

Methods in Guideline Development

HICPAC guidelines are now based on targeted systematic reviews of the best available evidence. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach is used to provide explicit links between the available evidence and the resulting recommendations.²⁷⁻³⁰ The guideline development process is outlined in *Figure 2*.

Development of Key Questions

Each HICPAC guideline begins with the drafting and refining of the key questions most critical to infection prevention and control personnel and providers for the given guideline topic. These questions then serve as a foundation for the guideline, and guide the systematic review of the evidence and the development of the guideline recommendations. To develop the key questions, the working group first conducts a search of medical literature databases and websites for all relevant guidelines and narrative reviews on the topic of interest, and then drafts key questions based on their review of these documents. Databases commonly searched include MEDLINE and the National Guideline Clearinghouse. Websites commonly searched include those of government technology assessment programs like the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom³¹, commercial payors like BlueCross/BlueShield in the United States³², or federal or state websites in the United States. Once a preliminary list of key questions is developed from an examination of the relevant guidelines and reviews identified in the search, the key questions are vetted and revised by the content experts, and then are presented to and finalized by HICPAC members.

Literature Search

Following the development of the key questions, search terms are developed for identifying the literature most relevant to those questions. For the purposes of quality assurance, these terms are compared to those used in relevant guidelines, reviews and seminal studies. These search terms are then incorporated into search strategies for the relevant electronic medical literature databases. Searches are commonly performed in MEDLINE, EMBASE, CINAHL and Cochrane, and the resulting references are imported into reference management software, where duplicates can be resolved. Cochrane reviews ultimately included in guidelines are checked for updates prior to completion of the first guideline draft.

Study Selection

In general, a best available evidence approach is used to review articles for inclusion. For example, if there are randomized controlled trials (RCTs) that address a therapy question, then evidence lower in the evidence hierarchy may not be considered. Inclusion and exclusion criteria that are general or specific to individual key questions are developed and used to review references, starting with titles and abstracts. Full text articles are reviewed using the same criteria and are retrieved if they meet inclusion criteria. Studies that are commonly included are those that are: 1) relevant to one or more key questions, 2) primary analytic research, a systematic review or meta-analysis, and 3) written in English. Disagreements between reviewers regarding whether an individual study meets inclusion/exclusion criteria are resolved by consensus of those reviewers.

Data Extraction and Synthesis

For those studies meeting inclusion criteria, data relevant to the evidence review and guideline development is extracted into evidence tables. This data commonly includes: the study author, year, design, quality, objective, population, setting, sample size, power, follow-up, and definitions and results of clinically relevant outcomes. Evidence tables are developed for each key question, with study data being extracted into the relevant evidence tables. Then, studies are organized by the common themes that emerge within each evidence table. Data are extracted by one or more authors, and disagreements are resolved by the remaining authors. Data and analyses are most often extracted as originally presented in the included studies. Meta-analyses are performed only where their use is deemed critical to a recommendation and only in circumstances where multiple studies with sufficiently homogenous populations, interventions, and outcomes can be analyzed. Systematic reviews may also be included in a guideline if there are a large number of relevant reviews available in the literature.³³ Otherwise, systematic reviews will be used as a source of primary references for the guideline. To ensure that all relevant studies are captured in the search, the bibliography is vetted by the content experts.

Grading of Evidence

First, the quality of each study is assessed using scales adapted from existing methodology checklists³⁴⁻³⁸, and scores are recorded in the evidence tables. Next, the quality of the evidence base is assessed using methods adapted from the GRADE Working Group.^{27-30, 39} In summary, GRADE tables are developed for each of the interventions or questions addressed within the evidence tables. Included in the GRADE tables are the intervention of interest, any outcomes listed in the evidence tables that

are judged to be clinically important by the working group, the quantity and type of evidence for each outcome, the relevant findings, and the GRADE of evidence for each outcome, as well as an overall GRADE of the evidence base for the given intervention or question. For therapy or harm questions, the initial GRADE of evidence for each outcome is deemed high if the evidence base includes an RCT or a systematic review of RCTs, low if the evidence base includes only observational studies, or very low if the evidence base consists only of descriptive studies (i.e., uncontrolled studies) or expert opinion. The initial GRADE is then modified by eight criteria. Criteria which can decrease the GRADE of an evidence base include poor quality of individual studies, inconsistent findings among studies, indirectness of study findings to the study question, imprecision of study estimates, and publication bias. Criteria that can increase the GRADE include a large magnitude of effect, a dose-response gradient, or inclusion of unmeasured confounders that would increase the effect size (Table 1). For questions regarding diagnostic measures (e.g., sensitivity or predictive values) or descriptive measures (e.g., prevalence or incidence), the initial GRADE of evidence can be high even if the evidence base only includes descriptive study designs, like cross-sectional studies.³⁰ The initial GRADE can then be modified by criteria similar to those used for therapy or harm questions. GRADE definitions are as follows^{27, 28}:

1. High - further research is very unlikely to change confidence in the estimate of effect
2. Moderate - further research is likely to affect confidence in the estimate of effect and may change the estimate

3. Low - further research is very likely to affect confidence in the estimate of effect and is likely to change the estimate
4. Very low - any estimate of effect is very uncertain

After determining the GRADE of the evidence base for each outcome of a given intervention or question, the overall GRADE of the evidence base for that intervention or question is calculated. The overall GRADE is based on the lowest GRADE for the outcomes deemed critical by the working group to making a recommendation.

Formulating Recommendations

Narrative evidence summaries are then drafted by the working group using the evidence and GRADE tables. One summary is written for each theme that emerges under each key question. The working group then uses the narrative evidence summaries to develop guideline recommendations. Factors determining the strength of a recommendation include²⁹: 1) the values and preferences of the working group when determining which study outcomes are critical²⁸, 2) the risks and benefits that result from weighing the critical outcomes, and 3) the overall GRADE of the evidence base for the given intervention or question (Table 2). If weighing the critical outcomes for a given intervention or question results in a "net benefit" or a "net harm", then a Category I Recommendation is formulated to strongly recommend for or against the given intervention respectively. If weighing the critical outcomes for a given intervention or question results in a "trade off" between benefits and harms, then a Category II Recommendation is formulated to recommend that providers or institutions consider the intervention when deemed appropriate. If weighing the critical outcomes for a given

intervention or question results in an "uncertain trade off" between benefits and harms, then No Recommendation is formulated to reflect this uncertainty.

For Category I recommendations, levels A and B represent the quality of the evidence underlying the recommendation, with A representing high to moderate quality evidence and B representing low to very low quality evidence but established standards (e.g., aseptic technique, education and training). For Category IB recommendations, although there may be low to very low quality evidence directly supporting the benefits of the intervention, the theoretical benefits are clear, and the theoretical risks are marginal. Category IC represents practices required by state or federal regulation, regardless of the quality of evidence. It is important to note that the strength of a Category IA recommendation is equivalent to that of a Category IB or IC recommendation; it is only the quality of the evidence underlying the Category IA recommendation that makes it different from a Category IB.

In some instances, multiple recommendations may emerge from a single narrative evidence summary. The updated HICPAC categorization scheme for recommendations is provided in Table 3.

Category I recommendations are defined as strong recommendations with the following implications²⁹:

1. For patients: Most people in the patient's situation would want the recommended course of action and only a small proportion would not. Patients should request discussion if the intervention is not offered.
2. For clinicians: Most patients should receive the recommended course of action.

3. For policymakers: The recommendation may be considered for policy in many situations.

Category II recommendations are defined as weak recommendations with the following implications²⁹:

1. For patients: Most people in the patient's situation would want the recommended course of action, but some may not.
2. For clinicians: Different choices will be appropriate for different patients and clinicians must help patients arrive at management decisions consistent with their values and preferences.
3. For policymakers: Policy making requires substantial debate and involvement of many stakeholders.

Our evidence-based recommendations are then cross-checked with those from guidelines identified in our original systematic search. In addition, recommendations from previous guidelines for topics not directly addressed by our systematic review of the evidence are included in a "Summary of Recommendations" if they are deemed critical to the target users of the guideline. Unlike recommendations informed by the literature search, these recommendations are not linked to a key question. Instead, these recommendations are agreed upon by expert consensus and are generally designated either Category IB if they represent a strong recommendation based on accepted practices (e.g., aseptic technique) or Category II if they are a suggestion based on a probable net benefit despite limited evidence.

We carefully select the wording of each recommendation to reflect the recommendation's strength.⁴⁰ We use the active voice when writing Category I

Updating the Guideline Methodology of HICPAC

recommendations - the strong recommendations. Phrases like "do" or "do not" and verbs without conditionals are used to convey certainty. A passive voice is used when writing Category II recommendations - the weak recommendations. Words like "consider" and phrases like "is preferable," "is suggested," or "is not suggested" are used to reflect the lesser certainty of the Category II recommendations. Rather than a simple statement of fact, each recommendation is actionable, describing precisely a proposed action to take.⁴¹

The category "No recommendation/unresolved issue" is most commonly applied to situations where either: 1) the overall quality of the evidence base for a given intervention is low to very low or 2) there is no published evidence on outcomes deemed critical to weighing the risks and benefits of a given intervention. If the latter is the case, those critical outcomes are noted at the end of the relevant evidence summary.

All recommendations are formulated to be consistent with policies from the United States Food and Drug Administration (FDA) and Environmental Protection Agency (EPA). All recommendations are approved by HICPAC members, and focus only on efficacy, effectiveness, and safety. Yet, the optimal use of these guidelines should include a consideration of the costs relevant to the local setting of guideline users.

Reviewing and Finalizing the Guideline

After a draft of the tables, narrative summaries, and recommendations is completed, the working group shares this draft with the content experts for review in depth. While the content experts are reviewing this draft, the working group completes

Updating the Guideline Methodology of HICPAC

the remaining sections of the guideline, including the executive summary, background, summary of recommendations, and recommendations for guideline implementation, audit, and further research. The working group then makes revisions to the draft based on feedback from the content experts, and presents the entire draft guideline to HICPAC for review. CDC then submits the guideline for clearance, and posts it on the Federal Register for public comment. After a period of public comment, the guideline is revised accordingly, and the final guideline is published and posted on the HICPAC website.

Updating the Guideline

Guidelines will be reassessed periodically, and general or targeted revisions to guidelines will be dictated by new research and technological advancements in the particular area of interest.⁴² Reassessments and updates will occur at the request of HICPAC.

Guideline Implementation

To improve the impact of guidelines on patient care quality and safety, multiple implementation initiatives are underway.⁵ In addition, future HICPAC guidelines will include an implementation and audit section. This section includes multi-modal implementation of specific recommendations or modules²⁵ that highlight the most critical recommendations in the guideline.²⁴ Besides being the focus of infection preventionists and healthcare epidemiologists, these recommendations may also be ripe for integration into computerized clinical decision support systems.⁴³ This section also includes performance indicators that can be used by healthcare facilities or regulators of such facilities to improve guideline adherence and ultimately patient care,

Updating the Guideline Methodology of HICPAC

and may be the focus of pay for performance contracts either locally or nationally.

These modules and performance indicators established by HICPAC are based on the evidence review and recommendations.