OPEN ACCESS

Original research

Which actionable statements qualify as good practice statements In Covid-19 guidelines? A systematic appraisal

Omar Dewidar ,[1,2] Tamara Lotfi,[3,4,5] Miranda Langendam,[6] Elena Parmelli,[7] Zuleika Saz Parkinson,[8] Karla Solo ,[3,4,5] Derek K Chu ,[3,9] Joseph L Mathew,[10] Elie A Akl,[3,11] Romina Brignardello-Petersen,[3,4] Reem A Mustafa,[3,12] Lorenzo Moja,[13] Alfonso Iorio,[3,4,9] Yuan Chi,[14,15] Carlos Canelo-Aybar,[16] Tamara Kredo ,[17,18] Justine Karpusheff,[19] Alexis F Turgeon,[20,21] Pablo Alonso-Coello,[16] Wojtek Wiercioch,[3,4,5] Annette Gerritsen,[17] Miloslav Klugar ,[22] María Ximena Rojas,[23] Peter Tugwell,[24,25] Vivian Andrea Welch ,[1,2] Kevin Pottie,[26] Zachary Munn,[27] Robby Nieuwlaat,[3] Nathan Ford,[28] Adrienne Stevens,[3,4] Joanne Khabsa ,[29] Zil Nasir,[3,4] Grigorios I Leontiadis,[4,9] Joerg J Meerpohl,[30,31] Thomas Piggott ,[3,4,5] Amir Qaseem,[32] Micayla Matthews,[3,4,5] Holger J Schünemann ,[3,4,5,9,33,34] the eCOVID-19 recommendations map collaborators

10.1136/bmjebm-2021-111866

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjebm-2021-111866).

For numbered affiliations see end of article.

Correspondence to:
Prof Holger J Schünemann, McMaster University Department of Health Research Methods Evidence and Impact, Hamilton, Canada; schuneh@mcmaster.ca

Linked

► http://dx.doi.org/10.1136/bmjebm-2022-111962

Check for updates

To cite: Dewidar O, Lotfi T, Langendam M, *et al*. *BMJ Evidence-Based Medicine* 2022;**27**:361–369.

Abstract

Objectives To evaluate the development and quality of actionable statements that qualify as good practice statements (GPS) reported in COVID-19 guidelines.

Design and setting Systematic review . We searched MEDLINE, MedSci, China National Knowledge Infrastructure (CNKI), databases of Grading of Recommendations Assessment, Development and Evaluation (GRADE) Guidelines, NICE, WHO and Guidelines International Network (GIN) from March 2020 to September 2021. We included original or adapted recommendations addressing any COVID-19 topic.

Main outcome measures We used GRADE Working Group criteria for assessing the appropriateness of issuing a GPS: (1) clear and actionable; (2) rationale necessitating the message for healthcare practice; (3) practicality of systematically searching for evidence; (4) likely net positive consequences from implementing the GPS and (5) clear link to the indirect evidence. We assessed guideline quality using the Appraisal of Guidelines for Research and Evaluation II tool.

Results 253 guidelines from 44 professional societies issued 3726 actionable statements. We classified 2375 (64%) as GPS; of which 27 (1%) were labelled as GPS by guideline developers. 5 (19%) were labelled as GPS by their authors but did not meet GPS criteria. Of the 2375 GPS, 85% were clear and actionable; 59% provided a rationale necessitating the message for healthcare practice, 24% reported the net positive consequences from implementing the GPS. Systematic collection of evidence was deemed impractical for 13% of the GPS, and 39% explained the chain of indirect evidence supporting GPS development. 173/2375 (7.3%) statements explicitly satisfied all five criteria. The guidelines' overall quality was poor regardless of the appropriateness of GPS development and labelling.

Summary box

What is already known about this subject?

⇒ Good practice statements (GPS) (ie, actionable statements about interventions that would do substantially more good than harm or vice versa) do not qualify for rating the certainty of evidence, but are important statements in guidelines. The GRADE Working Group developed five criteria to assess the appropriateness of issuing a GPS.

Summary box

What are the new findings?

⇒ Statements that qualify as GPS constitute more than half of the actionable statements in COVID-19 guidelines; there was rarely any appropriate labelling and a lack of transparency in the rationale for their development.

How might it impact clinical practice in the foreseeable future?

⇒ We provide a structured framework for GPS evaluation. Utilisation of this framework by researchers will help monitor the progress around GPS development and evaluate potential barriers slowing the uptake of available guidance by guideline developers.

Conclusions Statements that qualify as GPS are common in COVID-19 guidelines but are characterised by unclear designation and development processes, and methodological weaknesses.

Introduction

Several formal approaches have emerged to structure the process of developing health recommendations in guidelines.[1] Within guidelines, there are a variety of actionable statements for application by clinicians, consumers and other stakeholders.[2] These actionable statement can be further broken down into the categories of formal recommendations, informal recommendations and good practice statements (GPSs). Formal recommendations use the best available evidence and should be developed based on transparent and trustworthy methods.[3–6] Such recommendations are the central aim of guideline development. Informal recommendations resemble formal recommendations but they lack reporting or use of rigorous guideline development methods. GPSs, sometimes referred to as best practice statements, form a separate category of actionable statements that are considered important to issue for healthcare practice.[2] GPSs differ from formal and informal recommendations as they are not typically based on systematic reviews of the evidence and do not include a rating of the certainty of evidence using approaches such as Grading of Recommendations Assessment, Development and Evaluation (GRADE).[7 8] The GRADE approach is the most widely used tool for guideline developers to assess the certainty in effect estimates and subsequently translating the evidence into recommendations using a standardised and transparent evidence to decision framework.[7 9 10]

Due to the lack of international consensus guidance for GPS development and reporting, they are commonly confused with other GRADEd recommendations. For example, GPSs are frequently reported as strong recommendations with low or very low-quality evidence.[11–13] To clarify this confusion, GRADE proposed the following five criteria to assess the appropriateness of issuing a recommendation as a GPS and differentiate them from GRADEd recommendations[8] : (1) statement is clear and actionable, (2) message is necessary regarding healthcare practice, (3) implementation of the statement likely to result in large net positive consequences, (4) summarisation of evidence would be poor use of guideline panel's time and (5) the rationale connecting the indirect evidence used to support the statement is clear and explicit.

The prevalence and quality of GPS in guideline documents has not been empirically evaluated, particularly during the current COVID-19 pandemic where healthcare professionals, scientific societies and government agencies invested a substantial amount of time and resources in developing clinical practice guidelines to reduce information gaps and improve patient outcomes. Furthermore, the application of the GRADE criteria for GPS have neither been operationalised as guidance for those evaluating guidelines nor for developers of GPS. During the development of the global living map of COVID-19 recommendations and portal for contextualisation (eCOVID-19RecMap)[14 15] (https://COVID-19.recmap.org), we identified and evaluated GPS for their appropriateness for development to inform clinical practice.

Methods

Search

We systematically searched MEDLINE (PubMed) from 1 March 2020 to 24 September 2021 using a search string: ((practice guideline[PT]) OR (practice guidelines as topic*[MH])) NOT (comment[pt] or editorial[pt] or letter[pt] or interview[pt] or case reports[pt] or news[pt]), with no restrictions on the language of publication, as part of work to build the eCOVID-19RecMap.[15] We searched ECRI Clinical Guidelines, International Database of GRADE Guidelines (BIGG database), National Institute for Health and Care Excellence (NICE), the World Health Organization (WHO), Centers for Disease Control and Prevention (US CDC) and Guidelines International Network (GIN)'s libraries using an automated web scraping approach via Application Process Interfaces (API). We also manually searched MedSci and China National Knowledge Infrastructure (CNKI) databases to identify Chinese guidelines.

Additionally, we manually searched websites of the following guideline organisations: Public Health Agency of Canada (PHAC), Scottish Intercollegiate Guidelines Network (SIGN), Canadian Task Force on Preventive Health Care (CTFPHC), European Centres for Disease Control and Prevention (ECDC). We also contacted guidelines developers of all the above organisations to keep us apprised of any new or updated guidelines.[15]

Identifying COVID-19 guidelines

We included guidelines eligible for the eCOVID-19RecMap with the most recent guideline uploaded on 24 September 2021. These guidelines reported original or adapted recommendations and were consistent with the WHO definition of practice guidelines while addressing any topic regarding patients at risk for or infected with COVID-19.[16] Online supplemental table S1 describes the definition in detail. We selected guidelines for the eCOVID-19RecMap based on a prioritisation process developed within the eCOVID-19RecMap executive research team (https://COVID-19.recmap.org/about). A topic is a priority if it satisfies one of the following in COVID-19 context[1]: arises commonly in practice,[2] uncertainty in practice,[3] new evidence to consider,[4] existence of variations in practice,[5] important consequences for high resource use/cost,[6] not adequately addressed in existing guidelines.[17] The priority list was refined weekly according to the climate of the pandemic at the current point in time.

We did not restrict guideline eligibility by population group, organisation, country, guideline quality or language. However, we only extracted and evaluated non-English guidelines that could be translated to English by members of our multinational team. For guidelines with more than one version, we evaluated the most recent update. Guideline eligibility was determined by

Table 1 GRADE criteria for evaluating GPS modified from reference[8]*

Signalling question*	Description
Is the statement clear and actionable?	Specific statement that includes the specification of the population of interest.
Is the message really necessary in regard to actual healthcare practice?	Without the guidance provided by the statement, clinicians might fail to take the appropriate action. Knowledge of that practice among the clinicians who represent the target audience is suboptimal.
After consideration of all relevant outcomes and potential downstream consequences, implementing the good practice statement results in a large net positive consequence?	Certainty of benefits and harms are great; the values and preferences are clear; the intervention is cost saving; and the intervention is clearly acceptable, feasible and promotes equity.
Is collecting and summarising the evidence a poor use of a guideline panel's limited time, energy, or resources (opportunity cost is large)?	Poor use of a guideline panel's time and resources to collect and link the indirect evidence is an issue of opportunity cost and their time and energy better spent on other efforts to maximise the guideline's methodologic quality and over-all trustworthiness.
Is there a well-documented clear and explicit rationale connecting the indirect evidence?	The rationale should include an explicit statement of the chain of evidence that supports the recommendation.

*The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group developed these criteria for guideline developers (to designate GPS in their guidelines) and those evaluating the appropriateness of GPS. All five criteria should be fulfilled to designate a statement as GPS.

GPS, good practice statement.

two researchers independently, with consensus or arbitration for a final decision if needed.

Identifying actionable statements that qualify as GPS

We identified actionable statements from the included guidelines using the framework proposed by Lotfi *et al*.[2] In brief, statements that are actionable in isolation with an expected large net benefit, not GRADEd for strength or the certainty of evidence or accompanied by a citation for supporting evidence and the alternative of the stated statement were judged as illogical or did not conform with ethical norms were qualified as GPS.[2] Additionally, researchers extracted statements in the guidelines labelled as best practice or GPSs. We used this approach to identify GPS because there is no universally accepted approach for presenting GPS in guidelines and they are often inconsistently labelled.[13 18] Two researchers extracted the statements and experts in guideline development reviewed them as a quality control step. In addition, we extracted the source, topic (eg, infection prevention and control, vaccination) and intended user and applicable context of each guideline.

Evaluating GPS

We compared the appropriateness of issuing the GPS labelled by guideline developers with statements that qualified as GPS using the five GRADE criteria in table 1.[8] We piloted a form using answer options of 'yes', 'probably yes', 'probably no' and 'no' and developed instructions for how to use the form (online supplemental figure S1). Trained methodologists held weekly meetings to optimise these judgements by discussing examples from guidelines. We used the following approach for the judgements: researchers selected 'yes' and 'no' answers when information supporting or opposing the qualification of the statement as GPS, respectively, was explicit in the guideline (any primary document or supplements). We selected 'Probably yes' and 'Probably no' when the information supporting or opposing the qualification of the statement as GPS was implicit, respectively. For the statement to fulfil the GPS criteria, all the criteria ii–v must be answered 'probably yes' or 'yes'. We did not include criterion i as part of the assessment for appropriates of issuing the statement as GPS since it is a requirement for any recommendation.[8] Online supplemental table S3 presents examples of GPS. We then iteratively developed the explanations and signalling questions in table 2 and reordered the original GRADE criteria for the purpose of critical appraisal of GPS. We conducted all the evaluations in duplicate, and an expert

Table 2 Characteristics of included guidelines and good practice statements

	N (%)
Guideline Source (n=200 guidelines)	
WHO	128 (64)
Centers for Disease Control and Prevention	25 (13)
Public Health Agency of Canada	12 (6)
European Centre for Disease Prevention and Control	10 (5)
National Institute for Health and Care Excellence	2 (1)
Scottish Intercollegiate Guidelines Network	2 (1)
Other	21 (11)
Field (n=200 guidelines)	
Public health	160 (80)
Health policy and systems	88 (44)
Clinical practice	69 (35)
Health technology assessment	3 (2)
World region (n=200 guidelines)	
Global	100 (50)
North America	43 (22)
Europe-Central Asia	41 (21)
East-Asian Pacific	11 (6)
South Asia	3 (2)
Middle East-North Africa	2 (1)
Recommendation Topic (n=2375 statements)	
Infection Control	940 (40)
Vaccination	451 (19)
Health services and systems	446 (19)
Planning and monitoring	309 (13)
Treatment and rehabilitation	126 (3)
Diagnosis	52 (2)
Screening	51 (2)
Target users (n=2375 statements)	
Healthcare providers and professionals	894 (38)
Public health officials	845 (36)
General population	321 (14)
School administrations	258 (11)
Government	57 (2)

in guideline development validated them. We resolved disagreements by consensus in weekly group discussions.

Guidelines quality appraisal

To evaluate if the guidelines were developed with rigorous methods, we critically appraised their development process using the Appraisal of Guidelines for Research and Evaluation (AGREE) II tool for three out of six domains that were deemed important for guideline credibility: scope and purpose, rigour of development and editorial independence.[19] The other AGREE domains (stakeholder involvement, clarity of presentation domain and applicability) were not included in the evaluation as they are not as critical for determining the overall quality of the guideline. Two researchers independently conducted the evaluations of the guidelines and a guideline development expert subsequently reviewed them. The scores of each domain item were assessed on a seven-point scale; 0% if each reviewer scored a 1 (minimum value) and 100% for a score of 7 (maximum value) by both reviewers. We identified discrepancies when a difference of 3 points or more per item between the reviewers was found. We resolved these discrepancies by consensus or a third reviewer. The final score per item was calculated as the average of scores between reviewers after resolution of discrepancies if any. We extracted the information from the guidelines into the GRADEpro (www.gradepro.org) app through a new module that allows the creation of GPS. We then included the GPS in the RecMap (https://covid19.recmap.org/recommendations?recommendationFormality=gps).

Patient and public involvement statement

We partnered with public representatives from the Cochrane Consumer network in the development and conduct of the eCOVID-19RecMap project. The representatives participated in weekly calls of the project executive team where this project was reviewed for relevence of content and provided contextual feedback. The representatives were not involved in the extraction and evaluation of the GPS. The larger eCOVID-19RecMap investigator team also reviewed the design and conduct of the project and provided feedback accordingly.

Statistical analysis

Characteristics of the included guidelines and judgements for each of the GPS evaluation criteria were summarised as percentages. Univariate ORs were used to examine the association between guideline and statement characteristics with issuing of GPS. AGREE II scores were calculated according to the AGREE II manual and reported using the median and IQR. All analyses and figures were conducted with R V.4.1.1 software. GPS evaluation and AGREE II scores were stratified by labelling of GPS by guideline developers.

Results

Characteristics of eligible guidelines

We identified 4533 records through PUBMED, MedSci, handsearching and 11 guideline databases and websites. We excluded 1401 (31%) guidelines after deduplication and title screening, and a further 700 (25%) after screening at full text. Of the identified COVID-19 guidelines, 412 were related to care in the context of COVID-19 and 1746 pertained directly to COVID-19. The guidelines pertaining directly to COVID-19 were eligible for publishing on the eCOVID-19RecMap. Of those guidelines, 253 were extracted and evaluated since the formal launch in November 2020 to September 2021 (figure 1). We identified 2375 of 3726 (64%) statements that qualified as GPS in 200 of 253 (79%) guidelines included on the eCOVID-19RecMap (online supplemental table S2). Those 200 guidelines were included in our analysis. On average, 82% of the statements per guideline (range from 2% to 100%) qualified as GPS.

Characteristics of GPs

Table 3 shows that 64% of the guidelines were published by WHO and 13% by the CDC. One hundred and sixty (80%) guidelines were in the field of public health and 50% were produced for global use. Forty per cent of the GPS provided guidance on infection control while the remaining were on a variety of topics including vaccination, planning and monitoring health services, screening, diagnosis and treatment. The GPS targeted a range of users: 38% were nominally intended for healthcare providers and professionals and 36% targeted public health officials. The remaining GPSs were intended to be used by individuals outside the healthcare setting, patients, caregivers and the public. One guideline was translated from French to English while the remaining guidelines were published in English.

Issuing GPS according to guideline characteristics and statement topic

Figure 2 presents the associations between issuing GPS based on the guideline organisation, field, region, and recommendation topic. Guidelines published in the field of clinical practice were less likely to publish statements that qualify as GPS as compared with formal/informal recommendations, while guidelines in health systems and public health were more likely. Guidelines published by WHO, CDC, PHAC, ECDC and SIGN were more likely to issue statements as GPS with varying strengths of association. GPS were more frequently issued in guidelines published for European-Central Asian use (OR 2.01, 95% C.I 1.54 to 2.62). In contrast, guidelines published for global and North American use were less likely to issue statements as GPS. Issuing GPS was more common in statements regarding infection control (OR 1.63, 95% C.I 1.37 to 1.93), planning and monitoring (OR 1.32, 95% C.I 1.03 to 1.71) and health services and systems (OR 3.05, 95% C.I 2.30 to 4.05). Statements considering diagnosis (OR 0.40, 95% C.I 0.27 to 0.61), treatment and rehabilitation (OR 0.16, 95% C.I 0.12 to 0.20) and screening (OR 0.32, 95% C.I 0.22 to 0.47) were less likely to be issued as GPS. Statements concerning vaccination were also associated with being issued as GPS (OR 1.24, 95% C.I 1.00 to 1.53).

Evaluation of development process of the GPS

Only 27/2375 (1%) of the identified statements that qualified as GPS were actually labelled as GPS by the guideline developers. Of those, 23/27 (85%) statements satisfied all the GPS criteria (ii–v) with implicit and explicit rationales for development. 'Clear and actionable' was judged as 'yes' in 89%, 2% were judged as 'probably yes' and 3.7% were judged as 'probably no' (figure 3). For the criterion 'necessity of the message for healthcare', 63% of the GPS were judged as 'yes' and 37% were judged as 'probably yes'. Eleven per cent of those GPS were judged as 'yes' for the criterion relating to net positive consequences from implementing the statement, while 82% were judged as 'probably yes'. For the criterion relating to usefulness of collection and summarisation of evidence, 4% of the GPS were judged as 'yes', 82% as 'probably yes' and 15% as 'probably no'. Fifty-six per cent provided an explicit statement explaining the chain of indirect evidence supporting the development of the GPS and were judged as 'yes' for this criterion. Judgements 'probably yes' was assigned to 56% of the GPS for this criterion.

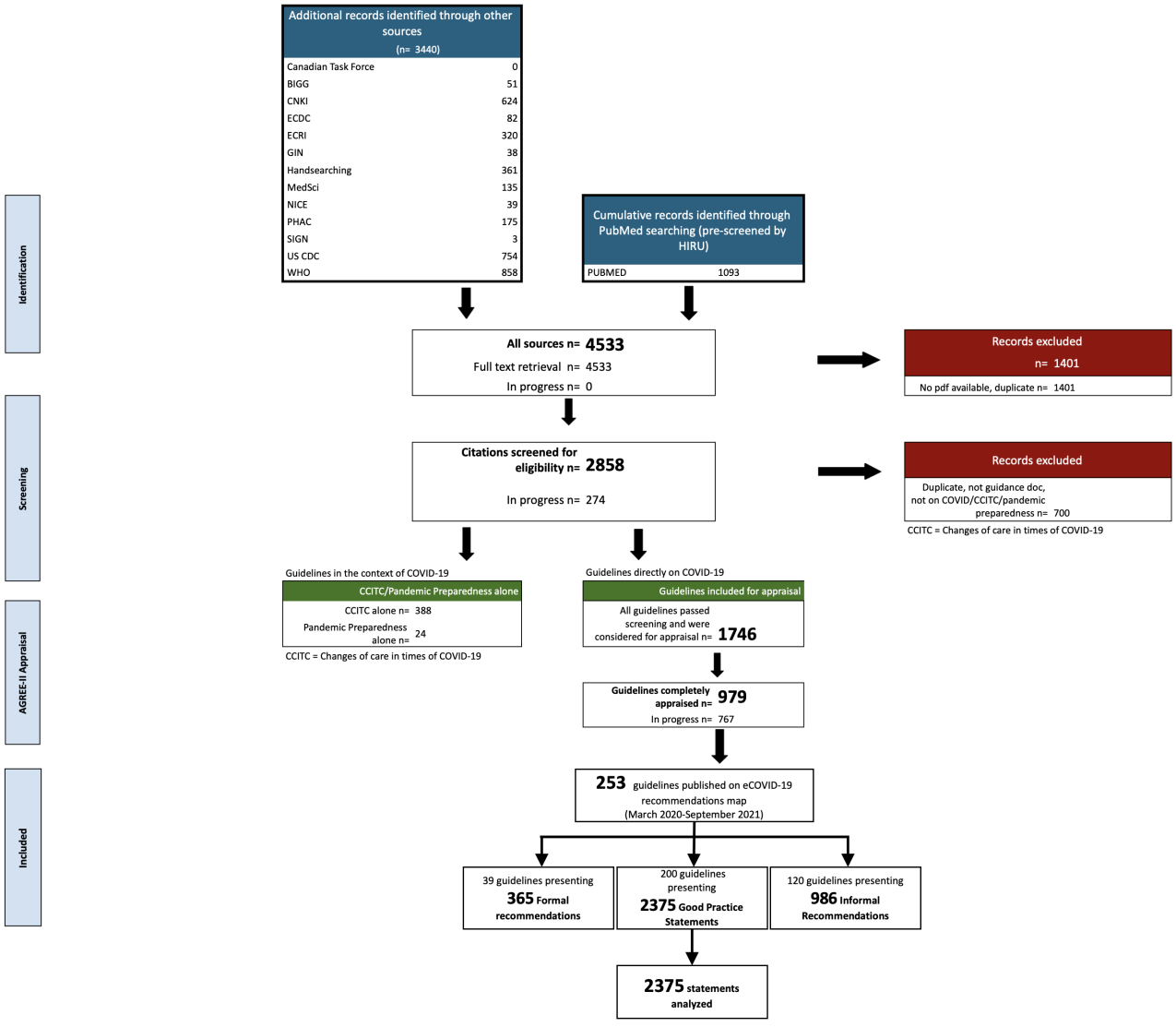

Figure 1 PRISMA chart for guidelines eligible for the eCOVID-19RecMap. BIGG, International Database of Grade Guidelines; CCITC, Changes of Care in Times of COVID-19; CDC: Centers for Disease Control and Prevention; ECDC, European Centres for Disease Control and Prevention; GIN, Guidelines International Network, NICE, National Institute for Health and Care Excellence; PHAC, Public Health Agency of Canada; SIGN, Scottish Intercollegiate Guidelines Network; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses

The reporting of implicit or explicit rationales supporting the development of statements that qualified as GPS (n=2348) was generally similar to those statements labelled as GPS by guideline developers. Of those, 2205/2348 (94%) statements satisfied all the GPS criteria (ii–v) with implicit and explicit rationales. Notable differences in proportion of statements supported with an explicit rationale were found for criteria 'statement leads to large net positive consequence' and 'summarising evidence is a poor use of a guideline development group's time', with more frequent reporting for statements reported as GPS. In contrast, explicit rationales explaining the chain of indirect evidence supporting the development of the GPS was more common for statements not reported as GPS, compared with statements reported as GPS (56% vs 39%, respectively).

Quality of guidelines reporting GPS

The AGREE II evaluation of the six guidelines reporting statements labelled as GPS based on the three domains of interest showed that the overall quality of these guidelines was limited; none of the guidelines scored over 60% for all three domains. Figure 4 shows that the six guidelines with labelled GPS scored a median of 81% (IQR 64–85) in the domain 'Scope and purpose', but only 9.4%, (IQR, 8.3–27 for the domain 'methodological rigour' and 0% (IQR) 0–0) for the domain 'editorial independence'. The 194 guidelines reporting statements that qualified as GPS scored similarly. Two of those guidelines scored over 60% for all three domains.

Discussion

Our evaluation of COVID-19 recommendations using a novel classification that anatomises guidelines into actionable statements[2] shows that guideline developers include advice that frequently qualifies as GPS, (64% of our eligible statements of which 94% satisfied all the GPS criteria ii–v with implicit and explicit rationales) although developers rarely label them as GPS. Accordingly, the evaluation of GPS development processes proved challenging. Statements were more likely to be issued as GPS in European-Central Asia guidelines in the field of public health, specifically statements concerning infection control, planning and monitoring

Table 3 Improving the good practice statement evaluation framework

Evaluation questions	Explanation and signalling questions	Judgement
Is collecting and summarising the evidence a poor use of a guideline panel's limited time and energy (opportunity cost is large)?	▶ Would the investigation of the effect of intervention result only in high certinty indirect evidence? that is, cannot directly investigate the effect of the intervention by comparing to the alternative of the intervention as it would not be sensible/ethical) *Answer 'Yes'* ▶ Does the evaluator believe that the alternative of the intervention is highly unlikely to be chosen due to ethical and human right issues? *Answer 'Probably yes'*	Y/PY/PN/N
Is the message really necessary in regard to actual healthcare practice?	▶ Do the authors provide a rationale in the text of the guideline to why this message is necessary? *Answer 'Yes'* ▶ Does the evaluator believe that the statement is relevant to healthcare practice? *Answer 'Probably yes'*	Y/PY/PN/N
After consideration of all relevant outcomes and potential downstream consequences, does implementing the good practice statement likely results in a large net positive consequence?	▶ Is there any information referenced that the implementation of the good practice statement would have a net positive impact on health outcomes, as well as on relevant Evidence to Decision criteria (eg, equity)? *Answer 'Yes'* ▶ Does the evaluator believe that the implementation of the good practice statement would have a net positive impact on health outcomes, as well as on relevant Evidence to Decision) criteria? *Answer 'Probably yes'*	Y/PY/PN/N
Is there a well-documented clear and explicit rationale connecting the indirect evidence?	▶ Is there a description in the guideline text of the chain of linked indirect evidence, used to infer the net desirable consequences (mainly large health benefits) on the implementation of the good practice statement? *Answer 'yes'* ▶ Does the evaluator believe that there is a chain of linked indirect evidence that can infer the net desirable consequences (mainly large health benefits) on the implementation of the good practice statement? *Answer 'Probably yes'*	Y/PY/PN/N
Is the statement clear and actionable?	▶ Does the statement specify what actions are needed while specifying population or setting in the standard PIC format? *Answer 'Yes'* ▶ Does the statement specify what action is needed while specifying population or setting but not in the standard PIC format? *Answer 'Probably yes'*	Y/PY/PN/N

Outcome is not relevant for the actionable statement as not all outcomes can be addressed in an actionable statement. Outcomes are also not typically part of a recommendation.
PIC, Population, Intervention, Comparator.

and health systems. We found only a few GPS that were supported by rationales for their development regardless of how the guideline developers labelled them. Overall, the quality of most guidelines including formal and informal recommendations was poor and, similar to GPS, the recommendations were often not supported by rationales for their development. Particularly, the reported editorial independence of the guidelines was very low, which could question their trustworthiness. Guidelines to overcome the COVID-19 pandemic would serve healthcare professionals and services better if included GPS were clearly identified and developed through an explicit process. If GPSs are not transparently reported by developers, it is likely that they can be misinterpreted. Thus, in the accompanying article[20], we provide operationalised and structured implementation of GRADE guidance for the development of GPS. Our findings suggest that significant changes are needed in the way guideline developers conduct GPS development. The high prevalence of GPS may be explained by the uncertainty and rapid spread of COVID-19, leading to a lack of direct evidence and immediate need for guidance, reducing the rigour of the guideline development process.

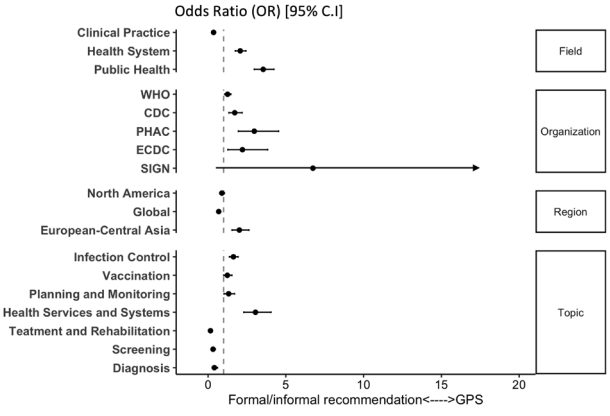

Figure 2 Association of guideline and statement characteristics with issuing statements that qualify as good practice statements. Reference was issuing actionable statements other than good practice statements. Dashed line corresponds to univariate OR of 1.00. We were not able to evaluate associations for guideline regions: South Asia and East Asian Pacific and NICE guideline organisation with issuing good practice statements due to absence of other types of statements. CDC, Centers for Disease Control and Prevention; ECDC, European Centres for Disease Control and Prevention; GPS, good practice statement; NICE, National Institute for Health and Care Excellence; PHAC, Public Health Agency of Canada; SIGN, Scottish Intercollegiate Guideline Network.

Our evaluation shows that the most poorly described criteria were the net consequences of implementing the statement and the usefulness of summarising and collecting the evidence. For the former, many rationales are presumed to be 'straightforward' and based on general knowledge, hence guideline developers may have been reluctant to document this rationale for each statement. For example, in statements regarding infection control (approximately 50% of the statements), the interventions aim to prevent transmission. Although net consequences are not often stated, it is implicitly clear that new cases (and deaths) might be prevented. However, for the latter criterion, the judgement rests on the belief of a guideline panel that they have high confidence in the indirect evidence. A formal documentation is needed to ensure that these statements should truly be issued.

Strengths and limitations

The strengths of this study include the first systematic evaluation of a large sample of COVID-19 GPS irrespective of language, topic, publication source or date of development. We used criteria previously proposed by the GRADE Working Group for GPS but created explanations and signalling questions in addition to response options, which allowed us to differentiate between statements explicitly or implicitly supported by a proper rationale (table 2). All judgements were conducted in duplicate and reviewed by an expert in guideline development after developing guidance for this approach.

Our work has several limitations. First, we did not assess if statements GRADEd as low or very low certainty were GPS rather than formal recommendations. It has been shown that GPS are often incorrectly GRADEd,[12 18] therefore, despite their abundance in COVID-19 guidelines, the actual proportion of GPS may be even higher. Second, despite the use of the most recent version of each guideline, this evaluation is limited by its cross-sectional nature. Temporal changes in the quality of GPS can be assessed in the future as more updated versions of guidelines and recommendations become available. Third, this is the first time this approach to identifying GPS is used and, despite

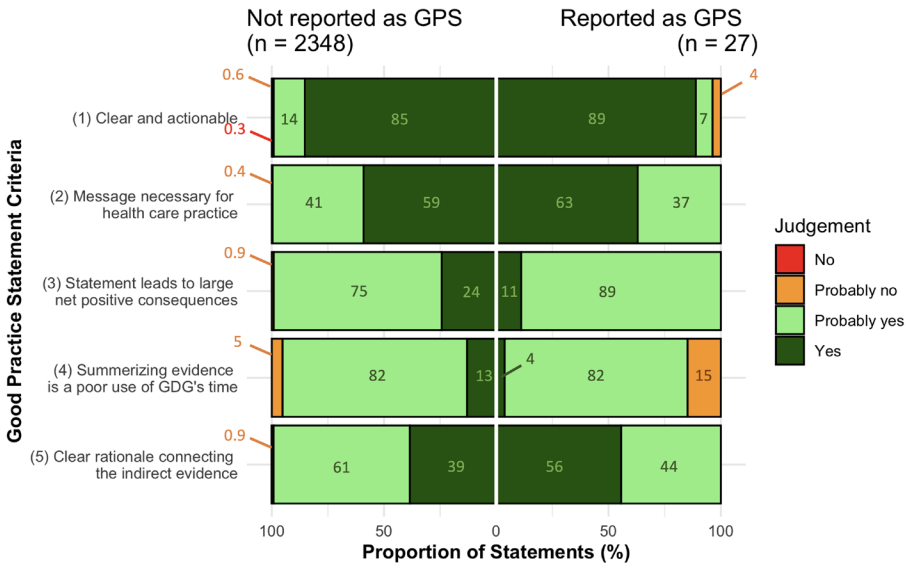

Figure 3 Distribution of judgements for good practice statement (GPS) criteria. Annotations correspond to percentage of statements with their respective judgement. GDG, guideline development group.

face validity using established criteria[8] and the rigorous methods applied (eg, duplicate judgements by extensively trained raters and validated by experts in guideline development), further validation is required. Fourth, our assessment depended on the completeness of reporting in the guidelines and not necessarily the guideline conduct or methods. Fifth, we acknowledge that the nature of the judgement is contingent on a judgement informed by the expertise and knowledge of the evaluator, which may have been variable. To increase confidence, all judgements were completed by two trained reviewers and verified by an expert in guideline development to validate the decisions methodologically. Our multidisciplinary team also includes content experts of various clinical knowledge who were engaged when needed.

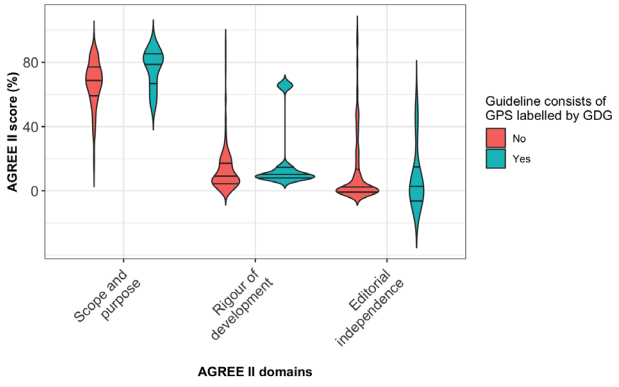

Figure 4 AGREE II assessment (three domains) of guidelines stratified by labelling of good practice statements by guideline developers. Guidelines containing statements labelled by guideline developers as GPS (n=6) and guidelines containing statements that qualify as GPS (n=194). The thickness of the plot represents the kernal density estimation to show the distribution shape of the data. The three lines represent the median and lower (25%) and upper (75%) quartiles based on density estimates. Wider sections of the plot represent a higher probability that guidelines will take on the given value; the slimmer sections represent a lower probability. AGREE, Appraisal of Guidelines for Research and Evaluation; GDG, guideline development group; GPS, good practice statement.

Comparison with other work

Previous work reported that GPS are commonly issued in non-COVID-19 guidelines.[12 18] A retrospective evaluation of discordant recommendations (low or very low confidence in the estimate of effect) in WHO guidelines identified 29 (18%) as GPS. Similarly, a study produced by the Endocrine Society found 43 (35.6%) of discordant statements were GPS, further indicating that GPS are prone to misjudgement.[12 18] Our findings show that GPS are prevalent in guidelines and may be even more commonly used during public health emergencies. The COVID-19 crisis may have impacted developers' ability and capacity to produce more rigorous guidance, forcing them to balance methodological rigour with speed.

Implications for guideline users and developers

First, our study shows that guideline developers should explicitly report the use of GPS in the guideline development process. When not explicitly labelled, two approaches using signalling questions on whether a GPS is justified for development were proposed in prior work.[8] The first involves identifying that the alternative of the statement is absurd or does not conform with ethical norms. The phrasing of the statement may present a source of confusion when identifying the alternative. Hence, may be unreliable when identifying GPS. The second method involves acknowledging that the collection of high-certainty indirect evidence to review and support the statement would be a time-consuming process (criterion iv: summarisation of evidence would be poor use of guideline panel's time). The latter method requires more expertise and familiarity with the field of the statement. In turn, users can assess if GPS were appropriately developed using our methodology.

Second, most of the guidelines were produced for global use but guidelines developed in regions other than high income countries (North America and Europe) were scarce. Thus, implementing the GPS in other settings, especially in low-income settings, may not be feasible. For example, GPS recommending increasing surveillance for farm workers and their close contacts or maintaining humidity level indoors

between 30% and 50% is heavily dependent on resources and influenced by organisational aspects.

Third, adherence to our updated guidance for the operationalisation and implementation of GPS development[20] may improve the transparency in the process of developing and reporting of GPS and help direct guideline developers' resources and efforts to what is needed and avoid the inappropriate issuing of GPS. For example, the European Commission Initiative of Breast Cancer Guidelines on Breast Cancer Screening and Diagnosis[21] reported their GPS in a supplementary document and provided detailed descriptions of the rationales supporting them.

Implications for research

We evaluated the GPS primarily through information provided in the guideline and judgement of the evaluators. Our evaluation of COVID-19 GPS using the previously published five criteria for GPS provided us with insight that improvements to the GPS framework are required to ensure reproducible and valid future evaluations of GPS. Our suggested framework for evaluating GPS builds on our incorporation of judgements with response options that we applied in our evaluation. We also provide a specific order, explanations and signalling questions for using the criteria for GPS evaluation (table 2). For example, the assessment if the statement is actionable and clear was placed at the end of the evaluation as it is not specific to GPS and does not impact on the appropriatness of the rationale for its development. Furthermore, it is not specific to GPS, but is relevant for all actionable statements. We found that using the criterion summarising evidence would be poor use of guideline panel's time as the first criterion for the evaluation, helps with differentiating the GPS from other types of actionable statements although is sometimes a difficult judgement to make. Further testing of this framework by other research teams is required, along with specific GRADE guidance for the development and evaluation of GPS.

Conclusions

The large number of GPS in COVID-19 guidelines emphasises their importance in guidelines especially during public health emergencies, when there is a need for urgent guidance and there is a lack of direct evidence to inform decision making. Our evaluation shows that improvements are needed in the presentation, transparent reporting and the rationale for GPS development beyond the existing GRADE guidance. Furthermore, we need studies to monitor the progress around GPS development and evaluate potential barriers slowing the uptake of available guidance by guideline developers.

Author affiliations
[1]Methods Centre, Bruyère Research Institute, Ottawa, Ontario, Canada
[2]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
[3]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
[4]Michael G DeGroote Cochrane Canada and McMaster GRADE Centres, McMaster University, Hamilton, Ontario, Canada
[5]WHO Collaborating Center for Infectious Diseases, Research Methods and Recommendations, McMaster University, Hamilton, Ontario, Canada
[6]Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[7]Joint Research Centre, European Commission, Ispra, Italy
[8]Instituto de Salud Carlos III, Agencia de Evaluación de Tecnologías Sanitarias, Madrid, Spain
[9]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[10]Department of Pediatrics, Post Graduate Institute of Medical Education and Research, Chandigarh, India
[11]Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
[12]Internal Medicine, Division of Nephrology and Hypertension, University of Kansas School of Medicine, Kansas City, Kansas, USA
[13]Department of Health Product Policy and Standards, World Health Organization, Geneve, Switzerland
[14]Yealth Network, Beijing Yealth Technology Co., Ltd, Beijing, China
[15]Cochrane Campbell Global Ageing Partnership, London, UK
[16]Iberoamerican Cochrane Center, Biomedical Research Institute Sant Pau-CIBERESP, Barcelona, Spain
[17]Cochrane South Africa, South African Medical Research Council, Cape Town, Western Cape, South Africa
[18]Clinical Pharmacology, Department of Medicine, Stellenbosch University, Stellenbosch, Western Cape, South Africa
[19]National Institute for Health and Care Excellence, London, UK
[20]Centre de Recherche du Centre Hospitalier Affilié Universitaire de Québec (CHA), CHA-Hôpital de l'Enfant-Jésus, Université Laval, Quebec, Quebec, Canada
[21]Department of Anesthesiology and Critical Care Medicine, Université Laval, Québec City, Québec, Canada
[22]Czech National Centre for Evidence-Based Healthcare and Knowledge Translation, Institute of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic
[23]Department of Clinical Epidemiology and Public Health, Institut d'Investigació Biomèdica Sant Pau IIB Sant Pau, Barcelona, Spain
[24]Department of Medicine, University of Ottawa Faculty of Medicine, Ottawa, Ontario, Canada
[25]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[26]Deparatment of Family Medicine, University of Ottawa, Ottawa, Ontario, Canada
[27]Joanna Briggs Institute, University of Adelaide, Adelaide, South Australia, Australia
[28]Department of HIV, Hepatitis and Sexually Transmitted Infections, World Health Organization, Geneva, Switzerland
[29]Clinical Research Institute, American University of Beirut, Beirut, Lebanon
[30]Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany
[31]Medical Center - University of Freiburg, Institute for Evidence in Medicine, Freiburg, Germany
[32]American College of Physicians, Philadelphia, Pennsylvania, USA
[33]Department of Biomedical Sciences, Humanitas University, Milan, Italy
[34]Cochrane Canada, Hamilton, Ontario, Canada

Correction notice This article has been corrected since it first published. ORCID has been added for Miloslav Klugar.

Twitter Omar Dewidar @okdewidar

Acknowledgements We would like to acknowledge the research collaborators that were involved in the screening of guidelines and the extraction and evaluation of the GPS included in the eCOVID-19 recommendations map.

Contributors OD, TL, ML, ZSP, EP and HJS contributed to the study conception and design. KS designed and ran the literature searches. OD, JK, ZN and MM screened literature and conducted the data extraction and evaluation. Authors provided feedback on the conceptual approach used in this study. All authors provided critical review, interpretation and approval of the final manuscript. Other members of the eCOVID-19 recommendations map contributed to the screening, data extraction and data validation but did not meet authorship criteria. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. HJS acts as guarantor and accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding CIHR (FRN VR4-172741 & GA3-177732) for COVID-19 recommendation mapping. AFT is the Chairholder of the Canada Research Chair in Critical Care Neurology and Trauma.

Competing interests HJS, AS, VAW report grants from Canadian Instituites of Health during the conduct of the study—FRN VR4-172741 & GA3-177732. RAM reports grants from WHO, grants from ASH, grants from ACR, other from Boehringer ingelheim international, grants from NIDDK, outside the submitted work. EA, PA-C and HJS report contribution to the development of the original five criteria for assessing the appropriateness of issuing good practice statements. The remaining authors have nothing else to declare.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information. All relevant data included.

ORCID iDs
Omar Dewidar http://orcid.org/0000-0001-6420-887X
Karla Solo http://orcid.org/0000-0001-6134-9140
Derek K Chu http://orcid.org/0000-0001-8269-4496
Tamara Kredo http://orcid.org/0000-0001-7115-9535
Miloslav Klugar http://orcid.org/0000-0002-2804-7295
Vivian Andrea Welch http://orcid.org/0000-0002-5238-7097
Joanne Khabsa http://orcid.org/0000-0002-4336-3501
Thomas Piggott http://orcid.org/0000-0003-1643-5386
Holger J Schünemann http://orcid.org/0000-0003-3211-8479

References

1 **Woolf S**, Schünemann HJ, Eccles MP, *et al*. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. *Implement Sci* 2012;**7**:61.
2 **Lotfi T**, Hajizadeh A, Moja L, *et al*. A taxonomy and framework for identifying and developing actionable statements in guidelines suggests avoiding informal recommendations. *J Clin Epidemiol* 2022;**141**:161-171.
3 **Oxman AD**, Fretheim A, Schünemann HJ, *et al*. Improving the use of research evidence in Guideline development: introduction. *Health Res Policy Syst* 2006;**4**:12.
4 **Qaseem A**, Forland F, Macbeth F, *et al*. Guidelines international network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012;**156**:525–31.
5 **Schünemann HJ**, Best D, Vist G, *et al*. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003;**169**:677.
6 **Schünemann HJ**, Wiercioch W, Etxeandia I, *et al*. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 2014;**186**:E123–42.
7 **Guyatt G**, Oxman AD, Akl EA, *et al*. Grade guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;**64**:383–94.
8 **Guyatt GH**, Alonso-Coello P, Schünemann HJ, *et al*. Guideline panels should seldom make good practice statements: guidance from the grade Working group. *J Clin Epidemiol* 2016;**80**:3–7.
9 **Guyatt GH**, Oxman AD, Kunz R, *et al*. What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008;**336**:995–8.
10 **Guyatt GH**, Oxman AD, Vist GE, *et al*. Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;**336**:924–6.
11 **World Health O**. Chapter 14: Strong recommendation when the evidence is low quality. In: *Who Handbook for Guideline development*. **2014**. 2 edn. Geneva: World Health Organization, 2014.
12 **Brito JP**, Domecq JP, Murad MH, *et al*. The endocrine Society guidelines: when the confidence CART goes before the evidence horse. *J Clin Endocrinol Metab* 2013;**98**:3246–52.
13 **Ponce OJ**, Alvarez-Villalobos N, Shah R, *et al*. What does expert opinion in guidelines mean? A meta-epidemiological study. *Evid Based Med* 2017;**22**:164–9.
14 COVID19 recmap. Available: https://covid19.recmap.org
15 **Lotfi T**, Stevens A, Akl EA, *et al*. Getting trustworthy guidelines into the hands of decision-makers and supporting their consideration of contextual factors for implementation globally: recommendation mapping of COVID-19 guidelines. *J Clin Epidemiol* 2021;**135**:182–6.
16 **Organization WH**. Who guidelines, 2021. Available: https://www.who.int/publications/who-guidelines
17 **Wiercioch W**. *Priority topics for panel engagement in health Guideline development McMaster University*, 2020.
18 **Alexander PE**, Brito JP, Neumann I, *et al*. World Health organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with grade guidance. *J Clin Epidemiol* 2016;**72**:98–106.
19 **Brouwers MC**, Kho ME, Browman GP, *et al*. Agree II: advancing Guideline development, reporting and evaluation in health care. *CMAJ* 2010;**182**:E839–42.
20 **Dewidar O**, Lotfi T, Langendam MW. Good or best practice statements: proposal for the operationalisation and implementation of grade guidance. *BMJ Evid Based Med* 2022. [Epub ahead of print: 15 April 2022]. doi:10.1136/bmjebm-2022-111962
21 **European Commission of breast cancer (ECIBC)**. Organising breast cancer screening programmes, 2021. Available: https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/organisation-of-screening-programme