Strength of recommendations in WHO guidelines using GRADE was associated with uptake in national policy


Abstract

Objectives

This study assesses the extent to which the strength of a recommendation in a World Health Organization (WHO) guideline affects uptake of the recommendation in national guidelines.

Study Design and Setting

The uptake of recommendations included in HIV and TB guidelines issued by WHO from 2009 to 2013 was assessed across guidelines from 20 low- and middle-income countries in Africa and Southeast Asia. Associations between characteristics of recommendations (strength, quality of the evidence, type) and uptake were assessed using logistic regression.

Results

Eight WHO guidelines consisting of 109 strong recommendations and 49 conditional recommendations were included, and uptake assessed across 44 national guidelines (1,255 recommendations) from 20 countries. Uptake of WHO recommendations in national guidelines was 82% for strong recommendations and 61% for conditional recommendations. The odds of uptake comparing strong recommendations and conditional recommendations was 1.9 (95% confidence interval: 1.4, 2.7), after adjustment for quality of evidence. Higher levels of evidence quality were associated with greater uptake, independent of recommendation strength.

Conclusion

Guideline developers should be confident that conditional recommendations are frequently adopted. The fact that strong recommendations are more frequently adopted than conditional recommendations underscores the importance of ensuring that such recommendations are justified.
Key findings

Uptake of World Health Organization recommendations in national guidelines is high and associated with strength of recommendation and evidence quality. A higher level of evidence quality was associated with greater uptake of the recommendation, independent of strength.

What this adds to what was known?

Conditional recommendations are frequently adopted, although less frequently than strong recommendations.

What is the implication and what should change now?

Guideline developers should be confident that conditional recommendations are frequently adopted. The fact that strong recommendations are more frequently adopted than conditional recommendations underscores the importance of ensuring that such recommendations are justified.

1. Introduction

Guidelines issued by the World Health Organization (WHO) aim to help policy makers, health care providers, and patients make evidence-informed decisions [1]. Although global in intent, WHO guidelines are primarily focused on the needs and resources of low- and middle-income countries, where they are often used to develop national guidelines. The potential impact of WHO recommendations is far reaching, including influencing individual clinician practice, health service organization, the procurement policies of international donors, and potentially the global demand for health technologies.

WHO adopted the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) [2] system in January 2009 to support the formulation of evidence-based recommendations. Since that time, there has been a steady increase in the number of recommendations issued by WHO that are based on the GRADE approach [3], [4] and [5].

According to GRADE, the strength of the recommendation is designated as either “strong” or “weak” (commonly termed “conditional” in WHO guidelines), with strong recommendations generally based on high-quality evidence from randomized controlled trials, except in certain circumstances [6]. If the GRADE framework is followed without any bias on the part of guideline developers, then it would be reasonable to expect that conditional recommendations would be adopted in fewer national guidelines than strong recommendations, as their scope is by definition more limited. However, there has been no assessment to date of the extent to which WHO recommendations are adopted in national policy, nor whether recommendation characteristics (ie, strength; quality of supporting evidence) influence uptake in national guidelines. Such an assessment would provide evidence to inform guideline development groups about the potential influence of recommendation strength on uptake in national guidelines.

This study aimed to assess the association between the strength of WHO recommendations and uptake in national policy.

2. Methods

2.1. Data sources

To assess the relationship between the characteristics of WHO recommendations and their uptake in national guidelines, a database of both WHO and national guidelines was constructed. HIV and TB
guidelines had previously been found to show a high adherence to the GRADE process [4], and exploratory Web searches found a number of repositories that could serve as sources for national guidelines. All HIV and TB guidelines published by WHO from January 2008 to December 2013 were retrieved from a WHO repository (http://who.int/publications/guidelines/en/) and cross-checked with experts at WHO to check that no guidelines were missed. To be included, WHO guidelines needed to have been developed according to the GRADE approach.

Four principle online resources were used to identify national guidelines: The USAID AIDSTAR-One National Treatment Database for HIV [7], AIDSspace Guideline Repository [8], the WHO Database of National HIV and TB guidelines 2005–2011 [9], and the UNAIDS Global Database of HIV/AIDS guidelines [10]. National guidelines had to fulfill the following criteria: published in English, low- or middle-income setting as determined by World Bank income grouping [11], and published between January 1, 2009 (when WHO first adopted the GRADE process), and December 31, 2013.

National guidelines were tabulated, assigned unique identification codes, and classified as follows: adult and adolescent HIV care, pediatric HIV care, prevention of mother to child transmission, HIV opportunistic infection care, HIV/TB coinfection, and TB management.

To be included, national guidelines had to be published after a corresponding WHO guideline to ensure that national guideline developers would have had access to the relevant WHO recommendations when formulating policy. Where this was uncertain (eg, when WHO and national guidelines were published within the same year), national guidelines were only included if they specifically referenced the corresponding WHO guideline.

2.2. Data management

Recommendations were extracted from WHO guidelines by one reviewer (S.M.U.N.), then tabulated, and coded in MS Excel (Microsoft, Redmond WA, USA). The following characteristics were recorded for each recommendation: strength of recommendation (strong or conditional); quality of evidence (high, moderate, low, or very low); domain (HIV, HIV/TB, or TB); target population or condition (general, adults and adolescents, maternal health, and pediatrics); and type of intervention (prevention, diagnostic test, and treatment). Quality of evidence in the GRADE paradigm refers to the confidence or uncertainty in the estimates of effect [12].

For each national guideline, corresponding WHO recommendations were sought by examining the full text of each guideline and searching for key words. Each national recommendation was considered to adhere to a specific WHO recommendation only if the national policy adhered to all the key features of the corresponding WHO recommendation. Nonadherence was determined when the national recommendation explicitly contradicted the corresponding WHO recommendation, reflected some but not all its key features, or adapted the WHO recommendation beyond its intended scope.

2.3. Statistical analysis

The association between the strength of a recommendation and its uptake in national guidelines was explored using descriptive statistics with differences assessed by the $\chi^2$ test. Univariate and multivariate logistic regression analyses were used to determine the strength of any association (the dependent variable) after controlling for the following variables that were defined a priori: quality of the evidence, type of guideline, and type of recommendation. The potential for clustering by country, guideline type, and intervention type was assessed by Spearman's rank correlation coefficient ($\rho$). All reported $P$-values are exacted and two tailed, and for each analysis, a $P$-value less than 0.05 was considered significant. All analyses were carried out in Stata, version 12 (StataCorp LP, TX, USA).

The full data set is available on request from the corresponding author (N.F.).
3. Results

3.1. Characteristics of guidelines and recommendations

Eight eligible WHO guidelines were included in this study, all published between 2009 and 2012 [13], [14], [15], [16], [17], [18], [19] and [20].

Five of these focused on HIV care, two focused on TB care, and one focused on HIV/TB comanagement. The total number of recommendations in these eight guidelines was 158, of which 109 were strong recommendations and 49 were conditional recommendations. The distribution of these recommendations is summarized in Table 1.

Table 1.
Summary of guideline recommendations, by strength of recommendation and quality of evidence

<table>
<thead>
<tr>
<th>Quality of evidence</th>
<th>Strong</th>
<th>Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WHO guidelines</td>
<td>National guidelines</td>
</tr>
<tr>
<td></td>
<td>WHO guidelines</td>
<td>National guidelines</td>
</tr>
<tr>
<td>High quality</td>
<td>27</td>
<td>138</td>
</tr>
<tr>
<td>Moderate quality</td>
<td>37</td>
<td>327</td>
</tr>
<tr>
<td>Low quality</td>
<td>37</td>
<td>505</td>
</tr>
<tr>
<td>Very low quality</td>
<td>8</td>
<td>283</td>
</tr>
</tbody>
</table>

Abbreviation: WHO, World Health Organization.

N.B. data are the number of recommendations.

The number of national guidelines that met the inclusion criteria was 44, containing 1,255 recommendations. Twenty countries were represented: nine lower income, seven lower-middle income, and four upper-middle income (Table 2).

Table 2.
Summary of national guidelines

<table>
<thead>
<tr>
<th>Income stratum</th>
<th>Region</th>
<th>Countries</th>
<th>Number of guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower income</td>
<td>Africa</td>
<td>Ethiopia, Kenya, Malawi, Tanzania, Uganda, Zimbabwe</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Asia</td>
<td>Bangladesh, Cambodia, Nepal</td>
<td>7</td>
</tr>
<tr>
<td>Lower-middle income</td>
<td>Africa</td>
<td>Ghana, Lesotho, Nigeria, Zambia</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Asia</td>
<td>India, Sri Lanka, Vietnam</td>
<td>7</td>
</tr>
<tr>
<td>Upper-middle income</td>
<td>Africa</td>
<td>Botswana, Namibia, South Africa</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Asia</td>
<td>Malaysia</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>Asia</td>
<td></td>
<td>44</td>
</tr>
</tbody>
</table>
As defined by the World Bank.

The 158 WHO recommendations were compared with the 1,255 recommendations reported in national guidelines. Overall, 76% of WHO recommendations were taken up by national guidelines: 81.9% (n = 715) of strong recommendations and 61.2% (n = 234) of conditional recommendations in WHO guidelines were adopted as national policy recommendations (P < 0.01). None of the national guidelines reported the strength of the WHO recommendations or the quality of the body of evidence.

In univariate logistic regression, strong recommendations were found to be almost three times more likely to be taken up by national guidelines compared with recommendations that were conditional [odds ratio (OR) 2.9; 95% confidence interval (CI): 2.2, 3.8]. This association remained statistically significant in multivariate logistic regression after controlling for the quality of the evidence supporting a recommendation (OR 1.9; 95% CI: 1.4, 2.7).

Among strong recommendations, the uptake of a recommendation varied according to the quality of the evidence supporting the recommendation. Compared with strong recommendations based on low- or very low-quality evidence, there was a statistically significantly greater odds of uptake for recommendations based on high or moderate quality evidence (OR 2.0; 95% CI: 1.4, 2.8).

There was no statistically significant difference in uptake between a strong recommendation based on very low quality of evidence and conditional recommendations in general (P = 0.3). Nor was there a statistically significant difference in uptake according World Bank income bracket or among population subgroups (all P > 0.05).

There was weak clustering of strength of recommendation by guideline (ρ 0.19, P ≤ 0.01) and by country (ρ 0.11, P ≤ 0.01) and even weaker clustering by intervention (ρ 0.01, P ≤ 0.03). However, these associations were not considered strong enough to significantly influence the main results.

**4. Discussion**

This study assessed the uptake of 158 recommendations contained within eight WHO guidelines by national guidelines from 20 countries and found that overall 76% of WHO recommendations were taken up by national guidelines. Strong recommendations were, on average, adopted 82% of the time, whereas conditional recommendations were taken up 61% of the time. After adjusting for quality of the evidence, there was a near twofold higher odds of uptake of strong recommendations compared with conditional recommendations. For strong recommendations, higher levels of quality of the evidence were associated with increased uptake in national guidelines; strong recommendations based on very low evidence quality were not more likely to be adopted than were conditional recommendations.

A recent evaluation of WHO guidelines reported that over half of strong recommendations were based on low or very low evidence quality [21]. The perception that conditional recommendations are infrequently adopted has been put forward as an explanation for why guideline developers may opt to make a recommendation strong despite low-quality evidence. Our study challenges this assumption: conditional recommendations are often taken up in national guidelines, although appropriately less frequently than are strong recommendations. These results should help to assuage any doubts within guideline development groups regarding the utility of conditional recommendations. The high average uptake of WHO recommendations, the disparity in uptake between strong and conditional recommendations, and the relatively lower uptake of strong recommendations based on weak evidence indicate that national policy makers are interpreting GRADE-based WHO guidelines appropriately.
We found that independent of the strength of recommendations, the quality of the body of evidence was also a factor related to uptake of national recommendations. This finding has important implications for WHO guideline development and presentation. The GRADE assessment of the quality of evidence must be carefully reasoned and presented in an explicit and transparent manner, given that end users take this assessment into consideration. The quality of the evidence for critical outcomes could be presented in each guideline's executive summary, and a summary of the key domains leading to the quality assessment should be provided in the text of the guideline and not buried in appendices. Of note, none of the national guidelines reported the strength of recommendation or the quality of the body of evidence, which has been put forward as a key criterion for trustworthiness by the US Institute of Medicine [22].

Finally, our study suggests that any disparity between recommendation strength and evidence quality should be carefully considered, given that strong recommendations based on very low evidence quality appear to be adopted no more frequently than conditional recommendations overall.

This study has a number of strengths, including examination of an extensive data set across a range of recommendation types and countries. A careful assessment was carried out to ensure that recommendations in national guidelines were preceded by, and corresponded with, the most recent relevant recommendations issued by WHO. Our study is limited, however, as we were limited in the number of national guidelines that we were able to retrieve, and the ones that we did identify may not be representative of the entire cohort of relevant guidelines. In addition, we examined only two diseases (HIV and TB), and the applicability of our findings to other conditions or interventions is unclear. We cannot rule out the possibility that some of the national guidelines may not have been based on WHO guidelines, although we consider this unlikely as most guidelines made explicit reference to WHO guidelines, and a criterion for inclusion in this study was that the national guideline was published after the corresponding WHO guideline. Last, although this assessment provides an insight into uptake of recommendations of national guidelines, no inferences can be drawn regarding the implementation of these recommendations in practice.

The results of this study represent a first step in the analysis of the uptake of WHO recommendations at the country level based on the strength of the recommendation. Further research with an expanded data set would better quantify the relationship between recommendation characteristics and uptake in national guidelines, including non-English language guidelines and guidelines in other topic areas. Such quantitative assessments could be complemented with qualitative research to understand the priorities and perceptions of policy makers.

References


22. Anon. Clinical Practice Guidelines We Can Trust, Institute of Medicine, Washington (2011)