Journal of Clinical Epidemiology 147 (2022) 69–75

Journal of Clinical Epidemiology

OTHER GRADE PAPERS

Using Explicit Thresholds were valuable for judging Benefits and Harms in partially contextualized GRADE Guidelines

Ignacio Neumann[a,b,c,*], Eduardo Quiñelen[b], Paula Nahuelhual[b], Pamela Burdiles[b], Natalia Celedón[b], Katherine Cerda[b], Paloma Herrera-Omegna[b], Patricia Kraemer[b], Karen Dominguez Cancino[b,c,d], Juan Pablo Valenzuela[b], Dino Sepúlveda[b], Gian Paolo Morgano[c], Elie A. Akl[c,e], Holger J. Schünemann[c]

[a]*Department of Internal Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*
[b]*Department of Health Technology Assessments, Ministerio de Salud de Chile, Santiago, Chile*
[c]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada*
[d]*Escuela de Enfermería, Universidad Científica del Sur, Lima, Peru*
[e]*Department of Internal Medicine, American University of Beirut, Beirut, Lebanon*

Accepted 21 March 2022; Published online 30 March 2022

Abstract

Objectives: Guideline panels must assess the magnitude of health benefits and harms to develop sensible recommendations. However, they rarely use explicit thresholds. In this paper we report on the piloting and the use thresholds for benefits and harms.

Study Design and Setting: We piloted the use of thresholds in a Chilean COVID-19 living guideline. For each of the critical outcomes, we asked panelists to suggest values of the thresholds for large, moderate, small, or trivial or no effect. We collected this information through a survey and an on-line discussion.

Results: Twelve panelists decided on thresholds for three critical outcomes (mortality, need for mechanical ventilation and serious adverse events). For all outcomes, an absolute risk reduction was considered larger with more than 50 events, moderate with less than 50 events, small with less than 25 events, and trivial with less than 10 events. Having these a priori thresholds in place significantly impacted on the development of recommendations.

Conclusion: Explicit thresholds were a valuable addition to the judgment of the certainty in the evidence, to decide the direction and strength of the recommendation and to evaluate the need for update. We believe this is a line of research worth perusing. © 2022 Elsevier Inc. All rights reserved.

Keywords: Clinical guidelines; GRADE; COVID-19

1. Introduction

Guideline panels must assess the magnitude of the potential health benefits and harms to develop sensible recommendations. Structured and transparent processes, for example, application of the GRADE Evidence to Decision (EtD) Framework [1], was designed to make transparent judgements about the size of the benefits and the size of the harms. The EtD framework categorizes these effects

Conflict of interest: This work had no specific funding. All authors declare no conflict of interest.

Author Contribution: Ignacio Neumann: Conceptualization, investigation, methodology, writing (original draft); Eduardo Quiñelen: investigation, methodology, writing (review & editing); Paula Nahuelhual: investigation, methodology, writing (review & editing); Pamela Burdiles: investigation, methodology, writing (review & editing); Natalia Celedón: investigation, methodology, writing (review & editing); Katherine Cerda: investigation, methodology, writing (review & editing); Paloma Herrera-Omegna: investigation, methodology, writing (review & editing); Patricia Kraemer: investigation, methodology, writing (review & editing); Karen Dominguez Cancino: investigation, methodology, writing (review & editing); Juan Pablo Valenzuela: investigation, methodology, writing (review & editing); Dino Sepúlveda: investigation, methodology, writing (review & editing); Gian Paolo Morgano: Conceptualization, methodology, writing (review & editing); Elie A. Akl: Conceptualization, methodology, writing (review & editing); Holger J. Schünemann: Conceptualization, methodology, writing (review & editing).

* Corresponding author. Department of Internal Medicine, Pontificia Universidad Católica de Chile, Avenida Libertador Bernardo O'Higgins 340, Santiago, Chile. Tel.: +56-2-354-3130; fax: +56-2-354-3130.

E-mail address: Ignacio.neumann@gmail.com (I. Neumann).

https://doi.org/10.1016/j.jclinepi.2022.03.017
0895-4356/© 2022 Elsevier Inc. All rights reserved.

What is new?

Key findings?

- We piloted the use of explicit thresholds for benefit and harms in a real GRADE living guideline.
- Given the lack of empirical data about thresholds for benefit and harms in COVID-19, we obtained them surveying panelists

What this adds to what was known?

- Guideline panels must assess the magnitude of the potential health benefits and harms to develop sensible recommendations. However, they rarely use explicit thresholds.

What is the implication and what should change now?

- Explicit thresholds for benefits and harms were a valuable addition to the judgment of the certainty in the evidence, to decide the direction and strength of the recommendation and to evaluate the need for update.

sizes as large, moderate, small, or trivial, or no effect. However, guideline panels, until now, rarely use explicit thresholds for these categories. Rather, panelists reach an implicit agreement through discussion and consensus.

Thresholds may be also needed to judge the certainty of evidence [2]. In the context of guidelines, panelists may choose to use a partially contextualized approach, in which the judgment of the certainty of the evidence reflects the confidence that the true effect lies within a specific range (large, moderate, small, or trivial. or no effect).

In Chile, since August 2020, the Ministry of Health and several National Scientific Societies have been developing and maintaining a GRADE living guideline about management options for COVID-19 [3]. Over the span of 18 months, the guideline panel has made recommendations and updated them in the light of new evidence. Maintaining consistency in judgments over an extended period was a challenge, and thus, we developed thresholds for the different criteria in the EtD Framework to anchor future discussions. In this paper, we report on the piloting and the use thresholds for benefits and harms.

2. Methods

Chilean COVID-19 living guideline is an on-going national effort. It currently has 25 recommendations: 21 about pharmacologic interventions, two about diagnostic test and two about non-pharmacologic interventions. The guideline panel has 12 content experts (from relevant scientific societies and front-line clinicians), two methodological coordinators and two representatives from payers and administrators. Content experts comes from different disciplines: Five are specialists in infectious diseases, two in respiratory diseases, two in emergency medicine, one in critical care, one in general internal medicine and one in palliative care.

In addition, the work of the guidelines is supported by a team of six methodologists who conduct and update systematic reviews. Half of the content experts had participated in an evidence-based guidelines before, and half were new to the process. At the time of this report, neither of the participants, content experts nor methodologists, had conflict of interests with any of the recommendation addressed in the guidelines.

After the initial nine recommendations, we obtained estimates of the thresholds for large, moderate, small, or trivial or no effect for benefits and harms from the twelve content experts.

First, using surveymokey (www.surveymonkey.com), we designed a survey and asked each panelist what magnitude of the effect they considered large, moderate, small, or trivial, or no effect for each of the critical outcomes: mortality, need of mechanical ventilation and serious adverse events. The magnitude of such outcomes were considered independently following the GRADE partially contextualized approach.

We provided four independent questions for each outcome, asking a specific category, following the formula ‘‘in your opinion, what magnitude of the effect corresponds to a large/moderate/small/trivial effect?’’. We framed these questions as neutral statement without introducing a specific direction.

For the outcomes mortality or need of mechanical ventilation, we did not provide a detailed description, since our respondents were clinicians and were aware of what those outcomes entail. However, for serious adverse events, we developed a standard description of the outcome for consistency purposes. We described it as follows: negative effects of the intervention that are serious enough to discontinue the treatment but resolves spontaneously or with specific treatment after discontinuation.

The response options were sliding bars going from one per 1,000 to 200 per 1,000. In each question, panelists had to select the magnitude they believed corresponded to effect size that was being asked. (e.g., for the question what magnitude of the effect corresponds to a trivial effect? They have to select a unique number, between 1 per 1,000 and 200 per 1,000, that they considered a trivial effect). The extremes of the response options were chosen based on empirical data: the largest effect of a pharmacological intervention ever observed in COVID-19 patients is a reduction of 150 events per 1,000 treated [4].

The second step in our process was to reach consensus based on the survey results. We averaged the responses of the panelist and presented them in a formal meeting. Through discussion and consensus, the panel agreed on

the final threshold for all the critical outcomes: for simplicity they decided to use the multiple of five that was closer to the average of the survey. In addition, they decided to place the same value in all the critical outcomes and hence, to use the same thresholds for the three of them.

3. Results

The results are presented in Table 1 (and conceptualized in Fig. 1). As can be observed, for different outcomes, the magnitude of the effect that corresponds to each category were different. However, during the panel meeting, though discussion, panelists decided to use the same thresholds for the three critical outcomes.

3.1. How thresholds impacted recommendations?

3.1.1. Judgments about imprecision

In the context of GRADE partially conceptualized approach, panelists must judge for each outcome, what is the confidence that the true effect of an intervention lies within a specific range (large, moderate, small, or trivial or no effect). Having explicit thresholds facilitated judgments about precision.

For example, our meta-analysis of three randomized trials (n = 4,628) showed that in patients with COVID-19, the use of colchicine may reduce mortality (RR 0.47, 95%CI 0.18-1.23) and the need of mechanical ventilation (RR 0.47, 95% CI 0.24-0.94) [3]. By focusing on the relative estimates, a particular panel may decide to rate down the certainty of the evidence of mortality by imprecision, given that the confidence interval (CI) includes no effect and potential harm. Using this approach, the need of mechanical ventilation likely would be considered precise. However, since mortality is a critical outcome, rating it down for imprecision would lead to a lower overall certainty of the evidence.

The absolute effects of colchicine are depicted in Figure 2. The point estimates of mortality and need of mechanical ventilation are within the boundaries of what panelists defined a priori as trivial effect. Furthermore, the entire 95% CI for both estimates are within these thresholds. Considering this, in addition to the large number of patients randomized, panelists may choose to not rate down

Table 1. Explicit Thresholds for Benefits and Harms used in a COVID-19 guideline per 1,000

Critical outcomes	Trivial	Small	Moderate
Survey results (average)			
Mortality	10	25	51
Mechanical Ventilation	14	37	68
Serious adverse events	12	28	51
Selected through consensus			
All critical outcomes	10	25	50

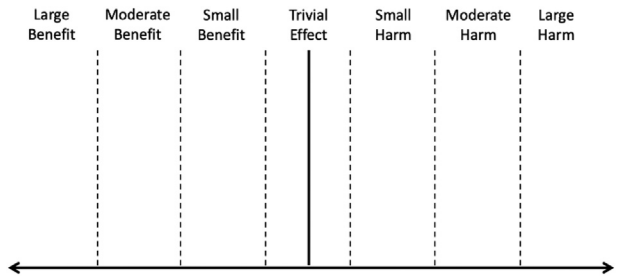


Fig. 1. Conceptualization of the thresholds for a Forest Plot of absolute effects.

by imprecision the outcome mortality, as it happened in our guideline.

Another example is the use of budesonide in non-hospitalized COVID-19 patients. Our metanalysis of two randomized trials showed that budesonide did not impact on outcomes such as mortality or need for mechanical ventilation. However, it did reduce the chance of hospitalization (RR 0.71, 95% CI 0.53-0.95) [3]. The patients enrolled in the trials, however, were in general at high risk of hospitalization: older than 65 years and with significant comorbidity. This contrasts with the usual practice, where most patients with COVID-19 do not require hospitalization. We contextualized this finding using appropriate baseline risks. To do this, we took as moderate risk of hospitalization the national average during the pandemic: 8% [5]; and to illustrate high and low risk, we multiplied the average by two for high risk, and divided it by four for low risk (although arbitrary, panelists considered it an appropriate range, given the lack of more precise data).

As we see in Figure 3, these baseline risks led to CIs around the absolute effects of different width. For lower risk patients, the CI around the absolute effect (from 1 to 9 fewer per 1,000) was entirely within what our panel considered a priory a trivial effect. For moderate (average) risk patient, the CI (4 to 38 fewer per 1,000) crossed the trivial, small, and moderate effect. Finally, for high-risk patients, the CI crossed the entire range of potential benefits.

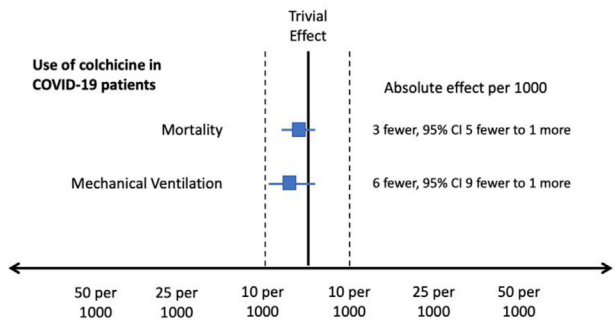


Fig. 2. Use of explicit threshold to label the trivial effect; Other contextual factors. Significant increase of adverse events (Diarrhea: 70 more per 1,000; 95% CI 46 to 97 more) Inexpensive drug.

This may be interpreted that there is more uncertainty regarding the effect of budesonide in high or moderate-risk patients than in low-risk patient. The benefit seems to exist in all, however, the CIof the absolute estimate in low-risk patient is precise enough to be confident that the benefit is trivial in magnitude. The CI for high-risk patients, on the other hand, is so wide that may be appropriate to further penalize it by imprecision: for different baseline risks, the certainty of the evidence seems to be different, even within the same outcome. An upcoming GRADE guidance paper will provide rules to operationalize the judgements about precision using explicit thresholds.

3.1.2. Guidance for the direction and strength of recommendations

In the examples presented on the previous section, the observation that the use of colchicine likely will lead to a trivial benefit, in addition to the increment in adverse events, led to a recommendation against its use. In the e.g, of budesonide, given the differences in absolute effects, the panel decided to make two separate conditional recommendations: one against its use in low-risk individuals and one in favor in moderate and high-risk patients (Table 2).

In another example, our meta-analysis of 10 randomized trials (n = 6,700) showed that the use of tocilizumab may be associate with a lower mortality (RR 0.84, 95% CI 0.75-0.94) [3]. Here, there were no serious concerns regarding risk of bias, inconsistency, indirectness, or publications bias. With this evidence, is perfectly reasonable for a panel to make a recommendation in favor of using tocilizumab in settings where tocilizumab is accessible and affordable.

In our case, however, there were some concerns with access and affordability, and hence, also with equity. We decided to explore further the effect of tocilizumab considering different groups of patients with different baseline risks.

The largest study evaluating tocilizumab vs usual care until now is the RECOVERY trial [6]. Investigators categorized patients according to the degree of respiratory support at randomization: Non-ventilatory support, non-invasive ventilation, and invasive mechanical ventilation. Mortality

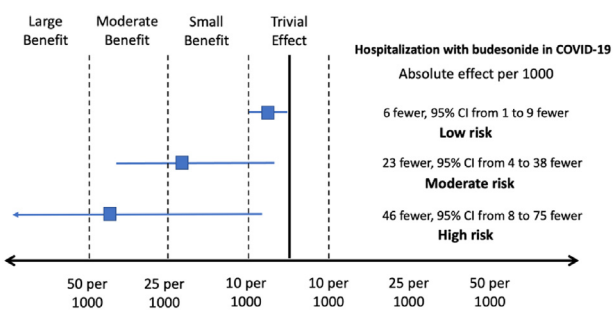


Fig. 3. Different imprecision judgments for different baseline risks; Other contextual factors. No significant increase of adverse events Inexpensive drug widely available.

for those groups in the control arm were 23%, 42% and 51%, respectively. No interaction between the severity of the disease at randomization and the effect of tocilizumab was detected. Therefore, the authors appropriately concluded that the benefit of tocilizumab was observed in a wide range of patients regardless of the level of respiratory support.

One immediate concern with these data is the unexpected high mortality in patients with no respiratory support. Thus, we decided to use a more conservative estimate from a systematic review of observational studies, which reported mortality in hospitalized COVID-19 patients of 5% [7]. With these data, we estimated that in hospitalized patients with a mild COVID-19 (death risk of 5%), the use of tocilizumab would result in a trivial effect (8 fewer deaths, 95% CI from 3 to 13 fewer). In contrast, in severe patients (death risk of 51%) the effect is large (82 fewer deaths, 95% CI from 31 to 128 fewer).

The use of tocilizumab did not significantly increase the risk of adverse events (RR 0.93, 95% CI 0.78-1.10). So basically, the direction of the recommendation had to balance the size of the benefits of the intervention with its cost and the existence of a limited stock. Having the explicit thresholds in place greatly facilitated discussion and increased the transparency of the decision. A trivial benefit in hospitalized patients with no respiratory support was not enough to justify the use of tocilizumab in a context of resources, feasibility, and equity concern. However, the situation was different for patients who may get a large mortality reduction. This analysis led to recommendations with opposite directions for different group of patients (Fig. 4 and Table 2).

Another example is use of dexamethasone in COVID-19 patients, which results in a significant reduction of mortality (RR 0.90, 95% CI 0.83-0.98, three RCTs n = 6,774). However, the RECOVERY trial [4], the largest of the three studies available, showed compelling evidence of an interaction between the effect of dexamethasone and the severity of patients. The relative effect of the intervention had a gradient with the severity of the disease and the interaction test was statistically significant [8]. Considering this interaction, and the influence of baseline risk, we estimated that in severe patients (death risk 44%) the use of dexamethasone will result in 120 fewer deaths per 1,000 patients (95% CI from 58 to 174 fewer). In moderate patients, however, with a death risk 26%, the use of dexamethasone will prevent 29 deaths per 1,000 (95% CI 0−55 fewer). There were no concerns regarding risk of bias, inconsistency, indirectness, or publications bias. However, as before, the precision of the estimates for different baseline risk was not the same: in severe patients, the confidence interval was entirely within the boundaries of a large effect. In contrast, for moderate patients, the CI crossed the entire range of benefits. Accordingly, we rated down the certainty of the evidence for this group by imprecision.

Table 2. Precision judgments, size of the effect and final recommendation of examples

Example	Precision judgment	Other concerns	Size of the effect	Final recommendation	Other contextual factors
Colchicine	Precise	None	Trivial	Strong against	Significant increase of adverse events
Budesonide					
Low risk	Precise	Risk of bias	Trivial	Conditional against	No significant increase of adverse events
High risk	Imprecise		Moderate	Conditional in favor	Inexpensive drug
Tocilizumab					
Mild patients	Precise	None	Trivial	Strong against	No significant increase of adverse events Moderate resources required
Severe patients	Precise		Large	Conditional in favor	Feasibility and Equity concerns
Dexamethasone					
Moderate patients	Imprecise	None	Moderate	Strong in favor	No significant increase of adverse events
Severe patients	Precise		Large	Strong in favor	Inexpensive drug
Remdesivir					
Before update	Imprecise	Risk of bias	Moderate	Conditional in favor	No significant increase of adverse events Large resources required
After update	Imprecise		Trivial	Conditional against	Feasibility and Equity concerns

Dexamethasone in relative low doses and for a short period of time was associated with mild adverse effects (mainly hyperglycemia). Also, is not an expensive drug, and it is widely available and accessible. Considering that at the time was the only intervention that provided a mortality reduction, the threshold for a strong recommendation in favor of dexamethasone in COVID-19 patients was relatively small. In the case of severe patients, this threshold was clearly achieved, and the panel quickly decided a strong recommendation in favor. However, given the striking differences in the absolute effect (120 vs 29 fewer deaths per 1,000) and the differences in the certainty of the evidence, panelists struggled with the decision for moderate patients. At the end, given that the point estimate was still in what they a priory considered a moderate effect they issued a strong recommendation also for moderate patients (Fig. 5 and Table 2). However, it worth noting that the recommendation in favor of dexamethasone in severe patients is based in high certainty evidence and is ‘‘stronger’’ than the recommendation for moderate patients, which in is based in moderate certainty evidence.

3.1.3. Decision about when to update a recommendation

Continuously presenting small changes in the Summary of Findings tables to a panel it is simply not feasible. Among the criteria we develop for when to present new evidence to the panel, it was ‘‘a significant change in the effect estimates’’ for any of the critical outcomes. Here, having the thresholds defined a priori was very useful.

For example, before the publication of solidarity trial [9], our meta-analysis about the use of remdesivir in COVID-19 patients suggested a benefit in mortality but with imprecise values (RR 0.76, 95% CI 0.57−1.01, three trials n = 1,882) [3]. Using the median of baseline risk observed in the control groups (12%, to standardize calculations) we estimated an absolute effect of 29 fewer deaths per 1,000, (95% CI 52 fewer to 1 more). This effect, according to our pre-defined thresholds was considered moderate, and the certainty of the evidence low (there were some concerns regarding the risk of bias of included trials and the results was judged imprecise, since included trivial benefit and harm). With the publication of the Solidarity trial, the pooled estimate changed to RR 0.93, 95% CI 0.81−1.06. Using the same baseline risk as before, the absolute effect moved from a moderate to a trivial benefit: eight fewer deaths per 1,000, 95% CI 23 fewer to seven more) (Fig. 6). The certainty of the evidence did not change, since the new estimate was considered still imprecise because included a small benefit. However, the change in the point estimate was enough to trigger a recommendation update.

This is similar to what has happened with other living guidelines when the effect estimates changes. For example, the American Society of Hematology guidelines on the use

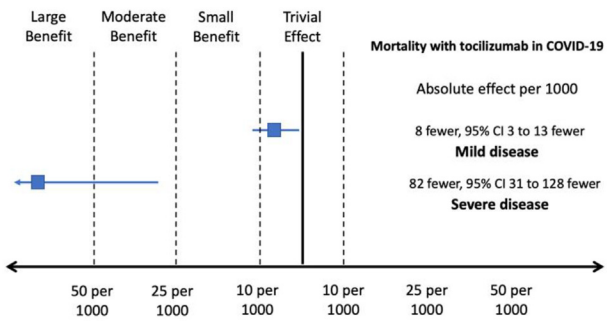


Fig. 4. Use of explicit threshold and the direction of recommendations; Other contextual factors. No significant increase of adverse events Moderate resources required Feasibility and Equity concerns.

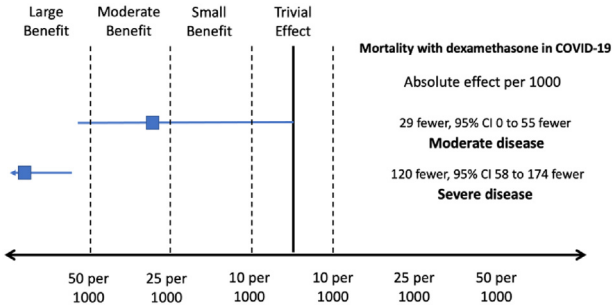


Fig. 5. Use of explicit threshold and the strength of recommendations; Other contextual factors May lead to hyperglycemia (28 more per 1,000) Negligible resources required No feasibility or Equity concerns.

of anticoagulation for thromboprophylaxis in patients with COVID-19, changed their recommendations as consequence of an evidence update [10]. Having the thresholds in place from the beginning may help to expedite the process.

4. Discussion

In this pilot work, using explicit thresholds for benefit and harms provided insights to the panel to judge the precision of estimates, to decide direction and strength of the recommendation and to evaluate the need for update.

Until now in GRADE guideline, panelists started with no pre-defined thresholds for benefits and harms, and they reach them by discussion and consensus. Typically, the way it happens is by making judgements through recommendations one-by-one and landing in an implicit agreement about what constitutes a large, moderate, or small effect. This method has the merits of eliciting the value that panelists may place in different outcomes in an iterative way, tailoring the decision to the specific clinical situation being analyzed. In the GRADE partially contextualized approach, however, having implicit thresholds have some limitations. The major conceptual limitation is that panelists may use their values, rather than patients', to judge the relative importance of each outcome. The major practical limitation, on the other hand, is that having implicit thresholds may not be functional for maintaining consistence in how the benefits and harms are judge in living guidelines developed months apart, or when different panels are making recommendation for competing interventions (e.g., when deciding which new intervention to fund).

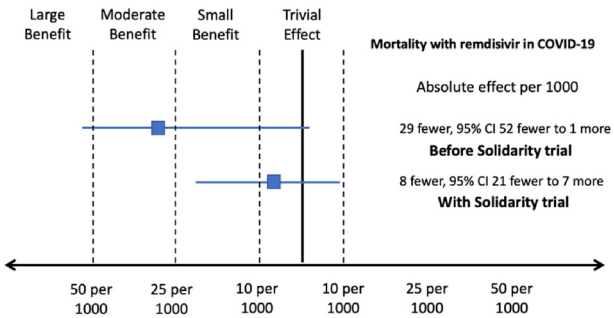


Fig. 6. Use of explicit threshold to update recommendations; Other contextual factors. No significant increase of adverse events Large resources required Feasibility and Equity concerns.

Ideally, when judging the size of benefits and harms in a partially contextualized approach, thresholds should come from empirical data: what users of guidelines may consider a small, moderate, or large benefit or harm. A survey of guideline panelists and users is underway [11]. Given the emergency of the COVID-19 pandemic, we took a pragmatic approach to identify empiric thresholds and estimated them directly from panelists appreciation.

One particularity of our pilot experience is that panelists placed the same value on all the critical outcomes. This is unlikely to be the case for most guideline recommendations since evidence point that different outcome has not the same importance for patients [12]. It is interesting how panelists decided to match the thresholds for mortality and mechanical ventilation, even when the survey results suggested that a priory, they placed more value in avoiding mortality. One potential explanation is the unique circumstances of the pandemic, where the excessive number of severe patients quickly overwhelmed critical care units and there was an actual shortage of mechanical ventilators. In these extreme circumstances, considering avoiding mechanical ventilation equally important than mortality is a sensible decision from a public health perspective. However, this is likely an exception.

Another particularity of our pilot experience is we assumed that the thresholds for benefits and harms were symmetrical. Although pragmatical and easy to implement, this may not be the case if guideline users place a different value in potential benefits and harms within the same outcome. For example, patients likely will place more value in avoiding an increase of mortality, thus, lower thresholds on the "harm side" may be appropriate. In fact, for some outcomes it may be not necessary to quantify the size of the harms; once a critical threshold of harm is surpassed, the intervention may be no longer acceptable for patients and clinicians.

One limitation of our work is we did not assess the utilities of critical the outcomes. To stablish thresholds in a sensible way, panels should consider the dyad of magnitude and the relatively importance of the effect. Making these values explicit may even lead to more quantitative approaches, like estimating the net benefit or the net harm [13]. Another limitation is our sample size: only one panel. However, our pilot experience was very positive: the use of explicit thresholds for benefits and harms enormously facilitated panel discussions and the decisions about recommendations update. Also, it enhanced consistency among recommendations developed month apart.

5. Conclusions

Using explicit thresholds for benefits and harms in a real panel helped to maintain consistency and to enhance transparency in living guidelines. Furthermore, they were a valuable addition to the judgments of the certainty in the evidence, to decide the direction and strength of the recommendation and to evaluate the need for update. We believe this is a line of research worth perusing.

References

[1] Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. BMJ 2016;353:i2089.

[2] Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol 2017;87:4−13.

[3] MINSAL. Evidence based recommendations for management of COVID-19 2021. Available at www.minal.cl. Accessed June 30, 2021.

[4] Horby P, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with COVID-19. N Engl J Med 2021;384:693−704.

[5] MINSAL. 122° Epidemiological report COVID-19 2021. Available at www.minal.cl. Accessed June 30, 2021.

[6] RECOVERY-Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet 2021;397:1637−45.

[7] Abate SM, Checkol YA, Mantedafro B. Global prevalence and determinants of mortality among patients with COVID-19: A systematic review and meta-analysis. Ann Med Surg (Lond) 2021;64:102204.

[8] Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. CMAJ 2020;192(32):E901−6.

[9] Pan H, Pan H, Peto R, Henao-Restrepo AM, Preziosi MP, Sathiyamoorthy V, et al. Repurposed antiviral drugs for COVID-19 - interim WHO solidarity trial results. N Engl J Med 2021;384:497−511.

[10] Cuker A, Tseng EK, Nieuwlaat R, Angchaisuksiri P, Blair C, Dane K, et al. American Society of Hematology 2021 guidelines on the use of anticoagulation for thromboprophylaxis in patients with COVID-19. Blood Adv 2021;5(3):872−88.

[11] Morgano G, Mbuagbaw L, Santesso N, Xie F, Brozek JL, Siebert U, et al. Defining decision thresholds for judgments on health benefits and harms using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks: a protocol for a randomized methodological study. BMJ Open 2022;12:e053246.

[12] Etxeandia-Ikobaltzeta I, Zhang Y, Brundisini F, Florez ID, Wiercioch W, Nieuwlaat R, et al. Patient values and preferences regarding VTE disease: a systematic review to inform American Society of Hematology guidelines. Blood Adv 2020;4(5):953−68.

[13] Alper BS, Oettgen P, Kunnamo I, Iorio A, Ansari MT, Murad MH, et al. Defining certainty of net benefit: a GRADE concept paper. BMJ Open 2019;9(6):e027445.