Assessing context suitability (generalizability, external validity, applicability or transferability) of findings in evidence syntheses in healthcare—An integrative review of methodological guidance

Alina Weise | Roland Büchter | Dawid Pieper | Tim Mathes

Institute for Research in Operative Medicine, Faculty of Health—School of Medicine, Witten/Herdecke University, Cologne, Germany

Correspondence
Alina Weise, Institute for Research in Operative Medicine, Faculty of Health—School of Medicine, Witten/Herdecke University, Ostmerheimer Str. 200, Building 38, 51109 Cologne, Germany. Email: alina.weise@uni-wh.de

Background: Evidence syntheses provide the basis for evidence-based decision making in healthcare. To judge the certainty of findings for the specific decision context evidence syntheses should consider context suitability (ie, generalizability, external validity, applicability or transferability). Our objective was to determine the status quo and to provide a comprehensive overview of existing methodological recommendations of Health Technology Assessment (HTA) and Systematic Review (SR) producing organizations in assessing context suitability of evidence on effectiveness of health care interventions. Additionally, we analyzed similarities and differences between the recommendations.

Methods: In this Integrative Review we performed a structured search for methods documents from evidence synthesis producing organizations that include recommendations on appraising context suitability in effectiveness assessments. Two reviewers independently selected documents according to predefined eligibility criteria. Data were extracted in standardized and piloted tables by one reviewer and verified by a second reviewer. We performed a thematic analysis to identify and summarize the main themes and categories regarding recommended context suitability assessments.

Results: We included 14 methods documents of 12 organizations in our synthesis. Assessment approaches are very heterogeneous both regarding the general concepts (eg, integration in the evidence synthesis preparation process) and the content of assessments (eg, assessment criteria).

Conclusion: Some heterogeneity seems to be justified because of the need to tailor the assessment to different settings and medical areas. However, most differences were inexplicable. More harmonization is desirable and appears possible.

Keywords
applicability, context suitability, evidence synthesis, generalizability, transferability

This manuscript contains original unpublished work and is not being submitted for publication elsewhere at the same time. There were no previous submissions.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Research Synthesis Methods published by John Wiley & Sons Ltd.
Evidence syntheses such as Systematic Reviews (SR) or Health Technology Assessments (HTAs) provide the basis for evidence-based decision making in healthcare.\textsuperscript{1,2} Users of evidence syntheses may include practitioners, decision makers in health policy or patients.\textsuperscript{3} When synthesizing evidence, researchers should consider context suitability to enable users to make a judgement about whether the evidence fits their decision context.\textsuperscript{4} If there are reasons that results from an evidence synthesis may not fully apply to specific decision-making contexts, users of the evidence synthesis can incorporate these into their decision making process and make a judgement on how differences in context may affect the certainty in the evidence.

Various concepts relevant to context suitability have been suggested. Some of the most common concepts and terms used in the literature are external validity, generalizability, applicability and transferability, but there exists no general definition of these concepts and sometimes, the terms are used interchangeably.\textsuperscript{1,5-7} According to the definitions of Burford and colleagues\textsuperscript{5} researchers may assess whether the results provide a correct basis for generalizations to other circumstances (generalizability/external validity), whether it is feasible to implement an intervention in a specific context (applicability) or whether a similar level of effectiveness could be achieved if an intervention would be implemented in another specific context (transferability). To summarize these concepts and associated terms, we use context suitability as a generic term in the following.

Context suitability is not a binary concept (suitable vs nonsuitable), but should rather be regarded as a continuum (more or less context suitable).\textsuperscript{8} Moreover, in contrast to internal validity, context suitability is not static (ie, it is not always the same for one body of evidence) but depends on the specific point of reference (ie, the situation to which the results should be applied).\textsuperscript{9} Therefore, the same body of evidence might be more suitable for one decision context but less suitable for another decision context.

Context suitability of research findings can be affected by temporal changes (eg, technology changes), geographical factors (eg, healthcare infrastructure differences) or aspects of study design (eg, strict inclusion criteria).\textsuperscript{3,10} Regarding the study design different study types can be distinguished, namely “pragmatic” and “explanatory”. “Pragmatic trials” focus on clinical practice conditions while “explanatory trials” focus on idealized settings to demonstrate proof of concept. Thereby, it is assumed, that “explanatory trials” tend to have less generalizability due to homogenous populations or small sample size while “pragmatic trials” tend to be more affected by local clinical practice.\textsuperscript{8}

Context suitability assessments in evidence syntheses (eg, the assessment of applicability or transferability) received increasing interest in the last years.\textsuperscript{1,5} Several tools for assessing context suitability have been developed and suggested, but there are no widely agreed approaches for systematic reviews or health technology assessments.\textsuperscript{1,3,5,9,11-13} There are several possible explanations for this. First, several different ways exist to design an assessment, for example, regarding the assessment level (assessment of individual studies or a body of evidence) or the standardization of the tool (standardized vs nonstandardized approaches). Second, there is a lack of

**Highlights**

- **What is already known**
  - In the literature, there exist a diverse terminology regarding context suitability terms (e.g. generalisability, applicability, transferability, external validity)
  - Several tools to assess concepts (e.g. applicability) of context suitability in the field of healthcare have been developed and suggested, but there are no widely agreed approaches to assessing these concepts

- **What is new**
  - We provide a comprehensive overview of the applied methods to assessing context suitability of evidence on effectiveness of healthcare interventions recommended by evidence synthesis producing organisations with national and international target audiences
  - We found little consistency in the terminology used between the organisations. There are various concepts to assess context suitability (e.g. assessing each study individually vs. a body of evidence as a whole). Moreover, we found heterogeneous recommendations regarding assessment criteria and a lack of guidance on how to apply them
  - Potential impact for Review Synthesis Methods
    - Authors of evidence syntheses should be aware of the different concepts to assessing context suitability of evidence on effectiveness of healthcare interventions and use clearly described and transparent methods to advance the field and help readers in applying results to their context

- **Concepts**
  - Generalizability
  - Applicability
  - Transferability
  - External validity
empirical evidence on factors that affect context suitability,\cite{12} (eg, in comparison with risk of bias), for example from meta-epidemiological studies. Third, context suitability highly depends on the decision context and thus should always be adapted to the research question at hand.\cite{15-17}

Our objective was to provide an overview of existing methodological recommendations of HTA organizations and Systematic Review organizations (evidence synthesis producing organizations) in assessing the context suitability of evidence on effectiveness of healthcare interventions. This includes all methodological recommendations for assessing aspects of context suitability, irrespective of type, scope or content. In addition, we analyze similarities and differences between the methodological recommendations. For this purpose, we performed an integrative review\cite{18} of applied methodological guidance on the assessment of context suitability in the field of healthcare.

This integrative review differs from previous research, as we did not review published checklists—regardless of whether they are used in practice or not—but methods for assessing context suitability as a whole, which are actually recommended to be used in practice. In addition, we provide a comprehensive overview, on both assessment approach concepts (including assessment standardization, assessment level and integration of assessment in the quality of evidence rating or preparation process) and assessment criteria.

This is the first part of a larger research project from our team on this topic. In a second part, it is planned to review and compare methodological recommendations in assessing context suitability of evidence on cost-effectiveness in evidence syntheses. In a third part, we aim to provide an overview on how context suitability of evidence on effectiveness and cost-effectiveness is assessed in evidence syntheses and which recommended methods are applied in practice.

2 | METHODS

There is no published protocol for this integrative review. Unless otherwise indicated, all described methods were prespecified. We performed an integrative review\cite{18}, “a specific review method that summarizes past empirical or theoretical literature to provide a more comprehensive understanding of a particular phenomenon or healthcare problem.”\cite{19} We do not provide a quality assessment of included documents as, to our knowledge, no tool exists for assessing the quality of documents containing methodological recommendations.

This review is reported—as far as applicable to methodological research—according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).\cite{20}

2.1 | Search strategy

Our synthesis is based on the methodological recommendations of evidence synthesis producing organizations, defined as institutions that conduct evidence syntheses, including but not limited to SRs, HTAs or rapid reviews in the field of healthcare. We focused on organizations which make concrete recommendations for conducting evidence syntheses, meaning we excluded general methodological papers. Many evidence synthesis producing organizations exist worldwide, but it can be anticipated that the methodological guidance documents of these organizations are mostly not listed in bibliographic databases (eg, Cochrane Handbook, manuals of HTA organizations). Therefore, the documents had to be searched manually.

We performed a structured hand search on webpages of evidence synthesis producing organizations with national (eg, AHRQ) and international (eg, EUnetHTA) target audiences to identify potentially relevant methodological guidance documents. We identified the organizations through publicly available member lists of HTA umbrella organizations. The member lists of the following HTA umbrella organizations were included: European Network of Health Technology Assessment Agencies (EUnetHTA), Health Technology Assessment International (HTAI), International Agency of Health Technology Assessment (INAHTA) and Red de Evaluación de Tecnologías en Salud de las Américas (RedETSA). In total we searched 158 webpages of identified organizations (see Data S1). In addition, we searched the webpages of the following Systematic Review organizations: The Cochrane Collaboration, the Joanna Briggs Institute (JBI), the Centre for Reviews and Dissemination (CRD) and the Institute of Medicine (IoM). We chose this approach for two reasons: First, it was not feasible to search for all organizations (universities, governmental organizations etc.) that could potentially provide a methodological guidance document for every country worldwide. Second, the chosen approach is structured and transparent.

Two reviewers independently searched for methodological guidance documents on the webpages. To allow for a thorough search on webpages with different and sometimes complex site-structures we checked each section of the webpages carefully to identify all potentially relevant documents. Moreover, we translated foreign language websites using machine based and browser translation tools to identify English or German language
documents. We performed the searches from January to March 2019. Each reviewer saved the identified documents independently. The lists of saved documents were compared and synchronized manually using excel. In cases were several versions of the document existed, we only considered the latest version. Duplicates were removed manually.

2.2 Eligibility criteria and screening

We screened all identified and potentially relevant documents against the following predefined eligibility criteria:

1. Publication type: Methodological guidance documents for the preparation of evidence syntheses (e.g., handbooks, manuals, guidelines, standard operation procedures)

2. Documents include recommendations for assessing the context suitability of quantitative studies (i.e., not exclusively qualitative studies) on the effectiveness of health technologies (defined as interventions “developed to prevent, diagnose or treat medical conditions; promote health; provide rehabilitation; or organize healthcare delivery”)

3. The process of context suitability assessment is specified, for example, in the form of concrete methods, questionnaires, tools or assessment criteria

4. Recommendations on context suitability related terms like external validity, generalizability, extrapolation, transferability or applicability are considered (according to the definitions of Bufford et al)

5. Publication language: English, German

All documents that did not fit our inclusion criteria were excluded. Two researchers screened all identified documents independently. Discrepancies were resolved in a discussion until consensus was reached. The terminology in the research field is highly heterogeneous. Therefore, we considered all sections of the identified documents that could be potentially relevant for the assessment of context suitability, carefully, that is, did not focus on specific terms.

2.3 Data extraction and synthesis

We performed data extraction using a standardized piloted data extraction sheet. In a first step, we preliminarily reviewed the methods documents and developed the data extraction sheet inductively. In a second step, we piloted the data extraction sheet: One reviewer extracted data from a sample of included methods documents (from five organizations) and a second reviewer checked it for accuracy. Afterwards, we discussed and modified the data extraction sheet and re-checked methods documents that have already undergone data extraction. In a third step, one reviewer extracted data carefully and a second reviewer checked it for accuracy. Discrepancies were resolved in a discussion until consensus was reached.

When performing data extraction initially, we took the formulations as literally as possible to avoid interpretation bias. We extracted and aggregated data according to the following issues: the target audience of the evidence synthesis product (specific jurisdiction vs generic), the types of considered interventions, the underlying terminology (e.g., applicability or external validity) and its definition, information on assessment approach concepts (assessment level, assessment standardization and integration of assessment in the preparation process and in the quality of evidence rating) and the relevant assessment criteria.

We performed a thematic analysis to identify and summarize the main themes (main topics to organize and structure the analysis) and relating categories (lower hierarchical unit of themes used for further categorization) regarding context suitability assessments across the recommendations of included organizations. We developed the themes (terminology, assessment approach concepts and assessment criteria) for the analysis deductively. Related categories for the terminology (e.g., applicability) were taken from Bufford et al. The categories for the assessment approach concepts (e.g., assessment level) were developed inductively, using an iterative procedure. The categories for the assessment criteria (e.g., population, setting) were largely taken from the AHRQs recommended assessment approach and adapted inductively. To develop a set of category-specific items (lower hierarchical unit of categories, which capture details of those categories such as demographic characteristics and comorbidities) of the assessment criteria, we extracted all reported items literally and assigned them to the corresponding category. Further, we grouped the extracted items by developing a set of category-specific items (e.g., we grouped “comorbidity,” “comorbidities,” “co-morbid conditions” as “comorbidity”). One reviewer developed the themes, categories and category-specific items and a second reviewer checked them for plausibility. A glossary of developed themes and related categories can be found in
### TABLE 1  Terms and themes with definition/explanation

| Terms and themes                        | Definition/Explanation                                                                                                                                                                                                 |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| **Target audience**                     | Describes who is informed by the evidence synthesis product or to what context it relates (here: differentiation between national and international target audiences)                                                   |
| **Type of considered intervention**     | Describes whether the guideline relates to a specific type of intervention (pharmaceuticals, medical devices, medical services/procedures, diagnostics, public health interventions) or whether it is generic                                       |
| **Original terminology and definition** | Describes which context suitability term was used in the guideline (eg, external validity, generalizability, applicability, transferability) and how this was defined by the organization                                         |
| **Harmonized terminology and definition** | Due to the high heterogeneity in the terminology we decided to reassign the different definitions in the guidelines to the corresponding term of Burford et al (2013) to achieve a uniform terminology                      |
| **Applicability according to Burford et al** | “Whether the findings of a review can be applied in a particular context or population. This includes the consideration of the feasibility of implementing the intervention and variation in intervention fidelity, population characteristics, context, culture, values, and preferences.” |
| **External validity/generalizability according to Burford et al** | “The extent to which results provide a correct basis for generalizations to other circumstances”                                                                                                                        |
| **Transferability according to Burford et al** | “Whether when implementing an intervention in a particular setting or population, the level of effectiveness of the intervention (ie, the effect size) will be similar to that observed in the systematic review.” |
| **Assessment approach concepts**        | Describes how the context suitability assessment is operationalized. This includes the integration of assessment in the preparation process, the assessment level, the assessment standardization and the integration of assessment in the quality of evidence rating |
| **Integration of assessment in the preparation process** | Context suitability can be integrated at different steps of the evidence synthesis preparation process. It can be considered in study selection, in study assessment, or both                                  |
| In study selection                      | Consideration of context suitability, when defining and applying eligibility criteria                                                                                                                                   |
| In study assessment                     | Assessment of context suitability of included evidence                                                                                                                                                                 |
| **Assessment level**                    | Describes, whether the context suitability assessment is applied to each individual study, only to the whole body of evidence, or both                                                                                   |
| Individual study assessment             | Context suitability is assessed for each included study individually                                                                                                                                                 |
| Body of evidence assessment             | Context suitability is assessed for all included studies as a whole                                                                                                                                                   |
| **Assessment standardization**          | Context suitability assessment may follow a clear structure or not. We differentiate between standardized and nonstandardized approaches. The first is further divided into previously developed assessment approaches and approaches developed by the organization itself |
| Nonstandardized assessment approach     | A nonstandardized assessment approach does not follow a clear structure                                                                                                                                              |
| Own standardized assessment approach    | An own standardized assessment approach follows a clear structure and may comprise different steps, checklists or questions that have to be followed or rated. It was developed by the guidance producer itself                          |
| Previously standardized assessment approach | A previously standardized assessment approach can be defined as an approach which was not developed by the guidance producer itself, but refers to a standardized assessment approach of other researchers              |
| **Integration of assessment in the quality of evidence rating** | Context suitability can be assessed as a part of other quality of evidence appraisals or not. We differentiate between standalone and integrated assessment approaches                                      |
| Standalone assessment                   | In a standalone assessment approach, context suitability is assessed independently from other aspects for assessing the quality of evidence, in particular the assessment of internal validity |
| Integrated assessment                   | In an integrated assessment approach, context suitability is combined with other criteria to assess the quality of the body of evidence, in particular the assessment of internal validity                 |
| **Assessor**                            | Context suitability can be assessed by the evidence synthesis producing organization itself, or by the reader of the evidence synthesis, or both. If the reader is the assessor, the information that is necessary for a judgement regarding context suitability should be provided in the evidence synthesis |
in Table 1. We describe our findings in a structured narrative manner and visually present extracted data using tabulations.

3 | RESULTS

3.1 | Literature search

We identified 155 potentially relevant methods documents on the webpages of four organizations with an international target audience. After a manual duplicate removal (some documents were found on more than one webpage) a total of 155 methods documents from 67 organizations remained for screening. Finally, we included 14 methods documents from 12 organizations in our synthesis. In addition, we identified five potentially relevant methods documents on the webpages of four organizations with an international target audience. After a manual duplicate removal (some documents were found on more than one webpage) a total of 155 methods documents from 67 organizations remained for screening. Finally, we included 14 methods documents from 12 organizations in our synthesis.10,15-17,23-32 The document selection process is illustrated in Figure 1. A list of excluded publications can be found in the appendix (Data S2).

Ten eligible publications were obtained from nine organizations with a national target audience (Agency
## Table 2: Guidance characteristics

| Organisation, Country                  | Title, year                                                                 | Terminology                  | Definition                                                                                                                                                                                                                                                                                                                                 | Harmonized terminology according to Burford et al (2013) | Target audience |
|----------------------------------------|------------------------------------------------------------------------------|------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|-----------------|
| ACE Agency for care effectiveness, Singapore | Medical Technologies Evaluation Methods and Process Guide Version 1.0, 2018 | Applicability                | The extent to which the published evidence reflects the decision problem in the local context                                                                                                                                                                                                                                                  | NA*                                                        | National        |
| AHRQ Agency for Healthcare Research and Quality, USA | Methods Guide for Effectiveness and Comparative Effectiveness Reviews, 2014 Methods Guide for Medical Test Reviews, 2012 | Applicability                | The extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under “real-world” conditions                                                                                                                                               | Transferability                                          | National        |
| AHTAPol The Agency for Health Technology Assessment and Tariff System, Poland | Health Technology Assessment Guidelines Version 3.0, 2016 | External validity, applicability or generalisability | External validity means the possibility of generalising conclusions from a study to the target population for a given health technology                                                                                                                                                                                                 | External Validity/Generalisability*                       | National        |
| Cochrane Collaboration                  | Cochrane Handbook for Systematic Reviews of Interventions Version 6.0, 2019 | Applicability, generalisability, external validity or transferability | All of these terms describe one overarching theme: whether or not available research evidence can be directly used to answer the health and healthcare question at hand, ideally supported by a judgement about the degree of confidence in this use                                                                 | NA                                                        | International   |
| CRD Centre for Reviews and Dissemination, United Kingdom | Systematic Reviews – CRD’s guidance for undertaking reviews in health care, third edition, 2009 | Applicability                | The extent to which public health interventions are expected to work in other contexts                                                                                                                                                                                                                                                   | Applicability                                             | National and international |
|                                         |                                                                              | Generalisability or external validity | The degree to which the results of a study hold true in other settings                                                                                                                                                                                                                                                                      | Transferability                                           |                 |
| Organisation, Country | Title, year | Terminology | Definition | Harmonized terminology according to Burford et al (2013) | Target audience |
|-----------------------|-------------|-------------|------------|--------------------------------------------------------|----------------|
| **EUnetHTA** European Network for Health Technology Assessment, Europe | HTA Core Model Version 3.0, 2016 | Applicability or external validity, generalisability, transposability | The extent to which the effects observed in clinical studies are likely to reflect the expected results when specific intervention is applied to the population of interest | Transferability | International (Europe) |
| | Levels of Evidence – Applicability of evidence in the context of a relative effectiveness assessment, adapted version 2015 | Transferability | Transferability is about the ability to apply information from one report into a user’s target setting | | |
| **GÖG Gesundheit Österreich GmbH, Austria** | Methodenhandbuch für Health Technology Assessment Version 1.2012, 2012 | External validity | External validity refers to the applicability of study results to a different population, than the study population | External Validity/ Generalisability | National |
| **LBI Ludwig Boltzmann Institut, Austria** | Guidelines for Evaluating the Clinical Effectiveness of Health Technologies in Ireland, 2018 | External validity | The extent to which the findings of the trial can be generalised to other settings or populations | External Validity/ Generalisability | National |
| **HIQA Health Information and Quality Authority, Ireland** | Finding what works in health care – Standards for Systematic Reviews, 2011 | Applicability, relevance, directness or external validity | The extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under ‘real-world’ conditions | Transferability | National |
| **IoM Institute of Medicine of the National Academies, USA** | (Internes) Manual – Abläufe und Methoden, 2007 | External validity or generalisability | NR | | National |
| **LBI Ludwig Boltzmann Institut, Austria** | Diagnostics Assessment Programme manual, 2011 | Applicability, external validity or Generalisability | How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered | Applicability | National |
| **NICE National Institute for Health and Care Excellence, England and Wales** | Assessment of methods in healthcare – a handbook, 2018 | Generalisability | The extent to which the scientific evidence is relevant and can be applied to Swedish conditions | Applicability | National |

*NA: Not assignable.

*Generalizability and external validity have the same definition according to Burford et al (2013).

*NR: Not reported.
for care effectiveness [ACE],27 Agency for Healthcare Research and Quality [AHRQ],16,28 The Agency for Health Technology Assessment and Tariff System [AHTAPol],29 Gesundheit Österreich GmbH [GÖG],15 Health Information and Quality Authority [HIQA],23 Institute of Medicine [IoM],26 Ludwig Boltzmann Institute [LBI],15,17 National Institute for Health and Care Excellence [NICE],24 Swedish Agency for Health Technology Assessment and Assessment of Social Services [SBU].

The additional four eligible publications were obtained from three organizations with an international target audience (Cochrane Collaboration,10 Centre for Reviews and Dissemination [CRD],31 European Network for Health Technology Assessment [EUnetHTA]).25,32

### 3.2 Analysis of guidance for assessing context suitability

#### 3.2.1 Types of considered interventions

Most guidelines (9/14) are not restricted to a specific type of intervention.10,16,17,23,25,26,30-32 One guideline is specific to pharmaceuticals,29 one to nonpharmaceuticals27 and two to diagnostics24,28 while one guideline could not be assigned to a specific type of intervention and thus can be assumed to be generic.15

Despite the provision of generic methodological recommendations regarding the assessment of context suitability, some organizations refer to some specifications for certain interventions (eg, Public Health Interventions [“a broadly defined set of activities that aim to protect, promote, and restore the health of all people”31; PHI]).31

#### 3.2.2 Terminology

Table 2 provides an overview of the used terminology and definitions as well as other guidance characteristics like the organization, country and target audience.

The 12 included evidence synthesis producing organizations use the terms generalizability, applicability, external validity, transferability, transposability, relevance and/or directness with different, missing or sometimes conflicting definitions. According to the definitions of Burford et al5 three organizations use the term applicability,24,30,31 four the terms external validity/generalisability10,15,23,29 and four organizations the term transferability.16,25,26,28,31,32 Some terms could not be reassigned to the terminology of Burford et al5, because the corresponding definitions were not reported or inconclusive.10,17,27

### 3.3 Assessment approach concepts

All assessment approach concepts are defined in Table 1. An overview of the assessment approach concepts recommended by the organizations is provided in Table 3. Further details on the recommended assessment approaches are available in Data S3. We did not structure the results according to the different context suitability related terms, because we could not find a relation between the used terminology and the recommended assessment approach concepts. However, Table 3 depicts the results in accordance with the harmonized terminology of Burford et al5.

Context suitability considerations can be applied in the study selection process, when defining and applying eligibility criteria, or in a later step, when assessing context suitability of included research findings. Moreover, a combination of both concepts is possible. Four organizations recommend such a combined approach.15,16,28,31 According to AHRQ16,28 and GÖG/LBI15 researchers should consider excluding studies with characteristics that would limit context suitability too much or make adaptations impossible. CRD31 and GÖG/LBI15 also point out that too narrowly defined eligibility criteria in evidence syntheses might affect context suitability.

While four organizations additionally consider context suitability in study selection, all organizations recommend a study assessment of context suitability.10,15-17,23-32 These study assessments can vary regarding the assessment level, the assessment standardization and the integration of assessment in the quality of evidence rating.

Regarding the assessment level five organizations recommend an individual assessment of context suitability for each included study (individual study assessment),15,17,23,24,29 two recommend assessing context suitability only for the whole body of evidence (body of evidence assessment)27,30 and five organizations recommend both.10,16,25,28,31,32

Regarding the assessment standardization, seven organizations recommend a nonstandardized approach,10,23,24,26,27,29,31 four developed a standardized approach15-17,28,30 and two recommend a previously developed standardized approach.25,31,32 CRD31 proposes the “List of questions to ask in determining applicability and transferability” from Wang et al2 and EUnetHTA25,32 recommends AHRQ’s16 “Four Specific Steps in Assessing and Reporting Applicability.”

Regarding the integration of assessment in the quality of evidence rating, six organizations recommend assessing context suitability independently from other aspects of quality of evidence (standalone assessment),15-17,23-25,28 while five organizations recommend an integrated assessment. Specifically, they recommend...
| Terminology\(^a\) | Organizations | Integration of assessment in the preparation process | Assessment level | Assessment standardization | Integration of assessment in the quality of evidence rating | Assessor |
|-----------------|---------------|-----------------------------------------------|----------------|--------------------------|-----------------------------------------------|----------|
| Generalizability/External Validity | AHTAPol\(^{29}\) | X                     | X               | X                        | X                                             | X        |
|                     | GÖG/LBI\(^{15}\) | X                     | X               | X                        | X                                             | X        |
|                     | HIQA\(^{23}\)   | X                     | X               | X                        | X                                             | X        |
| Applicability      | NICE\(^{24}\)  | X                     | X               | X                        | X                                             | X        |
|                     | SBU\(^{30}\)    | X                     | X               | X                        | X                                             | X        |
|                     | CRD (PHI)\(^{31}\) | X                 | X               | X                        | X                                             | X        |
| Transferability    | AHRQ\(^{16-28}\) | X                     | X               | X                        | X                                             | X        |
|                     | CRD (generic)\(^{33}\) | X                     | X               | X                        | X                                             | X        |
|                     | EUneHTA\(^{21,32}\) | X                 | X               | X                        | X                                             | X        |
|                     | IoM\(^{26}\)    | X                     | X               | X                        | X                                             | X        |
| Unclear            | ACE\(^{27}\)    | X                     | X               | X                        | X                                             | X        |
|                     | Cochrane\(^{10}\) | X                     | X               | X                        | X                                             | X        |
|                     | LBI\(^{17}\)    | X                     | X               | X                        | X                                             | X        |

\(^a\)Harmonized terminology according to Burford et al.\(^{5}\)
| Aspects                          | Generalisability/External Validity<sup>a</sup> | Applicability<sup>a</sup> | Transferability<sup>a</sup> | Unclear<sup>a</sup> |
|---------------------------------|-----------------------------------------------|--------------------------|----------------------------|------------------|
|                                 | AHTA-Pol<sup>29</sup>  | GÖG/LBI<sup>15</sup> | HIQA<sup>23</sup> | NICE<sup>b</sup> | SBU<sup>30</sup> | AHRQ<sup>16,28</sup> | CRD<sup>31</sup> | EUnet-HTA<sup>25,32</sup> | IoM<sup>26</sup> | ACE<sup>27</sup> | Cochrane<sup>10</sup> | LBI<sup>17</sup> |
| Population                      | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Eligibility Criteria            | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Demographics                    | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Socio-demographic status        | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Socio-economic status           | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Severity/stage of illness       | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Comorbidities                   | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Risk factors                    | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Genetic makeup                  | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Biological factors              | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Event rates                     | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Prevalence of disease           | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Aetiologies of disease          | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Proportion of screened patients enrolled | X                          | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Clinical parameters             | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Exclusion rate in run in period | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Sample size                     | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Intervention                    | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Intervention performance<sup>c</sup> | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Treatment regimen<sup>d</sup>   | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Clinical relevance              | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Relevance for current practice  | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Co-interventions                | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Adherence                       | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Mechanics of intervention and implementation process | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |
| Feasibility of intervention in real life settings | X                                             | X                        | X                          | X                | X                  | X                      | X                      | X                  | X                | X                  | X                  |

<sup>a</sup> Indicates the assessment criteria for the context suitability assessment.
| Aspects                | Generalisability/External Validity<sup>a</sup> | Applicability<sup>a</sup> | Transferability<sup>a</sup> | Unclear<sup>a</sup> |
|-----------------------|-----------------------------------|------------------------|--------------------------|-------------------|
|                       | AHTA-Pol<sup>29</sup> | GÖG/LBI<sup>15</sup> | HIQA<sup>23</sup> | NICE<sup>b</sup><sup>24</sup> | SBU<sup>30</sup> | AHRQ<sup>16,28</sup> | EUnet-HTA<sup>25,32</sup> | IoM<sup>26</sup> | ACE<sup>27</sup> | Cochrane<sup>10</sup> | LBI<sup>17</sup> |
| Comparator            | Comparator in general            | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
| Dosage                |                                   |                        |                         |                  |                   |                   |                     |                   |                   |                     |                   |
| Appropriateness       |                                   |                        |                         |                  |                   |                   |                     |                   |                   |                     |                   |
| Relevance             |                                   |                        |                         |                  |                   |                   |                     |                   |                   |                     |                   |
| Outcomes              | Composite outcomes               |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       |                                   |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Patient relevant (no surrogate or short-term) outcomes |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Definition of outcomes            |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Treatment trends                  |                         |                         |                  |                   |                   |                     |                   |                   |                     | X                  |
|                       | Length of follow up               | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
| Safety                | Side effects                      | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Evaluation of adverse events      | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Safety in vulnerable groups       |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Factors that may contribute to the occurrence of adverse events: |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | - Patient characteristics          |                         |                         |                  |                   |                   |                     |                   |                   |                     | X                  |
|                       | - Type of disease                  | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | - Severity of disease              | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | - Comorbidities                    | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | - Clinical setting                 |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
| Setting/context       | Setting/context in general        | X                      | X                       |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Study design                       | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Study conduction                  | X                      |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Study county                       | X                      | X                       |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Participation rate                 |                         |                         |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Care pathways                      |                         |                         |                  |                   |                   |                     |                   |                   |                     | X                  |
|                       | Available technologies             |                         |                         |                  |                   |                   |                     |                   |                   |                     | X                  |
|                       | Standards of care                  | X                      | X                       |                  |                   |                   |                     |                   |                   |                     |                   |
|                       | Specialty population               | X                      | X                       |                  |                   |                   |                     |                   |                   |                     |                   |

(Continues)
| Aspects                                      | Generalisability/External Validity<sup>a</sup> | Applicability<sup>a</sup> | Transferability<sup>a</sup> | Unclear<sup>a</sup> |
|----------------------------------------------|-----------------------------------------------|---------------------------|-----------------------------|---------------------|
|                                              | AHTA-Pol<sup>29</sup> GÖG/LBI<sup>15</sup>        | HIQA<sup>23</sup> NICE<sup>b</sup> 24 | SBU<sup>30</sup> AHRQ16,28 CRD<sup>31</sup> EUnet-HTA<sup>25</sup> 32 | IoM<sup>26</sup> ACE<sup>27</sup> Cochrane<sup>10</sup> LBI<sup>17</sup> |
| Level of care                                | X                                             | X                         | X                           | X                   |
| Primary, secondary, tertiary care             | X                                             | X                         | X                           | X                   |
| Everyday practice                            | X                                             | X                         | X                           | X                   |
| Context factors within host organizations     |                                               |                           |                             |                     |
| Fee or payment structure                     |                                               |                           |                             |                     |
| Insurance system                             |                                               |                           |                             |                     |
| Variation in values and preferences          |                                               |                           |                             |                     |
| Characteristics of the population:           |                                               |                           |                             |                     |
| -Cultural and linguistic diversity           |                                               |                           |                             |                     |
| -Socioeconomic position                      |                                               |                           |                             |                     |
| -Rural/urban setting                         |                                               |                           |                             |                     |
| Implementation of PHI in a specific setting<sup>e</sup> |                                               |                           | X                           |                     |
| -Responsibilities                            |                                               |                           | X                           |                     |
| -Implementation barriers                     |                                               |                           | X                           |                     |
| -Provider skills                             |                                               |                           | X                           |                     |
| -Resource availability                       |                                               |                           | X                           |                     |

<sup>a</sup>Harmonized terminology according to Burford et al.<sup>5</sup>

<sup>b</sup>Specific for diagnostics.

<sup>c</sup>Intervention performance: How the intervention is delivered (eg, everyday conditions vs study conditions, visit frequency not used in typical practice).

<sup>d</sup>Treatment regimen: eg, dose, schedule, duration of intervention.

<sup>e</sup>In case of PHI applicability instead of transferability is meant.
considering context suitability along with study quality,\textsuperscript{26,27,29} evidence grading\textsuperscript{26,30} and/or when formulating conclusions.\textsuperscript{10} CRD\textsuperscript{31} distinguishes between a generic context suitability assessment and a specific assessment for PHI. For the generic assessment an integrated assessment is recommended, while for PHI a standalone assessment should be performed.

### 3.4 | Assessment criteria

An overview of all relevant assessment criteria that are considered by the organizations, when assessing context suitability of evidence on effectiveness data, is provided in Table 4. More details regarding the assessment criteria can be found in the appendix (see Data S4). We did not structure the results for the assessment criteria according to the different context suitability related terms, because we could not find a relation between the used terminology and the recommended assessment criteria. However, Table 4 depicts the relevant assessment criteria in accordance with the associated terminology according to Burford et al.\textsuperscript{5}

### 3.5 | Population

All included organizations recommend taking population characteristics into consideration, when assessing context suitability of evidence on effectiveness of healthcare interventions. The most commonly recommended aspects are demographic characteristics (eg, age, sex, race, ethnicity),\textsuperscript{10,16,23,25,26,28,31} severity or stage of illness,\textsuperscript{10,16,23,25,26,28,31} comorbidities\textsuperscript{10,16,23-26} and the study’s eligibility criteria.\textsuperscript{15,17,25,26,28,31} In addition, three organizations\textsuperscript{16,25,26} recommend considering event rates and the exclusion rate within run-in period as a population-based item. Moreover, AHRQ,\textsuperscript{16} GÖG/LBI\textsuperscript{15} and LBI\textsuperscript{17} recommend the consideration of sample size. The AHRQ\textsuperscript{16} and the NICE\textsuperscript{24} also take genetic factors and etiology of disease into account. Other relevant items are variations between subgroups,\textsuperscript{10} biological variations,\textsuperscript{10,16} risk factors\textsuperscript{23} and the proportion of screened patients enrolled.\textsuperscript{16,25}

### 3.6 | Intervention

With exception of ACE,\textsuperscript{27} all organizations consider interventional aspects in their methodological recommendations for the context suitability assessment. Treatment regimen\textsuperscript{10,15-17,23,25,26,28} intervention performance\textsuperscript{10,15,16,25,26,28,30} and adherence\textsuperscript{10,15,16,23,25,26} are the most frequently reported aspects. Moreover, four organizations consider the clinical relevance of the intervention in their guidelines.\textsuperscript{15-17,25} Co-interventions and the interventions’ relevance to current practice are considered in the methodological recommendations of the AHRQ,\textsuperscript{16} EUnetHTA\textsuperscript{25} and IoM.\textsuperscript{26} GÖG/LBI\textsuperscript{15} also consider the treatment duration in their assessment.\textsuperscript{15} In addition, mechanics of the intervention and implementation process (of PHI) are mentioned by the CRD,\textsuperscript{31} while the Cochrane Collaboration\textsuperscript{10} recommends evaluating the feasibility of the intervention implementation in practice settings.

### 3.7 | Comparator

Six out of 12 organizations include comparator interventions as a relevant aspect in their methodological recommendations.\textsuperscript{10,16,21,23,25,26,28,30} AHRQ,\textsuperscript{16,28} EUnetHTA\textsuperscript{25} and the IoM\textsuperscript{26} describe inadequate dosing of the comparator drug and the use of a substandard alternative therapy as conditions that may limit context suitability. The SBU\textsuperscript{30} recommends assessing the control group’s relevance. Moreover, the Cochrane Collaboration\textsuperscript{10} recommends collecting relevant details of comparator interventions to help review users in assessing the context suitability to their target audience.

### 3.8 | Outcome

Seven out of 12 organizations include the outcome as a relevant aspect in their context suitability assessment approach.\textsuperscript{15-17,23,25,26,28,29} All of them differentiate between surrogate endpoints and clinically/patient relevant endpoints, when assessing context suitability. In addition, the AHRQ,\textsuperscript{16} EUnetHTA\textsuperscript{25} and the IoM\textsuperscript{26} define composite outcomes, which mix outcomes of different significance, as a condition that may affect context suitability. Moreover, AHRQ\textsuperscript{28} recommends the consideration of treatment trends\textsuperscript{28} and HIQA,\textsuperscript{23} AHRQ\textsuperscript{16} and LBI\textsuperscript{17} the length of follow-up.

### 3.9 | Safety

Safety items are included in the methodological guidelines of EUnetHTA,\textsuperscript{32} GÖG/LBI,\textsuperscript{15} LBI\textsuperscript{17} and CRD.\textsuperscript{31} According to the CRD\textsuperscript{31} several factors such as patient or disease characteristics may contribute to the occurrence of adverse events and therefore may limit the context suitability of study results.\textsuperscript{31} EUnetHTA,\textsuperscript{32} however, points out that safety data may not be suitable for vulnerable groups including people with comorbidities, children
or pregnant women. Moreover, a valid evaluation of adverse events is a crucial point in the assessment methods of GÖG/LBI\textsuperscript{15} and LBI.\textsuperscript{17}

3.10 | Setting

With exception of AHTAPol,\textsuperscript{29} all organizations make recommendations regarding the consideration of settings, when assessing context suitability. Care pathways,\textsuperscript{27} availability of health technologies,\textsuperscript{27} care sector\textsuperscript{15,17,23} or routine clinical practice\textsuperscript{15,31} are some of the recommended items for the context suitability assessment. Others include aspects of study design features such as study design,\textsuperscript{15,16} study conduct\textsuperscript{15,23} or study country\textsuperscript{15,24} in their assessment approaches. Moreover, the guidelines of AHRQ,\textsuperscript{16} EUnetHTA\textsuperscript{25} and IoM\textsuperscript{26} stipulate a comparison of standards of care, level of care and specialty population of the study setting and the setting of interest. The Cochrane Collaboration\textsuperscript{10} names many different contextual factors, which might be relevant to the context suitability assessment. Some of these factors are context factors within host organizations, fee or payment structure, insurance system or variation in values and preferences. The recommendations of CRD\textsuperscript{31} refer to PHI. Researchers should consider how closely a study reflects routine practice and should find out how the intervention could be implemented in the target setting. Here, relevant items are responsibilities, implementation barriers, provider skills or resource availability.

4 | DISCUSSION

4.1 | Summary of findings and interpretation

Our integrative review summarizes the applied methodological recommendations on context suitability assessments of 12 organizations with national and international target audiences. It depicts the status quo and provides a comprehensive overview of methods for considering context suitability.

However, the few hits of our structured search for applied methodological recommendations of evidence synthesis producing organizations show that the topic still receives little attention in practical application. From 162 organizations searched and 155 documents screened only 12 organizations (14 documents) made recommendations on assessing context suitability of evidence on effectiveness. This is surprising considering that most of these organizations perform evidence syntheses to develop recommendations. In future work, we aim to investigate whether and how context suitability is addressed in evidence syntheses in practice.

4.2 | Terminology

Our analysis of different context suitability terms and definitions reassigned to the terminology of Burford et al\textsuperscript{5} agreed with the inconsistent use of these terms in the literature.\textsuperscript{1} Only 7 out of 16 terms matched the corresponding definition of Burford et al.\textsuperscript{5} Moreover, three terms could not be reassigned to the terminology, because the corresponding definitions were reported or inconclusive.

The terms generalizability, applicability and transferability (according to Burford et al\textsuperscript{5}) have different scopes. Therefore, one would expect organizations focusing on a specific target audience to recommend the assessment of applicability or transferability, as these concepts of context suitability can only be assessed in relation to a specific context or population.\textsuperscript{12} In contrast, one would expect organizations with an international target audience to focus on generalizability, as they have no specific target audience and it is impossible to assess applicability or transferability for all imaginable contexts and populations. But, only one organization with an international target audience focused on generalisability.\textsuperscript{10}

Furthermore, there are different conditions that have to be satisfied to fulfil the respective definitions of generalizability, applicability and transferability. For example, the term applicability focuses on aspects which might affect the implementation of an intervention, while transferability also implies that comparable levels of effectiveness can be achieved (eg, effect sizes).\textsuperscript{5} Therefore, it could be expected that the depth of the assessment depends on the underlying concept. However, we did not find such a relation. For example, from three organizations focusing on applicability (according to Burford et al\textsuperscript{5}) only one organization specifically includes assessment criteria for implementation (eg, “Are essential resources for implementing the intervention available in the local setting?”\textsuperscript{31}). There are different possible reasons for this finding. First, the comprehensiveness of assessment criteria may depend more on the importance the organizations place on this topic than the specific terms and definitions used. Second, the depth of assessment could also be related to the standardization of the approach rather than to the underlying terminology.

4.3 | Assessment approach concepts

Our comparison of the different methodological recommendations shows strong heterogeneity regarding the
different assessment approach concepts, which confirms prior research.1,3,9,33

The assessment approaches are integrated at different steps in the evidence synthesis producing pathway. Eight organizations exclusively recommend considering context along with study assessments, while four organizations additionally take aspects of context suitability into consideration when defining and applying eligibility criteria. However, the recommendations for considering context suitability in study selection are limited to basic considerations and do not include specific recommendations for operationalizations.

Half of the assessment approaches are based on integrated assessments, that is, as part of the assessment of study quality, the assessment of the quality of the body of evidence or the discussion and interpretation of results. The other half is based on standalone assessments. AHRQ even advises against using context suitability assessments as a part of the evidence grading.16 This view is shared by other researchers.3 According to Munthe-Kaas et al3 the (exclusive) consideration of context suitability as a part of evidence grading (eg, GRADE) can be problematic, because of a lack of standardization and frequent use of ad hoc assessments. Interestingly, the GRADE system integrates context suitability (namely “[in]directness”) with other criteria that affect the confidence in an effect estimate (risk of bias, imprecision, inconsistency, publication bias), thus mixing internal validity with context suitability concerns.34 To a large extent it remains unclear, however, how findings from context suitability assessments should be considered within the GRADE framework.15,17,23,26 Regarding the assessment level most organizations recommend a context suitability assessment of individual studies or a combined approach, while only two organizations recommend the exclusive assessment of a body of evidence.

In most cases, the context suitability assessment is based on a nonstandardized approach, while six organizations recommend a standardized approach. Few of the recommended assessment approaches seem to be systematically developed or tested. In particular, the nonstandardized approaches are less structured and provide few instructions for users, which may lead to unsystematic or unreproducible assessments. However, too standardized assessments based on formalized checklists or questionnaires may not perfectly fit the specific purposes (eg, contain irrelevant assessment criteria).

Assessing context suitability is particularly challenging for organizations that address international target audiences with a large range of readers in different contexts, because it seems impossible to address all imaginable contexts. In this case one specific approach is the provision of sufficient information on PICO and setting, which allows end-users to conduct their own assessment. This approach is described in the literature as an alternative or addition to context suitability assessments.14 However, from four organizations with an international target audience only the Cochrane Collaboration10 recommends this approach.

Five organizations differentiate between explanatory and pragmatic studies instead of or as a part of the context suitability assessments.15-17,23,32 Interestingly, the organizations make different judgements about the consequences of a pragmatic trial design. Four of them consider them to be more suitable to routine practice settings,15-17,23 while EUnetHTA32 considers pragmatic trials to have reduced context suitability, as they are affected by local clinical practices. This disagreement between the recommendations seems to rest on two opposing lines of arguments regarding the context suitability of pragmatic trial designs. With their focus on clinical practice conditions, pragmatic studies may, on the one hand, be more context suitable than explanatory studies, as they better reflect reality.8,35,36 On the other hand, this focus may reduce context suitability in cases where clinical practice conditions differ from the context of interest.32 Similar opposing lines of arguments can be found with regard to the a priori consideration of context suitability in study selection. On the one hand broad eligibility criteria may better reflect reality. On the other hand context suitability might decrease, when studies with very specific characteristics that do not fit many other contexts, are included. This suggests that evidence synthesists that consider broad eligibility criteria should consider using specific tools to disentangle the possible effects of context relevant issues (eg, difference in population), using methods such as subgroup analyses, meta-regressions or matrix based techniques.37

4.4 | Assessment criteria

Most organizations either provide examples or a list of relevant assessment criteria or specific questions to consider when assessing context suitability. Burchett et al,9 who mapped and studied applicability assessment tools (for PHI) question the usability and usefulness of tools in checklist style and strict assessment criteria. The same is true for AHRQ16—that developed a formalized assessment approach which is also recommended by other organizations (EUnetHTA,32 IoM36) - based on the rationale that context suitability considerations always depend on the research question and the clinical area, as well as the intervention of interest. Therefore, the first step of AHRQ’s assessment approach is to determine the most important factors that might affect context suitability using the PICOS framework. AHRQ also provides a
list of factors potentially affecting context suitability. Moreover, it recommends exclusively considering known effect modifiers, focusing on items, which might affect outcomes.\(^{16}\) The Cochrane Collaboration\(^{10}\) and EUnetHTA\(^{25,32}\) also focus on effect modification.\(^{10,25}\)

There is high heterogeneity regarding the content of recommended assessment criteria. While most organizations use a generic concept, some organizations refer to specific considerations for certain types of interventions such as PHI or diagnostics. Population, intervention and setting are the most frequently and most comprehensively reported categories. Similarities between study setting and target setting regarding population characteristics such as demographics or stage of illness and intervention characteristics including treatment regimen, intervention performance or adherence are frequently recommended assessment criteria. Furthermore, the healthcare setting is an important aspect in most of the recommended assessment approaches. Such healthcare setting characteristics include different standards of care or the availability of technologies. Furthermore, consideration of the patient-relevance of outcomes is often part of the assessment approaches.

We found no indication, that the high heterogeneity regarding the recommended assessment criteria is related to the different terms (eg, applicability or generalizability) used by the organizations. The assessment criteria vary within and between the different terms and definitions used and most assessment criteria seem not to be term/concept-specific. Only the recommended assessment criteria of AHRQ,\(^{16}\) EUnetHTA\(^{25}\) and IoM\(^{26}\) show a significant agreement. However, this can be attributed to the fact that EUnetHTA and IoM adopted a large part of the assessment criteria, as well as the “context suitability” definition exactly from AHRQ. Moreover, it was not possible to find other reasons for the heterogeneity regarding the assessment criteria, as only few organizations (AHRQ,\(^{16}\) Cochrane\(^{10}\) or GÖG/LBI\(^{15}\)) provide explanations why certain assessment criteria were selected or why they might be relevant to assessing context suitability. The lack of justifications for including certain assessment criteria might be contributed to the fact that little empirical evidence on assessment criteria for context suitability exists.\(^{14}\) In accordance, Dyrvig and colleagues,\(^{33}\) who conducted a SR on checklists for external validity, found out that none of the reported assessment criteria were justified by empirical data.

Moreover, in many cases it remains unclear how the recommended assessment criteria should be applied, when judging context suitability. There are several reasons for this. First, most assessment approaches are generic and include relatively unspecific questions and criteria. Second, most methodological guidance documents do not provide any instructions or examples for users on how to use or operationalize the assessment tools. Missing instructions and guidance were also reported by Burchett et al.\(^{9}\) Third, we could not find any recommendations on how to weight different context suitability issues. This lack of instructions on weighting issues is confirmed by Burford et al.\(^{5}\) who analyzed context suitability assessment tools for complex interventions. All these restrictions and missing information on how to apply context suitability tools could have a negative impact on the reliability of assessments.

### 4.5 Other research

Prior research reviewing methodological guidelines for SR on health economic evaluations\(^{13}\) and HTAs of public health interventions\(^{12}\) also show high heterogeneity regarding the proposed context suitability assessment. The heterogeneity includes aspects of terminology, assessment approach concepts and/or relevant assessment criteria. Moreover, a variety of instruments and checklist to assess context suitability have been developed and reviewed.\(^{2,3,5,9,33,38,39}\) There are varying recommendations also within the same scope (eg, context suitability of economic evaluations, public health interventions, systematic reviews) or the same terminology (eg, external validity, applicability) and there seems to be no tool with wide acceptance in any area.\(^{1,3-5,16}\) Our review confirms these studies, as none of the included evidence synthesis producing organizations (except CRD\(^{31}\) for PHI) recommends one of these previously developed assessment approaches without adaptions. A promising solution might be a more flexible approach guided by stakeholders that is separated from other aspects of assessing the validity of the body of evidence.\(^{3}\)

### 4.6 Limitations

Our integrative review has some limitations. First, in some cases, it was unclear whether the statements in the guidelines could be understood as recommendations or rather as general scientific statements on context suitability. Second, many aspects left plenty of scope for user interpretation, as they were not described in detail or the guidelines only contained some examples. Third, we excluded some documents due to language restrictions. Fourth, we had some difficulties regarding the hand searching on the webpages of organizations as some webpages had confusing web-site structures and we did not contact HTA organizations or SR organizations for unpublished methods documents. Moreover, the
methodological recommendations included in this review are restricted to those specific organizations, which could be identified in a structured way by using the member lists of international HTA umbrella organizations and major SR organizations. Sixth, the methods of an integrative review are not fully applicable for our purpose. Therefore we had to adapt these methods.

5 | CONCLUSION

A large number of different approaches exist on how to consider concepts of context suitability, when producing evidence syntheses. However, the few hits of our structured search for applied methodological recommendations of evidence synthesis producing organizations show that the topic still receives little attention in practical application. The included methodological recommendations vary widely regarding the underlying terminology, the assessment approach concepts and the assessment criteria. Further, the provision of nonstandardized assessment approaches and missing instructions on how to use the assessment tools or how to judge the different assessment criteria might lead to nontransparent and heterogeneous assessment performances. It appears justified that the assessment of context suitability is somewhat heterogeneous because of the need to tailor the assessment to different target audiences/decision contexts and the varying relevance of aspects in relation to the medical area. Consequently, a single universal rating system seems inappropriate. Nevertheless, many differences between the recommendations seem unexplained. More harmonization is desirable and appears possible.

Our work provides a basis for developing or adapting current context suitability assessment approaches that suite specific requirements. A strict set of defined assessment criteria and the use of rigid checklists or questionnaires seem inappropriate for the application to different research questions, medical areas, decision situations and concepts of context suitability. Instead, the consideration of “basic” categories which can be adapted for each research question is more appropriate.9,16,28 One challenge in developing new tools is to find a good balance between standardized, reproducible procedures (apart from checklists and questionnaires) and flexible adaptable assessment criteria. Such a “semi-standardized” assessment approach should be accompanied by clear user-instructions. This may include a core/basic part for the context suitability assessment relevant for most intervention types, decision contexts and all concepts of context suitability (generalizability/external validity, applicability, transferability) and the possibility for supplementing additional parts, which might be necessary for specific intervention types, decision contexts or concepts of context suitability (e.g., assessing implantation issues in case of applicability assessment; assessing potential impact on the effect size in case of transferability assessment). Here it would be important to focus less on differences and similarities between the study context and the decision context, but rather to consider the mechanism of action9 or effect modifiers. This means focusing on differences that are likely to have an impact rather than all that could possibly have an impact. Another challenge in the development of a flexible tool will be balancing validity (all context suitability problems identified and adequately assessed) and reliability because the first one will probably increase with additional flexibility but the latter one will decrease.

ACKNOWLEDGEMENTS

None. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interests.

AUTHOR CONTRIBUTIONS

Alina Weise; Institute for Research in Operative Medicine, Faculty of Health—School of Medicine, Witten/Herdecke University, Ostmerheimer Str. 200, Building 38, 51 109 Cologne, Germany; e-mail: alina.weise@uni-wh.de.

Roland Büchter; Institute for Research in Operative Medicine, Faculty of Health—School of Medicine, Witten/Herdecke University.

Dawid Pieper; Institute for Research in Operative Medicine, Faculty of Health—School of Medicine, Witten/Herdecke University.

Tim Mathes; Institute for Research in Operative Medicine, Faculty of Health—School of Medicine, Witten/Herdecke University.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ETHICS STATEMENT

Our manuscript does not report on or involve the use of any animal or human data or tissue.
REFERENCES

1. Munthe-Kaas H, Nokleby H, Nguyen L. Systematic mapping of checklists for assessing transferability. Syst Rev. 2019;8(1):22. https://doi.org/10.1186/s13643-018-0893-4.

2. Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. Health Promot Int. 2006;21(1):76-83. https://doi.org/10.1093/heapro/da025.

3. Munthe-Kaas H, Nokleby H, Lewin S, Glenton C. The TRANSFER approach for assessing the transferability of systematic review findings. BMC Med Res Methodol. 2020;20(1):11. https://doi.org/10.1186/s12874-019-0834-5.

4. Schloemer T, Schröder-Bäck P. Criteria for evaluating transferability of health interventions: a systematic review and thematic synthesis. Implement Sci. 2018;13(1):88. https://doi.org/10.1186/s13021-018-0751-8.

5. 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(11):1251-1261. https://doi.org/10.1016/j.jclinepi.2013.06.017.

6. Boulenger S, Nixon J, Drummond M, Ulmann P, Rice S, de PG. Can economic evaluations be made more transferable? Eur J Health Econ. 2005;6(4):334-346. https://doi.org/10.1007/s10198-005-0322-1.

7. Cambon L, Minary L, Ridde V, Alla F. Transferability of interventions in health education: a review. BMC Public Health. 2012;12:497. https://doi.org/10.1186/1471-2458-12-497.

8. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:h2147. https://doi.org/10.1136/bmj.h2147.

9. Burchett HED, Blanchard L, Kneale D, Thomas J. Assessing the applicability of public health intervention evaluations from one setting to another: a methodological study of the usability and usefulness of assessment tools and frameworks. Health Res Policy Syst. 2018;16(1):88. https://doi.org/10.1186/s12961-018-0364-3.

10. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Eds). Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Cochrane; 2019.

11. Craig P, Ruggiero E, Frohlich KL, Mykhalovskiy E, White M, on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group. Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research. 2018. Southampton: NIHR Evaluation, Trials and Studies Coordinating Centre. https://doi.org/10.17863/CAM.26129.

12. Mathes T, Antoine S-L, Prengel P, Bühn S, Polus S, Pieper D. Health technology assessment of public health interventions: a synthesis of methodological guidance. Int J Technol Assess Health Care. 2017;33(2):135-146. https://doi.org/10.1017/S0266462317000228.

13. Mathes T, Walgenbach M, Antoine S-L, Pieper D, Eikermann M. Methods for systematic reviews of health economic evaluations: a systematic review, comparison, and synthesis of method literature. Med Decis Making. 2014;34(7):826-840. https://doi.org/10.1177/0272989X14526470.

14. Ahmad N, Bourtin I, Dechartres A, Durieux P, Ravaud P. Applicability and generalisability of the results of systematic reviews to public health practice and policy: a systematic review. Trials. 2010;11:20. https://doi.org/10.1186/1745-6215-11-20.

15. Gesundheit Österreich GmbH. Methodenhandbuch für Health Technology Assessment. Version 1. Vienna: Gesundheit Österreich GmbH; 2012.

16. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2014. www.effectivehealthcare.ahrq.gov.

17. Ludwig Boltzmann Institut für Health Technology Assessment. (Internes) Manual: Abläufe und Methoden. [Teil 2]. Vienna, Austria: Ludwig Boltzmann Gesellschaft GmbH; 2007.

18. Schick-Makaroff K, MacDonald M, Plummer M, Burgess J, Neander W. What synthesis methodology should I use? A review and analysis of approaches to research synthesis. AIMS Public Health. 2016;3(1):172-215. https://doi.org/10.3934/publichealth.2016.1.172.

19. Whittemore R, Knaff K. The integrative review: updated methodology. J Adv Nurs. 2005;52(5):546-553. https://doi.org/10.1111/j.1365-2648.2005.03621.x.

20. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097.

21. Adelaide Health Technology Assessment. HTA Glossary. http://htaglossary.net/health+technology. Accessed June 30, 2020.

22. Popay J, Roberts H, Sowden A, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme. Version 1; 2006. https://www.researchgate.net/publication/233866356_Guidance_on_the_conduct_of_narrative_synthesis_in_systematic_reviews_A_product_from_the_ESRC_Methods_Programme. Accessed February 9, 2020.

23. Health Information and Quality Authority. Guidelines for Evaluating the Clinical Effectiveness of Health Technologies in Ireland. Dublin, Ireland: Health Information and Quality Authority (HIQA); 2018.

24. National Institute for Health and Care Excellence. Diagnostics Assessment Programme Manual. Manchester, England: National Institute for Health and Clinical Excellence (NICE); 2011.

25. European Network for Health Technology Assessment. Levels of Evidence: Applicability of Evidence for the Context of a Relative Effectiveness Assessment. Adapted Version; 2015.

26. The National Academies Press. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press; 2011.

27. Ministry of Health S. Medical Technologies Evaluation Methods and Process Guide. Version 1.0. Singapore: Ministry of Health, Singapore; 2018.
28. Agency for Healthcare Research and Quality. Methods Guide for Medical Test Reviews. Rockville, MD; 2012 www.effectivehealthcare.ahrq.gov/reports/final.cfm.

29. The Agency for Health Technology Assessment and Tariff System. Health Technology Assessment Guidelines. Version 3.0. Warsaw, Poland: The Agency for Health Technology Assessment and Tariff System; 2016.

30. Swedish Agency for Health Technology Assessment and Assessment of Social Services. Assessment of Methods in Health Care: A Handbook; 2018.

31. Centre for Reviews and Dissemination. Systematic Reviews: CRD Guidance for Undertaking Reviews in Health Care. York: York publishing Services Ltd; 2008.

32. European Network for Health Technology Assessment. Joint Action 2, Workpackage 8: HTA Core Model®. Version 3; 2016. www.htacoremodel.info/BrowseModel.aspx.

33. Dyrvig A-K, Kidholm K, Gerke O, Vondeling H. Checklists for external validity: a systematic review. J Eval Clin Pract. 2014;20(6):857-864. https://doi.org/10.1111/jep.12166.

34. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. J Clin Epidemiol. 2011;64(12):1303-1310. https://doi.org/10.1016/j.jclinepi.2011.04.014.

35. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials. 2009;10:37. https://doi.org/10.1186/1745-6215-10-37.

36. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Clin Epidemiol. 2009;62(5):499-505. https://doi.org/10.1016/j.jclinepi.2009.01.012.

37. Bridget Candy, Michael King, Louise Jones, Sandy Oliver. Using Qualitative Synthesis to Explore Heterogeneity of Complex Interventions.

38. Cuijpers P, de GI, Bohlmeijer E. Adapting and disseminating effective public health interventions in another country: towards a systematic approach. Eur J Public Health. 2005;15(2):166-169. https://doi.org/10.1093/eurpub/cki124.

39. Cambon L, Minary L, Ridde V, Alla F. A tool to analyze the transferability of health promotion interventions. BMC Public Health. 2013;13:1184.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Weise A, Büchter R, Pieper D, Mathes T. Assessing context suitability (generalizability, external validity, applicability or transferability) of findings in evidence syntheses in healthcare—An integrative review of methodological guidance. Res Syn Meth. 2020;11:760–779. https://doi.org/10.1002/jrsm.1453