fully appreciating the approach are further open questions unrelated to the role of evidence. (For example, the Socratic approach identifies questions of its own to identify ethical issues: In what relation do the questions identified by the core model and the Socratic approach stand?)
		Conducting an ethical analysis "consist[s] of using structured methods to expose the relevant, often competing, moral values in the HTA, and to weigh their relative merits" (p. 257). Various methods are identified for conducting ethical analysis. It is assumed that combining various methods will often be necessary:	
		(1) Casuistry	
		(2) Coherence analysis	
		(3) Interactive, participatory HTA approach	
		(4) Principlism	
		(5) Social shaping of technology	
		(6) Wide reflective equilibrium	
		(7) The 'triangular model' based on the human person-centred approach	
		(8) Axiological (Socratic) approach	
		The choice of method of ethical analysis will depend on various factors: most importantly, the technology being evaluated, the role and authority of the	

REIGN

Appendix

		HTA organization, the time and resources available and the methodological expertise available.	
		Ethical analysis should be informed by evidence. Evidence can be collected via various strategies:	
		(1) Systematic literature searching (What kind of information should be searched is not clearly explained, but the answer seems to be normative and empirical literature. Either way, it is argued that the literature sources that will have to searched must be broader than reviews targeting effectiveness or cost data. Droste et al [2003, 2010] is quoted as providing guidance with regard to searching the literature in ethical contexts.) (2) Professional guidelines (3) Expert opinion (4) Patient/service user opinion (5) Views of organizational stakeholders	
International Network of Agencies for Health Technology Assessment (**INAHTA)**	INAHTA's Working Group on Handling Ethical Issues – Final Report	This work sets out to more broadly discuss how to handle ethical issues in health technology assessments. It accordingly spends considerable time reflecting on the various roles ethics can have in HTAs: (1) Reflection of the values underlying the choice of question and process of conducting an HTA (e.g., what endpoints are chosen) (2) Reflection on the ethical implications of implementing and developing the technology (3) Stakeholder analysis to identify stakeholder values The focus will be on the second role, as this role comes closest to what happens in ethics guideline development. With regard to this role, the document also stipulates (implicitly) a two-step process, in which first at the topic refinement phase, ethical issues to be addressed by the HTA will have to be identified. Hofmann's [2005] list of questions (but possibly other approaches) will be useful in identifying relevant ethical issues. The technology in question might be identified as "business as usual" when no explicit ethical analysis is necessary, but recommendations can be given based on established measures (e.g., effectiveness) that are legitimized by an assumed normative consensus [Grunwald 2004]. An explicit ethical analysis will be necessary only in cases of "moral conflict". If an ethical analysis has to be conducted, there are several methods to do so, and none is prescribed (as the choice will depend on the particular	**Relevant**: The report discusses how to arrive at ethics recommendations in the context of an HTA and what role (research) evidence should play in this. However, the reflections on the role of (research) evidence are very limited, provide only marginal guidance, and leave many questions unanswered (including what kind of information is needed and what kind of information should accordingly be searched and collated systematically).

| | | organization's role in the healthcare system and stance on HTA). Either way, the agency will have the choice to just describe the relevant dimensions of the ethical issue or to also try to provide recommendations, which should be based on participatory approaches. Principlism [Beauchamp/Childress 2009] is mentioned several times as a promising approach to analysing ethical issues. It is preferred over some other normative doctrines (e.g., utilitarianism or Christianity) because it rests on commonly accepted mid-level principles, but there might be other helpful methods (e.g., virtue-based ethics). Irrespective of the method chosen, it will also be important to have relevant (empirical) context information to conduct the ethical analysis.

The ethical analysis should be supported by a literature search strategy (and possibly by content experts within the project team). "A literature search strategy of relevant sources will identify primary studies and reviews on ethical aspects for the specific technology under assessment and related technologies. The study findings, derived from a qualitative analysis of the relevant ethical issues and resultant policy implications would need to be phrased within the context of the health care system" (p.15). | |
| Health Technology Assessment International **(HTAi)** | Different resources, mostly slides from conferences/meetings: https://www.htai.org/interest-groups/ethics/ethics-resources/

Particularly interesting:
• Bond, Ken: Appraising the primary ethic literature
• Bond, Ken: Introduction to Ethics in Health Technology Assessment

Available at:
https://www.htai.org/filead-min/HTAi_Files/ISG/Ethics/ Bond_HTAi2017_Engag-ing_with_ethics_lit_June2017.pdf | Both presentations discuss which literature will be needed to answer ethical questions regarding development/implementation of technologies. These presentations emphasize that both descriptive (qualitative and quantitative) and normative studies will be informative and needed. Quality appraisal for both types of literature with a focus on qualitative literature is addressed. For normative literature, certain arguments (historic facts, majority opinions, justifying actions ethically just because they are permitted by law, mere opinion, biologic truths, relativistic reasoning ("there is no right or wrong answer") are in accordance with McCullough et al [2004] disqualified. A publication discussing quality assessment in ethics analysis is mentioned [Scott et al 2016]. It is pointed out that identifying relevant empirical literature for ethical analysis is difficult because, among other things, authors do not self-identify as addressing ethical issues. Databases for searching relevant publications are listed [Droste et al 2010]. | **Limitedly relevant** |

	https://www.htai.org/fileadmin/HTAi_Files/ISG/Ethics/HTAi-Oslo2015-Bond_Critical_appraisal_of_primary_literature_FINAL.pdf		
	• Scott, Anna Mae & Sacchini, Dario: Reporting on ethics in HTA: Methods, results and interpretation Available at: https://www.htai.org/fileadmin/HTAi_Files/ISG/Ethics/AMS_DaS_-_HTAi_2017_-_Ethics_workshop_-_16_June_2017.pdf)	The talk sets out to discuss why and for whom guidelines for ethics integration in HTA would be beneficial. The talk points out that methods have been developed to consider ethical issues systematically [Assasi et al 2014]. Examples of systematic reviews on normative questions are provided [Droste et al 2011, McCullough 2007] and tips given on how to report and interpret the results of such reviews.	**Limitedly relevant:** Although limitedly relevance, this talk provides helpful references.
	• Stoklosa, Anna & Bond, Ken: Workshop on Methodology in Ethics for Health Technology Assessment: Assessing the Need For and Quality of Ethics Analyses in HTA Available at: https://www.htai.org/fileadmin/HTAi_Files/ISG/Ethics/Ethics_Checklist_Workshops_-Edmonton_and_Cologne_-_FINAL_1_Apr_2014.pdf	The documents provide a rough summary of discussions during two workshops (held consecutively in Canada and Germany). The discussions focused on two topics: (1) Tools for <u>assessing the need for an ethical analysis</u> as part of an HTA. As this need is already established when ethics guidelines are to be developed, the summary of this topic will be kept to a minimum. However, it should be remembered that these tools can also possibly be used to structure ethical analysis or help identify ethical issues (as pointed out by the workshop participants). The tool that features most prominently consists of checklists. Several checklists used by various institutions (SBU in Sweden, HTA in France, IQWiG in Germany, and OSTEBA in Spain) are introduced. One checklist that has been used as orientation point for institutional approaches is Hofmann's list [2005]. However, some institutions use much shorter lists. HAS, for example, checks three things to identify a need for an ethical analysis: specific features of the technology, conflicts between the technology and fundamental rights, and whether (public) controversies around the technology exist. It is unclear what components should be included in these checklists. How the following ethical analysis (if triggered) is conducted is also quite different across institutions. HAS, for example, structures the ethical analysis as follows: ethics analysis consists of identifying ethical arguments through	**(Partially) relevant**: The discussions surrounding quality of ethical arguments/analyses are particularly helpful for developing our framework.

literature review and discussion among stakeholders, presenting the ethical arguments in the review (often by using a principlism framework [Beauchamp & Childress 2009]) and identifying the main ethical disagreements. Various additional questions were also addressed (but not always answered), such as whether the questions should be open ended, how the questions are to be answered (e.g., should they be answered by conducting a literature review) or whether there should be different checklists for different contexts.

(2) Critical (quality) appraisal in/of ethics analyses: First, it was pointed out that one probably needs different criteria of quality if one evaluates ethical arguments, the ethical literature (all the arguments) or the whole ethics analysis. Unfortunately, it was not always clear what participants were addressing, and the following should therefore be read as a (somewhat subjective) reconstruction.

Participants agreed that one will have to refer to empirical and normative papers to conduct an ethical analysis. However, it will not be always clear to which group a paper belongs, and it was therefore argued that it will make more sense to group items into those that are more descriptive, more normative/prescriptive or both. Either way, one will need different criteria to assess the quality of empirical or normative papers/arguments. As quality criteria are established for empirical literature, the discussion centred on normative literature. For normative arguments/literature, two quality criteria are introduced and discussed:

(a) the validity/soundness of the argument according to formal and informal logic

(b) the thoroughness/comprehensiveness of the arguments identified

It is argued that checklists can probably be developed to check for the validity/soundness of the arguments. These arguments should probably refer to formal and informal logic that is well known in the philosophical discourse but not so much in the HTA world. It is discussed whether introducing such checklists would be reasonable. Advantages of such a tool would be transparency, clarity, methodological guidance, standardization of methods, and management of expectations of and improved communications with users of HTA reports. However, there is a risk that using checklists might just become a box-ticking exercise and prevent further reflection. A checklist and scoring system for the normative ethics literature is furthermore proposed. It is meant as a starting point for further debates as the tool leaves open

		many questions, e.g., how to operationalize "reasonableness" (of premises and conclusions), which is used as a criterion.	
		To ensure that all (relevant) arguments are captured (thus satisfying comprehensiveness), a systematic review can be conducted, and a checklist using, e.g., the four principles of Beauchamp & Childress [2009], can be used to check whether relevant aspects are considered and/or a public consultation/stakeholder interviews can be conducted. With regard to systematic reviews (using the methods developed by Strech and Sofaer [2011] as a blueprint), it is argued that one should be wary of bias (especially if focusing on English literature). In these cases, additional public consultation can be more important. Further challenges with this approach are identified: What training do reviewers need to conduct such a review? Is a systematic review feasible (in terms of time and resources)? It is also pointed out that Strech and Sofaer's approach does not include a clear approach to quality appraisal of the single argument. A systematic review of normative literature can also provide only the "raw data" for further ethical analysis and thus does not serve the same function as systematic reviews on scientific evidence.	
		It is furthermore pointed out that checking for the validity/soundness of an argument should not be confused with checking the "ethical relevance and force of the argument" (these two activities are not the same). When the quality of the whole ethical analysis is considered, the relevance and force of the arguments will also have to be considered. It is furthermore argued that the ethical analysis will have to be context sensitive. Which norms/theoretical lens for ethical analysis should be adopted is furthermore debated.	
		Lastly, reporting guidelines for ethics analyses are discussed.	

Table 14: Summary and evaluation of relevance of institutional documents

Appendix B: Overview Academic Discourses Screened

Scoping Review Strategy

For the scoping reviews, due to time constraints and workload, only the databases BELIT, EthicsWeb, and PubMed were searched. The literature found was first screened for relevance by title and then abstract level; sometimes, the full text was also briefly assessed, if accessible. If a paper was found obviously irrelevant in one of these two (or three) steps, it was excluded. Generally, as with the screening strategy for institutions/organizations (see Appendix A), the strategy tended to be overinclusive rather than too narrow in its selection.

Search String and Hits Found/Considered Relevant

Evidence-based Ethics/Evidence and Ethics

Database:	**BELIT**
Search string:	title:evidenz* AND (title:ethi* OR title:bioethi*)

Database:	**EthicsWeb**
Search string:	title:(ethic* OR bioethic* AND proof) OR title:("evidence based ethics") OR title:("evidence based bioethics")

Database:	**PubMed**
Search string:	((ethic*[TI] OR bioethic*[TI]) AND proof[TI]) OR evidence based ethic*[TI] OR evidence based bioethic*[TI]
Sort by:	Best Match

Results total:	73
Considered relevant total:	**17**

Guideline Development in Ethics/for Ethical Topics

Database:	**BELIT**
Search string:	#1: title:Guideline* AND (title:polic* OR title:ethic* OR title:bioethic*) AND (title:establish* OR title:develop*) #2 (title:Ethik OR title:Bioethik) AND title:Leitlinie*

Database:	**EthicsWeb**
Search string:	title:(Guideline*) AND (title:(Ethic*) OR title:(Polic*)) AND title:(establish* OR develop*)

Database:	**PubMed**
Search string:	(((Guideline*[TI] AND (ethic*[TI] OR bioethic*[TI])) OR (("Guidelines as Topic/standards"[MESH]) AND "Ethics"[MESH])) OR ((Guideline*[TI] AND polic*[TI]) OR (("Guidelines as Topic/standards"[MESH]) AND "Policy"[MESH])) AND (establish*[TI] OR develop*[TI])) NOT ("clinical practice"[TI] OR "practice guidelines"[TI])

Results total:	186
Considered relevant total:	**18**

Integrating Ethical Issues in Guidelines or HTA Reports

Database:	**BELIT**
Search string:	title:("Leitlinien") AND (title:ethi* OR title:bioethi*)

Database:	**EthicsWeb #1**
Search string:	#1: title:("clinical guidelines" OR "clinical guideline") AND title:(bioethic* OR ethic*) #2: title:("medical guidelines" OR "medical guideline") AND title:(bioethic* OR ethic*) #3: title:("health technology assessment" OR "health technology assessments" OR hta) AND title:(bioethic* OR ethic*) #4: title:("professional guideline" OR "professional guidelines") AND title:(ethic* OR bioethic*)

Database:	**PubMed**

Search string:	(medical guideline*[TI] OR clinical guideline*[TI] OR CPG[TI] OR clinical practice guideline*[TI] OR health technology assessment*[TI] OR HTA[TI] OR health technolog*[TI] OR healthcare technolog*[TI] OR practice guideline*[TI] OR professional guideline*[TI]) AND (ethic*[TI] OR bioethic*[TI]) OR ("Biomedical Technology/ethics"[MESH] AND "Ethical Analysis/methods"[MESH])

Results total:	297
Considered relevant total:	**54**

Methods/Concepts for (Systematic) Reviews in Ethics

Database:	**PubMed**
Search string:	((("Review Literature as Topic"[MAJR] OR systematic review[TI]) AND Ethics[MAJR]) OR (("Review Literature as Topic"[MESH] OR "Periodicals as Topic"[MESH]) AND "Ethical Analysis"[MESH])) NOT "Conflict of Interest"[Mesh:noexp]

Results total:	96
Considered relevant total:	**11**

Examples of Relevant Literature in the Four Academic Discourses

Evidence-based ethics/Evidence and ethics
Bonneux L (2007) From evidence based bioethics to evidence based social policies. *European Journal of Epidemiology* 22:483-485
Frize M, Walker RC, Ennett CM (2003) Development of an Evidence-Based Ethical Decision-Making Tool for Neonatal Intensive Care Medicine. *Proceedings of the 25th Annual International Conference of the IEEE EMBS Cancun, Mexico, September 17-21, 2003*
Goldenberg MJ (2005) Evidence-based ethics? On evidence-based practice and the "empirical turn" from normative bioethics. *BMC Medical Ethics* 6:11
Halpern SD (2005) Towards evidence based bioethics. *British Medical Journal* 331:901-903
Jansen R (1997) Evidence-based ethics and the regulation of reproduction. *Human Reproduction* 12:2068-2075
Kim SY (2004) Evidence-based ethics for neurology and psychiatry research. *NeuroRx* 1:372-377
Major-Kincade TL, Tyson JE, Kennedy KA (2001) Training pediatric house staff in evidence-based ethics: An exploratory controlled trial. *Journal of Perinatology* 21:161-166
Mertz M (2011) *Evidenzbasierte Klinische Ethik. Philosophische Untersuchungen zur Verwendung von Empirie und Evidenz in der (Medizin-)Ethik.* GRIN Verlag, München/Ravensburg
Orešković S (2016) Evidence-based bioethics: delineating the connections between science, practice, and values in medicine. *Croatian Medical Journal* 57:307-310
Petrini C (2008) From evidence based bioethics to evidence based social policies. *European Journal of Epidemiology* 23:75
Reiter-Theil S, Mertz M, Schürmann J, Stingelin N, Meyer-Zehnder B (2011) Evidence – competence – discourse: the theoretical framework of the multi-centre clinical ethics support project METAP. *Bioethics* 25(7):403-412
Roberts LW (2000) Evidence-based ethics and informed consent in mental illness research. *Archives of General Psychiatry* 57:540-542
Rubenfeld GD, Elliott M (2005) Evidence-based ethics? *Current Opinion in Critical Care* 11:598-599
Salloch S (2012) "Evidenzbasierte Ethik"? Über hypothetische und kategorische Handlungsnormen in der Medizin. *Ethik in der Medizin* 24:5-17
Sieber JE (2009) Evidence-based ethical problem solving (EBEPS). *Perspectives on Psychological Science* 4: 26-27
Strech D (2008) Evidence-based ethics – what it should be and what it shouldn't. *BMC Medical Ethics* 9:16
Strech D (2008) *Evidenz und Ethik. Kritische Analysen zur Evidenz-basierten Medizin und empirischen Ethik*. Lit Verlag, Berlin
Strech D (2008) Evidenz-basierte Ethik. Zwischen impliziter Normativität und unzureichender Praktikabilität. *Ethik in der Medizin* 20:274-286
Tyson JE (1995) Evidence-based ethics and the care of premature infants. *Future Child* 5(1):197-213
Tyson JE, Stoll BJ (2003) Evidence-based ethics and the care and outcome of extremely premature infants. *Clinics in Perinatology* 30:363-387

Guideline development in ethics/for ethical topics
Cho MK (2014) Ethics and empiricism in the formation of professional guidelines. Editorial. *American Journal of Bioethics* 14:1-2

Hashemi A, Yeketaz H, Asghari F (2015) Developing an ethical guideline for clinical teaching in Tehran University of Medical Sciences. *Journal of Medical Ethics and History of Medicine* 8:5

Jox R (2014) Entwicklung einer ethisch-rechtlichen Klinik-Policy. *Hessisches Ärzteblatt* 5:268-281

Kangasniemi M, Arala K, Becker E, Suutarla A, Haapa T, Korhonen A (2017) The development of ethical guidelines for nurses' collegiality using the Delphi method. *Nursing Ethics* 24:538-555

Mertz M (2011) *Evidenzbasierte Klinische Ethik. Philosophische Untersuchungen zur Verwendung von Empirie und Evidenz in der (Medizin-)Ethik*. GRIN Verlag, München/Ravensburg

Neitzke G, Riedel A, Brombacher L, Heinemann W, Herrmann B (2015) Empfehlungen zur Erstellung von Ethik-Leitlinien in Einrichtungen des Gesundheitswesens. *Ethik in der Medizin* 27(3):241-248

Reiter-Theil S, Mertz M, Albisser Schleger H, Meyer-Zehnder B, Kressig RW, Pargger H (2011) Klinische Ethik als Partnerschaft – oder wie eine Leitlinie für den patientengerechten Einsatz von Ressourcen entwickelt und implementiert werden kann. *Ethik in der Medizin* 23(2):93-105

Riedel A (2017) Ethik-Leitlinien als Verfahren der Ethikberatung – Stellenwert und Beitrag zur ethischen Reflexion und Entscheidungsfindung in der Behindertenhilfe. *EthikJournal* 4:1 http://www.ethikjournal.de/fileadmin/user_upload/ethikjournal/Texte_Ausgabe_1_06-2017/Riedel_Ethik-Leitlinien_als_Verfahren_der_Ethikberatung_EthikJournal_4_2017_1.pdf (19.10.2017)

Winkler E (2005) The ethics of policy writing: how should hospitals deal with moral disagreement about controversial practices? *Journal of Medical Ethics* 31:559-566

Integrating ethical issues in guidelines or HTA reports

Assasi N, Schwartz L, Tarride J-E, Campbell K, Goeree R (2014) Methodological guidance documents for evaluation of ethical considerations in health technology assessment: a systematic review. *Expert Review of Pharmacoeconomics & Outcomes Research* 14(2):203-220

Assasi N, Tarride JE, O'Reilley D, Schwartz L (2016) Steps toward improving ethical evaluation in health technology assessment: a proposed framework. *BMC Medical Ethics* 17:34

Autti-Rämö I, Mäkelä M (2007) Ethical evaluation in health technology assessment reports: An eclectic approach. *International Journal of Technology Assessment in Health Care* 23:1-8

Bombard Y, Abelson J, Simeonov D, Gauvin FP (2011) Eliciting ethical and social values in health technology assessment: A participatory approach. *Social Science & Medicine* 73:135-144

Burls A, Caron L, Cleret de Langavant G, Dondorp W, Harstall C, Pathak-Sen E, Hofmann B (2011) Tackling ethical issues in health technology assessment: A proposed framework. *International Journal of Technology Assessment in Health Care* 27:230-237

Daniels N, van der Wilt GJ (2016) Health technology assessment, deliberative process, and ethically contested issues. *International Journal of Technology Assessment in Health Care* 32(1/2):10–15

Droste S, Gerhardus A, Kollek R (2003) Methoden zur Erfassung ethischer Aspekte und gesellschaftlicher Wertvorstellungen in Kurz-HTA Berichten – eine international Bestandsaufnahme. DIMDI: Köln (http://portal.dimdi.de/de/hta/hta_berichte/hta62_bericht_pub_de.pdf)

Duthie K, Bond K (2011) Improving ethics analysis in health technology assessment. *International Journal of Technology Assessment in Health Care* 27(1):64-70

Heintz E, Lintamo L, Hultcrantz M, Jacobson S, Levi R, Munthe C, Tranæus S, Östlund P, Sandman L (2015) Framework for systematic identification of ethical aspects of healthcare technologies: the SBU approach. *International Journal of Technology Assessment in Health Care* 31(3):124-130

Hofmann B (2005) Toward a procedure for integrating moral issues in health technology assessment. *International Journal of Technology Assessment in Health Care* 21:312-318

Hofmann B, Droste S, Oortwijn W, Cleemput I, Sacchini D (2014) Harmonization of ethics in health technology assessment: A revision of the Socratic approach. *International Journal of Technology Assessment in Health Care* 30:3-9

Hofmann B, Lysdahl KB, Droste S (2015) Evaluation of ethical aspects in health technology assessment: more methods than applications? *Expert Review of Pharmacoeconomics & Outcomes* Research 15:5-7

Lehoux P, Williams-Jones B (2007) Mapping the integration of social and ethical issues in health technology assessment. *International Journal of Technology Assessment in Health Care* 23:9-16

Lysdahl KB, Mozygemba K, Burns J, Chilcott JB, Brönneke JB, Hofmann B (2016, eds) Guidance for assessing effectiveness, economic aspects, ethical aspects, socio-cultural aspects and legal aspects in complex technologies [Online]. Available from: http://www.integrate-hta.eu/downloads/

Lysdahl KB, Oortwijn W, van der Wilt GJ, Refolo P, Sacchini D, Mozygemba K, Gerhardus A, Brereton L, Hofmann B (2016) Ethical analysis in HTA of complex health interventions. *BMC Medical Ethics* 17:16

Mertz M, Strech D (2014) Systematic and transparent inclusion of ethical issues and recommendations in clinical practice guidelines: a six-step approach. *Implementation Science* 9:184

Mittelstadt B, Stahl B, Fairweather B (2013) PHM-Ethics and ETICA: complementary approaches to ethical assessment. *Studies in Health Technology and Informatics* 187:117-135

Saarni SI, Hofmann B, Lampe K, Lühmann D, Mäkelä M, Velasco-Garridod M, Autti-Rämöa I (2008) Ethical analysis to improve decision-making on health technologies. *Bulletin of the World Health Organization* 86:617-623

Saarni SI, Braunak-Mayer A, Hofmann B, Jan G (2011) Different methods for ethical analysis in health technology assessment: an empirical study. *International Journal of Technology Assessment in Health Care* 27:305-312

Samuli SI, Hofmann B, Lampe K, Lühmann D, Mäkelä M, Velasco-Garridod M, Autti-Rämöa I (2008)

Sacchini D, Virdis A, Refolo P, Pennacchini M, de Paula IC (2009) Health technology assessment (HTA): ethical aspects. *Medicine, Health Care and Philosophy* 12:453-457

Sandman L, Heintz E (2014) Assessment vs. appraisal of ethical aspects of health technology assessment: can the distinction be upheld? *GMS Health Technology Assessment* 10:Doc05

Scott AM, Bond K, Gutiérrez-Ibarluzea I, Hofmann B, Sandman L (2016) Quality assessment of ethics analyses for health technology assessment. *International Journal of Technology Assessment in Health Care* 32:362-369

Scott AM, Hofmann B, Gutiérrez-Ibarluzea I, Lysdahl KB, Sandman L, Bombard Y (2017) Q-SEA – a tool for quality assessment of ethics analyses conducted as part of health technology assessments. *GMS Health Technology Assessment* 13:Doc02

Tantivess S (2014) Social and ethical analysis in health technology assessment. *Journal of the Medical Association of Thailand* 97: S81-S86

Wahlster P, Brereton l, Burns J, Hofmann B, Mozygemba K, Oortwijn W, Pfadenhauer l, Polus S, Rehfuess E, Schillin I, van Hoorn R, van der Wwilt GJ, Baltussen R, Gerhardus A (2016) Guidance on the integrated assessment of complex health technologies – The INTEGRATE-HTA Model [Online]. Available from: http://www.integrate-hta.eu/downloads/

Methods/concepts for (systematic) reviews in ethics
Droste S (2008) Systematische Gewinnung von Informationen zu ethischen Aspekten in HTA-Berichten zu medizinischen Technologien bzw. Interventionen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 102:329-343
Droste S, Dintsios CM, Gerber A (2010) Information on ethical issues in health technology assessment. How and where to find them. *International Journal of Technology Assessment in Health Care* 26:441-449
McCullough LB, Coverdale JH, Chervenak FA (2004) Argument-based medical ethics: A formal tool for critically appraising the normative medical ethics literature. *American Journal of Obstetrics and Gynecology* 191:1097-1102
McCullough LB, Coverdale JH, Chervenak FA (2007) Constructing a systematic review for argument-based clinical ethics literature: the example of concealed medications. *Journal of Medicine and Philosophy* 32:65-76
McDougall R (2014) Systematic reviews in bioethics: types, challenges, and value. *Journal of Medicine and Philosophy* 9:89-97
McDougall R (2015) Reviewing literature in bioethics research: increasing rigour in non-systematic reviews. *Bioethics* 29(7):523-528
Mertz M, Sofaer N, Strech D (2014) Did we describe what you meant? Findings and methodological discussion of an empirical validation study for a systematic review of reasons. *BMC Medical Ethics* 15:69
Mertz M, Kahrass H, Strech D (2016) Current state of ethics literature synthesis: a systematic review of reviews. *BMC Medicine* 14:152
Mertz M, Strech D, Kahrass H (2017, accepted) What methods do reviews of normative ethics literature use for search, selection, analysis and synthesis? In-depth results from a systematic review of reviews. *Systematic Reviews*
Mertz M (2017) Qualitätsbewertung in systematischen Übersichtsarbeiten normativer Literatur. Eine Problemanalyse. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* OnlineFirst
Strech D, Synofzik M, Marckmann G (2008) Systematic reviews of empirical bioethics. *Journal of Medical Ethics* 34: 472-477
Sofaer N, Strech D (2012) The need for systematic reviews of reasons. *Bioethics* 26(6):315-328
Strech D, Sofaer N (2012) How to write a systematic review of reasons. *Journal of Medical Ethics* 38(2):121-126

Table 15: Examples of relevant literature in the four academic discourses

Appendix C: Systematic Reviews for Synthesizing Normative Evidence

With regard to systematic reviews for normative evidence (SRNEs), no methodological consensus has been reached that can be compared to the developments in the empirical fields (see PRISMA guidelines [Moher et al 2009] or publications by the Cochrane Collaboration [e.g., Higgins et al 2019]) because SRNEs are a rather new phenomenon with methodological discussion truly just starting to gather momentum. A recently conducted review of SRNEs by one of the collaborating researchers on this discussion paper showed that until 2015, only 84 SRNEs (or more precisely, SRs of normative literature) had been published [Mertz/Kahrass/Strech 2016]. Accordingly, it is impossible to offer more than an overview of some methodological issues that arise in SRNEs and to do more than point out where SRNEs might (have to) diverge from established systematic review methodology for collecting and synthesizing empirical research. In doing so, no comprehensiveness is claimed, and it is acknowledged that it is a somewhat subjective reconstruction of the state of the debate. By going through the different steps of a systematic review, those points of deviation and contention that appear most critical to the authors are highlighted:

Searching for Relevant Literature

Limits of PICO Strategies

Those conducting systematic reviews of empirical questions are often urged to use the PICO framework to formulate search questions. PICO stands for *patient/population, intervention, comparator, and outcome*, thus specifying the components of the question that need to be explicitly defined [Higgins/Green 2008]. The PICO framework is designed to serve the needs of systematic reviews focusing on intervention effectiveness, and it might therefore not necessarily work well in a normative-ethical context. One of the reasons is that PICO is aimed at questions regarding *causal* relationships (e.g., effectiveness) that can be researched by experimental designs, while the questions of interest in normative-ethical context are *conceptual* in nature (e.g., the latter questions target values and principles, concepts, ethical issues or arguments). Some SRNEs, however, have employed the PICO framework; for example, McCullough et al [2007] posed the following question: In patients with mental disorders (schizophrenia, dementia, etc.) (P), is concealing medication in food or drink (I) rather than prescribing medications in the usual way or forcibly administering them (C) ethically justifiable (O)?

"Ethics plus Context" Strategies

Questions addressed by SRNEs often seem to consist of two things: (a) a specification of the normative information unit of interest (e.g., ethical issues, arguments, frameworks, or recommendations) and (b) a specification of the context or topic of interest (e.g., post-trial access to drugs, patient safety research, or authorship in multicentre trials). An example would be the following question: What is the nature of ethical issues experienced in nursing homes [Preshaw et al 2015]? A similar approach was proposed by Droste et al [2008, 2010], who, however, divided the context/topic into

two parts that might have to be specified as follows: (a) the problem (for example, breast cancer) and (b) the intervention (for example, stem cell transplantation). Whether such an "ethics plus context" ("EPC") approach, the PICO framework or another framework for question formulation does most justice to the specificities of SRNEs will need further academic attention; additionally, the suitable format for the search strategy might depend on the research interest or the guideline topic. The authors would, however, argue that there are only a few topics where an adaptation of the PICO framework makes sense. The framework will not be applicable in the case of most topics in ethics, and researchers are well advised to use an "ethics plus context" approach instead.

Need for Broad Strategies

Generally, it is better to use broader strategies than narrower ones, as in ethics – currently – not much standardization regarding the search for normative terms has been established, and even less so regarding "study types". It is impossible, for example, to search only for "argumentative" papers (i.e., normative literature) by using adequate search terms or search restrictions, as is the case regarding, e.g., RCTs or cohort studies in the biomedical realm. Normative literature can often be identified and distinguished from empirical literature only when screening publications (although a possible strategy is to use ethics and philosophy databases and search engines, as these, in general, do not index empirical literature; see also below). Furthermore, as discussed in chapter 3, normative evidence can also be contained in empirical literature, but which publications are of interest can most often be decided only when screening titles and abstracts (or even when screening full texts). Additionally, ethical issues, arguments or values and principles are often formulated on a more general level and are much more transferable to different specific cases, as is the case with, e.g., the results of effectiveness studies. Therefore, one may not find any (or any useable) literature when searching too narrowly. For example, searching ethical issues in palliative breast cancer therapy using mistletoe extracts will result in barely any literature; therefore, one should consider implementing a broader search strategy for ethical issues of complementary medicine in the context of breast cancer therapy, or even of ethical issues of complementary medicine in general, irrespective of the kind of therapy or disease.

Searching for Publications

When searching for publications on ethical issues, it will be necessary to search databases and journals that are less or not important in classical review endeavours and therefore less well known; however, most SRNEs use *PubMed* or *MEDLINE* [Mertz/Strech/Kahrass 2017] as at least one of several databases, thereby indicating the usefulness of this database for SRNEs. Particularly important will be databases that focus on ethics, philosophy, or social science. Droste [2008, 2010] and Mertz/Strech/Krahrass [2017] provide an overview of potentially interesting databases that those conducting SRNEs might find helpful. The *Kennedy Institute of Ethics at Georgetown University* is an additional source of information on relevant databases with an ethics focus (see https://bioethics.georgetown.edu/explore-bioethics/online-bioethics-resources/). In addition to disciplinary or interdisciplinary databases or search engines, *Google Scholar* and *Google Books*, or *Google Search*

itself, can be used as *additional* sources for SRNEs, although limitations caused by their opaque and personalized search algorithms must be considered [Piasecki/Waligora/Dranseika 2018].

In her articles, Droste [2008, 2010] provides ample helpful advice on how to search, including advice on identifying synonyms and thesaurus terms for the ethical search component. She also points out that MEDLINE has created a bioethics subset (usable via "bioethics[sb]") that was developed to help identify bioethics articles and can potentially facilitate building search strategies [Droste 2008]. It might quite obviously be used for the "ethics" part in the "ethics plus context" strategy. However, the massive number of terms reflected in this subset can lead to many hits, with most being irrelevant.

Many characteristics one is interested in when searching literature can be identified only when selecting publications. This implies that the selection step might be even more important in SRNEs than in traditional SRs but also more time-consuming, as more time has to be invested to decide if a publication is useful or irrelevant.

Selecting Publications

Definition of the Normative Information Unit

A prerequisite for selecting publications is a clear understanding or definition of the normative information unit of interest (e.g., ethical issue, argument or principle); otherwise, making inclusion/extraction decisions will be problematic. Because of a lack of resources, guidance cannot be provided for all potential units of interests (however, see table 9 for some possible information units), and the focus will instead be on one that has received considerable attention in the literature and that might be more contentious than others: ethical issues (also referred to as problems, challenges, conflicts, or dilemmas).

In the literature, ethical issues have often been defined using principlism [Beauchamp/Childress 2009]. According to principlism, four *prima facie* binding moral norms (meaning binding as long as they do not conflict with other relevant obligations) guide decision-making in the healthcare context: (a) respect for autonomy, (b) beneficence, (c) non-maleficence, and (d) justice. These principles or norms are then used by various researchers to define and identify ethical issues [see Caplan/Hoffecker/Prochazka 2008; Chung/Pushman/Bellfi 2009; Seitzer et al 2016]. Seitzer et al, for example, assume an ethical issue arises "(a) because of inadequate consideration of one or more (contextualised) ethical principles (for example: insufficient consideration of patient preferences in ALS care decisions) or (b) because of conflicts between two or more (contextualised) ethical principles (for example, balancing benefits, harms, and respect for autonomy in decision making for or against mechanical ventilation)" [Seitzer et al 2016, p. 202].

Principlism does not necessarily have to be the theoretical underpinning, however. Other theoretical approaches (e.g., consequentialism) can also be used (e.g., in a certain reading of

consequentialism, ethical issues can be defined as arising when a person is harmed in any substantial way). However, in the context of the WHO's work, a more pluralist approach to defining ethical issues would probably seem more warranted, e.g., the Socratic approach of Hofmann for HTA purposes [Hofmann 2005; Droste et al 2011] can be adapted to the context of ethics guideline development. Such an approach would not favour one particular theoretical framework over another but would include various theoretical lenses. Researchers might now be inclined to argue that an intuitive approach – basically embracing a "you know an ethical issue when you see it" attitude that renounces the necessity of defining ethical issues – is most inclusive. However, such an approach should not be taken because it is problematic for various reasons. A "you know it when you see it" attitude is detrimental to being systematic and will make the conducted review less reproducible. Most likely, this approach would lead to researchers simply applying their particular framework (e.g., principlism or utilitarianism) instead of one that is prespecified without being transparent. This approach thereby probably increases the risk of relevant issues going unnoticed. A pluralist approach would simply be very inclusive of relevant principles that could ground a claim for an ethical issue to arise and should, as stated already, probably be favoured at least in the context of the WHO's work.

Inclusion and Exclusion Criteria

Apart from a clear definition of ethical issues (to stick with the example), also in SRNEs, clear inclusion and exclusion criteria need to be defined, as in every systematic review. Obviously, the inclusion/exclusion criteria will vary by subject of interest. SRNEs will have in common that they have to somehow introduce a "filter" to ensure publications are truly discussing the normative information unit of interest. For the first-order inclusion decision based on title/abstract screening, generally two ways are open: (a) include when the authors of a publication *self-identify* the publication as discussing ethical issues (as, e.g., Whicher et al [2014] writing on patient safety research put it: "Articles were retained if they were related to safety or quality improvement and the abstract included the term ethical issues or ethics issues or referenced specific ethical issues") or (b) include when the researchers conducting the SRNE understand from the title/abstract that the publication discusses ethical issues in the prespecified way (e.g., according to principles), even if the authors of the publication do not explicitly say so. The latter strategy was, for example, chosen by one of the authors of this discussion paper in their review [Klingler et al 2017]. The advantage of also including publications that – from the viewpoint of the authors of the publications – do not (therefore only *implicitly*) discuss ethical issues is that more publications that contain ethically relevant information can be included. Additionally, when using this strategy, one may face the challenge that almost every article should be included because the article might address ethical issues at least as quasi "incidental" findings. Using such a strategy might therefore come at a cost of reading many full-text articles that do not bring many new insights regarding ethical issues. This might be particularly true, as publications meant to further the normative-ethical discourse are most likely to provide a comprehensive assessment of the relevant issues and arguments. Those whose writings are part of the

normative-ethical discourse will in most cases also self-identify their publications as discussing ethical issues, problems, or challenges to ensure they are being read by the intended target audience. Both strategies might be defendable; however, they should be explicitly chosen after reflecting on the advantages and disadvantages of each.

Quality Appraisal

As of yet, no detailed, let alone consented criteria or method for appraising the quality of normative literature in the context of SRNEs have been formulated [see, e.g., Sofaer/Strech 2012; McDougall 2014; Mertz 2017; Mertz/Strech/Kahrass 2017; and Stoklosa/Bond 2013 and the summary of the discussion of the HTAi workshop in table 14 in Appendix A]. Therefore, it is also impossible to provide a definitive answer on how a quality appraisal can or should be conducted (but see chapter 4c and esp. table 10 for some proposals and possible criteria).

The general discussion about quality appraisals of normative literature (in the context of SRNEs) can be split into three (related) questions [see also Scott et al 2016; Mertz 2017]: (a) is a quality appraisal necessary (especially if the aim of the SRNE is only to descriptively show what ethical arguments or issues are discussed in the literature); (b) what exactly is appraised for quality (the whole publication/study or the concrete normative information, e.g., the ethical argument or issue); and (c) which criteria and methods exactly can or should be used when a quality appraisal is conducted? It is to be expected that for different subtypes of SRNEs (see Table 9), one will have to give (slightly) different answers. For example, conducting a quality appraisal of ethical concepts will at least somehow differ from conducting a quality appraisal for publications discussing ethical issues or providing ethical arguments.

In a paper by one of the collaborating researchers, Mertz [2017] outlines three types of quality criteria and respective methods that have or could be applied, with all having advantages and disadvantages: (a) *reporting criteria* (e.g., methods should be comprehensively reported); (b) *procedural quality* securing criteria (e.g., an article should be published in a peer-reviewed journal or by a respected scientific publisher); (c) *content-related quality criteria* (e.g., no argumentative (i.e., formal or informal logical) fallacies are made in an ethical argumentation or in assessing the ethical relevance of the issues, etc.). Most often at the moment, procedural criteria are used to exclude certain publications or simply to describe the quality of the included publications. However, it is questionable whether peer review is truly a good indicator of quality. The same is true for reporting criteria that are aimed at the whole publication, not the specific information that one is interested in. Content-related criteria should preferably be used, but it will be a future task to develop practically useful tools where such criteria are (further) operationalized (see chapter 4c).

In the meantime, the least that can be done is reporting for SRNEs if a quality appraisal has been conducted or not (and arguments for not doing so), what was critically appraised, and how the appraisal was conducted.

Analysing and Synthesizing Normative information

As for selecting publications (see above), a clear definition of the normative information that is sought is also indispensable for analysis. Regularly, both definitions will – and must – be the same, although it might be possible (or even necessary) to detail or further operationalize the definition for the analysis to identify text passages in the publications where for example, ethical issues are presented or discussed. This means that the main method of analysing and, later, synthesizing is a version of *qualitative analysis* or a close reading and text interpretation (i.e., hermeneutical) method [cf. Mertz/Kahrass/Strech 2016; Mertz/Strech/Kahrass 2017]. However, sometimes quantitative methods are also used to synthesize and present the information (mainly in the form of descriptive statistics of how many times a certain ethical issue was mentioned, etc.). Nevertheless, synthesis methods are mainly qualitative.

Category-building in Analysis and Synthesis

Where qualitative methods are chosen to analyse and synthesize the evidence (and, as already said, this is most often the case), analysis and synthesis generally are closely connected and cannot be completely separated from each other.

As part of the analysis process, relevant text passages have to be clearly identified and extracted; however, that will generally be considered insufficient. Several qualitative methods can potentially be used to further analyse and synthesize the information (e.g., *thematic analysis* [Thomas/Harden 2008], *qualitative content analysis* [Schreier 2012] or, to a lesser degree, grounded theory [Charmaz 2014]). The goal of each of these methods will generally be (at least in this context) to build categories or better: a coding scheme (including higher- and lower-order categories) presenting an overview of ethical issues (or arguments, etc.) identified in publications. Referring particularly to qualitative content analysis, categories of the coding scheme can be predefined by either a normative framework (e.g., principlism) or a more process-oriented framework (e.g., the phases of public health surveillance [see Klingler et al 2017]). These categories can also be developed inductively from the material found. Furthermore, a mixed strategy can be used when the (final) categories are in part determined deductively (predefined) and in part inductively (on the basis of, e.g., the issues or arguments found in the literature).

To provide one example, the four principles (beneficence, non-maleficence, justice and respect for autonomy) can provide the prespecified categories that help sort various ethical issues found in publications (deductive category development). However, the lower-level categories (e.g., harming patients by late diagnosis of dementia as a sub-category of non-maleficence) would still have to be developed from the data (inductive category development).

For analysis and synthesis, tables or mind maps are often used; qualitative content analysis software, such as MAXQDA, can also be used to especially support the analysis process.

Final Synthesis and Stakeholder-Orientation

For the final synthesis, the categories are often again refined (e.g., some categories are eliminated due to being redundant, or some information is subsumed under a new category because it became clear that the original subsumption was not convincing). Additionally, it is possible to introduce further characteristics (possibly as higher order categories) for the issues or arguments identified, e.g., on which level of decision-making in the health care system they are located (micro, meso or macro level) [e.g., Kahrass/Strech/Mertz 2017]. A leading ideal of the synthesis process should be what can be called "stakeholder orientation" [Mertz/Strech/Kahrass 2017]. This means that the way the normative information found is condensed and presented should be oriented to the intended target audience of the SRNE. The synthesis of an SRNE addressing professional ethicists might use other (final) categories and characteristics as one addressing health care practitioners. However, criteria about what is "good" stakeholder orientation for SRNEs are still lacking. Comparable to quality appraisal, it should at least be reported how synthesis was achieved, and whom the results are addressing in which way.

Inclusion of a Recommendation

A further open question regarding synthesis in SRNEs is whether the SRNE should include a recommendation about how to handle ethical issues or which course of action is ethically the "best". [cf. Scott et al 2016]. Alternatively, the goal of the SRNE can be more descriptive, just providing the state of normative evidence. This question is subject to increased discussion when arguments are the normative information unit of interest [see, e.g., Strech/Sofaer 2012] because traditionally in ethics, arguments are appraised and balanced to reach a final recommendation. In conceptualizing the role of evidence chosen in the REIGN framework, it makes no sense for the review group to attempt balancing arguments, as this should be left to the GDG. This issue might, however, have to be addressed if SRNEs are conducted in or for other contexts.

References

Beauchamp TL, Childress JF (2009) *Principles of Biomedical Ethics*. 6th edition. Oxford University Press, New York/Oxford

Caplan L, Hoffecker L, Prochazka AV (2008) Ethics in the rheumatology literature: A systematic review. *Arthritis & Rheumatism* 59(6):816-821

Charmaz K (2014) Constructing Grounded Theory. 2nd edition. Sage, London

Chung KC, Pushman AG, Bellfi LT (2009) A systematic review of ethical principles in the plastic surgery literature. *Plastic and Reconstructive Surgery* 124(5):1711-1718

Droste S (2008) Systematische Gewinnung von Informationen zu ethischen Aspekten in HTA-Berichten zu medizinischen Technologien bzw. Interventionen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 102:329-343

Droste S, Dintsios CM, Gerber A (2010) Information on ethical issues in health technology assessment. How and where to find them. *International Journal of Technology Assessment in Health Care* 26:441-449

Droste S, Herrmann-Frank A, Scheibler F, Krones T (2011) Ethical issues in autologous stem cell transplantation (ASCT) in advanced breast cancer: a systematic literature review. *BMC Medical Ethics* 12:6

Higgins JPT, Green S (2008) Cochrane Handbook for Systematic Reviews of Interventions. Wiley-Blackwell, Chichester

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2019) *Cochrane Handbook for Systematic Reviews of Interventions* version. 6th edition. Cochrane. Available from www.training.cochrane.org/handbook

Hofmann B (2005) Towards a procedure for integrating moral issues in health technology assessment. *International Journal of Technology Assessment in Health Care* 21:312-318

Klingler C, Silva DS, Schuermann C, Reis A, Saxena A, Strech D (2017) Ethical issues in public health surveillance: a systematic qualitative review. *BMC Public Health* 17(1):295

McCullough LB, Coverdale JH, Chervenak FA (2007) Constructing a systematic review for argument-based clinical ethics literature: the example of concealed medications. *Journal of Medicine and Philosophy* 32:65-76

McDougall R (2014) Systematic reviews in bioethics: types, challenges, and value. *Journal of Medicine and Philosophy* 9:89-97

Mertz M (2017) Qualitätsbewertung in systematischen Übersichtsarbeiten normativer Literatur. Eine Problemanalyse. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 127-128:11-20

Mertz M, Kahrass H, Strech D (2016) Current state of ethics literature synthesis: a systematic review of reviews. *BMC Medicine* 14:152

Mertz M, Strech D, Kahrass H (2017) What methods do reviews of normative ethics literature use for search, selection, analysis and synthesis? In-depth results from a systematic review of reviews. *Systematic Reviews* 6:261

Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine* 151:264-269

Piasecki J, Waligora M, Dranseika V (2018) Google search as an additional source in systematic reviews. Letter. *Science and Engineering Ethics* 24(2):809-810

Preshaw DH, Brazil K, McLaughlin D, Frolic A (2016) Ethical issues experienced by healthcare workers in nursing homes: Literature review. *Nursing Ethics* 23(5):490-506

Schreier M (2012) *Qualitative Content Analysis in Practice*. SAGE, Los Angeles/London/Washington

Scott AM, Bond K, G Gutiérrez-Ibarluzea I, Hofmann B, Sandman L (2016) Quality assessment of ethics analyses for health technology assessment. *International Journal of Technology Assessment in Health Care* 32(5):362-369

Seitzer F, Kahrass H, Neitzke G, Strech D (2016) The full spectrum of ethical issues in the care of patients with ALS: A systematic qualitative review. *Journal of Neurology* 263(2):201-209

Sofaer N, Strech D (2011) Reasons why post-trial access to trial drugs should, or need not be ensured to research participants: a systematic review. *Public Health Ethics* 4(2):160-184

Stoklosa A, Bond K (2013) *Workshop on Methodology in Ethics for Health Technology Assessment: Assessing the Need For and Quality of Ethics Analyses in HTA*.
https://htai.org/wp-content/uploads/2018/02/Ethics_Checklist_Workshops_-Edmonton_and_Cologne_-_FINAL_1_Apr_2014.pdf (24.10.2019)

Strech D, Sofaer N (2012) How to write a systematic review of reasons. *Journal of Medical Ethics* 38(2):121-126

Thomas J, Harden A (2008) Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology* 8:45

Whicher DM, Kass NE, Audera-Lopez C, Butt M, Jauregui IL, Harris K, Knoche J, Saxena A (2015) Ethical issues in patient safety research: A systematic review of the literature. *Journal of Patient Safety* 11(3):174-184

Appendix D: The REIGN Toolkit

The REIGN Toolkit

Dr. hum. biol. Corinna Klingler, MSc (Faculty of Health Sciences Brandenburg, University of Potsdam)
Dr. phil. Marcel Mertz (Institute of Ethics, History & Philosophy of Medicine, Hannover Medical School)

About the Toolkit

The **REIGN framework** (Use of **R**esearch **E**vidence to **I**nform **G**uidance regarding **N**ormative-ethical Topics) is a first attempt at structuring how **evidence** can and should be incorporated in **ethics guideline development**. The framework applies predominantly to guidelines developed for the (inter)national level (e.g., the WHO ethics guidelines).

This document summarizes the **central tenets of the framework** and provides **conceptual tools** to guide ethics guideline development. This document is part of the **full REIGN report**, which also includes further theoretical background information. Much of the information presented here only briefly is elaborated in the report, and the elaborated information should be consulted when questions arise.

While the framework will help ethics guideline developers think carefully about evidence integration, the methodological discussion in this field is still in its infancy. Accordingly, this document provides not a set of recipes but, rather, **tools to carefully consider the essential questions of the development process**. All flow charts, tables and further content are to be considered **preliminary**.

Target Group

- The toolkit targets primarily the **guideline development group (GDG)** and not those collecting evidence – the **review group (RG)** in WHO terminology. The REIGN framework emphasizes that it is the GDG's responsibility to decide on further evidence collection.
- For evidence collection, (additionally) the **WHO Handbook**[*] for guidance on empirical evidence or the **REIGN report** for guidance on normative evidence must be consulted.

[*] World Health Organization (2014) *WHO Handbook for Guideline Development*. 2nd Edition. Geneva

## Content

- Theoretical background to REIGN (pp. **1, 2, 5**)
- A focused flow chart of ethics guideline development using an evidence-based approach (pp. **2**)
- Descriptions of the process (pp. **2, 4, 5**)
- A checklist to assist decision-making with regard to evidence collection (p. **3**)
- An overview of the different sources and methods for collecting so-called normative evidence (p. **4**)
- A summarizing/overall flow chart (p. **6**)

Basic Premises of REIGN

Definition of evidence: Evidence is understood broadly and incorporates various types of information, not just aggregated quantified data. Evidence is assumed to describe **(a body of) information that is more or less qualified to support holding a specific statement true, plausible or right** (or false, implausible, or wrong) in the context of decision-making or of directing actions.

Normative vs. empirical evidence: Empirical evidence consists of empirical information (e.g., whether a phenomenon exists or in what way it is perceived) based on qualitative or quantitative analysis. Normative evidence describes normative phenomena (such as ethical principles, challenges or arguments) and addresses what should be done or what is valuable. Mostly, empirical evidence will be collected from empirical literature (e.g., social science studies), and normative evidence from normative literature (e.g., philosophical papers). Value judgments or arguments might also be found in empirical literature, although normative literature might be most appropriate for providing normative evidence.

Role of evidence: Evidence in ethics guideline development **can only inform—but never determine—**ethics recommendations. Arriving at recommendations involves weighing and balancing different ethical requirements or arguments. This task is assigned to the GDG and cannot be substituted by evidence collection, collation and analysis.

```
Evidence  --informs-->  (Deliberation among) GDG  --decides-->  Ethics Recommendation
```
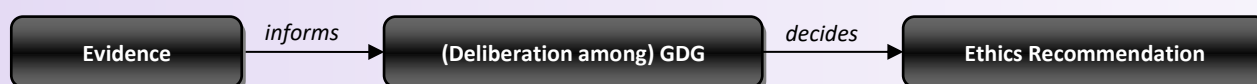
(0) Flow Chart Depicting Ethics Guideline Development as envisioned by the REIGN framework

See box on the lower right and p. 3 (Checklist) · *See p. 4* · *See p. 5* · *See pp. 5-6*
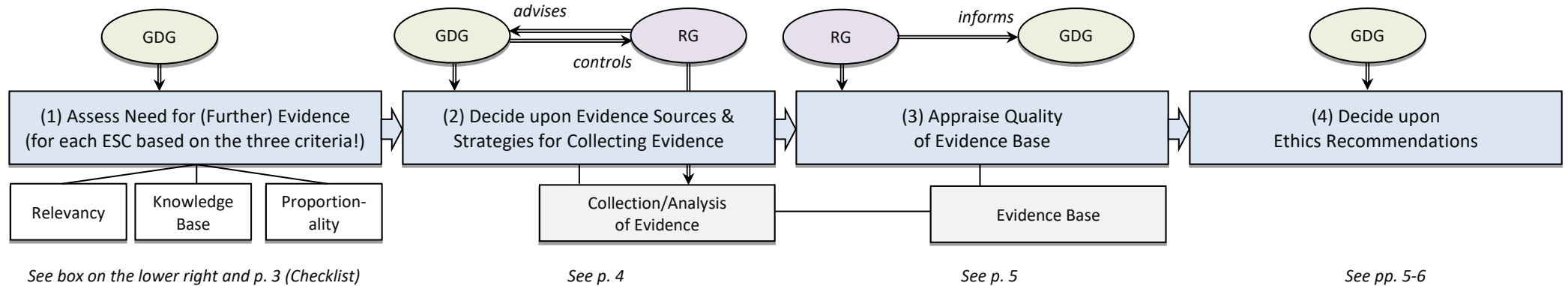
Evidential Support Components (ESC)

- ESCs are **five complexes of questions related to evidence** for distinctive **justificatory aspects** that underlie ethics recommendations; these components have to be **addressed by the GDG** to arrive at recommendations.

- For each ESC, a **main normative question** is identified that has to be answered by the GDG (see also checklist on p. 3):

    o ESC 1 – Value Base: *What basic normative principles should guide action and serve as orientation points for the topic of the guideline?*

    o ESC 2 – Conceptual Disambiguation*: What terms (e.g., abortion or foeticide) should be used for the main topics discussed in the guideline, and how should they be defined?*

    o ESC 3 – Need for Action: *What ethical issues should be addressed by the guideline?*

    o ESC 4 – Strategies for Addressing Needs: *Which strategies for addressing ("solving") the identified ethical issues should be considered by the guideline?*

    o ESC 5 – (Hypothetical) Arguments for Action: *Why should specific strategies be recommended by the guideline, and what further aspects have to be considered when following this strategy?*

- The checklist (p. 3) indicates what kind of **normative [NE] and empirical [EE] evidence** may support the GDG in addressing these questions.

(1) Assess Need for (Further) Evidence

- To assess whether there is a need to collect and analyse (further) evidence, three questions need to be posed **to each ESC** (in this order!):

    Relevancy of posing the question:

    o *Is the main normative question of the ESC* (see box on the left) *relevant to the work (meaning NOT externally answered or clearly answered for other reasons)?*

    Existence of a **Knowledge Base**:

    o *Can the GDG answer this question by itself because relevant expertise is available or can the GDG access existing evidence bodies (e.g., published systematic reviews)?*

    Proportionality of evidence collection:

    o *Are the financial and time costs associated with evidence collection justified considering the expected benefits?*

    The answer to this question should consider the different available evidence collection strategies (see p. 4) and associated costs!

- If the question on relevancy or proportionality is answered negatively or the knowledge base question answered positively, no further evidence collection is necessary. Otherwise, further evidence should be collected and considered.

---

(1) Assess Need for (Further) Evidence: Checklist for the GDG

ESC 1 – Value Base	Relevant	Knowledge Base	Proportional
Main Question: What basic normative principles should guide action and serve as orientation points for the topic of the guideline?	☐ Yes ☐ No	☐ Yes ☐ No	Not applicable
Evidence to support decision-making: ▪ Overview of normative principles commonly used in the context [**NE**]	Not applicable		☐ Yes ☐ No

ESC 2 – Conceptual Disambiguation	Relevant	Knowledge Base	Proportional
Main Question: What terms (e.g., abortion or foeticide) should be used for the main topics discussed in the guideline, and how should they be defined?	☐ Yes ☐ No	☐ Yes ☐ No	Not applicable
Evidence to support decision-making: ▪ Overview of terms in use for the main topics discussed [**NE/EE**] ▪ Overview of the definitions provided for the main topics [**NE/EE**] ▪ Overview of ethical implications of/reasons for choosing particular terms/definitions [**NE**]	Not applicable		☐ Yes ☐ No ☐ Yes ☐ No ☐ Yes ☐ No

ESC 3 – Need for Action	Relevant	Knowledge Base	Proportional
Main Question: What ethical issues should be addressed by the guideline?	☐ Yes ☐ No	☐ Yes ☐ No	Not applicable
Evidence to support decision-making: ▪ Overview of the ethical issues associated with the topic of the guideline [**NE**] ▪ Overview of data on the urgency of ethical issues (prevalence, consequences, etc.) [**EE**] ▪ Overview of (further) reasons for prioritizing ethical issues [**NE**] ▪ Overview of regulatory documents addressing (certain) ethical issues to see whether additional guidance is needed [**EE/NE**]	Not applicable		☐ Yes ☐ No ☐ Yes ☐ No ☐ Yes ☐ No ☐ Yes ☐ No

ESC 4 – Strategies for Addressing Need	Relevant	Knowledge Base	Proportional
Main Question: Which strategies for addressing ("solving") the identified ethical issues should be considered by the guideline?	☐ Yes ☐ No	☐ Yes ☐ No	Not applicable
Evidence to support decision-making: ▪ Overview of the strategies for addressing prioritized ethical issues [**EE**]	Not applicable		☐ Yes ☐ No

ESC 5 – (Hypothetical) Arguments for Action	Relevant	Knowledge Base	Proportional
Main Question: Why should specific strategies be recommended by the guideline, and what further aspects have to be considered when following this strategy?	☐ Yes ☐ No	☐ Yes ☐ No	Not applicable
Evidence to support decision-making: ▪ Overview of the (hypothetical) arguments given for choosing a particular strategy [**NE**] ▪ Overview of data on the (probable) consequences of choosing a particular strategy (to substantiate hypothetical (consequentialist) arguments) [**EE**]	Not applicable		☐ Yes ☐ No ☐ Yes ☐ No ☐ Yes ☐ No

(2) Decide upon Evidence Sources & Strategies for Collecting Evidence

- Concerning collecting, analysing and reporting **empirical evidence**, the *WHO Handbook* provides advice that should also be used by guideline developers in the context of ethics guidelines.

- No guidance, on the other hand, has been developed for **normative evidence**.

- Normative evidence can and possibly should be collected **from various sources**: not just from the academic debates but also through further stakeholder involvement.

- There exist **different strategies for collecting normative evidence** (see box to the right).

- When choosing an approach to evidence collection, the goal of **thematic or argumentative saturation** should be considered (ideally, evidence collection should result in a comprehensive overview of the topic of interest, for example, *all* ethical issues).

- **How the strategy is implemented** (e.g., how many databases are searched) will also impact how far thematic or argumentative saturation can be reached. However, this issue will have to be addressed by those responsible for evidence collection and analysis (**the RG**).

- The GDG has to decide which strategy should be implemented while also considering associated resource investments; **however, the GDG should take advice from the RG** regarding value, feasibility and limitations of the different possible strategies.

- In the box to the right, a list of strategies for collecting evidence is provided; short descriptions are also included.

- **No hierarchy of strategies is intended**, with the possible exception of the strategies for reviewing academic literature. One reason is that the context of interest/the topic of the guideline will also impact the fitness of the strategy to reach thematic or argumentative saturation (e.g., for an under-researched topic, it might be more important to involve additional stakeholders).

Sources	Evidence collection strategy	Explanation
External Research Evidence (Academic Literature*)	Systematic Review	A literature review that methodically follows ex ante defined steps to identify, synthesize and present relevant research (see also Appendix C, REIGN report).
	Unsystematic or Narrative Literature Review	A literature review that identifies, synthesizes and presents relevant research without following a clearly explicated process.
	Several Single Papers	A convenience sample of papers that supplies the evidence base.
	Single Paper (n=1)	A single paper that supplies the evidence base.
External Research Evidence (Academic Experts)	Consensus Process	Consensus among academic experts is built regarding the topic of interest by using, for example, Delphi methods.
	Workshop	A face-to-face meeting allowing various experts to present their research and discuss findings among themselves (and with the GDG).
	Commissioned Theory Application	A researcher is asked to analyse the question of interest (e.g., ethical issues in a given context) using specific theoretical lenses (principlism, consequentialism, etc.).
	Consultation (written or verbal)	Academic experts are asked to present their positions on a specific topic or question in writing or verbally during a meeting.
External Evidence from Additional Stakeholders	Interviews/Focus Groups	Stakeholders share their views in interviews or groups discussion.
	Opinion Survey	Stakeholders are asked to share their views in a (postal or online) survey. Compared with interviews/focus groups, a survey allows more people to be approached; however, no deeper engagement with their positions is possible.
	Consensus Process	Consensus among stakeholder representatives is built regarding the topic of interest by using, for example, Delphi methods.
	Workshop	A face-to-face meeting allowing various stakeholder representatives to present their positions and discuss findings among themselves (and with the GDG).
	Consultation (written or verbal)	Stakeholder representatives or the public are asked to present their positions on a specific topic or question in writing or verbally during a meeting.

* Similar strategies for evidence collection can be used for accessing written sources (e.g., policy documents) from additional stakeholder groups.

(3) Appraise Quality of Evidence Base

- Quality appraisal is an important part of evidence collection. Standards have been developed for **empirical evidence** (see *WHO Handbook*). **For normative evidence, no standards have been established yet**, though one may rely on criteria stemming from informal and formal logic, critical thinking and philosophy in general.

- For normative evidence, REIGN stipulates that both the **quality of individual information units** (e.g., arguments or ethical issues) and the **quality of the body of evidence** (in each ESC) have to be assessed.

- Below, exemplary questions and orientation points are provided:

- **Quality of individual information units:**

  - *Are the arguments valid and sound (deductive arguments), are they very strong (inductive arguments) or do they have considerable explanatory power (abductive arguments)? Are ethical issues relevant to the topic, well-described and justified (e.g., by referring to normative frameworks/moral theories)?*

  - *General information-critical approach: reflecting upon the trustworthiness, relevance and completeness of the information (and its sources!) that is used to inform decisions – why should the information be used, and what legitimates the information unit as being used for informing ethics recommendations?*

- **Quality of the body of evidence (in each ESC):**

  - *The academic (or public) discourse might be biased in various ways, and therefore, certain perspectives (and accordingly relevant principles, issues or arguments) might be missing; the discourse might also be incomplete for other reason; how well does the body of evidence fulfil the criterion of argumentative/thematic saturation?*

  - *How do the following impact saturation?*

    (a) the attributes of the chosen evidence collection strategy (e.g., systematic review vs. single paper or focus group vs. workshop);

    (b) contextual factors (e.g., new technology, scarcely any related research, "perspective bias" from specific disciplines); and

    (c) the actual implementation of the strategy (e.g., how many databases are used in a systematic review, how diverse regarding background and interests are members of a consensus process)?

(4) Decide upon Ethics Recommendations

- The **task of the GDG is to prioritize issues/balance arguments** and thereby arrive at final ethics recommendations.

- A good **choice of participants and fair moderation** of the process should ensure that the discussion is not dominated by certain strong opinions but stays oriented to finding and acknowledging the "best" rational argument(s).

Working with a Review Group (RG)

- **The RG should be involved in the development process as soon as possible** to advise the GDG and align expectations.

- While the necessary skill set to be represented among the RG will depend on which strategies are chosen, experience with ethical discourses are indispensable.

- The RG should work independently from the GDG to allow unbiased assessments.

- The RG may want to consult **Appendix C** when systematic reviews for normative evidence (SRNEs) prove to be the methods of choice for evidence collection.

Open Questions

The discussion of how to develop ethics guidelines is still in its infancy. Accordingly, in advancing REIGN, the authors had to make many conceptual decisions that could not be based on a widespread methodological consensus in the academic community. The authors therefore **strongly encourage those involved in ethics guideline development to advance the development of actual methodological guidance manuals**. The REIGN framework will be a useful information base for building consensus among experts in the field.

Furthermore, while explicitly considering evidence might improve decision-making in guideline development, other aspects might be equally important, namely: (a) **Who participates in the process?** (b) **How is the process structured?** or **(c) What resources are available?**

It is still open to debate how decision-making in ethics guideline development should best be structured. It is therefore particularly important for the GDG to transparently **report on methods employed** not just in terms of evidence collection, but also regarding consensus building to learn from experience and to be able to constantly improve the underlying methods and procedures.

Tasks of the Guideline Development Group (GDG)	Tasks of the Review Group (RG)

controls

ESC 1: Value Base → If required: Evidence collection for ESC 1

In its work, the group adheres to

- ✓ **Methodological standards** of the chosen method
- ✓ **Quality criteria in reporting** the group's work

Definition of basic **principles**

ESC 2: Conceptual Disambiguation → If required: Evidence collection for ESC 2

As part of its work, the group

- ✓ Identifies relevant information sources
- ✓ Collects relevant information
- ✓ Analyses/Synthesizes relevant information
- ✓ Appraises the quality of the evidence base

Definition/use of basic **concepts**

ESC 3: Need for Action for...

| ...Concept #1 | ...Concept #2 | ...Concept #3 |

decides whether and how → If required: Evidence collection for ESC 3

Definition of relevant **problems** ← *feeds back to GDG*

If applicable: Please also consult **Appendix C** of the REIGN report for practical tips on conducting systematic reviews for normative evidence.

ESC 4: Strategies for addressing...

| ...Problem #1 | ...Problem #2 | ...Problem #3 |

→ If required: Evidence collection for ESC 4

Definition of relevant **solutions**

ESC 5: (Hypothetical) Arguments for...

| ...Solution #1 | ...Solution #2 | ...Solution #3 |

→ If required: Evidence collection for ESC 5

Definition of relevant **arguments**

Final Ethics Recommendation