From severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) to coronavirus disease 2019 (COVID-19): a systematic review of the quality and responsiveness of clinical management guidelines in outbreak settings [version 1; peer review: 1 approved with reservations]

Samuel Lipworth¹*, Ishmeala Rigby¹*, Vincent Cheng²*, Peter Bannister¹, Eli Harriss³, Karen Cook⁴, Erhui Cai¹, Mais Tattan¹, Terrence Epie¹, Lakshmi Manoharan¹, Kate Lambe⁵, Melina Michelen¹*, Anna Vila-Gilibets⁵, Peter Hart⁶, Helen Groves⁶, Andrew Dagens¹, Louise Sigfrid¹, Peter Horby¹

¹Nuffield Department of Medicine, University of Oxford, Oxford, UK
²University of Bristol, Bristol, UK
³Bodleian Library, University of Oxford, Oxford, UK
⁴No affiliation, NA, UK
⁵Brighton and Sussex Medical School, Brighton, UK
⁶Wellcome Trust, 215 Euston Road, London, NW1 2BE, UK

* Equal contributors

Abstract

Background: Clinical management guidelines (CMGs) can be useful tools to guide clinician’s decision making and enable consistent evidence-based high-quality care. Here, we assessed whether their objective quality has improved over time by considering CMGs for severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) and from different timepoints for coronavirus disease 2019 (COVID-19).

Methods: We performed a rapid literature review, quality assessment and focus group consultation. The Appraisal of Guidelines for Research and Evaluation (AGREE-II) tool was used to evaluate the quality of the CMGs. In total, six COVID-19 treatments were selected to assess the responsiveness of a subset of guidelines and their updates to 20th November 2020. We ran two sessions of focus groups

Open Peer Review

Reviewer Status

Invited Reviewers

version 1
02 Jul 2021

1

Daniela O. de Melo
Universidade Federal de São Paulo, São Paulo, Brazil

Any reports and responses or comments on the article can be found at the end of the article.
with patient advocates to elicit their views on guideline development. **Results:** We included 37 COVID-19, six SARS, and four MERS CMGs. Evidence appraisals in CMGs generally focused on novel drugs rather than basic supportive care; where evidence for the latter was provided it was generally of a low quality. Most CMGs had major methodological flaws and there was no evidence of improvement in quality over time. CMGs scored lowest in the following AGREE-II domains: scope and purpose, editorial independence, stakeholder engagement, and rigour of development. Of the COVID-19 CMGs, only eight included specific guidance for the management of elderly patients and only ten for high-risk groups; a further eight did not specify the target patient group. Early in the pandemic, multiple guidelines recommended unproven treatments and whilst in general findings of major clinical trials were eventually adopted, this was not universally the case. **Conclusions:** The quality of most CMGs produced in coronaviridae outbreaks is poor and we have found limited evidence of improvement over time, highlighting that current development frameworks must be improved. **PROSPERO registration:** CRD42020167361 (17/02/2020)

**Keywords**
Clinical Management Guidelines, Quality, Inclusivity, responsiveness, COVID-19, SARS, MERS, AGREE-II

This article is included in the Coronavirus (COVID-19) collection.
Corresponding author: Samuel Lipworth (s.i.lipworth@gmail.com)

Author roles: Lipworth S: Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Rigby I: Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; Cheng V: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Bannister P: Data Curation, Writing – Review & Editing; Harriss E: Methodology, Writing – Review & Editing; Cook K: Investigation, Writing – Review & Editing; Cai E: Data Curation, Writing – Review & Editing; Tattan M: Data Curation, Writing – Review & Editing; Epie T: Data Curation, Writing – Review & Editing; Manoharan L: Data Curation, Writing – Review & Editing; Lambe K: Data Curation, Investigation, Writing – Review & Editing; Michelen M: Data Curation, Investigation, Writing – Review & Editing; Vila-Gilibets A: Data Curation, Writing – Review & Editing; Hart P: Conceptualization; Groves H: Conceptualization; Dagens A: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Writing – Review & Editing; Sigfrid L: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Writing – Review & Editing; Horby P: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [215091]. The results presented have been obtained with the financial support of the EU FP7 project PREPARE [602525]. SL is an MRC Clinical Research Training Fellow [MR/T001151/1]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Lipworth S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Lipworth S, Rigby I, Cheng V et al. From severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) to coronavirus disease 2019 (COVID-19): a systematic review of the quality and responsiveness of clinical management guidelines in outbreak settings [version 1; peer review: 1 approved with reservations] Wellcome Open Research 2021, 6:170 https://doi.org/10.12688/wellcomeopenres.16735.1

First published: 02 Jul 2021, 6:170 https://doi.org/10.12688/wellcomeopenres.16735.1
**List of abbreviations**

CMG- Clinical Management guidelines  
SARS-CoV-1-Severe Acute Respiratory Syndrome Coronavirus-1  
MERS-CoV- Middle East Respiratory Syndrome Coronavirus  
COVID-19- Coronavirus Disease-19  
SARS-CoV-2- Severe Acute Respiratory Syndrome Coronavirus-2  
PREPARE-Platform for European Preparedness Against (Re-)emerging Epidemics  
PROSPERO- International Prospective Register of Systematic Reviews  
AGREE-II - Appraisal of Guidelines for Research and Evaluation II  
HCID- High Consequence Infectious Disease  
ISARIC- International Severe Acute Respiratory and emerging Infection Consortium  
IQR- Interquartile Range  
NIV- Non-invasive ventilation  
IV- Intravenous  
RECOVERY- Randomised Evaluation of COVID-19 Therapy  
CDC- Centers for Disease Control and Prevention  
IDSA- Infectious Diseases Society of America  
WHO- World Health Organisation  
REMAP-CAP- Randomised, Embedded, Multi-factorial, Adaptive Platform Trial for Community- Acquired Pneumonia  
RCT- Randomised Control Trial  
ACTT-1- Adaptive COVID-19 Treatment Trial  
ARDS- Acute Respiratory Distress Syndrome

**Introduction**

Clinical management guidelines (CMGs) are useful tools to help clinicians provide quality, evidence-based care to patients. Their utility is potentially even greater in an outbreak setting when clinicians are faced with the challenges of managing a new pathogen combined with increased pressures on healthcare services and redeployment to areas in which they have limited experience. Outbreaks are however also associated with significant time pressure and high levels of uncertainty, making production of methodologically rigorous guidelines difficult.

The coronavirus disease 2019 (COVID-19) pandemic has highlighted the disproportionate impact of infectious disease outbreaks on vulnerable (e.g. the elderly and immunosuppressed) and socioeconomically disadvantaged groups in society. Infectious diseases often present differently in these populations and yet most CMGs produced early in the pandemic did not provide specific advice for the management of these groups. As knowledge about new diseases increases as time elapses, the inclusivity, quality and usefulness of CMGs should also improve. Pandemics such as COVID-19 are likely to occur with increasing frequency throughout the 21st Century and a failure to improve the processes by which clinical practice learns and responds will ultimately lead to unnecessary morbidity and mortality.

In this manuscript, we track the evolution of clinical management guidelines across three coronaviridae pan/epidemics: severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We particularly focus on whether the rigour of development of guidelines and inclusivity of vulnerable groups has improved between these outbreaks and over the course of the COVID-19 pandemic. We aim to identify the strengths and weaknesses of guidelines produced in these settings and to evaluate whether lessons from previous outbreaks have been learnt. For a subset of guidelines in the current SARS-CoV-2 pandemic we also examine how responsive these CMGs are in incorporating new evidence from the latest clinical trials. In doing so we ask the bigger question of how clinical management guidelines can be improved as health professionals continue to manage large numbers of COVID-19 patients and for future pandemics.

**Methods**

This review is an update of a rapid review and part of a wider project evaluating the availability, quality and inclusivity of clinical management guidelines for high consequence infectious diseases (HCID). The Preferred Reporting in Systematic Reviews and Meta-Analyses (PRISMA) checklist was used to construct this review (Figure 1). The protocol for this study has been registered at PROSPERO (CRD42020167361, 17/02/2020).

**Search strategy**

In a previous review, we found that most CMGs were not published in peer-reviewed journals and rarely indexed in the electronic databases. We therefore focussed our efforts on extensive hand-searches of the grey literature using a combination of systematic Google and Google Scholar searches and by specifically searching Ministry of Health, national public health agency institutions, World Health Organisation (WHO), Centres for Disease Control and other infectious disease society websites with pre-defined keywords (Extended data, supplementary file 1.0)). This was complemented by utilising the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) network to contact clinical networks in regions where limited numbers of CMGs where initially identified. Finally, we searched reference lists of included CMGs. We aimed to identify a globally representative sample of CMGs, focusing on international and national
CMGs in this review for feasibility and because these likely inform the development of locally developed guidelines at a hospital/regional level. The full search strategy is shown in the Extended data (supplementary file 1.0,1.1). The search was completed on the 6th June 2020.

Inclusion/exclusion criteria
COVID-19, SARS, and MERS CMGs that included recommendations intended to optimise patient care were included. Guidelines that were substantially local policy documents or focused primarily on infection control/diagnostics were therefore excluded. A standardised data extraction template (Extended data, supplementary file 2) was developed by the systematic review team consisting of infectious disease clinicians and researchers with experience in systematic review methodologies and clinical management guidelines. There were no language restrictions.

Screening
Records identified from searches were independently screened, first by title and abstract, followed by full text, by IR and MT. Any disagreements were resolved by PB. Authors of this paper with knowledge of the language the CMG was written in were used; where this was not possible translations were produced with Google Translate.

Data extraction
We utilised a standardised form to extract data (Extended data, supplementary file 2). Data was extracted by one reviewer and checked by a second reviewer (PB, EC, MT, TE, KL, LM, IR, SL, AD, MM, VC and AVG).

Quality assessment
The quality of each CMG was assessed using the Appraisal of Guidelines, Research and Evaluation version II (AGREE-II). The tool consists of 23 questions (scored on a 7-point scale, from 1 (strongly disagree) to 7 (strongly agree) across six key domains (scope and purpose; stakeholder involvement; rigour of development; clarity of presentation; applicability; editorial independence). All CMGs were assessed independently using AGREE-II by reviewers PB, EC, MT, TE, KL, LM, IR, SL, AD, MM, VC and AVG. CMGs where there was
significant discordance in the reviewers’ assessments were identified by calculating Cohen’s Kappa; a threshold of 0.4 was used to trigger further discussion between reviewers to resolve major disagreements. We considered three measures of whether a CMG was high quality: an overall weighted score ≥ 0.7 (threshold suggested by the AGREE-II developers), weighted score ≥ 0.7 on domains 3 and 5 (rigour of development and applicability, previously shown to be most predictive of overall score\(^9\)) and reviewers’ overall assessment of whether they would recommend use of the CMG. Weighted scores were calculated according to the formula presented in the AGREE-II manual\(^{10}\):  

\[
\text{Obtained score} = \frac{\text{Maximum possible score} - \text{Minimum possible score}}{\text{Minimum possible score}}
\]

**Responsiveness/quality over time (subset analysis)**

We tracked a subset of 11 COVID-19 CMGs (selected because they also featured in our earlier rapid review\(^1\)) over time to assess their responsiveness to key results from randomised clinical trials (RCTs) for six treatments (hydroxychloroquine, convalescent plasma, lopinavir-ritonavir, remdesivir, dexamethasone and tocilizumab). For each CMG in this subset, we also compared the AGREE-II scores to those given to earlier versions at the beginning of the pandemic in our previous review.

**Patient and public involvement**

Members of the public were invited to comment on the results and interpretation of our study via a COVID-19 research involvement group on Facebook. There were 14 participants involved, the members were self-selected members of the public, and no incentives were given. Two one-hour semi-structured focus groups, facilitated by two authors (SL, IR), were held via a teleconference call on the afternoon of 22\(^{nd}\) November 2020. Participants worked with review authors (SL, IR) to comment on the methodology and inform the interpretation and presentation of results. The interview questions and field notes can be found as extended data.

**Ethical approval**

We sought the opinion of the University of Oxford ethics committee who opined our involvement of a patient group constituted patient-public involvement and thus did not require ethical review.

**Statistics**

Statistical analysis was performed in the R language for statistical computing\(^{11}\) version 4.0.2 with the ggplot library used to produce graphics\(^{12}\). Wilcoxon Rank-Sum tests were used to compare AGREE scores between groups and p values (Bonferroni-adjusted where indicated) considered significant at 0.05 threshold.

**Results**

In the main searches completed on 6\(^{th}\) June 2020, we identified 47 CMGs (Figure 1)\(^{13-20}\). 37 covered clinical management of COVID-19, four of MERS and six of SARS. Most COVID-19 CMGs were developed by government agencies and published on the websites as standalone documents or acquired via the ISARIC\(^{C}\) network. Although we attempted to ensure that there were at least five national COVID-19 CMGs per continent, we found fewer guidelines produced in Australasia (n=1), South America (n=3) and Africa (n=6), compared to North America (n=7), Europe (n=12), and Asia (n=15). By the World Bank definition\(^{11}\), most guidelines were produced in high (n=21) and upper middle (n=14), followed by lower middle (n=8) and low-income countries (n=1). Three CMGs were produced by international agencies\(^{13-15}\). Additional characteristics of the guidelines can be found in the extended data (files 3 and 3.1).

**Quality evaluation**

Most CMGs were not high quality by any of the three objective measures we used. For example, (27%) 10/37 of COVID-19 CMGs had an overall weighted AGREE-II score of 0.7 or above compared to 2/4 (50%) MERS and 0/6 (0%) SARS\(^5\). Only one guideline scored 0.7 or more for domains 3 (clarity of presentation) and 5 (rigour of development) (Korean Society of Infectious Disease MERS-CoV guideline\(^9\)); notably no COVID-19 guidelines met this standard. In total 25/47 CMGs were recommended for use by both reviewers though there were only six (two MERS-CoV and four COVID-19) where both reviewers agreed no modification was desirable. The highest score of these were COVID-19 CMGs developed by the Infectious Diseases Society of America (IDSA)\(^{13}\), Surviving Sepsis Campaign\(^{14}\), and a MERS CMG developed by the Korean Society of Infectious Disease\(^{15}\). These were notable for their clear expression of clinical questions which were answered rigorously using a defined methodology and were presented to a high standard.

Considering all included CMGs, quality was not equal across the domains of the AGREE-II tool (Kruskall-Wallis p<0.001) and there was a wide distribution of scores within domains (Figure 2). Editorial independence’ (median weighted score 0 interquartile range (IQR) 0-0.08) and rigour of development (median weighted score 0.23 (IQR 0.13-0.35) were the lowest scoring domains. The low scores for editorial independence were generally because there was no statement about the role of the funding body and many lacked conflicts of interest declarations. Low scores for rigour of development reflected the absence of a description of a systematic evidence search methodology, a lack of explicit links to supporting evidence and unclear methods for selecting key recommendations. CMGs scored better for the ‘Clarity of Presentation’ domain (median weighted score 0.67 (IQR 0.47-0.81)). There was weak evidence of a difference in the overall scores of guidelines produced by academic societies vs public health agencies (median 4.5 (IQR 3.5-5.5) vs. median 3.8 (IQR 3.0-4.5), Wilcoxon p=0.06).

**No improvement in quality over time**

To evaluate whether the quality of CMGs improved over time, we appraised CMGs from three coronaviridae outbreaks (SARS 2002-2004, MERS 2012 and COVID-19 2020). There was no evidence that overall scores were different between these outbreaks (SARS median 0.47 (IQR 0.33-0.61), MERS median 0.54 (IQR 0.27-0.83), COVID-19 median 0.57 (IQR 0.50-0.71), Kruskal-Wallis p=0.35). Notably there was
also no evidence of improvement in any of the six domains measured by the AGREE-II tool between the initial and updated COVID-19 guidelines (Bonferroni adjusted paired Wilcoxon rank sum p>0.1 in all cases, Figure 2).

Inclusivity of CMGs
Many CMGs were not specific in their description of the target population. This was reflected in the fact that only 34% (16/47) of all CMGs scored five or more in this AGREE-II question. Most guidelines included general advice for the management of adults, pregnant women, and children, but older and other high-risk groups patients (e.g., immunosuppressed) were notable omissions from many guidelines (Extended data, Tables S1, S6). There were however some examples where this was done well for example in the WHO CMG which includes specific sections relating to the care of older people and pregnant women with COVID-19 as well as guidance on palliative care.

Supportive care recommendations
Nearly all CMGs gave recommendations for aspects of basic supportive care though there was generally little or no supporting literature cited. Most suggested target saturations and methods of oxygen delivery in hypoxic patients, but there were often no links to or discussion of relevant studies. For example, the WHO CMG notes that there is no evidence based guidance for the use of high flow nasal cannula (HFNC) in this setting and recommends that selected patients with COVID-19 and mild acute respiratory distress syndrome (ARDS) be considered for a therapeutic trial of Non-invasive ventilation (NIV)\(^1\). No literature is provided to support this recommendation and the criteria for selecting patients for such a trial are unclear. Similarly the Surviving Sepsis COVID-19 CMG recommends the use of HFNC over NIV but notes that the quality of evidence is weak\(^1\). As a further example, (62%) 23/37 COVID-19 guidelines recommended the use of antimicrobial therapy if bacterial superinfection was clinically suspected (table S5). However most did not give guidance as to how this decision should be made nor give clear criteria for stopping (table S5). Three guidelines recommended the use of procalcitonin to guide antimicrobial use though did not provide specific thresholds\(^18\). Some stratified recommendations for initiating antibiotics by severity of presentation\(^1\)\(^3\).

---

Figure 2. Weighted scores for the six domains of the AGREE-II tool for the four groups of clinical management guidelines (CMGs) included. Dagens et al.\(^1\) refers to CMGs published in the early part of the coronavirus disease 2019 (COVID-19) pandemic. The boxplots show median and interquartile range (IQR) with the upper/lower whiskers showing the position of 1.5*IQR; individual datapoints are represented by black dots.
Recommendations prior to the availability of high-quality evidence

CMGs varied markedly in their approach to uncertainty of therapeutic efficacy. Some noted the presence of ongoing clinical trials but made no comment on whether an agent should be used whilst others explicitly stated that no recommendation either way could be made. There were several instances where CMGs recommended that where such uncertainty existed, individual case-by-case decisions should be made based on clinical judgement (e.g. COREB for remdesivir/hydroxychloroquine/lopinavir-ritonavir and the Korean Society of Infectious Diseases for Intravenous (IV) immunoglobulin). Others (e.g. US CDC, IDSA and WHO) stated that where there was a lack of evidence, agents should only be used in the context of a clinical trial (Figure 3).

Responsiveness to emerging evidence

We followed a group of COVID-19 guidelines and tracked their recommendations on six treatments between January and November 2020 (Figure 3). Of the COVID-19 CMGs, 6/11 (55%) changed their guidance on the use of Dexamethasone in response to the results of the Randomised Evaluation of COVID-19 Therapy (RECOVERY) trial and the Embedded, Multi-factorial, Adaptive Platform Trial for Community-Acquired Pneumonia (REMAP-CAP) trial. Four CMGs initially recommended the use of lopinavir/ritonavir and/or hydroxychloroquine, and all except one, a Russian CMG which noted anecdotal success with its use and recommended use in moderate cases, recommended against its use after the publication of the RECOVERY/SOLIDARITY trials. In the case of Remdesivir, 5/11 (45%) CMGs recommended its use prior to the publication of the results of the Adaptive COVID-19 Treatment Trial (ACTT-1). A similar theme was apparent in the SARS/MERS guidelines where 4/10 (40%) recommended the use of corticosteroids either absolutely or on a case-by-case basis, despite a lack of evidence.

Stakeholder engagement

In our AGREE-II evaluations, CMGs were consistently poorly rated for their involvement of patient groups in their development (median score 0, IQR 0-0). Whilst our patient group acknowledged the need for speed in the development of the CMG in an outbreak setting, they unanimously and strongly believed that public involvement in the production of CMGs for COVID-19 would have been desirable to ensure that the patient perspective is incorporated. For example, whilst specialists are understandably focused on acute and critical care, the group felt that patient involvement might have highlighted the need for better integration with primary care.

Figure 3. Adoption of evidence from clinical trials by clinical management guidelines (CMGs) over time. Intersecting vertical lines show the publication of key clinical trials either as peer-reviewed articles/pre-prints or press-releases. Dots show the publication of CMGs by bodies shown on the y axis coloured according to the recommendation made.
care and the potential utility of ambulatory monitoring (e.g. pulse oximetry). Involving patients early in a pandemic is understandably challenging, nevertheless, development of a pre-identified group that can quickly be available when needed in future was suggested. The patient group were of the opinion that stakeholder involvement, rigour of development and editorial independence (the worst performing domains in our AGREE-II evaluation) were important and that a compromise in their quality was not acceptable despite the mitigating consideration of a pandemic setting. All participants agreed that making CMGs more accessible to a lay audience is something they would value. A few individuals proposed the use of guideline summaries written in plain English, or in the form of infographics and videos. Participants felt that this would better enable patient centred care by facilitating informed discussions with health care professionals.

**Discussion**

This review and responsiveness evaluation of CMGs in MERS, SARS and COVID-19 demonstrates that, as was the case earlier in the pandemic, many CMGs have substantial methodological flaws and there has been little or no improvement between outbreaks/within the COVID-19 pandemic. The substantial heterogeneity observed in therapeutic recommendations at the beginning of the pandemic did however narrow as reliable evidence from clinical trials became available. The rationale for recommendations around supportive care was often unclear and the quality of evidence used to inform these was notably poor. Many CMGs recommended treatments despite them being non-evidence based or even having demonstrated futility. Despite a body of literature now available highlighting atypical presentations of COVID-19, particularly in elderly patients (e.g. less fever, more delirium, falls and diarrhoea), and risk of more severe diseases, most guidelines did not provide specific advice for management of this patient group.5,21,22,36–39

**Unanswered questions and future research**

Our review highlighted that recommendations on supportive care made by CMGs are often underpinned by limited and/or low-quality evidence. Where CMGs did conduct a systematic evidence review, this was usually primarily focussed on antiviral or immunomodulatory therapy. General aspects of supportive care (e.g. timing of intubation vs. a trial of NIV, target oxygen saturations, whether to give antibiotics, fluid balance decisions and thromboprophylaxis dose/agent/post-discharge regimen) are applicable to all viral infectious diseases with pandemic potential and especially important for emerging infections when the evidence base for pathogen specific therapy is limited. These issues should be addressed in living syndromic systematic reviews which would highlight knowledge gaps to be addressed in clinical trials and aid the rapid production of rigorous pathogen specific guidelines. Significant investment in the evidence base surrounding basic supportive care would likely yield great rewards in future and be globally applicable, especially given the relatively greater accessibility and lower cost of these interventions. At the onset of outbreaks, guideline committees could then identify pathogen specific clinical questions for which pragmatic RCTs could be established.

These results demonstrate the need for a better framework for the development of CMGs in outbreak settings. CMGs can still be useful and developed in a rigorous manner even when the quality and quantity of evidence available is minimal. Dissemination of expert opinion may be useful where there is no better option but should be clearly signposted as such and the rationale for recommendations needs to be clearly and transparently presented. We suggest that at least the initial methodology used to produce CMGs is subjected to a more transparent review process and ideally that these reviews should also be published. This is particularly pertinent given that the quality of CMGs did not appear to improve over time, and updated versions use a near identical format to the original. This need not slow their release which could initially be noted as interim guidance having not yet undergone such review (in a similar manner to preprints).

CMGs would benefit from incorporating succinct summaries, with decision making tools such as flowcharts and algorithms to aid rapid decision making on the front line. Patient groups should be involved in the development of CMGs from the beginning and lay summaries should be produced to enable patients to take a proactive and informed role in their care. Whilst this is more challenging in the initial phase of the pandemic, it would be feasible and desirable to have a pool of lay volunteers on standby who could be recruited at short notice to provide input into both guideline development and clinical research. As the COVID-19 pandemic has evolved, a variety of different issues have emerged, including atypical COVID-19 presentations, post COVID-19 syndrome and difficulty accessing medical care during lockdowns. Continuous engagement with all stakeholders would help to identify these issues and ensure that guidelines are responsive to them.

We observed substantial variation in the way that CMGs approach uncertainty when making recommendations on the basis of little and/or low-quality evidence. There are several examples where guidelines either recommended an unproven agent for use in particular patient groups or on a case-by-case basis. There is always a temptation for “compassionate use” of biologically plausible agents for individual patients in extremis with no proven treatment option. If however all patients who were treated with steroids/hydroxychloroquine/remdesivir/convalescent plasma had been randomised into trials from the beginning of the pandemic, we would have known whether these agents are beneficial (or indeed harmful) much sooner and more patients could have benefitted from these results. The success of pragmatic trials such as SOLIDARITY/RECOVERY have demonstrated the feasibility of this even in pandemic settings.

**Strengths and weaknesses of the study**

The inclusion of CMGs from a wide range of countries and organisations over a period of time is a strength of this study. This allowed us to evaluate the response of guideline committees to new emerging evidence. Our review is skewed towards countries in higher income classifications and we only identified one CMG from a low-income country (LIC). The AGREE-II tool is not specifically designed to appraise...
infectious disease CMGs produced during a pandemic, which may have caused us to underestimate the quality of some guidelines.

Conclusions and policy implications
In conclusion, the quality of guidelines has not improved over time and despite publication of data from key clinical trials, some CMGs continue to recommend the use of agents found to be ineffective in RCTs. Existing guideline development frameworks which have successfully improved the quality of CMGs in general, have had minimal effect on those produced in response to epidemics and pandemics. This highlights a need for a CMG development framework to produce timely, evidence based, resource conscious, locally adaptable and inclusive CMGs in response to emerging outbreaks. Vulnerable groups and in particular the elderly continue to be disproportionately overlooked and the relevant specialties (e.g. geriatrics) are underrepresented in CMG development groups. Given that COVID-19 has had such a profound impact on so many people’s lives and that such a vast quantity of public money has been spent, involvement of patients and the public in outbreak preparedness and response, including in CMG development is an area that needs to be urgently improved and must not be neglected in the future.

Data availability
Underlying data
Figshare: AGREE Scores for COVID19/MERS/SARS CMGs.

This project contains the following underlying data:
- raw_agree.tsv (Raw AGREE-II scores)

Extended data
Figshare: AGREE Scores for COVID19/MERS/SARS CMGs.

This project contains the following extended data:
- Supplementary file C2 06.04.2021.docx (Contains the search strategy, data extraction form, guidelines included in the review, AGREE-II domain and average scores, supplementary Tables S1, S2, S3, S4, S5, S6, S7)
- PPI 29.03.2021.docx (Patient-public focus group notes)

Reporting guidelines

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Transparency statement
The lead author affirms that the manuscript is an honest, accurate and transparent account of the study being reported; no important aspects of the study have been omitted; any discrepancies from the study as originally planned have been explained.

Authors’ contributions
SL, IR and VC wrote the first draft of the manuscript, with input from LS and AD. EH performed the search strategy and executed the database search. SL performed the analysis of the AGREE-II scores and SL and IR created the figures. PB, EC, MT, TE, KL, LM, IR, SL, AD, MM, VC, and AVG screened the references, assisted with data extraction and interpretation. KC provided additional comments from a lay perspective and helped to draft sections relating to PPI. LS and AD conceptualised the protocol and study. LS and PH provided overall supervision of the project. All authors reviewed and approved the final content for publication.

The corresponding author attests that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Acknowledgements
We would like to express our gratitude to the ISARIC network of Infectious disease physicians/public health practitioners who were valuable in searching and providing clinical management guidelines used in this review. Notably, the Ministry of Health Seychelles, Christine Williams, Dr Desmond Oppong at the Greater Accra Regional Hospital, Ghana and Prof. Simon Anderson at the University of the West Indies. We thank our patient group for their suggestions and feedback. We thank the ISARIC global coordinating center for their help and support. Thanks to Sarah Dawson for providing advice regarding the formatting of the references. The authors also thank Julian Higgins for his critical reading of the manuscript.

An earlier version of this article can be found on medRxiv (https://doi.org/10.1101/2021.01.12.21249654)


Open Peer Review

Current Peer Review Status: 

Version 1

Reviewer Report 26 July 2021

https://doi.org/10.21956/wellcomeopenres.18456.r44774

© 2021 Melo D. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Daniela O. de Melo

Department of Pharmaceutical Sciences, Institute of Environmental, Chemical and Pharmaceutical Sciences, Universidade Federal de São Paulo, São Paulo, Brazil

I appreciate the opportunity to contribute to improving the paper of this interesting work. Here are some questions/suggestions/comments.

Search

1. The authors make it clear that this is an update of an earlier paper and that they have chosen to only search the gray literature. Why wasn't a search performed in guideline repositories such as the G-I-N (https://g-i-n.net/international-guidelines-library/) or the ECRI (https://guidelines.ecri.org/), for example?

2. Recently a paper that also assessed quality of COVID guidelines found 188 guidelines, it would be important to cite and even discuss differences from that paper to the one just published (https://www.jclinepi.com/article/S0895-4356(21)00077-9/fulltext).

3. It is not clear to the reader at what point the 10 guidelines from the previous work entered the search (figure 1). Were they all new versions? When we see the quality score for the 47 guidelines from this paper, are the 10 new versions of these guidelines also included?

4. When there was an updated version of a guideline that was not included in the previous paper, did you consider both versions (old and new) or just the most recently published one?

AGREE II assessment

1. How many appraisers actually assessed each guideline? The authors only indicate the total number of appraisers, but do not clarify the division between guidelines and appraisers.

2. “further discussion between reviewers to resolve major disagreements” - How was this process? Was it considered item by item or the domain in this concordance analysis? Did the
appraisers discuss until they reached consensus?

3. In addition to the guidelines, were the supplementary documents to the guidelines evaluated when applying AGREE II or only the main document?

4. “an overall weighted score $\geq 0.7$ (threshold suggested by the AGREE-II developers)” - AGREE’s manual does not establish a cutoff and I am unaware of any publications by its developers establishing a threshold to determine which guidelines would be of high quality. What is the reference for this sentence?

5. “score $\geq 0.7$ on domains 3 and 5 (rigour of development and applicability, previously shown to be most predictive of overall score” – For overall quality, only domain 3 was predictive. Domains 3 and 5 were predictive of whether or not the authors would recommend the guideline, which is different of quality.

Results and Discussion

1. “Our review highlighted that recommendations on supportive care made by CMGs are often underpinned by limited and/or low-quality evidence” - Some results appeared in results and discussion sections without being described in methods what would be done and in what way... this is one example.

2. The analysis about the population the guideline is aimed at and about the recommendation of drugs was very interesting, but it lacked an explanation that it would do this in methods. The authors also talk about the quality of the evidence supporting the recommendation without explaining how this analysis was performed.

Conclusion
Both the discussion and the conclusion need to be more related to the findings of this study rather than extrapolating the considerations.

Is the topic of the review discussed comprehensively in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Is the review written in accessible language?
Yes

Are the conclusions drawn appropriate in the context of the current research literature?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Clinical practice guidelines

I confirm that I have read this submission and believe that I have an appropriate level of
expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.