International Journal of Technology Assessment in Health Care, 33:2 (2017), 135–146.

doi:10.1017/S0266462317000228

HEALTH TECHNOLOGY ASSESSMENT OF PUBLIC HEALTH INTERVENTIONS: A SYNTHESIS OF METHODOLOGICAL GUIDANCE

Tim Mathes
Institute for Research in Operative Medicine
tim.mathes@uni-wh.de

Sunya-Lee Antoine
Peggy Prengel
Stefanie Bühn
Stephanie Polus
Dawid Pieper
Institute for Research in Operative Medicine

Objectives: The evaluation of public health interventions poses some challenges. As a consequence, health technology assessment (HTA) methods for public health interventions (PHI) have to be adapted. This study aimed to summarize the available guidance on methods for HTA of PHI.
Methods: We systematically searched for methodological guidance on HTA of PHIs. Our focus was on research synthesis methods to evaluate effectiveness. Relevant information was synthesized narratively in a standardized way.
Results: Only four guidance documents were identified specifically for HTAs of PHI. The approaches used for HTAs of PHIs are broader and more flexible than those for medical interventions. For this reason, there is a tendency to identify the intervention components and context factors that influence the effectiveness and transferability of an intervention rather than to assess its effectiveness in general. The details in the guidance vary without justification. Unjustified heterogeneity between the different guidance approaches is most pronounced for quality assessment, assessment of applicability, and methods to integrate qualitative and quantitative evidence. Descriptions for the assessment of integrity, heterogeneity, sustainability, context factors, and applicability are often vague.
Conclusions: The heterogeneity in approaches indicates that there is currently no consensus on methods to deal with the challenges of the PHI evaluations. A possible explanation for this may be that the methods are not sufficiently developed, and advantages and disadvantages of a certain method in relation to the research question (e.g., broad/focused) have not yet been sufficiently evaluated.

Keywords: HTA, Public health, Health promotion, Evidence synthesis

The majority of health technology assessments (HTAs) still focus on clinical medicine, particularly on pharmaceuticals (1), while HTAs on public health interventions (PHI) are rarely conducted (1). In 2002, David Sackett criticized the use of preventive interventions without an evidence base (2). A survey conducted in five countries in 2010 found that only 5 percent of HTAs focused on public health (3).

Conducting HTAs of PHIs poses some challenges compared with medical interventions. Randomized controlled trials (RCTs) are often not available in the field of public health as they are usually difficult to conduct (4). PHIs, such as the implementation of a school nurse, are highly complex, The standardization of interventions (e.g., nurses), various intervention components (e.g., medical and psychological), participants (e.g., different age groups), contextual factors (e.g., number of general practices nearby), and the number and variability of outcomes add to this complexity (5). As a result, PHIs often rely on study designs that are not at the top of the evidence hierarchy.

Our objective was to provide an overview of the existing methodological guidance on HTAs of PHIs. Furthermore, we analyzed the similarities and differences between the methodological recommendations.

METHODS

We systematically searched the Web pages of international organizations for systematic reviews and HTA organizations. The HTA organizations were identified using the member lists of the HTA umbrella organizations: the International Network of Agencies for Health Technology Assessment (INAHTA), Health Technology Assessment International (HTAi), and the

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

European Network for Health Technology Assessment (EUnetHTA). In total, the Web pages of 135 organizations were screened. Furthermore, we searched the Web sites of Cochrane, the Centre for Reviews and Dissemination (CRD), and the Joanna Briggs Institute (JBI).

Searches were performed in March and April of 2015. We contacted all HTA organizations by e-mail to capture unpublished documents not available on the Web sites. We obtained contact information from the Web pages of INAHTA, HTAi, and EUnetHTA. We sent the initial e-mail on March 18, 2015, and a reminder on April 10, 2015. Replies were accepted until May 1, 2015.

All potentially relevant documents were screened according to the following *a priori* defined inclusion criteria (see Figure 1): (i) Methodological guidance for the preparation of HTAs for PHIs (handbooks, manuals, guidelines, etc.); (ii) Language: English, Spanish, German, Italian.

We focused on methodological guidance for the preparation of research synthesis to evaluate effectiveness. We did not consider the economic, legal, or organizational aspects of a PHI. Documents focusing on evaluation methods other than research synthesis (e.g., surveys) were not considered. We assumed that the challenges of HTAs of PHIs might require fundamentally different approaches to overcome. Each step in the preparation process of an HTA (e.g., literature search) is always interrelated with all other steps conducted in the HTA. We, therefore, focused on manuals that described the entire preparation process. We did not search medical databases, as it is very unlikely that scientific papers provide detailed methodological guidance on the entire preparation process. We included documents on prevention, screening, and vaccination if described in the context of population health analysis. Only the most recent document was included if different versions of the same document existed. Two reviewers independently screened the identified publications according to the inclusion criteria. Different judgements on inclusion were resolved in a discussion until a consensus was reached.

We prepared and piloted standardized tables for data extraction. All data describing the methodology for research synthesis to evaluate the effectiveness of PHIs were summarized. Data were extracted on *a priori* defined aspects (see headings in Table 1 and Table 2). We extracted data specific for HTAs of PHIs. Organizational aspects, general scientific descriptions, general methods for systematic reviews, and recommendations for reporting were not extracted. We also excluded specific review types (e.g., review of reviews, rapid reviews). The wording in the included documents was copied as closely as possible to avoid interpretation bias. The data were extracted by one reviewer and checked by a second reviewer to ensure that all relevant information was captured and to guarantee the accuracy of data extraction. Discrepancies were resolved in discussion until a consensus was reached.

RESULTS

Literature Search

The Web page search and e-mail inquiries (response rate 70 percent) of the 135 HTA organizations resulted in sixty-five potentially relevant publications provided by forty-three HTA organizations (some organizations provided more than one document). Furthermore, we searched the Web pages of Cochrane and CRD and identified two further potentially relevant publications. Of the forty-five organizations (43 HTA organizations plus Cochrane and CRD), forty-one organizations provided no eligible publications (63 excluded publications, list of excluded publications see Supplementary Material 1). Four organizations (four publications) were finally included in the analysis (6–9) (Table 1). One manual was only available in German (10). The selection process is illustrated in Figure 1.

Two publications were provided by national HTA organizations (Gesundheit Österreich [GOEG], National Institute for Health and Care Excellence [NICE]) (7;8) and two were provided by international systematic review organizations (Cochrane Collaboration, CRD)), which develop methods to gather research evidence and publish HTAs (6;9). The details on the methods are presented in Tables 1 and 2.

Scope/Definition of PHIs

The scope of public health is defined by three organizations. Two organizations use their own definitions and one is based on the World Health Organization's (WHO) definition of health promotion (7;9;10). The holistic view of health in the area of public health is mentioned by two organizations (7;9). These two organizations have a wide scope, including prevention, health protection, and health restoring and focus on population rather than individuals (7;9). One organization focuses exclusively on health promotion (8).

Review Type(s) (HTA Products)

The review types differ between the organizations. Cochrane describes the preparation of systematic reviews of effectiveness studies (6). The CRD also focuses on the preparation of systematic reviews of effectiveness studies, but in the manual it mentions that researchers may find realist synthesis useful (9). Both organizations consider including qualitative studies, depending on the research question (e.g., theoretical underpinning, description of patient experience). The two national organizations produce different products depending on the research question (7;8).

GOEG suggests the use of systematic reviews of effectiveness studies, realist reviews (6), or the interactive domain model (based on underpinnings, understanding of the environment and practice). These approaches are used separately or in combination. NICE has a broad spectrum of HTA products, including reviews of effectiveness studies, reviews of reviews,

https://doi.org/10.1017/S0266462317000228 Published online by Cambridge University Press

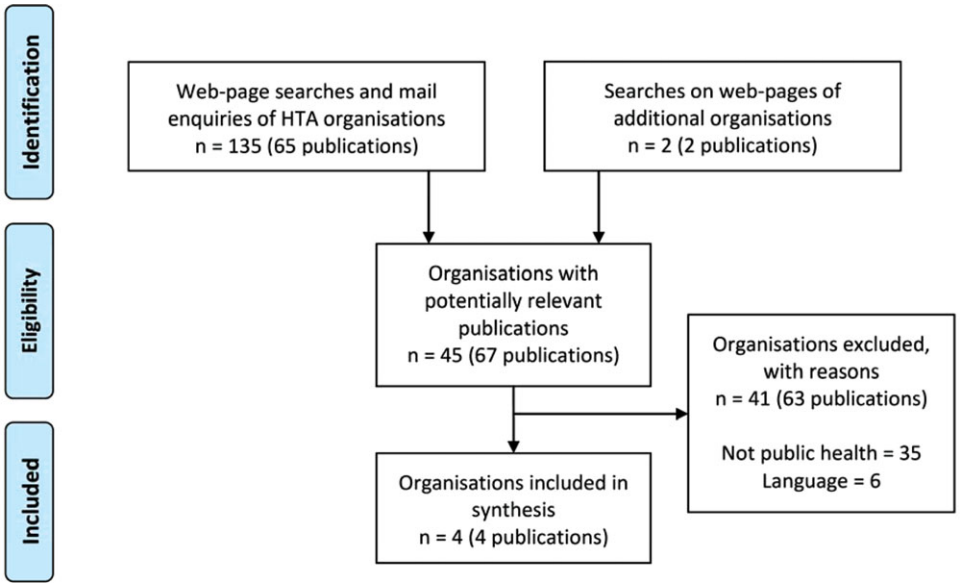

Figure 1. Flow-diagram of guidance selection.

reviews of cost-effectiveness studies, and reviews of epidemiological studies (7). In general, one effectiveness review is supplemented by other products.

Planning the Review

Three organizations recommend starting with the development of a conceptual framework and a definition of scope (6;7;9), including the formulation of the review question(s). All organizations suggest using the PICO scheme (patients, interventions, comparison, outcomes) to formulate the review questions. Three organizations advise the consideration of additional aspects, for example, the context and the setting (7–9). Two organizations (7;8) recommend preliminary searches to support the planning process. The order of the steps (scope, conceptual framework, review question, preliminary searches) varies between the organizations. The Cochrane Collaboration and the CRD emphasize the importance of perspective (narrow versus broad) for the scope to decide whether to lump or split the review question.

Study Designs to Include

A specification of relevant study designs is given by three organizations (6–8). It is emphasized that a wide range of study designs guided by the research question should be included, rather than the classic hierarchy of evidence, due to the broader scope and diversity of interventions as well as methods. All organizations agree that RCTs, cluster-RCTs, controlled before-and-after studies (CBA), or before–after studies and interrupted time series (ITS) should be considered for inclusion (6–8). Various other quantitative study types (e.g., non-RCTs, historically controlled studies), qualitative studies (e.g., focus groups, interviews), and economic evaluations are also quoted.

Searching for Literature

All organizations recommend searching a wide range of databases relevant for the addressed topic (e.g., education, transport, engineering) in addition to medical databases (6–9). The Cochrane Collaboration, CRD, and NICE recommend the use of free text words due to the heterogeneous terminology and poor indexing of public health concepts (6;7;9). All organizations recommend additional searches (e.g., hand searches, contacting experts, organizational Web sites).

Quality Assessment

Guidance on the quality assessment of included studies (quality assessment of the evaluation of the intervention [study quality]) is given by all organizations (6–9). The Cochrane Collaboration recommends the use of the "Quality Assessment Tool for Quantitative Studies" for RCTs and all nonrandomized study designs (6). None of the organizations suggest any tool for uncontrolled studies, while *prima facie* criteria are suggested for qualitative studies. GOEG refers to its own checklist for systematic reviews and cohort studies and also suggests criteria for qualitative studies. They do not provide recommendations for any other study design (8). CRD discusses aspects of risk of bias for different study designs but only presents quality criteria for RCTs (9). NICE uses its own checklists to assess the internal validity of quantitative as well as qualitative studies (7).

Data Extraction

CRD, NICE, and GOEG address data extraction in separate sections (7–9). All three organizations recommend extracting data according to the PICO scheme and setting. CRD and NICE recommend extracting data on context, the integrity of the intervention, and the theoretical underpinning of the primary

https://doi.org/10.1017/S0266462317000228 Published online by Cambridge University Press

Table 1. Information on Definition of Public Health and HTA Products

Organization	Title, year	Scope/definition of public health intervention	Review type(s) (HTA products)
Cochrane Collaboration (Guidelines for Systematic reviews in health promotion and public health taskforce): The Cochrane Collaboration (12)	Systematic Reviews of Guidelines for Systematic reviews of health promotion and public health interventions	NR	— Systematic review of effectiveness studies — Researchers may find are Realist synthesis useful
CRD (Centre for Reviews and Dissemination) (10)	CRD's guidance for undertaking reviews in health care, 2009	Activities that aim to protect, promote, and restore the health of all people Address change at the individual level, structural/policy—changing level intent, a wider population or community effect often focusing on the social, physical, economic, or legislative context	Systematic review of electiveness studies
GÖG/BIQG: Gesundheit Oesterreich GmbH (11)	Aufbereitung von Evidenz zu Gesundheitsförderung Band Nr. 10 aus der Reihe WISSEN (Teil 1 Handbuch), 2013	Health promotion as defined by the WHO Ottawa Charter for Health Promotion: "Health promotion is the process of enabling people to increase control over, and to improve, their health. To reach a state of complete physical, mental and social well—being, an individual or group must be able to identify and to realize aspirations, to satisfy needs, and to change or cope with the environment. Health is, therefore, seen as a resource for everyday life, not the objective of living. Health is a positive concept emphasizing social and personal resources, as well as physical capacities. Therefore, health promotion is not just the responsibility of the health sector, but goes beyond healthy life—styles to well—being." (WHO 1986)	— Systematic review of effectiveness studies — Realist review — Interactive domain model — Approaches can be combined
NICE: National Institute for Health and Care Excellence (13)	Methods for the development of NICE public health guidance (third edition), 2012	The subject matter of public health is broad and diverse. It involves disease prevention, health promotion, protecting individuals and populations from hazards, and it is concerned with health improvement. It has a population rather than an individual focus. It draws on social models of health as well as biomedical ones. The conceptual framework is based on a number of principles. These are as follows. First, that there are determinants of health and disease which are much broader than, but include, biomedical causes. Second, these determinants operate in highly patterned ways which reflect inequalities in society. Third, the determinants work through causal pathways to disease. Fourth, the causal pathways help to identify ways of preventing and ameliorating disease. Fifth, there are also causal pathways for the promotion of health. Sixth, positive and negative causal pathways cross physical, biological, social and psychological boundaries. Public health may be direct or indirect.	— (Cost—)Effectiveness reviews — Review of reviews — Epidemiology review — Correlate reviews — Review of qualitative evidence — Qualitative review — Mapping review — Hybrids of the review types above (at least one effectiveness review)

NR, not reported.

studies (7;9). NICE recommends the use of a combination of narrative summaries and evidence tables (7). CRD endorsed the development of data extraction forms according to the review question (9). The Cochrane Collaboration does not give explicit recommendations for data extraction in a separate section but mentions data extraction checklists and the importance of extracting data on PHI integrity (6).

Theoretical Framework

The organizations use theoretical frameworks in different ways. NICE uses a theoretical framework to identify relevant research questions (7). CRD proposes the use of a theoretical framework to group the interventions or include interventions that are based on one or more specific theories (inclusion criteria) (9). The theoretical framework of the primary studies is

https://doi.org/10.1017/S0266462317000228 Published online by Cambridge University Press

Table 2. Methodology for HTAs of Public Health HTs

Agency	Cochrane Public Health Group (12)	CRD (10)	GÖG/BIQG (11)	NICE (13)
Planning the review	Scope of the review: lumping versus splitting review question Research question: patients, HTs, outcomes	Determining the scope Development of a conceptual framework Review question according to PICO, context, setting Broad question can be split	Preliminary search for background information Specify the review question according to target population, setting, intervention, reference to compare effectiveness, outcomes, context factors, theoretical underpinnings, core values and principles of health promotion Development of a conceptual framework	Develop a topic—specific logic model based on the overarching conceptual framework based on the conceptual framework Preliminary searches Logic model is starting point for the scope Scope includes key questions based on PICO
Study designs to include	Preliminary search to be familiar with the types of study designs that may have been used. Primarily reflect the question/s being answered in the review, rather than any predetermined hierarchy RCT, cluster—RCT, non—RCT, CBA, ITS, qualitative studies, comparisons with historical controls/national trends	The choice of study design should be guided by the review question and the needs of the end users. The traditional hierarchy of evidence is relevant RCT, cluster—RCT, before—and after studies, ITS, uncontrolled studies, discontinuity, matched controlled designs, qualitative studies	NR	Rather than relying on the standard hierarchy of evidence, a wide range of study designs and methodologies should be used to answer these questions RCT, non—RCT, cluster—RCT, before—and—after study, case—control—study, cohort study, correlation study, cross—sectional study, ITS, document analysis, focus groups, interview study, observation, economic studies
Searching for literature	Sensitive search: text words and synonyms freely as there may be few, or no indexing terms related to your topic, and because terminology varies historically and culturally Databases covering a range of relevant disciplines Content pages of journal publications, gray literature and other sources Study design filters (if appropriate)	Databases according to the question being addressed (e.g., engineering) Public Health Language (PHL) Differences in terminology need to be compensated for by using free text terms and synonyms Supplement database search with internet search, scanning relevant organizational Web sites, contacting experts in the field, reference checking, gray literature sources, handsearching selected journals, snowballing	Databases (public health/ medical/health, social, economics, legal, education) Study registries, relevant journals, unpublished literature, additional information, opportunistic search for literature etc.	Core databases (relevant to public health topic), and other resources, Web sites, economic searches, snowballing, gray literature, handsearching, contacting experts, review—level searches Study—type limits or filters should be used with caution Thesaurus terms and free—text/keywords

Table 2. Continued

Agency	Cochrane Public Health Group (12)	CRD (10)	GÖG/BIQG (11)	NICE (13)
Quality assessment	SR: Health—evidence.ca—Tool, Critical Appraisal Skills program (Non—)RCT: Quality Assessment Tool for Quantitative Studies Qualitative studies: prima facie criteria	Evaluation quality (risk of bias): only important design features for different study designs are discussed. Guiding questions for RCT and reference to the Cochrane risk of bias tool Intervention quality (part of quality assessment): integrity and fidelity of intervention Intervention quality (quality assessment): definition of intervention (theoretical underpinning)	GOEG checklist for assessing sources of evidence, systematic reviews and meta—analyses, quality of cohort studies GOEG guiding questions for qualitative studies	NICE checklists (internal validity): checklist for intervention studies or that for correlation (quantitative studies), checklist for qualitative studies, checklist for economic studies
Data extraction	NR reported in a separate section Abstraction should contain information relating to the assessment of integrity	Broad format according to PICO, context, setting Adequate detail relating to the population Identify where theoretical information is absent from the primary studies Forms should always be developed according to the requirements of the individual review Recording of the potential impact of missing data	Relevant information: target population, setting, intervention (detailed), outcomes, context factors, theoretical underpinnings details of integrity of intervention and context	Narrative summaries and evidence tables According to patients, HTs, outcomes setting (concise detail) for qualitative and quantitative evidence
Theoretical framework	HTs can be group by their theoretical basis. The studies, according to different theories, may be tabulated, combined narratively, or statistical ly combined It may also be useful for authors to assess whether HTs have used a Program Logic or Program Theory of Action approach	Only include HTs based on one particular theory, or based on different theories (record theoretical underpinning). Theory can be used to group HTs and explore potential differences in effect	Identification of a theoretical model that elucidates the requirements, for whom and why a intervention is effective (part of data synthesis)	Topic—specific conceptual framework based on NICE public health conceptual framework Programme theory or theories of change or the logic model used to develop key questions
Integrity of intervention	Collection, assessment, and synthesis of information relating to integrity Information should be elicited regarding factors which influence the effectiveness of HTs	Not reported in a separate section (see quality assessment)	NR in a separate section (see data extraction)	NR

Table 2. Continued

Agency	Cochrane Public Health Group (12)	CRD (10)	GÖG/BIQG (11)	NICE (13)
Heterogeneity	Sources of heterogeneity: variability in study populations, HTs, settings, outcomes, study designs, analysis Subgroup analysis Consider the likely sources of heterogeneity as described above, and consider these as they synthesise and analyse the results, either narratively or in meta—analysis	Differences in study design, participants, context, and in processes/methods of implementation, theoretical underpinnings, outcomes and outcome measures Subgroup analysis to evaluate differential impacts across groups and inequalities, interactions between effects and the quality of the intervention Harvest plot as a method to combine aspect of the graphical directness of a forest plot with a narrative account	Strong heterogeneity should be considered Variability of the examined study participants, HTs, observed indicators and methods to measure effectiveness	The degree of heterogeneity in the data should be assessed to determine how the results have been affected by the circumstances in which studies were carried out. Random effects model for statistical heterogeneity Subgroup analysis Meta—regression
Integrating qualitative and quantitative studies	Multi—layered model of social determinants of health Three types of syntheses in the same review: meta—analysis, qualitative synthesis of views, mixed methods synthesis (effect sizes which matched implications for HTs from people's compared with those which do not, using sub—group analysis) Narrative synthesis	NR	Mixed method See also column HTA products in table 1	NR
Ethics, equity and Inequalities	Ethical implications of every decision made throughout the review process Definition of inequalities: disadvantage may be considered in terms of PROGRESS Define effectiveness of intervention in reducing health inequalities Locate studies which examine inequalities Subgroup analysis	Researchers might want to consider investigating differential outcomes according to varying levels of disadvantage Set of criteria for measuring disadvantage is PROGRESS	No recommendations for consideration of equity aspects	Specific issues in relation to groups identified in the Equality Act 2010 or groups who are particularly disadvantaged with respect to the topic under consideration throughout the whole review process

Table 2. Continued

Agency	Cochrane Public Health Group (12)	CRD (10)	GÖG/BIQG (11)	NICE (13)
Sustainability	If relevant consider what outcomes have been measured, over what period, and what is the pattern of outcomes over time Where sustainability has not been measured, explore the potential of the intervention outcomes to be sustained Particular features relevant to sustainability should be considered in addition to general aspects of study quality	Sustainable should be considered at the protocol development stage and again later during quality assessment Particular attention to the validity and reliability of interim or surrogate outcomes s, and to the extent to which they can actually predict the outcomes	Sustainability mentioned but not further specified	NR
Context	Cautious generalizations from one context to another Report on the presence of context—related information in intervention studies May alert investigators to the need to qualify their statements about "intervention" effects	Details of the context Assess whether the context is a contributor to the effectiveness of the intervention	Data on context should be extracted Analysis of mechanism of action and interaction with context	Assessment of context—specific aspects throughout the review
Applicability	Characteristics to be included (individual studies and body of evidence): applicability, relevance, appropriateness, feasibility, adverse effects, equitability, sustainability, transferability	Usually necessary to examine details of process and context Study findings are similar across a range of circumstances: confidence that the findings are transferable is increased; Effects vary: information is useful for understanding in which circumstances the evidence is likely to be applicable Assessment of applicability and transferability according to checklists	The following criteria can support judgement: comparability between populations, conditions under which the study was conducted and the intervention shall take place, societal and health system, context of study population and target population Cautious generalizations from one context to another Checklist to assess the applicability	NICE quality appraisal checklist for included studies (external validity) Applicability of evidence based on population characteristics, setting, intervention and outcomes

NR, not reported; HT, health technology; PICO, Patients, Intervention, Comparison, Outcome; PROGRESS, place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital.

part of the quality assessment (quality of the intervention). The Cochrane Collaboration suggests grouping the interventions according to their theoretical basis for tabulation, or for narrative or quantitative synthesis. GOEG applies the theoretical framework within the data synthesis to identify the components that determine effectiveness (8).

Integrity of Intervention

Integrity is addressed by three organizations (6;8;9). Integrity of the primary studies is either considered as part of the quality assessment of the intervention or data extraction (description of intervention). The Cochrane Collaboration recommends collecting (description of studies), analyzing (e.g., factors influencing the effectiveness), and synthesizing data based on integrity (6).

Heterogeneity

All of the organizations emphasize the importance of taking heterogeneity into account while reviewing PHIs. They describe the various sources of heterogeneity (e.g., methodological, statistical, PICO, setting, context). The most common suggestion for dealing with heterogeneity is to perform subgroup analysis (7;9). NICE recommends using a random-effects model for meta-analyses and performing a meta-regression if necessary (7). CRD suggests using Harvest plots.

Integrating Qualitative and Quantitative Studies

GOEG and the Cochrane Collaboration report on methods to integrate qualitative and quantitative evidence (6;8;9). GOEG refers to a mixed methods approach, but without further description (8). Moreover, the synthesis methods, realist reviews (6) and interactive domain model suggested by GOEG (8) can be considered as holistic methods to combine qualitative and quantitative studies. The Cochrane Collaboration recommends three different approaches (6): The multi-layered model of social determinates in health, including three types of syntheses in the same review (meta-analysis, qualitative synthesis of views, mixed methods), or narrative synthesis (6).

Ethics, Equity, and Inequality

All organizations recommend considering equity aspects in the assessment. The Cochrane Collaboration incorporates equity aspects in different review steps (locating studies, indicators for inequalities, subgroup analysis, applicability) (6). NICE requires addressing disadvantaged groups throughout the entire review process (7). GOEG mentions equity aspects in some areas of the publication, but there are no explicit recommendations on how to integrate these into the HTA. CRD considers equity aspects in relation to the outcome definition (9).

Sustainability

Two organizations address the sustainability of PHIs (6;9). The Cochrane Collaboration recommends seeking data on outcome patterns as well as contextual and project factors to assess sustainability. CRD indicates that, if long-term outcomes of the review are defined, attention should be paid to the validity of interim or surrogate outcomes (9). Both organizations evaluate sustainability as part of the quality assessment of included studies (6;9).

Context

All organizations recommend considering information on context (6–9). CRD, the Cochrane Collaboration, and GOEG emphasize that the context is an important aspect for the generalizability of the results. Furthermore, CRD and GOEG suggest analyzing the effect of the context on the effectiveness of the intervention (8;9).

Applicability, Transferability, External Validity, and Generalizability

Issues related to applicability are also addressed by all organizations (6–9). NICE assesses the external validity and applicability using their own checklists (7). The Cochrane Collaboration recommends a tool to assess the applicability and transferability (6;11). GOEG and CRD do not distinguish between the four terms (8;9). Both organizations recommend applying a checklist to assess the applicability of PHIs to different contexts (8;9;11).

DISCUSSION

All four HTA manuals for PHIs consider similar (additional) methodological aspects (e.g., assessment of context). The methodology of specific steps is basically similar. However, in detail, the recommendations often vary widely without an obvious justification. Unjustified heterogeneity concerns the quality assessment, assessment of applicability and integration of qualitative and quantitative evidence. Furthermore, detailed descriptions of most process steps (e.g., integrity, heterogeneity, sustainability, context, and applicability) are missing, although these descriptions are necessary to guide the preparation for less experienced reviewers. Heterogeneity and lack of comprehensiveness are probably caused by the diversity and complexity of PHIs, which makes a detailed description and standardization of methods suitable for all research questions almost impossible.

There is a tendency for HTAs of PHIs to have an extended scope compared with medical clinical interventions. Comprehensive groundwork before conducting literature synthesis is particularly important in the area of public health to account for the high complexity of most interventions and to optimally adjust the methodology to the research question (12). Therefore, most organizations suggest exploratory work to develop the scope and a conceptual framework. There are

https://doi.org/10.1017/S0266462317000228 Published online by Cambridge University Press

different approaches using the theoretical framework in the process of preparing the HTA. The guidance documents for all organizations recommend structuring certain steps of the process (research question, inclusion criteria, data preparation, data synthesis). This demonstrates the broad applicability of a conceptual framework through the entire process of preparation for complexity in general (13). Furthermore, the theoretical framework is important as a part of the background and discussion of findings (6).

All HTA methods are based on systematic reviews (6–9). None of the organizations focus only on RCTs (6–9). There seems to be a consensus that it is unreasonable to rely merely on evidence from RCTs (6–8). The included study designs are flexibly adapted to the research question to achieve the answer in the best possible way. On the one hand, this reflects the broad perspective and complexity of PHIs (reference). On the other hand, this approach takes into account that randomized designs are often not feasible or are unethical for the evaluation of public health interventions (14;15).

No standardized terminology has evolved to label a PHI. Index terms for public health concepts are poor or unavailable, and the bibliographic language is heterogeneous (6;7;9). Therefore, search strategies should cover a wide range of databases, text words should be used freely and additional searches are important. A further challenge is the use of study filters to restrict the study design because different study designs are considered, and study designs may be termed differently (e.g., difference in difference analysis versus controlled before-and-after study). Consequently, the decision between the sensitivity and the precision of the search strategy is especially important for PHIs (6;7;9), and a more pragmatic and iterative approach may be used (7) to optimally balance this conflict.

Internal validity assessments of quantitative studies differ between the organizations. In particular, the risk of bias assessment of observational studies and qualitative studies varies between the different organizations (6;16).

All organizations agree that data extraction should include detailed information on PICO and setting. Additional aspects of data extraction, such as context, theoretical underpinning, and integrity of the intervention, vary between the manuals, and the necessary data for the individual components are not further specified. The heterogeneous and imprecise reporting might be due to the diversity and complexity of PHIs. Integrity and context should be assessed by all organizations. Statements related to the intervention integrity refer mainly to the assessment of included studies. Indeed, the Cochrane Collaboration recommends analyzing and synthesizing information related to intervention integrity and describes relevant aspects. However, it is not elucidated how the assessment and synthesis should be performed such as methods suggested for other complex interventions (17).

Standard statistical methods are predominantly suggested to deal with heterogeneity. Only NICE suggests meta-regression, and CRD suggests Harvest plots to analyze the differential effects of the intervention (7;18). We found few suggestions on how to use the different methods and no information for what type of heterogeneity (content related, statistical, methodological) the individual method is suitable for, although detailed methodological literature from the area of complex interventions on analyzing heterogeneity quantitatively still exists (e.g., graphical methods or advanced meta-analysis methods) (19–21).

GOEG and the Cochrane Collaboration address the integration of qualitative and quantitative evidence (6;8). Both segregated and integrated designs are described. Segregated methods are characterized by a prior synthesis of qualitative and quantitative studies separately, and a subsequent synthesis of the two syntheses (22). Two different segregated methods are described in the Cochrane publication: narrative synthesis and combining the findings of meta-analysis with qualitative analysis of reviews. However, neither is described in detail. GOEG mentions mixed methods as a possibility. In integrated methods, the synthesis of qualitative and quantitative data is performed in parallel and integrated design (22). Three different integrated methods are suggested: the realist reviews (6), the interactive domain model (23) (GOEG), and the multi-layered model of social determinates (Cochrane Collaboration) (24).

All organizations consider the assessment of equity/ethical aspects as important, but do not provide any detailed descriptions on evaluation methods and integration of results in the review. Recent recommendations on the consideration of equity aspects in systematic reviews entail a checklist to assess the applicability to disadvantaged population groups (25–27). However, they were developed after the development of the method guides assessed in this study.

Assessing the sustainability of outcomes is important to evaluate the long-term effectiveness of interventions. Intervention characteristics, context, capacity, processes, and interactions can all influence the sustainability of the intervention outcome/effectiveness (28). The Cochrane Collaboration provides some support in assessing the sustainability by suggesting relevant factors and assessing the pattern of outcomes. Given the large number of factors that need to be measured and reported, it is difficult to assess the sustainability of primary studies. However, one way to evaluate the sustainability is to assess the trend of outcomes after implementation of the included studies and consider the body of evidence. Another way would be the use of surrogate or intermediated outcomes because their effect on long-term outcomes has been validated.

Information on context can be used for two purposes. On the one hand, context can feed the analysis of barriers and facilitators to the intervention. On the other hand, it can be applied for the assessment of the applicability to the context (e.g., the country in which the PHI should be implemented). The organizations describe the context primarily as a part of the assessment of applicability. The approaches to assess the applicability

https://doi.org/10.1017/S0266462317000228 Published online by Cambridge University Press

of the included studies vary. The main reasons include the differing definitions of applicability, transferability, and external validity (generalizability); furthermore, some organizations use these terms interchangeably (8;9;29). Most recommend checklists to assess the applicability of the intervention. Transferability and applicability of a complex intervention can only be assessed in relation to a particular population and context (30).

Therefore, one would expect that the organizations with international target audiences to focus on the external validity because an analysis of the applicability for all imaginable patients and contexts is not possible. Consequently, an assessment of applicability seems to be feasible only if it is performed by the end-user of the HTA. The initial provision of a detailed description of applicability information and subsequent performance of an example assessment of applicability for one target population and context that guides the assessments of the end-users may serve this purpose (30). An applicability assessment of the individual included studies can be supplemented by assessing the external validity of the body of evidence, i.e., the comparison of findings across studies.

The U.S. Preventive Services Task Force (USPSTF) published a manual after our systematic literature search (31). In contrast to the included manuals, this focuses on all (individual and population level) primary and secondary preventive services and not only on PHIs. As expected, the methodology is more similar to the assessment of medical interventions (e.g., focus on RCTs) than to the assessment of PHIs as described in the included manuals. Thus, only a few suggestions dealing with the complexity of PHIs can be found in the USPSTF manual. However, it is noteworthy that the USPSTF uses modelling for linking evidence (e.g., of a diagnosis-therapy-chain).

Our study has some limitations. First, we searched only for literature specifically related to HTAs of PHIs. Literature on other complex interventions (e.g., psychological interventions) and literature on complex interventions themselves were not included. Second, we included only manuals in German, English, Spanish, and Italian. Third, in the CRD guidance, the PHIs are only one section of a larger manual. Additionally, the Cochrane manual must be considered in connection with the Cochrane Handbook of Systematic Reviews (section 3–11) (6;9;32). It can be assumed that aspects that are not addressed in the public health specific documents should be performed as described in the "higher level" manuals of the respective institutions (e.g., Cochrane Handbook) (32).

CONCLUSION

The approaches used for HTAs of PHIs seem to be generally broader and more flexible than those for clinical interventions, which can be justified by the high complexity of PHIs. There is a tendency to identify the intervention components and other factors that influence the effectiveness and transferability of interventions/HTAs (complex perspective on a complex intervention) rather than to assess the effectiveness of an intervention in a more general way (simple perspective on a complex intervention) (33).

Research is needed to further develop methods for HTAs of PHIs and ascertain the advantages and disadvantages of different approaches (e.g., realist synthesis versus systematic review), taking into account the research questions (e.g., broad versus focused perspective). This would positively contribute to the harmonization of the methods and consequently the usability for end-users (e.g., decision makers) of HTAs of PHIs. The assessment of quality, integrity, sustainability, equity aspects, context, and applicability require detailed reporting, especially if the influence of individual factors is to be quantified. A premise for accomplishing this is the improvement of reporting of primary studies.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0266462317000228

CONFLICTS OF INTEREST

The authors have nothing to declare.

REFERENCES

1. Draborg E, Gyrd-Hansen D, Poulsen PB, Horder M. International comparison of the definition and the practical application of health technology assessment. *Int J Technol Assess Health Care*. 2005;21:89-95.
2. Sackett DL. The arrogance of preventive medicine. *CMAJ*. 2002;167:363-364.
3. Lavis JN, Wilson MG, Grimshaw JM, et al. Supporting the use of health technology assessments in policy making about health systems. *Int J Technol Assess Health Care*. 2010;26:405-414.
4. Petticrew M, Chalabi Z, Jones DR. To RCT or not to RCT: Deciding when 'more evidence is needed' for public health policy and practice. *J Epidemiol Community Health*. 2012;66(5):391-6. Epub 2011/06/10.
5. Espallargues M, Pons JM, Almazan C, de Sola-Morales O. [Health technology assessment in public health interventions: Is prevention better than cure?]. *Gac sanit*. 2011;25(Suppl 1):40-48. La evaluacion de tecnologias sanitarias en intervenciones de salud publica: mas vale prevenir que curar?
6. Armstrong R, Waters E. Guidelines for systematic reviews of health promotion and public health interventions. Version 2. 2007. The Cochrane Collaboration.
7. National Institute for Health and Clinical Excellence (NICE). *Methods for the development of NICE public health guidance (third edition)*. September 2012. Paris: NICE.
8. Haas S, Breyer E, Knaller C, Weigl M. Evidenzrecherche in der Gesundheitsförderung. Band Nr. 10 aus der Reihe WISSEN (Teil 2 Kurzanleitung). [electronic ressource]. Gesundheit Österreich GmbH/Fonds Gesundes Österreich (ed.). Retrieved on 05.06.2013.
9. Akers J, Aguiar-Ibáñez R, Baba-Akbari A. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination (CRD); 2009.
10. Haas S, Breyer E, Knaller C, Weigl M. Evidenzrecherche in der Gesundheitsförderung (Wissen 10 Teil 1 Handbuch), hg. v. GÖG/FGÖ. Gesundheit Österreich / Geschäftsbereich Fonds Gesundes Österreich. Wien; 2013.

11. Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. *Health Promot Int*. 2006;21:76-83.
12. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *Int J Nurs Stud*. 2013;50:587-592.
13. Anderson LM, Petticrew M, Rehfuess E, et al. Using logic models to capture complexity in systematic reviews. *Res Synth Methods*. 2011;2:33-42.
14. Osrin D, Azad K, Fernandez A, et al. Ethical challenges in cluster randomized controlled trials: Experiences from public health interventions in Africa and Asia. *Bull World Health Organ*. 2009;87:772-779.
15. West SG, Duan N, Pequegnat W, et al. Alternatives to the Randomized Controlled Trial. *Am J Public Health*. 2008;98:1359-1366.
16. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *Int J Epidemiol*. 2007;36:666-676.
17. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review – A new method of systematic review designed for complex policy interventions. *J Health Serv Res policy*. 2005;10(Suppl 1):21-34.
18. Ogilvie D, Fayter D, Petticrew M, et al. The harvest plot: A method for synthesising evidence about the differential effects of interventions. *BMC med res methodol*. 2008;8:8.
19. Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: An overview with suggestions for practice. *Res Synth Methods*. 2010;1:66-80.
20. Baker WL, White CM, Cappelleri JC, et al. Understanding heterogeneity in meta-analysis: The role of meta-regression. *Int J Clin Pract*. 2009;63:1426-1434.
21. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Stat Methods Med Res*. 2013;22:133-158.
22. Sandelowski M, Voils CI, Barroso J. Defining and designing mixed research synthesis studies. Research in the schools: A nationally refereed journal sponsored by the Mid-South Educational Research Association and the University of Alabama. 2006. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.567.748 (accessed March 14, 2017).
23. Kahan B, Goodstadt M. The interactive domain model of best practices in health promotion: Developing and implementing a best practices approach to health promotion. *Health Promot Pract*. 2001;2:43-67.
24. Dahlgren G, Whitehead M. *Policies and strategies to promote social equity in health. Background document to WHO-Strategy paper for Europe*. Stockholm: Institute for Futures Studies; 1991.
25. Tugwell P, Petticrew M, Kristjansson E, et al. Assessing equity in systematic reviews: Realising the recommendations of the Commission on Social Determinants of Health. *BMJ*. 2010;341:c4739.
26. Welch V, Petticrew M, O'Neill J, et al. Health equity: Evidence synthesis and knowledge translation methods. *Syst Rev*. 2013;2:43.
27. Welch V, Brand K, Kristjansson E, Smylie J, Wells G, Tugwell P. Systematic reviews need to consider applicability to disadvantaged populations: Inter-rater agreement for a health equity plausibility algorithm. *BMC Med Res Methodol*. 2012;12:187.
28. Wiltsey Stirman S, Kimberly J, Cook N, Calloway A, Castro F, Charns M. The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implement Sci*. 2012;7:17.
29. Burchett H, Umoquit M, Dobrow M. How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks. *J Health Serv Res Policy*. 2011;16:238-244.
30. Burford B, Lewin S, Welch V, Rehfuess E, Waters E. Assessing the applicability of findings in systematic reviews of complex interventions can enhance the utility of reviews for decision making. *J Clin Epidemiol*. 2013;66:1251-1261.
31. U.S. Preventive Services Task Force. *Procedure manual*. Rockville, MD: U.S. Preventive Services Task Force; 2015.
32. Higgins J, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 The Cochrane Collaboration; 2011. www.cochrane-handbook.org (accessed June 12, 2016).
33. Petticrew M, Anderson L, Elder R, et al. Complex interventions and their implications for systematic reviews: A pragmatic approach. *J Clin Epidemiol*. 2013;66:1209-1214.

https://doi.org/10.1017/S0266462317000228 Published online by Cambridge University Press