

# Methodological assessment of Mexican Clinical Practice Guidelines: Critical appraisal and GRADE framework adherence.

Luis Colunga-Lozano<sup>1</sup>, Vilma Gerardo-Morales<sup>2</sup>, Giordano Pérez-Gaxiola<sup>2</sup>, Alan Omar Vázquez-Alvarez<sup>3</sup>, Francisco Javier González-Torres<sup>3</sup>, Leonardo Perales-Guerrero<sup>3</sup>, Ekatherina Yanowsky-Ortega<sup>3</sup>, Ulises I. Martínez-Tolentino<sup>3</sup>, Sergio J. Sánchez-Villaseca<sup>3</sup>, Luz Angélica García-Macías<sup>3</sup>, Robby Nieuwlaat<sup>4</sup>, Gordon Guyatt<sup>4</sup>, and Ivan Florez<sup>5</sup>

<sup>1</sup>Hospital Civil de Guadalajara Dr Juan I Menchaca

<sup>2</sup>Hospital Pediatrico de Sinaloa

<sup>3</sup>Universidad de Guadalajara Centro Universitario de Ciencias de la Salud

<sup>4</sup>McMaster University

<sup>5</sup>Universidad de Antioquia

May 5, 2020

## Abstract

Background and Objective Clinical Practice Guidelines (CPGs) provide evidence-based recommendations to healthcare professionals, policy makers, patients and other stakeholders. Mexico is the biggest producer of CPGs in Latin America and Caribbean countries. The National Healthcare Technology Excellence Center (acronym in Spanish: CENETEC) is responsible for the CPG development, adaptation and update. The aim of this study was to assess the Mexican CPG quality and adherence to the GRADE framework. Study design We conducted a descriptive cross-sectional study of 86 CPGs representing all the CPGs produced by CENETEC between 2015 and 2017 and published in an online database called “Catalogo Maestro”. We performed quality assessment with the online AGREE II tool and assessed the reporting on the GRADE framework. Results Of the 86 CPGs, 34 were published in 2015, 21 in 2016 and 31 in 2017. The overall quality by AGREE II proved a median of 16.6% (Min 16.6%, Max 50%). Of the 86 CPGs, 25 (29%) used the GRADE framework; adherence to GRADE standards was, however, inconsistent and generally poor. Conclusion CPGs produced by CENETEC during this period had a low score by AGREE II standards and low adherence to the GRADE framework. A concerted initiative could rapidly improve CENETEC guidelines.

## Abbreviation list

AGREE = Appraisal of Guidelines for Research and Evaluation.

CENETEC = National Healthcare Technology Excellence Center (acronym in Spanish: CENETEC)

CPGs = Clinical Practice Guidelines.

EtD = Evidence to Decision

EBM = Evidence-based Medicine.

GRADE = Grading of Recommendations, Assessment, Development and Evaluation

NICE = National Institute for Health and Care Excellence.

OCEBM = Oxford Centre for Evidence-Based Medicine.

RS = Systematic Review.

SIGN = Scottish Intercollegiate Guidelines Network.

SoF = Summary of Findings.

## Introduction

Clinical Practice Guidelines (CPGs) are health-related documents created to provide evidence-based recommendations to healthcare professionals, policy makers, patients and other stakeholders.<sup>(1)</sup> In Mexico, CPGs are developed by the National Healthcare Technology Excellence Center (acronym in Spanish: CENETEC). CENETEC was founded in 2004 and includes the Mexican Healthcare institutions. To date, CENETEC has produced more than 760 CPGs, focusing on four categories: diseases, nursing, medical procedures and healthcare process.<sup>(2)</sup>

CPGs methodology has improved over time, from non-standardized “Good Old Boys Sitting Around a Table” (GOBSAT) approaches to standardized methods.<sup>(3)</sup> Different resources and methodologies have contributed to this improvement. Two of them particularly have been key in the guidelines methodology enterprise; The development of the Appraisal of Guidelines for Research and Evaluation (AGREE II) tool - which is the most widely used tool to assess the quality of CPGs - and the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) framework, which has become the reference methodological standard for the development of guidelines.<sup>(4-10)</sup>

CENETEC’s massive CPGs production may be impressive and admirable but raises the issue of the rigor with which its guidelines are produced, their applicability and usefulness. To date, few studies had appraised CENETEC guidelines with the AGREE tool,<sup>(11)</sup> however, there is no systematic assessment of the quality of the guidelines produced by the CENETEC aiming to include their entire production. The aim of the current study was to assess the quality of Mexican CPGs and to evaluate their adherence to the GRADE framework.

## METHODS

### Study design

We conducted a descriptive cross-sectional study of Mexican CPGs produced by the CENETEC.

### Selection criteria and sample size calculation

Our population of interest was all the CPGs, regardless of their scope, produced by CENETEC and published on their online database called “*Catálogo Maestro – Master Database*” between 2015 and 2017. (Website citation) The total population of CPGs at the time of this study design were 224 CPGs (Accessed on el 20/11/2017 <http://cenetec-difusion.com/gpc-sns/?cat=52>).

We considered the GRADE framework as the reference standard to frame the question, judge the quality of evidence, and move from evidence to recommendations.<sup>(9)</sup> Leyva de Los Rios et al. reported the frequency of the GRADE framework use in the CPGs from CENETEC as 10%.<sup>(12)</sup> Therefore, considering an alpha error of 0.05, and a 95% confidence interval, we determined that we needed to assess 86 CPGs. We chose a random sample from the total population of 224 CPGs using the software Epidat, version 4.2.<sup>(13)</sup> The CPGs were allocated randomly among the assessors (GMV, VAA, PGL, YOE, MTU, SVS, GML, GTF, LEC). The list of the included CPGs is reported in Table 1S.

### CPG selection and collected variables

Two team members (LEC, GTF) extracted in duplicate the CPGs characteristics, disagreements were solved by discussion. Recorded variables were: Year of publication, type of CPG (Full Update/Partial Update), librarian involvement, number of questions, adaptation strategy, type of studies included (International Guideline, Non-Cochrane and Cochrane Systematic Reviews), quality assessment for GCPs and Systematic Reviews, number of articles obtained by the search strategy, the use of GRADE approach, adherence to the

GRADE framework, recommendations systems, type of recommendations, number of recommendations and good clinical practice statements.

## AGREE II tool

We used the AGREE II tool to assess the quality of the CPGs development. (6) AGREE II is a validated instrument composed of 23 items grouped under 6 domains and one final item to evaluate the overall quality of the CPG. The domains are: Scope and Purpose (3 items), Stakeholder Involvement (3 items), Rigor of Development (8 items), Clarity of Presentation (3 items), Applicability (4 items), and Editorial Independence (4 items). For each item, each appraiser is asked to score based on the statement, using a Likert scale from 1 ((Strongly disagree) to 7 (Strongly agree). A score between 2 and 6 is assigned when the reporting of the AGREE II item does not meet the full criteria or considerations depending on the completeness and quality of reporting. Scores increase as more criteria are met and considerations addressed.<sup>(5, 14, 15)</sup>

Following the recommendations by the AGREE collaboration, each CPG was evaluated independently by three assessors. With the aim of reducing the variability among the assessors we performed an intensive training in the tool led by one of the investigators (IDF), expert in the use of AGREE II. After initial training, the assessors were calibrated by independently assessing two CPGs (one low quality and one high quality), followed by feedback from the trainer (IDF), and by a group discussion to solve any issues that arose.

We performed two rounds of assessment and feedback. We considered disagreement between the assessors when the differences among the scores were of 3 or more points, and they were resolved by discussion. If agreement was not achieved after the discussion, two reviewers (IDF, LEC) made the final assessment and decision. Final scores per domain were calculated by summing up scores from the 3 assessors per individual item in a domain and by scaling the total as a percentage of the maximum possible score for that domain. Therefore, each domain was scored in a range from 0 to 100%.

The AGREE collaboration does not define minimum domain scores or patterns of scores across domains to differentiate between high quality and poor-quality guidelines.<sup>(20)</sup> Instead, the recommendation for users is to use the scores to compare CPG among them and determine specific thresholds based on the context in which the guidelines are to be used. All the assessment were performed using the My AGREE-PLUS online tool provided by the AGREE collaboration (<http://www.agreetrust.org/resource-centre/agree-plus/>).

## GRADE framework adherence

We describe the use of the GRADE framework in the included CPGs. GRADE can be used to define the CPG questions, develop SR, identify the importance of the outcomes, prepare evidence profile, summary of findings (SoF), assess the overall quality of the evidence and decide on the direction and strength of the recommendation. Currently, there is not a validated tool to assess adherence to GRADE. We defined an appropriate use of GRADE framework if the CPGs reported the following items:<sup>(7-10)</sup>

- Questions with explicit statement for each PICO element.
- SoF with the following characteristics: Number of studies, number of participants, risk of bias domains, relative effect, absolute effect and certainty of the evidence.
- Explicit reference to all GRADE domains:
- **Rating down criteria** : Risk of bias, imprecision, indirectness, inconsistency and publication bias.
- **Rating up criteria** : Large magnitude of an effect, Dose-response gradient and effect of plausible residual confounding.

## Statistical Analysis

We calculated descriptive statistics to summarize our results. We explored the normal distribution of the continuous variables with the Shapiro-Wilk test. The results were express with mean and standard deviation (SD) if we fail to reject a normal distribution. The results were express with median and interquartile range (IQR) if normal distribution was rejected. We calculated proportions and frequencies for categorical

variables. All the analyses were performed using the software SPSS (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp). A Venn diagram was created with Eulerr web platform/site to display the combination of the grading systems (<http://eulerr.co/>).<sup>(16)</sup>

## RESULTS

### General description

We included 86 CPGs in our analysis. Table 1 summarizes the guideline characteristics. Thirty-four (39.5%) were published in 2015, 54 (62.8%) de novo guidelines, 79 (91.9%) provided therapeutic recommendations and 45 (52.3%) focus on specific diseases.

Table 2 summarizes the guideline development process. Librarian involvement was present in 32 CPGs (37.2%), international guidelines were included in 66 CPGs (76.7%), Cochrane SR in 36 CPGs (41.9%) and non-Cochrane SR in 59 CPGs (68.6%); no critical appraisal was reported. Several of the guidelines used more than one evidence grading system in the same document to provide their recommendations (48.8%). The systems implemented were the approaches by the National Institute of Health and Care Excellence (NICE) (42, 48.8%), Shekelle (27, 31.4%), GRADE (25, 29.1%), the Scottish International Guideline Network (SIGN) (19, 22.1%) and by the Oxford Centre for Evidence Based Medicine (OCEBM) (18, 20.9%) (figure 1). The authors used the NICE and SIGN approaches that had no adopted the GRADE framework.

### AGREE II critical appraisal.

The overall quality assessment obtained a median score of 16.6% (IQR 0.0) (Figure 2). Table 3 summarizes the details per domain by percentage of the AGREE II tool. Scope and Purpose and Clarity of presentation domains showed the highest scores (27.7% IQR 16.6 and 27.7% IQR 11.1, respectively). Individual domains are reported in table 4. The following items were not reported in any of the included CPGs: The views and preferences of the target population (patients, public, etc.) and the methods for formulating the recommendations.

### GRADE framework

Twenty-five (29.1%) of the CPGs reported using the GRADE framework for the quality of evidence assessment. However, most of the CPGs combined GRADE with others grading systems; only five GPGs used the GRADE framework exclusively. Regarding the GRADE adherence: None of the included CPGs reported the explicit statement of each PICO element, only one (4%) reported SoF tables and none reported the explicit reference to all GRADE domains.

## DISCUSSION

Our study found that CENETEC guideline production had a low score by the AGREE II tool. Using this instrument, we identified the following pitfalls in eligible guidelines i) Most of the health related questions were not created using the explicit elements of the PICO format ii) The target audience was excessively broad – *aiming to reach nurses, doctors and other healthcare professional, across the different levels of health care* – iii) The views and preferences of the target population were not included iv) The systematic review methods were inconsistent across the guidelines v) The CPGs mixed different grading systems, vi) no formal external peer review was conducted vi) The transition from the evidence to the recommendations was not transparent and clear vii) and the recommendations were sometimes were unclear and ambiguous viii)

We also found that guideline adherence to the GRADE framework was very low. Only one of the approximately 30% of the guidelines that purported to use GRADE included SoF tables, nor did they transparently address the GRADE domains related to reasons rating the quality of the evidence down or up.

The strengths of this study include the intensive training in the AGREE II tool provided to the reviewers, the experience of the team in used of the GRADE framework, and the relatively large number of CPGs included. Our study is limited by the following: although we did provide intensive training before the CPGs assessment, most of the reviewers were not experts in guideline development, and so possibilities of incomplete

understanding remain. A standardized and valid tool to assess adherence to the GRADE framework is unavailable. To address this issue, we created an ad hoc instrument, the reliability and validity of which remains untested.

In 2019 Cabrera et al. reviewed the guideline development process with respect to the GRADE framework in guidelines developed in Latin America and Caribbean countries.<sup>(17)</sup> The CPGs were developed by Colombia (68%), Peru (13%), Chile (9%), Argentina, Costa Rica, Brazil, Honduras and Dominican Republic. The authors focused their analysis in detecting the methods for grading the quality of evidence and topic prioritization without formally assessing the adherence to the GRADE framework. The authors conducted a literature search on MEDLINE, SciELO and Embase, also, they search within guideline repositories and other governmental websites. They did not, however, include the “Catalogo maestro”, and thus excluded all the CENETEC CPGs. Of the 1,370 CPGs the authors assessed, only 98 (8.9%) followed the GRADE framework. In our findings we identify more CPGs using the GRADE approach (29.1%), but found very poor adherence to GRADE, an issue unaddressed in this previous work.

In general, our findings shown that despite the impressive and admirable production of CPGs by CENETEC, they in general have a number of methodological limitations.

## CONCLUSION

CPGs produced by CENETEC during this period had a low score by AGREE II standards and low adherence to the GRADE framework. Improving the training of methodologists in the GRADE methodology and addressing the specific pitfalls in the guidelines development process that we have highlighted should be a priority for CENETEC guidelines process. Enhancing the developer’s capacity and the guidelines quality will result in better and more trustworthy guideline to be used in Mexico

**CONFLICT OF INTEREST:** The authors declared that they do not have financial conflicts of interest. Also, none of the authors are part of CENETEC organization. RN, IDF and GG are members of the GRADE working group. IDF is part of the AGREE collaboration.

## AUTHORS CONTRIBUTIONS

**LECL:** Refining the research idea, creating the research design, selecting an Instrument, selection of statistical tests, performing statistical analyses, interpretation of statistical analyses, performing the CPGs assessment, writing the manuscript and managing submission process.

**GPG:** Conceptualizing the research idea, creating the research design, selecting an instrument, interpretation of statistical analyses and manuscript feedback.

**GMV:** Conceptualizing the research idea, creating the research design, selecting an Instrument, performing the CPGs assessment, manuscript feedback.

**VAAO:** Performing the CPGs assessment and manuscript feedback.

**GTFJ :** Performing the CPGs assessment and manuscript feedback.

**PGL:** Performing the CPGs assessment and manuscript feedback.

**YOE:** Performing the CPGs assessment and manuscript feedback.

**MTUI:** Performing the CPGs assessment and manuscript feedback.

**SVSJ:** Performing the CPGs assessment and manuscript feedback.

**GMLA:** Performing the CPGs assessment and manuscript feedback.

**RN:** Interpretation of statistical analyses and manuscript feedback. **GG:** Interpretation of statistical analyses and manuscript feedback.

**IDF:** Selecting an Instrument, selection of statistical tests, interpretation of statistical analyses and manuscript feedback.

## References

1. World, Health, Organization. WHO Handbook for guideline development. 2nd ed. Switzerland (Geneva)2014.
2. CENETEC. CENETEC difusión 2018. Available from: <http://cenetec-difusion.com/gpc-sns/>.
3. Wollersheim H. Beyond the evidence of guidelines. *The Netherlands journal of medicine*. 2009;67(2):39-40.
4. Agree., Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality & safety in health care*. 2003;12(1):18-23.
5. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting, and evaluation in health care. *Preventive medicine*. 2010;51(5):421-4.
6. Dans AL, Dans LF. Appraising a tool for guideline appraisal (the AGREE II instrument). *Journal of clinical epidemiology*. 2010;63(12):1281-2.
7. GRADE, Working, Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) 2004 - 2018 [01/21/2019]. Available from: <http://www.gradeworkinggroup.org/>.
8. Group TGW. GRADE handbook for grading quality of evidence and strength of recommendations. Schünemann H BJ, Guyatt G, Oxman A, editor2013 (Updated October).
9. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology*. 2011;64(4):383-94.
10. Zhang Y, Akl EA, Schünemann HJ. Using systematic reviews in guideline development: The GRADE approach. *Research Synthesis Methods* [Internet]. 2018. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1313>.
11. Poblano-Verastegui O, Vieyra-Romero WI, Galvan-Garcia AF, Fernandez-Elorriaga M, Rodriguez-Martinez AI, Saturno-Hernandez PJ. [Quality and compliance with Clinical Practice Guidelines of Chronic Noncommunicable Diseases in primary care]. *Salud publica de Mexico*. 2017;59(2):165-75.
12. Leyva De Los Ríos CD P-GG, Izabal Delgado DE, Coronado Araujo J, Rodríguez Páez IV, Rodríguez Lizárraga JD. Use of Cochrane Evidence in Mexican clinical practice guidelines Gloval Evidence Summit Using Evidence Improving lives; Cope Town, South Africa2017.
13. Epidat. programa para análisis epidemiológico de datos. Versión 4.2. Consellería de Sanidade, Xunta de Galicia, España; Organización Panamericana de la Salud (OPS-OMS); Universidad CES, Colombia. 2016.
14. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2010;182(10):1045-52.
15. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2010;182(10):E472-8.
16. J L. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 5.1.0,. 2019.
17. Cabrera PA, Pardo R. Review of evidence based clinical practice guidelines developed in Latin America and Caribbean during the last decade: an analysis of the methods for grading quality of evidence and topic prioritization. *Globalization and health*. 2019;15(1):14.

Tables

**Table 3. AGREE II Critical appraisal overall domains (%), N = 86**

Variables	Estimates
Scope and Purpose, median (IQR)	27.7 (16.6)
Stakeholder Involvement, median (IQR)	22.2 (5.5)
Rigor of Development, median (IQR)	17 (6.9)
Clarity of Presentation, median (IQR)	27.7 (11.1)
Applicability, median (IQR)	16.6 (8.3)
Editorial Independence, median (IQR)	16.6 (0.0) <sup>\euro</sup>
Overall Assessment, median (IQR)	16.6 (0.0) <sup>\euro</sup>
\euro. Minimal value 16.6, maximal value 33.3, £ Minimal value 16.6, maximal value 50	

**Table 3. AGREE II Critical**

Estimates	
27.7 (16.6)	
22.2 (5.5)	
17 (6.9)	
27.7 (11.1)	
16.6 (8.3)	
16.6 (0.0) <sup>\euro</sup>	
16.6 (0.0) <sup>\euro</sup>	
\euro. Minimal value 16.6, maximal value 50	

**Table 4. AGREE II Critical appraisal - individual domains. N = 86**

Variables	Scores
Scope and Purpose, median (min – max)	
1. The overall objective(s) of the guideline is (are) specifically described.	3 (2 – 5)
2. The health question(s) covered by the guideline is (are) specifically described.	2 (2 – 5)
3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	2 (2 – 6)
Stakeholder involvement, median (min – max)	
4. The guideline development group includes individuals from all relevant professional groups.	3 (2 – 5)
5. The views and preferences of the target population (patients, public, etc.) have been sought.	1
6. The target users of the guideline are clearly defined.	3 (2 – 6)
Rigor of development, median (min – max)	
7. Systematic methods were used to search for evidence.	2 (2 – 4)
8. The criteria for selecting the evidence are clearly described.	2 (2 – 3)
9. The strengths and limitations of the body of evidence are clearly described.	2 (2 – 3)
10. The methods for formulating the recommendations are clearly described.	1
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	2 (1 – 4)
12. There is an explicit link between the recommendations and the supporting evidence.	3 (2 – 4)
13. The guideline has been externally reviewed by experts prior to its publication.	2 (1 – 3)
14. A procedure for updating the guideline is provided.	2 (2 – 5)
Clarity of presentation, median (min – max)	
15. The recommendations are specific and unambiguous.	3 (2 – 5)
16. The different options for management of the condition or health issue are clearly presented.	2 (2 – 5)
17. Key recommendations are easily identifiable.	3 (2 – 7)
Applicability, median (min – max)	
18. The guideline describes facilitators and barriers to its application.	2 (1 – 4)
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.	3 (2 – 4)
20. The potential resource implications of applying the recommendations have been considered.	2 (2 – 3)
21. The guideline presents monitoring and/or auditing criteria.	2 (1 – 5)
Editorial independence, median (min – max)	
22. The views of the funding body have not influenced the content of the guideline.	2 (2 – 4)
23. Competing interests of guideline development group members have been recorded and addressed.	2 (2 – 4)
Overall Assessment, median (min – max)	2 (2 – 4)

**Hosted file**

Figure 1.docx available at <https://authorea.com/users/292563/articles/420287-methodological-assessment-of-mexican-clinical-practice-guidelines-critical-appraisal-and-grade-framework-adherence>

## Hosted file

Figure 2.docx available at <https://authorea.com/users/292563/articles/420287-methodological-assessment-of-mexican-clinical-practice-guidelines-critical-appraisal-and-grade-framework-adherence>