DESCRIBING DIVERGENCE IN COMPARABLE COVID-19 RECOMMENDATIONS

AN EXPLORATORY DESCRIPTIVE STUDY EVALUATING DIVERGENCE IN THE JUDGMENT OF CLINICAL AND PUBLIC HEALTH RECOMMENDATIONS FOR THE MANAGEMENT OF THE NOVEL CORONAVIRUS DISEASE (COVID-19)

BY: ZIL NASIR, HBSc.

A Thesis Submitted to the School of Graduate Students in Partial Fulfilment of the Requirements for the Degree of Master of Public Health

McMaster University MASTER OF PUBLIC HEALTH (2021) Hamilton, Ontario (History)

TITLE: An Exploratory Descriptive Study Evaluating Divergence in the Judgment of Clinical and Public Health Recommendations for the Management of the Novel Coronavirus Disease (COVID-19)

AUTHOR: Zil-e-Huma Nasir, HBSc. (McMaster University)

SUPERVISOR: Professor Holger Jens Schünemann

NUMBER OF PAGES: xii,87

LAY ABSTRACT

Stakeholders use health guidelines to direct their decision-making on a range of clinical and public health issues. Divergence in the judgment or unexplained sub-group considerations in recommendations addressing the same intervention may lead to misunderstandings and possible mistrust in the guideline development process. Conversely, divergence can be valuable when recommendations have been contextualized for different settings. This thesis describes the frequency and types of divergence between comparable recommendations issued to prevent and treat the novel severe acute respiratory syndrome coronavirus 2. Furthermore, this thesis explains the differences in the methods and contextualization factors used to formulate a selected sample of diverging recommendations.

ABSTRACT:

Background: The emergence of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) pandemic has unravelled a global demand for rapid and reliable guidance at the clinical, systems, and policy levels. Therefore, to equip decision-makers with the appropriate knowledge and tools, organizations have published evidence-informed guidance for its prevention and treatment.

Objectives: Various organizations may produce comparable but diverging recommendations for the same intervention or health scope. Diverging recommendations are those that contain varying judgements in their strength, direction, or subgroup consideration associated with the intervention. Nonetheless, the extent of divergence between COVID-19 recommendations remains unknown. Consequently, the primary objectives of this study are to 1) describe the frequency and types of divergence between COVID-19 recommendations for the same intervention and 2) investigate differences in the guideline development process for a selected sample of diverging recommendations.

Methods: We screened guidelines for divergence using the digital COVID-19 Recommendations Catalogue (covid19.recmap.org). Diverging recommendations for the same intervention were grouped into clusters, and differences in their formal judgment of strength and direction were appraised. Additionally, we compared any differences between PICO criteria for comparable recommendations addressing the same health scope. Descriptive statistics were performed to assess the frequency and types of divergence. Finally, we applied deductive content analysis to evaluate differences in the methods for a sample of 12 recommendation clusters.

Results: Two-hundred twenty-three diverging recommendations resulted in the categorization of 66 clusters. Twenty-nine clusters contained clinical also stated as therapeutic recommendations, and 37 clusters contained public health recommendations. Each cluster had a range of 2-8 individual recommendations in divergence with at least one recommendation for the same intervention. Clinical recommendations were more likely to diverge in formal judgment than public health recommendations ($P < 0.001$). We identified differences in the date of publication, the interpretation of evidence, and in the judgments of the Evidence-to-Decision framework between comparable recommendations. Consequently, results from our study may have important implications for comparing duplicate recommendations and for making clinical practice decisions.

Conclusions: From our study, we have identified diverging recommendations for a range of COVID-19 related interventions. These recommendations may have important implications for clinical practice and public health decisions.

PREFACE

This master's thesis is structured as a "sandwich thesis." It contains three chapters, including a review of the literature, a draft manuscript that is intended for publication, and a conclusion which explains the implications of our work for the development of the digital COVID-19 recommendations catalogue.

ACKNOWLEDGEMENTS

I would first like to express my deepest gratitude to my supervisor, Dr. Holger Schünemann, for his support and care throughout the completion of the thesis. Dr. Schünemann has openly welcomed me to his team and provided me with various opportunities for learning and growth. He has shared his expertise as a clinical researcher and has given me unwavering encouragement.

I would also like to thank members of my committee, Dr. Dominik Mertz, Dr. Robby Nieuwlaat, and Dr. Nancy Santesso. They have been incredibly patient, kind, and understanding through the various changes that have been made in the thesis. Their wisdom, expertise in research methodology, and advice have served to better the outcomes of this project.

I also want to express my gratitude to Dr. Tamara Lotfi. Dr. Lotfi has been very generous with her time and knowledge. She has attended every committee and supervisory meeting, has stayed in touch with me throughout the year I have been working with the team, and has provided technical and practical advice to conduct my work. I would also like to thank Dr. Emma Apatu, all members of the MPH staff and faculty, my peers at McMaster University, and everyone who has taken the time to support me. This has been an incredible journey, and I am grateful for everyone I have learned from along the way.

Lastly, I would like to thank my family and friends. I would not be able to complete this work without the support and encouragement of my parents, who are the most selfless people I know, my sisters, Aamna and Maryam, who have humoured me during my ups and downs, and my brothers, Humayon and Rafay. They have been nothing short of patient and loving and I am blessed to have them in my life.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION *1*

1.1. The Novel Coronavirus Pandemic **1**

1.2 Guidelines for the Management of COVID-19 ..... **3**

1.2.1 The GRADE Approach 4

1.2.2 Current strategies for formulating timely COVID-19 guidelines 6

1.3 Recommendation Mapping **7**

1.3.1 The Living COVID-19 Recommendation Map 9

1.4 Review of Divergence in Existing Health Research **10**

1.5 Research Question **15**

1.5.1 Objectives 15

References **16**

CHAPTER TWO: MANUSCRIPT DRAFT *20*

2.1 Introduction **21**

2.2.1 Inclusion and Exclusion Criteria 23

2.3 Data Collection & Measurement: **24**

2.3.1 Searching and Screening for Diverging Recommendations 24

2.3.2 Measures: 24

2.3.3 Data Analysis and Handling 25

2.4 Statistical Analysis methods **26**

2.4.1 Quantitative Analysis 26

2.4.2 Qualitative Analysis 26

2.5 Quantitative Results: **27**

2.5.1 Summary of Guidelines 27

2.5.2 Summary of Recommendations 28

2.5.3 Types of Diverging Clusters 29

2.5.4 Types of divergence Between Interventions 29

2.6 Qualitative findings **30**

3.6.1 Convalescent plasma 31

2.6.2 Remdesivir 32

2.6.3 The Use of Face Masks in Children 33

3.4 Discussion **34**

3.4.1 Summary of findings 34

3.4.3 Limitations: 37

3.5 Conclusion **38**

Manuscript Tables and Figures **39**
Table One: Summary of Diverging Clusters 39
Table Two: Summary of Diverging Recommendation Clusters across Overarching Healthscope 40
Table Three: Differences in the Methods Between Diverging Recommendations 41

References: **44**

Appendix A: **47**
Summary of the Types of Divergence 47

Appendix B: Table of Intervention Clusters and Type of Divergence Between Recommendations within Clusters **49**

Appendix C: **56**
Therapeutic Interventions 57
Public Health Interventions 67

CHAPTER THREE: IMPLICATIONS AND CONCLUSION ***78***

3.1 Considerations for developers of the COVID19 Recommendations and Gateway to Contextualization Catalogue **78**

3.2 Considerations for Users of the COVID19 Recommendations and Gateway to Contextualization Catalogue **79**

3.3 Discussion **81**

ADDITIONAL TABLES AND FIGURES: **83**
Table One: Guideline Characteristics Extraction Table 83
Table Two: Categorization Matrix of Predetermined Themes for Content Analysis 84
Three: Detailed Frequencies of Divergence Within Clusters 85
Table Four: Chi-Square Table comparing the type of divergence between public health and therapeutic interventions 86
Figure One: Bar Graph of the Type of Divergence Identified within public health and therapeutic clusters 87

LIST OF TABLES AND FIGURES

Chapter Two: Manuscript Draft	**Pg.#**
Table One: Summary of Diverging Clusters	39
Table Two: Summary of Diverging Recommendation Clusters across Overarching Healthscope	40
Table Three: Differences in the Methods Between Diverging Recommendations	41
Appendix B	
Table 1B: Type of Divergence and Intervention for Recommendation Clusters	48
Chapter Three: Implications and Conclusions	
Table One: Guideline Characteristics Extraction Table	82
Table Two: Categorization Matrix of Predetermined Themes for Content Analysis	83
Table Three: Detailed Frequencies of Divergence Within Clusters	84
Table Four: Chi-Square Table comparing the type of divergence between public health and therapeutic interventions	85
Figure One: Bar Graph of the Frequency of Divergence Identified within Public Health and Therapeutic Clusters	86

LIST OF ABBREVIATIONS:

AGREE Checklist: The international Appraisal of Guidelines, Research and Evaluation

ACOEM: American College of Occupational and Environmental Medicine

Australia: Australian Clinical Taskforce

CDC: Centre for Disease Control and Prevention

COVID-19: The Novel Coronavirus 2019

ECDC: European Centre for Disease Control and Prevention

EtD Framework: Evidence to Decision Framework

GRADE: The Grading of Recommendations Assessment, Development and Evaluation

IDSA: Infectious Disease Society of America

NACI: National Advisory Committee on Immunization

NIH: National Institute of Health

PICO groups: Population, Intervention, Comparison, Outcome groups

PHAC: Public Health Agency of Canada

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

SoF Table: Summary of Findings Table

SSC: Surviving Sepsis Campaign

TB: Tuberculosis

WHO: World Health Organization

DECLARATION OF ACADEMIC ACHIEVEMENT

I, Zil Nasir, declare this thesis to be my own. I developed the study design with assistance from my supervisor and committee members. I also completed the data collection, performed all statistical tests, and completed the thesis and draft-manuscript write-up.

My supervisor, Dr. Holger Schünemann, and my committee members, Dr. Dominik Mertz, Dr. Robby Niewlaat, and Dr. Nancy Santesso, provided me with guidance and feedback throughout the completion of this work. Part of this work was in collaboration with the development of the COVID19 Recommendations and Gateway to Contextualization project, including the tools used to collect data and screen for studies. Some definitions were also adopted from the project.

Part of this work will also be submitted for publication in the future. Co-authorship may be granted to other members associated with the overall project.

CHAPTER ONE: INTRODUCTION

1.1. The Novel Coronavirus Pandemic

In the final months of 2019, a novel variant of the coronavirus disease was first identified in human populations. The virus, clinically termed SARS-CoV-2, manifests as the coronavirus disease (COVID-19), leading to severe acute respiratory tract infection in affected populations [1]. Like its other variants, SARS-CoV-2 transmits primarily from infected individuals through aerosols and respiratory droplets [1,2]. The magnitude of SARS-CoV-2, however, has significantly outweighed any outbreak attributed to preceding variants of this coronavirus. As a consequence of its airborne nature and high infectivity rate, SARS-CoV-2 rapidly spread throughout the globe from human-to-human transmission. Therefore, The World Health Organization (WHO) officially declared the start of the COVID-19 pandemic on March 11, 2020, when the global incidence of the virus in human populations surpassed 13-fold worldwide compared to its originating country [3].

As of May 2021, more than 170 million cases of COVID-19 have been confirmed worldwide, causing more than 3.54 million deaths [3]. Clinical manifestations of COVID-19 are variable, where some individuals remain asymptomatic, whereas others develop symptoms ranging from mild and moderate to severe. The most commonly reported symptoms for mild and moderate cases are non-specific and include fever, dry cough, and shortness of breath [4]. Conversely, severe cases are frequently characterized by additional clinical complications, including sepsis, liver dysfunction, and acute

respiratory distress syndrome [4,5]. These complications increase the risk of mortality, necessitating hospitalization to treat and manage symptoms [4,5]. Observational studies have identified older adults and people with comorbidities as high-risk [1-5]. For these individuals, preventative strategies can be the critical difference between life and death [4]. Consequently, to protect vulnerable populations and reduce overall transmission, several countries with persisting cases have implemented infection prevention and control (IPC) measures [3].

A variety of drug interventions have also been prescribed to reduce and manage symptoms of COVID-19 infection [6,7]. Despite this, available pharmacotherapy options are largely dependent on authorization from national government bodies. For example, there is consensus among global healthcare providers to use anti-inflammatory drugs such as dexamethasone to reduce symptoms associated with severe dispositions of COVID-19 [6,7]. However, the use of alternative drug options, such as remdesivir, remains divisive, with some countries adopting its use while others wait for more reliable clinical evidence before deciding to prescribe it to patients [6,8,9].

As a result of global efforts, several vaccines for COVID-19 have also been developed and authorized for distribution. Some vaccines, including the Pfizer-BioNTech and Moderna, require two complete doses, whereas others, such as the AstraZeneca, need one to be most efficacious [10]. Clinical trial data have demonstrated vaccine efficacy to range from 79-95% for preventing mild and moderate cases and nearly 95-100% for preventing severe cases [10,11]. Data from observational studies show that vaccine effectiveness is comparable but lower than efficacy rates [12]. Despite these promising

results, many countries face vaccine shortages due to the limited resources available to manufacture and distribute them [13]. To address this concern, most countries have adopted a risk-based prioritization model where healthcare workers, older adults, and other higher-risk groups are vaccinated before groups with a lower risk of developing severe COVID-19 [13]. At the time of this study, only 5.5% of the global population was vaccinated entirely [14]. Therefore, most countries have continued to implement IPC measures and will continue to do so until enough members of their population are vaccinated to reduce community transmission impactfully.

1.2 Guidelines for the Management of COVID-19

With a shared resolution to stop the spread of COVID-19, several organizations comprising guideline societies, health groups, government bodies, and academic assemblies have issued practice guidelines containing recommendations for the prevention, screening, diagnostics, and treatment of the coronavirus. Although we can consider various standards of what constitutes a health guideline, they are predominantly described as "statements that include recommendations intended to optimize patient care … informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options" [15]. Guidelines are intended to advise decision-makers or institutions on an assortment of interventions for different groups of populations [15]. Decision-makers typically include policymakers, program managers, clinicians, and other healthcare professionals, whereas institutions can extend from clinics and hospitals to larger public organizations and establishments [16]. When generating recommendations, a range of informal or formal methods can be applied by guideline developers. In this

context, standard methods using existing validated approaches are employed to produce evidence-informed recommendations.

1.2.1 The GRADE Approach

Endorsed by more than 100 organizations, The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach is a framework that involves using clinical evidence to transparently formulate practice recommendations [16-19]. GRADE specifically intends to judge recommendations on a scale, defining them as either strong or conditional (sometimes called weak) and favouring the intervention or the comparator intervention. This appraisal is primarily dependent on the quality of evidence used to formulate the recommendation and balance of the desirable and undesirable health effects associated with its implementation [16-19]. Accordingly, a strong recommendation for the intervention is issued when an expert panel is confident that desirable effects will outweigh any undesirable effects.

The GRADE process is prompted by formulating a health question for each intervention under assessment. The health question is modelled using the PICO (Population, Intervention, Comparison, Outcome (s)) ontology and answered by synthesizing data from relevant clinical trials or observational studies as systematic reviews or equivalent evidence synthesis designs [16-18]. Each PICO question assesses a range of outcomes for the intervention. To determine which outcomes are critical or important to the decision-making process, a systematic approach employing a 7-point scale rating each possible outcome is applied [16-19]. All included outcomes are then compiled into a Summary of Findings (SoF) table, which summarises the measured

estimates of effects determined by each included study across all outcomes [16-19]. The quality of included studies is simultaneously assessed using the GRADE approach. This approach contains five items that downgrade and three items that upgrade the certainty of evidence [16-19]. The synthesized evidence is then presented to a guideline panel, ideally consisting of a multidisciplinary group of experts on the guideline topic, who convene to review and translate the findings into practice recommendations. Panellists can also include consumers and other people affected by the recommendation. However, to participate in the process, panellists are required to formally declare potential conflicts of interest to ensure their judgment is unbiased.

Evidence-informed recommendations are essential tools for stakeholders. One cross-sectional study found that 86% of clinicians in a pool of 219 participants prefer to have explicit statements defining strong recommendations accompanying evidence summaries. Similarly, the study found 81% of clinicians in a pool of 248 participants also prefer to have definitive statements defining conditional recommendations accompanying evidence summaries [20]. Regardless, there are some limitations to the GRADE approach that relates to concerns of its objectivity. For instance, although GRADE is considered a standardized tool, empirical data suggests some experts and guideline developers are concerned with appraisers using different interpretations to grade each item, thus affecting its reliability [21-23]. Although some evidence does suggest the GRADE approach has higher interrater reliability than other methods, this possible limitation may not be a significant concern as GRADE does not put emphasis on reliability rather the transparency of the processes applied to formulate a judgement [16-19,24, 25].

Finally, guideline developers can also employ additional criteria to contextualize recommendations and facilitate their adoption or adaptation in different settings [26]. For instance, the Evidence-to-Decision (EtD) framework consolidates a range of evidence to answer unique PICO questions for a list of context-specific measures. Criteria used to contextualize recommendations include evaluating patient values and preferences, equity, resources, cost-effectiveness, acceptability and feasibility of the intervention for the setting it is considered for. [26]. Subsequently, guideline panels can review the evidence for each criterion to provide one singular context-specific judgment on the recommendations' strength and direction [24].

1.2.2 Current strategies for formulating timely COVID-19 guidelines

Although there is a general agreement among health organizations to employ systematic and explicit approaches, guideline developers may be nuanced by the steps required to compile, appraise, and interpret clinical evidence to formulate practice guidelines in a timely manner. This challenge is further exacerbated by additional time constraints and a scarcity of available evidence to issue high-quality recommendations relating to COVID-19. To circumnavigate these challenges, developers need to select the most feasible approach without compromising the integrity of the guideline development process. The production of rapid guidelines is one pragmatic solution that has resulted from this obstacle. This process entails omitting specific steps in the guideline development process, thus facilitating the production and uptake of recommendations within a shorter time frame [28]. Guideline developers can also choose to formulate

recommendations for priority issues only, thus expending more time and resources for evaluating interventions crucial for the setting [28].

Further issues can persist if there is a deficit in the availability of clinical evidence. This is especially relevant in consequence of the pandemic's unprecedented nature, where the absence of direct clinical research has faltered the production of high-quality evidence-informed recommendations [29]. Preliminary guidelines published during the early days of the pandemic bear witness to our limited initial understanding of SARS-CoV-2 [29]. This constraint meant using alternative resources such as knowledge obtained from case studies, previous outbreaks, and viruses with similar pathologies to formulate interim consensus guidelines while anticipating results from ongoing clinical and observational research. To account for these discrepancies, innovative approaches such as “living frameworks” have been adopted by guideline organizations [30]. This process is multidisciplinary and involves a combination of continuous surveillance to update the literature and timely consultations with a team of living guideline panellists to ensure existing recommendations are either reformed or better supported with higher certainty evidence [30]. Guideline users should consolidate living guidelines regularly to review the most updated recommendations. However, these resources should be publicly available in a modus that is both accessible and transparent.

1.3 Recommendation Mapping

Recommendation mapping is a process that draws on existing concepts within the framework of evidence mapping and computable guidelines to provide a digital curation of clinical and public health recommendations for a specific health condition or disease

[31]. This concept was formally introduced in 2019 as a case study capturing recommendations provided in WHO's Tuberculosis (TB) guidance documents to facilitate the uptake of TB-related recommendations by stakeholders [31]. The initial framework was developed by a team of guideline and GRADE methodologists from McMaster University and partnering groups, interaction designers from Evidence Prime, and staff from WHO's Global TB Programme [31]. Within this framework, individual recommendations pertaining to a specific health condition or disease, coupled with the evidence applied to formulate them, are extracted verbatim from pre-identified guidelines and uploaded to a searchable database.

Recommendations are presented as both map and list views on the platform. List view presents each recommendation on the digital platform in the form of a clickable list. In contrast, the map view conceptualizes a cross-tabulation of each recommendation with the populations corresponding to the horizontal rows and the interventions corresponding to the vertical columns [31]. Accompanying recommendations are any implementation considerations, the judgment of their strength and direction, and the certainty of the evidence for the assessed outcomes. Additionally, any completed SoF tables, EtD criteria, and supporting files are uploaded alongside the recommendations to provide users with a transparent overview of the evidence used to create them [31]. Lastly, the map is targeted for guideline users and decision-makers to access the appropriate recommendations to answer their unique health questions and facilitate their adoption.

1.3.1 The Living COVID-19 Recommendation Map

Modelled after the original recommendation map, existing and new members of the McMaster University team, in collaboration with partnering organizations, received a grant from the Canadian Institute of Health Research to develop a living digital recommendation catalogue of COVID-19 guidelines. The project was proposed with the consideration of increasing the uptake of evidence-informed COVID-19 recommendations in clinical and public health settings by centralizing them to one publicly available platform [32]. Since the project launched in June 2020, a systematic approach orchestrated by a team of global methodologists and guideline experts has been applied and modified to search, screen, appraise, and reconstruct existing guidelines as individually uploaded recommendations on the catalogue [32].

The catalogue is built within the same framework of evidence mapping as the original TB recommendation catalogue with slight modifications in its design and function. Consistent with the original map, each recommendation is categorized per a designated population and intervention code. Official codes are determined by trained extractors using the SNOMED Clinical Terms Database, The International Classification of Diseases 11 (ICD-11), and The International language for Drug Utilization Research from the WHO Collaborating Centre for Disease Statistics Methodology. The primary function of these codes is to structure recommendations with the same population and intervention together. Users are then able to view all possibly relevant recommendations for their clinical or public health question.

Revisiting the concept of "living frameworks," it is apparent that our clinical understanding of SARS-CoV-2 has grown significantly since the start of the pandemic.

Nonetheless, there may be some ambiguity about the effectiveness of certain prevention and treatment strategies. The recommendation catalogue has accounted for this limitation and is presented as a living digital platform [32]. Accordingly, a process has been established to monitor existing recommendations on the map, archive outdated or retracted recommendations, and replace them with new or updated recommendations based on emerging evidence [32].

Certain challenges have also materialized during the production of the catalogue. One limitation in the map is addressing gaps between comparable recommendations for the same intervention by different organizations. There are two possible outcomes when assessing comparable recommendations. Organizations can either make the same recommendation with the same judgment, which would support the conclusions made by guideline developers. Conversely, they can have diverging judgments, which may lead to different implications for policy and practice.

1.4 Review of Divergence in Existing Health Research

The concept of divergence in health research was first introduced in complementary study models evaluating discordance in reported findings from systematic reviews published by different author groups evaluating equivalent PICO questions. Systematic reviews pursue a comprehensive and iterative process to search, appraise and synthesize findings from primary studies, providing an exhaustive overview of existing evidence. [33]. Reviews can further apply meta-analytical statistical procedures to collate quantitative results from individual studies answering the same question to generate a singular outcome or assess heterogeneity between studies [33]. In doing so, systematic

reviews increase the power of outcomes and reduce the effects of random error and bias associated with individual study results [33].

As they are regarded as high-level evidence, systematic reviews should provide consistent results across multiple reviews evaluating the same PICO elements. Similarly, for the same review, other researchers can theoretically apply the same methodology used by the original review authors to reach the same conclusions. However, a growing body of literature suggests that results from duplicate reviews may not be as consistent as advocated, even when adjusting for each PICO criteria [34-37]. As a result, several tools have been developed by experts to help determine which systematic review to use when multiple exist for the same health scope or topic. Notably, the BMJ Best Practices provides a checklist of items for stakeholders to appraise when selecting the most appropriate review for their decision-making process. The checklist begins with more apparent criteria, such as using the most recent review when multiple systematic reviews for the same topic arrive at the same conclusion [34]. This process becomes more nuanced when reviews reach different conclusions, when authors include different inclusion criteria, or when they apply different methodological approaches. When this occurs, each option should be appraised critically by users to select the review most appropriate for their specific health question.

The dilemma of multiple systematic reviews having varying outcomes was first introduced by Jadad in 1997 [35]. This type of disagreement was defined as discordance and propelled a series of studies appraising systematic reviews with dissenting conclusions. In their paper, the authors argue that discordance between multiple reviews

evaluating the same PICO question can occur in either the results or in the conclusions made by interpreting the results [35]. The authors then provide a sequence of procedures for users to follow when deciding which review reaches a suitable conclusion. First, users are recommended to evaluate basic criteria such as confirming if reviews are 1) truly duplicate or if they vary in any PICO element, 2) if there are any differences in the inclusion criteria, and 3) if they apply different screening methods to capture individual studies [35]. If differences are attributed to differences in PICO elements, users are simply instructed to select the systematic review most relevant to their situation [35]. Furthermore, if systematic reviews include different clinical trials or appraise clinical trials differently, they should select the review with the most rigorous assessment of evidence quality. If systematic reviews are similar to the criteria mentioned above, a more comprehensive appraisal needs to be conducted. This includes assessing differences in the specific data extraction methods, tests of heterogeneity, and data synthesis [35].

More recently, Moja and colleagues outlined an updated methodology building on the concepts of Jadad's original framework to compare and contrast duplicate systematic reviews. Both studies concur that quantitative findings or results of systematic reviews can differ in three ways: the direction of effect, magnitude of effect, or the statistical significance of the outcomes [35,36]. Nevertheless, the implications for each type of discordance can vary considerably for decision-makers. In particular, Jadad argues that differences in the clinical benefits versus harms of the intervention between systematic reviews are more important than differences attributed to the magnitude of effect [35].

This is because disagreements on the benefits of an intervention compared to its alternatives will lead to different healthcare decisions [35].

By applying the Jadad framework through a series of chronological steps, Moja examined how frequent duplicate systematic reviews reach different results or conclusions, in addition to deciphering which characteristics can explain these differences [36]. Moja et al. iteratively screened and categorized eligible overlapping systematic reviews made for the same intervention to examine the similarities and differences across their results [36]. Systematic reviews containing the same PICO elements were first categorized into clusters. For studies containing meta-analytical methods, Moja applied the Jadad framework to evaluate the factors contributing to discordance in the direction and statistical significance of results. This framework has since been applied to multiple interventions, including systematic reviews for thrombolytic therapy and interventions related to myocardial infarctions. In their analysis, authors have generally reported discordance attributing to variations in study design, differences in inclusion criteria, and lack of precision of the results [37].

There are several plausible but unconfirmed explanations for divergence between recommendations. Guideline panellists can make different judgments about a recommendation if they consider different contextualization factors such as equity, feasibility, and the cost-effectiveness of interventions across different settings. Contrarily, differences in the methodological design and included evidence may be concerning if not adequately addressed. Although we have flagged a small number of recommendations containing diverging judgments on the COVID-19 recommendations catalogue during our

extraction process, the true extent of divergence between comparable recommendations is unknown. Furthermore, we are uncertain what is causing differences in the judgments between these guidelines.

1.5 Research Question

What is the frequency and types of divergence between COVID-19 recommendations for the same intervention or health scope?

1.5.1 Objectives

1. To describe the frequency of recommendations containing diverging judgments on the Digital eCOVID-19 Recommendations Catalogue and to evaluate whether divergence is attributed to the 1) formal judgment of strength, 2) formal judgment of direction, or 3) PICO element.
2. To describe the differences in guideline development methods for a sample of diverging recommendations by applying deductive comparative content analysis.

References

[1] Ouassou H, Kharchoufa L, Bouhrim M, et al. The Pathogenesis of Coronavirus Disease 2019 (COVID-19): Evaluation and Prevention. J Immunol Res. 2020;2020. doi:10.1155/2020/1357983

[2] Dhama K, Khan S, Tiwari R, et al. Coronavirus disease 2019–COVID-19. Clin Microbiol Rev. 2020;33(4):1-48. doi:10.1128/CMR.00028-20

[3] Timeline of WHO's response to COVID-19. Accessed February 12, 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[4] Alshukry A, Ali H, Ali Y, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) patients in Kuwait. Tizzoni M, ed. PLoS One. 2020;15(11):e0242768. doi:10.1371/journal.pone.0242768

[5] Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Heal. 2020;25(3):278-280. doi:10.1111/tmi.13383

[6] Treatments for COVID-19 - Harvard Health. Accessed April 25, 2021. https://www.health.harvard.edu/diseases-and-conditions/treatments-for-covid-19

[7] Coronavirus disease (COVID-19): Dexamethasone. Accessed April 25, 2021. https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-dexamethasone

[8] FDA Approves First Treatment for COVID-19 | FDA. Accessed May 25, 2021. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19

[9] WHO recommends against the use of Remdesivir in COVID-19 patients. (20 November 2020). Retrieved May 25, 2021, from https://www.who.int/news-room/feature-stories/detail/who-recommends-against-the-use-of-remdesivir-in-covid-19-patients

[10] *AZD1222 US Phase III trial met primary efficacy endpoint in preventing COVID-19 at interim analysis*. (2021). Retrieved April 25, 2021, from https://www.astrazeneca.com/media-centre/press-releases/2021/astrazeneca-us-vaccine-trial-met-primary-endpoint.html

[11] Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Frenck, R. W., Hammitt, L. L., … Gruber, W. C. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine, 383(27), 2603–2615. https://doi.org/10.1056/nejmoa2034577

[12] Dagan, N., Barda, N., Kepten, E., Miron, O., Perchik, S., Katz, M. A., Hernán, M. A., Lipsitch, M., Reis, B., & Balicer, R. D. (2021). BNT162b2 mRNA Covid-19 Vaccine

in a Nationwide Mass Vaccination Setting. *New England Journal of Medicine*, *384*(15), 1412–1423. https://doi.org/10.1056/nejmoa2101765

[13] Torjesen, I. (2021). Covid-19 vaccine shortages: what is the cause and what are the implications? *BMJ (Clinical Research Ed.)*, *372*, n781. https://doi.org/10.1136/bmj.n781

[14] *Covid World Vaccination Tracker - The New York Times*. (2021). Retrieved April 25, 2021, from https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html

[15] *Introduction - Clinical Practice Guidelines We Can Trust - NCBI Bookshelf*. (2011.). Retrieved May 25, 2021, from https://www.ncbi.nlm.nih.gov/books/NBK209546/

[16] Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.ad

[17] Guyatt, G. H., Oxman, A. D., Kunz, R., Falck-Ytter, Y., Vist, G. E., Liberati, A., & Schünemann, H. J. (2008). GRADE: Going from evidence to recommendations. In *BMJ* (Vol. 336, Issue 7652, pp. 1049–1051). British Medical Journal Publishing Group. https://doi.org/10.1136/bmj.39493.646875.AE

[18] Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is "quality of evidence," and why is it important to clinicians?. Bmj, 336(7651), 995-998.

[19] *GRADE handbook*. (2013). Retrieved April 25, 2021, from https://gdt.gradepro.org/app/handbook/handbook.html

[20] Neumann, I., Alonso-Coello, P., Vandvik, P. O., Agoritsas, T., Mas, G., Akl, E. A., Brignardello-Petersen, R., Emparanza, J., McCullagh, L., De Sitio, C., McGinn, T., Almodaimegh, H., Almodaimegh, K., Rivera, S., Rojas, L., Stirnemann, J., Irani, J., Hlais, S., Mustafa, R., … Schünemann, H. J. (2018). Do clinicians want recommendations? A multicenter study comparing evidence summaries with and without GRADE recommendations. *Journal of Clinical Epidemiology*, *99*, 33–40. https://doi.org/10.1016/j.jclinepi.2018.02.026

[21] Norris, S. L., & Bero, L. (2016). GRADE methods for guideline development: Time to evolve? In *Annals of Internal Medicine* (Vol. 165, Issue 11, pp. 810–811). American College of Physicians. https://doi.org/10.7326/M16-1254

[22] Rehfuess, E. A., & Akl, E. A. (2013). Current experience with applying the GRADE approach to public health interventions: An empirical study. *BMC Public Health*, *13*(1), 1–13. https://doi.org/10.1186/1471-2458-13-9

[23] Anttila, S., Persson, J., Vareman, N., & Sahlin, N. E. (2016). Conclusiveness resolves the conflict between the quality of evidence and imprecision in GRADE. Journal of clinical epidemiology, 75, 1-5.

[25] The Grading of Recommendations Assessment, Development and Evaluation Reliability Study (the GRADERS) | Colloquium Abstracts. (2012). Retrieved May 25, 2021, from https://abstracts.cochrane.org/2012-auckland/grading-recommendations-assessment-development-and-evaluation-reliability-study

[26] Hartling, L., Fernandes, R. M., Seida, J., Vandermeer, B., & Dryden, D. M. (2012). From the trenches: A cross-sectional study applying the grade tool in systematic reviews of healthcare interventions. In PLoS ONE (Vol. 7, Issue 4). Public Library of Science. https://doi.org/10.1371/journal.pone.0034697

[27] Moberg, J., Oxman, A. D., Rosenbaum, S., Schünemann, H. J., Guyatt, G., Flottorp, S., ... & Alonso-Coello, P. (2018). The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. Health research policy and systems, 16(1), 1-15.

[28] Morgan, R. L., Florez, I., Falavigna, M., Kowalski, S., Akl, E. A., Thayer, K. A., ... & Schünemann, H. J. (2018). Development of rapid guidelines: 3. GIN-McMaster Guideline Development Checklist extension for rapid recommendations. Health research policy and systems, 16(1), 1-12.

[29] Griffith, G. J., Morris, T. T., Tudball, M. J., Herbert, A., Mancano, G., Pike, L., ... & Hemani, G. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. Nature communications, 11(1), 1-12.

[30] Vogel, J. P., Dowswell, T., Lewin, S., Bonet, M., Hampson, L., Kellie, F., ... & Oladapo, O. T. (2019). Developing and applying a'living guidelines' approach to WHO recommendations on maternal and perinatal health. BMJ global health, 4(4), e001683.

[31] Hajizadeh, A., Lotfi, T., Falzon, D., Mertz, D., Nieuwlaat, R., Gebreselassie, N., ... & Schünemann, H. J. (2021). Recommendation mapping of the World Health Organization's guidelines on tuberculosis: A new approach to digitizing and presenting recommendations. Journal of Clinical Epidemiology.

[32] Lotfi, T., Stevens, A., Akl, E. A., Falavigna, M., Kredo, T., Mathew, J. L., ... & eCOVID Collaborators. (2021). 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. Journal of Clinical Epidemiology.

[33] Uman, L. S. (2011). Systematic reviews and meta-analyses. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20(1), 57.

[34] Multiple systematic reviews on the same question | BMJ Best Practice. (n.d.). Retrieved May 25, 2021, from https://bestpractice.bmj.com/info/toolkit/learn-ebm/multiple-systematic-reviews-on-the-same-question/

[35] Jadad, A. R., Cook, D. J., & Browman, G. P. (1997). A guide to interpreting discordant systematic reviews. Cmaj, 156(10), 1411-1416.

[36] Moja, L., Del Rio, M. P. F., Banzi, R., Cusi, C., D'Amico, R., Liberati, A., ... & Parmelli, E. (2012). Multiple systematic reviews: methods for assessing discordances of results. Internal and emergency medicine, 7(6), 563-568.

[37] Lucenteforte, E., Moja, L., Pecoraro, V., Conti, A. A., Conti, A., Crudeli, E., ... & Virgili, G. (2015). Discordances originated by multiple meta-analyses on interventions for myocardial infarction: a systematic review. Journal of clinical epidemiology, 68(3), 246-256.

[38] Riva, N., Puljak, L., Moja, L., Ageno, W., Schünemann, H., Magrini, N., & Squizzato, A. (2018). Multiple overlapping systematic reviews facilitate the origin of disputes: the case of thrombolytic therapy for pulmonary embolism. Journal of clinical epidemiology, 97, 1-13.

CHAPTER TWO: MANUSCRIPT DRAFT

Title: Factors Associated with Diverging Clinical and Public Health Recommendations for The Management of The Novel Coronavirus Disease (COVID-19)

Abstract:

Background: McMaster University and partnering organizations have developed a living COVID19 Recommendations and Gateway to Contextualization catalogue, a digital platform centralizing published guidelines for the prevention and treatment of the novel coronavirus disease. During this process, we have identified a number of comparable recommendations diverging in their formal judgement of strength and direction or additional consideration associated with the intervention.

Objectives: We conducted an exploratory, descriptive study to describe 1) the frequency and types of divergence between recommendations for the same health intervention and 2) the differences in the guideline development process between comparable recommendations.

Methods: We grouped diverging recommendations for the same intervention into clusters. Descriptive statistics was performed to measure the range, frequency, and proportion of divergence between interventions. We then used deductive content analysis to investigate differences in the guideline development process for a selected sample of recommendations. We used descriptive statistics and analytical statistical approaches to compare differences in the types of divergence between therapeutic and public health interventions.

*Results***:** We identified 223 recommendations in divergence with at least one comparable recommendation and categorized them into 66 individual clusters. Twenty-nine clusters contained clinical recommendations, and 37 clusters contained public health recommendations. Clinical recommendations were more likely to diverge in the formal judgment of strength and direction than public health recommendations ($P < 0.001$). We identified differences in the date of publication, the interpretation of evidence, and judgments of the Evidence-to-Decision framework between comparable recommendations.

Conclusions: The findings from our paper support the current work being completed on the development of the recommendations catalogue. Furthermore, we have provided a foundation to evaluate differences in the methodology between comparable recommendations to evaluate causes of divergence in future studies.

2.1 Introduction

Clinical practice guidelines are indispensable tools for clinicians, policymakers, and other stakeholders to maximize the adoption of evidence-informed recommendations. Many organizations have established procedures for formulating guidelines. In general, these procedures entail synthesizing the best available evidence and relying on a panel of multidisciplinary experts to provide judgment on clinical, public health or health policy interventions [1-3]. Additionally, guideline developers can provide explicit formal judgments of recommendations' strength and direction. This process involves evaluating the certainty of evidence and balancing the desirable and undesirable effects of relevant health outcomes [1-3]. By applying this system, recommendations can be categorized anywhere between a 'strong recommendation for the intervention' to a 'strong recommendation against the intervention' [1-3]. Guideline developers and users can also adopt strategies to contextualize recommendations, thus facilitating their adaptation. Notably, the concept of "adolopment" has been included in the terminology of epidemiology describing the process of adoption, adaptation or de novo creation of recommendations for different populations and settings using tools such as the Evidence-to-Decision (EtD) framework [4]. The objective of this procedure and the overall guideline development process is to improve efficiency in guideline development and in the delivery and outcomes of healthcare interventions.

To facilitate the uptake of guidelines, we have developed a digital platform centralizing all high-quality recommendations for the prevention and treatment of the novel coronavirus (COVID-19), covid19.recmap.org. Recommendations are categorized

corresponding to their population and intervention, allowing users to search for relevant recommendations related to their unique health question(s). The catalogue also publishes any relevant remarks, evidence synthesis, and contextualization factors such as completed EtD criteria, presenting users with a transparent overview of the decision-making process.

While developing the catalogue, a number of comparable recommendations addressing the same intervention but having diverging judgments in their strength and direction have been flagged. The concept of divergence in the results and conclusions of analogous evidence synthesis was first introduced by Jadad et al. in 1997. In their paper, the authors discuss the types of discordance identified in systematic reviews and provide a framework to select an appropriate review when multiple options are available [5]. Specifically, users should appraise a list of items. These items comprise comparing the study question, inclusion criteria, literary search methods to synthesis data, and the metanalytical processes of each systematic review with dissenting conclusions [5]. In 2013, Moja et al. provided an updated methodology to identify how often discordance occurs between comparable systematic reviews and examined a list of characteristics that can explain these differences by applying the Jadad framework [6]. Consequently, it may be feasible to apply approaches from these existing studies to evaluate the concept of divergence and how it corresponds to recommendations on the eCOVID-19 Recommendations Catalogue.

To our knowledge, no current publications have investigated differences in the judgments of comparable COVID-19 recommendations. Ergo, we conducted a descriptive exploratory study using a mixed-methods study design. Our primary objective is to

describe the frequency of COVID-19 recommendations with diverging formal judgments in their strength, direction, or additional subgroup considerations. Our second objective was to investigate differences in the methodological processes used to formulate a sample of diverging recommendations. Ultimately, the findings from our study are intended to inform developers and users of the digital COVID-19 recommendations about these differences to guide future work on the project.

2.2 Methods

2.2.1 Inclusion and Exclusion Criteria

We reviewed any self-reported guideline or guidance document containing recommendations for COVID-19. A self-reported guideline may contain formal recommendations, good practice statements, or additional guidance recommendations. Conversely, we excluded primary studies, including clinical trials, editorials & commentaries, case studies, literature reviews, and singular systematic reviews. Guideline documents from academic societies, formal guideline development groups, government and intergovernmental agencies, and health organizations were all included for review. For this study, we defined a cluster as a set of two or more diverging recommendations for the same intervention or health scope. This term was modified from Moja's work of discordance [6].

2.3 Data Collection & Measurement:

2.3.1 Searching and Screening for Diverging Recommendations

The eCOVID-19 Recommendations Catalogue (covid19.recmap.org) was used as the primary database to search for recommendations for the same health scope or intervention. In conjunction with this, we also assessed flagged guidelines awaiting extraction. Before their analysis, we reviewed guidelines for any updates to ensure we evaluate their most recent versions. Guidelines pre-screened for the eCOVID-19 recommendation map before May 2021 were included in this study for timely analysis. Older versions of updated guidelines and retracted guidelines or recommendations were excluded. Recommendations were searched using their designated population and intervention codes in the catalogue. These codes have been predetermined by extractors on the project. All recommendations issued for the same health scope were flagged for assessment of possible divergence. Guidelines were also individually reviewed to identify additional interventions that can be searched in the catalogue.

2.3.2 Measures:

After reviewing each uploaded guideline, we aggregated all diverging recommendations for the same intervention into alphabetized clusters on Microsoft Excel. The primary classifier of divergence was any difference in the formal judgment of a recommendation's strength and direction. We predominantly used definitions of formal judgment from the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to compare recommendations [3]. Ergo, all recommendations were required to be made for the same population, have the same

comparator group, and have at least one equivalent outcome of interest. However, we separately evaluated divergence in PICO (population, intervention, comparison, and outcome) elements between comparable recommendations. This could include differences in the population, the intervention, the comparator or variation in specific metrics or subgroup considerations associated with the interventions.

To organize our data, we created a characteristics table containing descriptive information pertaining to each flagged recommendation. In particular, we extracted the guideline name, data of most recent update, name of organization issuing the guideline, guideline title, author names, PICO elements, formal judgment, and whether divergence was observed in the strength, direction, or PICO element.

2.3.3 Data Analysis and Handling

We proceeded to code recommendations into nominal or ordinal variables for quantitative assessment using the following nominations: the type of recommendation, overarching health scope of the recommendation (e.g., therapeutic or public health intervention), and intervention subtype. Additionally, we coded the strength and direction of recommendations on a 5-point ordinal scale ranging from (-2) to (+2). Negative two indicates a strong recommendation was issued against the intervention, whereas (+2) indicates a strong recommendation issued for the intervention. Finally, we defined recommendations that deviated in population or intervention separately. Complete extractions and coding details can be found in the supplementary materials file.

2.4 Statistical Analysis methods

2.4.1 Quantitative Analysis

First, we used descriptive statistics to summarize the frequency of diverging recommendations identified from our search. The mean, median, and range of identified recommendations within each cluster were individually measured. Next, we calculated the proportion of diverging recommendation clusters within each overarching health scope and intervention subtype. Finally, we performed crosstabulation calculations to assess patterns of divergence between public health and therapeutic interventions. To evaluate if there is a statistically significant difference between these groups, we conducted a Chi-square and Fisher's exact test. Both tests were conducted in conjunction, because chi-square provides information regarding distributions whereas Fisher's exact test is more conservative. We used the statistical package for the social science (SPSS) to conduct the analyses.

2.4.2 Qualitative Analysis

We applied qualitative content analysis methods to evaluate variance in the guideline development process for a selected sample of diverging clusters. Clusters were selected as random. Furthermore, some clusters were selected using purposeful sampling methods. Recommendations were selected at random. Additionally, some recommendations were selected through purposeful sampling. The content for each included guideline containing diverging recommendations was reviewed twice by one appraiser and compared. Using this information, we developed a list of themes that represent individual methodological differences. To develop our themes, we applied a

deductive approach using existing indicators from 1) Jadad and Moja's framework for evaluating discordance, 2) The GRADE Handbook for grading the quality of evidence and the strength of recommendations, 3) The GRADE BMJ series on guideline development, and 4) The GIN-McMaster Guideline Development Checklist [5-6,7-8,9]. The following themes were incorporated for analysis: differences in the date of publication, differences in the included studies, differences in the interpretation of the results, differences in the EtD frameworks, which incorporates differences in the certainty of the evidence, and additional variances such as the use of experts to assist with final judgments.

The content analysis approach is a series of three chronological steps commencing at the preparation phase, followed by the organization phase, and concluding with the reporting phase [10]. The completed step-by-step guide for the qualitative analysis is available in **Appendix A**. The unit of analysis for our study was the individual recommendations that were assessed for divergence. Content analysis was conducted for six therapeutic interventions and six non-therapeutic or public health interventions. Recommendations were selected randomly using an online generator. In addition to this, we included a few clusters with confirmed recommendations from World Health Organization (WHO) guidelines.

2.5 Quantitative Results:

2.5.1 Summary of Guidelines

Evaluation was conducted on April, 2021. At this time, 73 guidelines were uploaded to the map, with 1330 extracted actionable statements including formal

recommendations, good practice statements, and additional guidance. Additionally, 65 guidelines awaiting extraction were also reviewed. The remaining guidelines were not assessed as a result of time constraints and incomplete screening. Notable organizations with diverging recommendations from other organizations included WHO, The Australian Clinical Task Force, the National Institutes of Health (NIH), and the Infectious Diseases Society of America (IDSA). In total, 35 organizations issued at least one guideline containing diverging recommendations.

2.5.2 Summary of Recommendations

We observed a total of 223 individual recommendations in divergence with at least one other recommendation for the same health scope or intervention. We identified 99 clinical recommendations and 124 public health recommendations. 77 (34.5% of the total sample) clinical recommendations were issued for drug therapies and 22 (9.9% of the total sample) recommendations were issued for other clinical treatments. Moreover, we identified 61 (27.4% of the sample) Infection Prevention and Control (IPC) recommendations, 37 school-related recommendations, and 26 vaccine-related recommendations. Ultimately, 115 recommendations in our sample diverged in judgment of strength or direction and 108 recommendations diverged in population or a specific consideration relating to the intervention.

We aggregated all recommendations into 66 unique clusters **(Appendix B)**. Each cluster consists of singular interventions and all the guidelines with diverging recommendations concerning that intervention. Forty-three percent (n=29) of the clusters contained clinical interventions and 56.1% (n=37) of the clusters contained public health

interventions. The range of individual recommendations in a cluster was 6 (min=2, max=8). Clusters encompassed a median of 3 recommendations (mean=3.38, s=1.57). Furthermore, 31.8% of clusters consisted of pharmaceutical recommendations (n=21) and IPC measures respectively (n=21). Finally, 12.3% of clusters contained additional clinical recommendations (n=8), vaccine-related recommendations (n=8), and school-related recommendations (n=8) respectively as well.

2.5.3 Types of Diverging Clusters

Thirty-four recommendation clusters diverged in the formal judgment of their strength or direction. Of these, 55.9% (n=19) diverged in strength only and 26.5% (n=9) diverged in direction only. Eighteen percent (n=6) also diverged in strength and direction of the recommendations. This meant at least one recommendation in the cluster diverged in strength, and at least one recommendation in the cluster diverged in direction.

Thirty-two recommendation clusters also diverged in PICO element, constituting 48.5% of our total sample. Within this group, 18.8% (n=6) of the recommendation clusters had diverging populations to which the intervention was recommended. Furthermore, 26 clusters had varying subgroup considerations for the same intervention. We have provided a detailed summary of descriptive findings in **Table One.**

2.5.4 Types of divergence Between Interventions

We also classified our clusters into two overarching health scopes and five primary interventions. Intervention classifications included drug therapy, non-pharmaceutical interventions, infection-prevention control measures, vaccine-related

measures, and school-related measures. **Table Two** provides an overview of the types of divergence observed for each overarching intervention. Of the 29 therapeutic interventions, 93.1% diverged in formal judgments of strength or direction, and 6.9% diverged in PICO elements. Furthermore, of the 37 public health recommendations, 81.1% diverged in PICO element. All school-related measures (n=8) diverged in PICO elements, and all non-pharmaceutical clinical recommendations (n=8) diverged in formal judgment only.

To assess if there is a statistically significant difference between therapeutic and public health recommendations, we performed a chi-square and fisher's exact test. To do this, we first formulated a contingency table comparing the types of overarching divergence (e.g., formal judgment or PICO element) to the health scope (e.g., therapeutic or public health). From our analysis, we found clinical recommendations were more likely to diverge in the formal judgment of recommendations. In contrast, public health interventions were more likely to diverge in population or intervention considerations ($P<0.001$) within our sample.

2.6 Qualitative findings

The following therapeutic interventions were selected for qualitative analysis: convalescent plasma, gelatin, intravenous immunoglobulins, ivermectin, remdesivir, and zinc. Additionally, the following public health interventions were appraised: delaying the second dose of COVID-19 vaccines, the frequency of times surfaces should be cleaned to reduce transmission of COVID-19, the use of face masks in children, the preferred mode

of birth for pregnant people with COVID-19, rooming newborns and mothers with COVID-19, and quarantining practices for vaccinated travellers.

A total of 20 guidelines across the 12 recommendation clusters were appraised in our content analysis. From our assessment, we identified three clusters containing recommendations with significant differences in the date of publication or last literature search, three clusters with recommendations formulated using different evidence and three clusters with recommendations developed using different interpretations of the evidence. Additionally, we identified six clusters containing recommendations with different judgments or interpretations of EtD criteria and five clusters containing recommendations that considered additional factors to formulate their final judgment. Differences in EtD criteria also includes variation in the ratings of the certainty of evidence, which was found across two clusters.

This section summarizes differences for three interventions, including convalescent plasma, remdesivir, and the use of face masks in children. These recommendations were selected because they diverge in different manners. However, the completed qualitative analysis for all twelve clusters can be viewed in **Table Three** and **Appendix C** of this paper.

2.6.1 Convalescent plasma

Four organizations issued recommendations concerning the use of convalescent plasma to treat patients with COVID-19. The primary source of divergence was observed in the strength of the recommendations. The Australian clinical task force and the NIH guidelines published strong recommendations against the intervention. In contrast, the

IDSA and the Surviving Sepsis Campaign (SSC) guidelines published conditional recommendations against the intervention [11-14]. The primary differences between guidelines were identified in their date of publication (or last literature search) and judgments of the Evidence-to-Decision criteria.

In particular, the SSC guideline was published in January 2021 [11]. Consequently, the guideline excluded results from the RECOVERY trial, which were published after the guideline was updated. Differences in the interpretation of evidence and EtD criteria were also observed between the Australian and IDSA guidelines [12,13]. Both guidelines applied the GRADE approach. However, the Australian guideline downgraded the certainty of evidence to 'moderate certainty,' whereas the IDSA guideline downgraded the certainty of evidence to 'low certainty.' The NIH guideline did not use the GRADE framework, but it issued a strong recommendation formulated using a system that describes high-quality evidence when there is evidence from at least one randomized control trial [12].

2.6.2 Remdesivir

Eight organizations issued recommendations for the use of remdesivir to treat adult patients with severe COVID-19 who do not require invasive mechanical ventilation. Divergence was primarily observed in the direction of the recommendation. The WHO and the Public Health Agency of Canada (PHAC) issued conditional recommendations against the intervention [15,16]. The remaining organizations issued conditional recommendations for the intervention **(Table 3)** [11-14,17-18]**.** There were observed

differences in the interpretation of the evidence, judgments of EtD criteria, and the use of experts between guidelines.

Most guidelines in this cluster applied similar clinical evidence to formulate their recommendation. WHO downgraded its certainty of evidence to low due to the risk of bias in the studies and imprecision in the results [15]. Conversely, The Australian and NICE guidelines downgraded their certainty to moderate due to imprecision resulting from wide-confidence intervals in the results [12,16.] We observed differences in additional EtD criteria. According to WHO, the current evidence on the use of remdesivir does not indicate any important effects for critical outcomes, including mortality, need for mechanical ventilation, and time to clinical improvement [15]. In opposition, the Australian, IDSA, and NICE guidelines both expected small net benefits or net benefits in the intervention [13,14,17]. The NICE guideline concurred with the WHO guideline in that substantial variability is anticipated for patients' values and preferences [17]. However, this did not influence the direction of its recommendation. Lastly, the Australian guideline differentiates from WHO on its evaluation of cost-effectiveness. Specifically, the panellists did not presume any challenges associated with cost in the Australian setting [13]. Other guidelines, including the American College of Occupational and Environmental Medicine, also acknowledged costs associated with the intervention but still issued conditional recommendations for its implementation [18].

2.6.3 The Use of Face Masks in Children

Three organizations issued recommendations concerning the use of non-medical face masks in children. All guidelines favour the intervention of masking but diverge on

the minimum recommended age of children who should mask. The Centres for Disease Control and Prevention (CDC) and PHAC recommend masking for children above the age of two, whereas WHO recommends masking for children above the age of five implying that they at least suggest against mask use below the age of 5 [15-17]. We observed differences in the included evidence and incomplete evidence between guidelines.

The CDC applied evidence from a single cohort study evaluating oxygenation levels in children wearing masks. The results from the study indicate no significant differences in oxygenation level for children younger or older than 24 months [18]. Nonetheless, CDC, WHO, and PHAC all consider factors such as discomfort, tolerance, and a child's ability to wear properly wear a face mask to determine that young children may not benefit from masking. WHO used evidence from three studies to formulate their recommendation. Of these, two studies were Randomized Control Trials (RCT) evaluating the use of facemasks and hand hygiene to prevent influenza transmission, and one was an observational study evaluating the feasibility of masks in children at elementary school during influenza season. PHAC does not provide any direct or indirect evidence for their recommendation. Likewise, CDC and WHO do not provide explicit rationale regarding how the numerical age was determined.

3.4 Discussion

3.4.1 Summary of findings

We evaluated the frequency of diverging recommendations on the COVID19 Recommendations and Gateway to Contextualization RecMap by applying an iterative

search and extraction process. In total, we found 223 recommendations that diverged from at least one other recommendation for the same intervention or overarching health scope. We identified similar frequencies of divergence in formal judgement and PICO element. However, in our sample, therapeutic interventions were more likely to diverge in the formal judgment of the recommendations' strength and direction. Conversely, public health recommendations were more likely to diverge in the population or concerns associated with how the intervention should be implemented.

In our sample, we perceived a larger frequency of guidelines diverging in strength when compared to direction. In total, we identified 15 recommendation clusters diverging in direction. The impact on clinical practice and policy, and health and other outcomes may be more significant for this type of divergence. This is partly because divergence in direction implies different desirable or undesirable consequences. Therefore, this will likely result in different healthcare decisions [5,6]. Our study has identified divergence in the direction for 15 interventions. These included both therapeutic and public health recommendations which have been presented in detail in our Appendix.

3.4.2 Differences in the Methodology between Diverging Recommendations

To consider possible factors that may contribute to divergence, we conducted a qualitative analysis evaluating differences in the methodological processes and contextualization factors used to formulate diverging recommendations. We identified six clusters at random and six clusters with guaranteed recommendations from WHO which was sufficient to achieve information saturation. The findings from our qualitative

evaluation do not assume causation. They are meant to describe differences in the guideline development processes between guidelines with different judgments for a selected sample of clusters.

Within our sample of diverging recommendations, we found differences in the interpretation of evidence, as well as differences in the EtD criteria. When comparing EtD criteria, differences in judgments of the quality or certainty of the evidence, and differences in the interpretation of the balance of desirable and undesirable effects, were observed more frequently than other EtD criteria. Some guidelines used the certainty of evidence to make their final judgment on the strength of a recommendation. Among other factors, a strong recommendation was issued if the desirable effects were judged to significantly outweigh the undesirable effects, especially if the certainty of the evidence for the undesirable effects was low. We also observed differences in the ratings of the certainty of evidence for the same studies between diverging recommendations, even when using the same frameworks to assess quality. Other guidelines in our sample did not formally use EtD criteria to formulate their recommendations or inform other judgements. Nonetheless, they inadvertently compared equivalent criteria to contextualize recommendations for their setting or region. For example, the CDC's recommendation against quarantining fully vaccinated people returning from international travel takes into account the attitudes and preferences of the US population, as well as the desire to increase vaccine uptake in their country **(Table 3)**.

In addition to the predetermined themes, we identified additional differences within clusters. For guidelines with insufficient details regarding their guideline

development process, it is unknown whether pragmatic factors or undisclosed evidence were used to make the final judgment. Furthermore, multiple differences can exist within guidelines. When comparing the Australian guideline to other organizations with diverging recommendations for remdesivir, for example, we identified notable differences in the interpretation of evidence, judgments of various EtD criteria, and the use of experts on the panel for consultation. Lastly, we identified pragmatic differences such as variation in the date of publication or latest literature searches and differences in the included studies used to guide their judgment.

3.4.3 Limitations:

A single reviewer searched for comparable recommendations on the COVID-19 Recommendations Catalogue and filtered guidelines. Furthermore, guidelines that were not included for upload but had passed the initial screening stage were excluded given the large volume of guidelines to review. Although the latter might not be a direct limitation in the study design, it does imply some guidelines containing diverging recommendations were likely not captured in our study. Moreover, a single reviewer was used to evaluate the factors contributing to divergence. As a result, we did not conduct any tests to evaluate the inter-rater reliability of the findings from our content analysis. Finally, because our study is descriptive, we are unable to draw any causal interpretations and focus on associations.

3.5 Conclusion

To our knowledge, this is the first study evaluating divergence between COVID-19 recommendations to this scale. This study was exploratory and described the frequency of divergence between recommendations for the same intervention or health scope. This study also described differences in guideline development processes that may be used to evaluate causes of divergence in future studies. Future research should also compare the frequencies of guidelines with agreement to the frequencies of guidelines with disagreement to gain a better understanding of the true magnitude of divergence between COVID-19 recommendations.

Manuscript Tables and Figures

Table One: Summary of Diverging Clusters

Overall Summary of Diverging Clusters	**(N)**	**(%)**
Number of total diverging clusters	66	100
Number of clinical clusters	29	43.9
Number of non-clinical clusters (i.e., public health)	37	56.1
Summary of Type of Diverging Clusters		
Diverging in the formal judgment of strength **only**	19	28.8
Diverging in the formal judgment of direction **only**	9	13.6
Diverging in the formal judgment of strength **and** direction[a]	6	9.1
Diverging in population	6	9.1
Diverging in subgroup consideration of intervention	26	39.4
Summary of Diverging Clusters Across Intervention Groups		
Pharmacological Intervention	21	31.8
Other clinical care interventions	8	12.3
Infection Prevention and control measures	21	31.8
Vaccination related measures	8	12.3
School-Related Measures	8	12.3

[a] At least one recommendation in the cluster diverges in strength, and at least one recommendation in the same cluster diverges in direction.

Table Two: Summary of Diverging Recommendation Clusters across Overarching Healthscope

Health scope	Types of Diverging Clusters					
	Judgment		PICO element		Total	
	N	%	N	%	N	%
Divergence across overarching health scope						
Therapeutic Interventions	27	93.10[a]	2	6.90[a]	29	43.94
Public Health Interventions	7	18.92[b]	30	81.08[b]	37	56.06
Divergence across Intervention groups						
Pharmacological Intervention	19	90.10	2	9.50	21	31.81
Other clinical care interventions	8	100.00	0	0.00	8	12.31
Infection Prevention Control measures	5	23.81	16	76.19	21	31.81
Vaccination related measures	2	25.00	6	75.00	8	12.31
School-Related Measures	0	0.00	8	100.00	8	12.31
Total	**34**	**51.52**	**32**	**48.48**	**66**	**100.00**

[a] Percentage of diverging clusters of all therapeutic interventions
[b] Percentage of diverging clusters of all public health interventions

Table Three: Differences in the Methods Between Diverging Recommendations

Intervention	Organization	Methodological Differences
		Therapeutic interventions
Convalescent plasma	Australia IDSA NIH SSC	● **Theme 1: Differences in the date of publication** ○ The SSC guideline was updated before results from the RECOVERY trial were published whereas the Australian, NIH, and IDSA were updated using results from the trial ● **Theme 4: Differences in the EtD criteria** ○ The IDSA downgraded the certainty of evidence to low whereas the Australian guideline downgraded the certainty of evidence to moderate. The NIH guideline appraised the quality of evidence as high for at least one clinical trial
Gelatin	SSC WHO	● **Theme 4: Differences in EtD criteria** ○ Both WHO and SSC agree on the balance of desirable and undesirable effects and the cost-effectiveness of the intervention. However, interpretation of this criteria varied. SSC issued a conditional recommendation and WHO issued a strong recommendation against the intervention despite having similar judgements.
Intravenous IG	Australia NIH SSC	● **Theme 1: Differences in the date of publication** ○ The Australian guideline most recently updated its recommendation using evidence from one single RCT. The NIH and SSC guideline did not update their recommendation, despite updating their guidelines. ○ SSC guideline's initial guideline acknowledges no data on efficacy. Their updated guideline did not include any changes to the recommendation ● **Theme 4: Differences in the EtD criteria** ○ The Australian panel has concerns regarding harms ○ All guidelines had the same rating for certainty of evidence
Ivermectin	Australia IDSA NIH	● **Theme 3: Differences in the interpretation of results and assessments of quality** ○ Ivermectin may have small net benefits, however, the uncertainty of evidence leads the Australian panel to issue a strong recommendation against the evidence and the IDSA to make a conditional recommendation against the intervention. ○ The NIH does not issue a recommendation for or against the intervention as a result of insufficient/low quality evidence.

Remdesivir	Australia ACOEM IDSA NIH NICE PHAC SSC WHO	• **Theme 4: Differences in the EtD criteria** ○ Australia, NIH NICE, and IDSA appraise certainty of evidence as moderate whereas WHO appraise certainty of evidence as low ○ Additional differences in other components of EtD criteria, notably for the balance of desirable and undesirable effects and cost-effectiveness were observed between WHO, NICE, and the Australian guidelines ○ ACOEM also compares desirable and undesirable effects implicitly to formulate a conditional recommendation but also acknowledge costs of intervention ○ WHO and NICE both agree that substantial variability is expected for patient preferences whereas Australia does not expect substantial variability. ● **Theme 5: Additional factors - experts for consolidation** ○ The Australian guideline explicitly indicates using experts to help formulate their final judgment including virologists and clinicians. ○ PHAC issued their recommendation using evidence from WHO's guideline
Zinc	Australia NIH	● **Theme 3: Differences in the interpretation of evidence and assessments of quality** ○ Australian guideline asserts the risk of bias and imprecision from the limited evidence should constitute a strong recommendation against the intervention whereas the NIH guideline asserts no judgment for or against the intervention as a result of the limited evidence.
		Public Health Interventions
Delaying the second dose of COVID-19 vaccines	CDC ECDC PHAC	● **Theme 2: Differences in the included studies/evidence** ○ PHAC uses evidence from observational study to recommend delaying the second dose, but the study does not provide direct evidence of how long the second dose should be delayed ● **Theme 3: Differences in the interpretation of evidence and assessments of quality** ○ CDC does not recommend delaying the second dose because of the limited available evidence on the effectiveness of the intervention ● **Theme 5: Additional factors - contextualization factors** ○ PHAC considers delaying the second dose because of shortages of vaccine supplies in their setting ○ ECDC considers delaying the second dose to reduce the risk of variants in the community

Frequency of times to clean surfaces	CDC PHAC WHO	● **Theme 1: Differences in the date of publication** ○ WHO published their recommendation in June 2020, whereas CDC updated their evidence in April 2021. ● **Theme 5: Additional factors - incomplete evidence** ○ CDC uses evidence from QMRA studies whereas PHAC does not provide details about their recommendation development process and the evidence used to formulate it
Masks in children	CDC PHAC WHO	● **Theme 2: Differences in the included studies** ○ WHO applied evidence from three separate studies evaluating mask use in children during Influenza season whereas CDC applied evidence from a single observational study assessing oxygenation levels in children wearing masks ● **Theme 5: Additional factors - incomplete evidence** ○ CDC and WHO do not provide direct evidence for their recommendation in terms of how the age cut-off was determined ○ PHAC does not provide any evidence for its recommendation
Mode of birth for COVID-19+ mothers	Australia WHO	● **Theme 2: Differences in the included evidence/studies** ○ WHO and Australia use evidence from different systematic reviews ● **Theme 4: Differences in the EtD criteria** ○ WHO believes the known desirable effects significantly outweigh the unknown undesirable effects, therefore issued a strong recommendation
Rooming COVID-19+ mothers and newborns together	Australia WHO	● **Theme 4: Differences in the EtD criteria** ○ WHO believes desirable effects significantly outweigh undesirable effects, therefore, issued a strong recommendation ○ Australia issued a conditional recommendation as a result of the low quality of evidence available. The Australian guideline acknowledges benefits as well.
Quarantining for vaccinated travellers	CDC PHAC WHO	● **Theme 5: Additional factors - contextualization considerations** ○ CDC makes their recommendation considering studies on population attitudes and behaviours towards vaccination in addition to using data evaluating vaccine effectiveness. To increase uptake in the American population, they reduced restrictions for vaccinated individuals. ○ PHAC and WHO acknowledge limited evidence in vaccine effectiveness and vaccine protection against variants so makes a recommendation against the intervention

References:

[1] Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.ad

[2] Guyatt, G. H., Oxman, A. D., Kunz, R., Falck-Ytter, Y., Vist, G. E., Liberati, A., & Schünemann, H. J. (2008). GRADE: Going from evidence to recommendations. In *BMJ* (Vol. 336, Issue 7652, pp. 1049–1051). British Medical Journal Publishing Group. https://doi.org/10.1136/bmj.39493.646875.AE

[3] Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is "quality of evidence" and why is it important to clinicians?. In *BMJ*, 336(7651), 995-998.

[4] Tugwell, P., & Knottnerus, J. A. (2017). Adolopment – a new term added to the Clinical Epidemiology Lexicon. In *Journal of Clinical Epidemiology* (Vol. 81, pp. 1–2). Elsevier USA. https://doi.org/10.1016/j.jclinepi.2017.01.002

[5] Jadad, A. R., Cook, D. J., & Browman, G. P. (1997). A guide to interpreting discordant systematic reviews. *CMAJ*, 156(10), 1411-1416.

[6] Moja, L., Del Rio, M. P. F., Banzi, R., Cusi, C., D'Amico, R., Liberati, A., ... & Parmelli, E. (2012). Multiple systematic reviews: methods for assessing discordances of results. *Internal and emergency medicine*, 7(6), 563-568.

[7] Guyatt, G. H., Oxman, A. D., Kunz, R., Falck-Ytter, Y., Vist, G. E., Liberati, A., & Schünemann, H. J. (2008). GRADE: Going from evidence to recommendations. In *BMJ* (Vol. 336, Issue 7652, pp. 1049–1051). British Medical Journal Publishing Group. https://doi.org/10.1136/bmj.39493.646875.AE

[8] *GRADE handbook*. (2013). Retrieved April 25, 2021, from https://gdt.gradepro.org/app/handbook/handbook.html

[9] Schünemann, H. J., Wiercioch, W., Etxeandia, I., Falavigna, M., Santesso, N., Mustafa, R., Ventresca, M., Brignardello-Petersen, R., Laisaar, K. T., Kowalski, S., Baldeh, T., Zhang, Y., Raid, U., Neumann, I., Norris, S. L., Thornton, J., Harbour, R., Treweek, S., Guyatt, G., … Akl, E. A. (2014). *Guidelines 2.0: Systematic development of a comprehensive checklist for a successful guideline enterprise.* CMAJ, 186(3). https://doi.org/10.1503/cmaj.131237

[10] Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. J*ournal of advanced nursing,* 62(1), 107-115.

[11] Alhazzani, W., Evans, L., Alshamsi, F., Møller, M. H., Ostermann, M., Prescott, H. C., Arabi, Y. M., Loeb, M., Ng Gong, M., Fan, E., Oczkowski, S., Levy, M. M., Derde,

L., Dzierba, A., Du, B., MacHado, F., Wunsch, H., Crowther, M., Cecconi, M., … Rhodes, A. (2021). Surviving Sepsis Campaign Guidelines on the Management of Adults with Coronavirus Disease 2019 (COVID-19) in the ICU: First Update. Critical Care Medicine, E219–E234. https://doi.org/10.1097/CCM.0000000000004899

[12] The Australian Clinical Taskforce. *National COVID-19 Clinical Evidence Taskforce*. (2021). Retrieved April 24, 2021, from https://covid19evidence.net.au/

[13] Bhimraj, A., Morgan, R. L., Shumaker, A. H., Lavergne, V., Baden, L., Cheng, V. C., Edwards, K. M., Gandhi, R., Muller, W. J., O'Horo, J. C., Shoham, S., Murad, M. H., Mustafa, R. A., Sultan, S., & Falck-Ytter, Y. (2020, November 4). *IDSA Guidelines on the Treatment and Management of Patients with COVID-19*. 4/11/2020. https://www.idsociety.org/practice-guideline/covid-19-guideline-treatment-and-management/

[14] COVID-19 Treatment Guidelines Panel. *Coronavirus Disease 2019 (COVID-19) Treatment Guidelines.* National Institutes of Health. from https://www.covid19treatmentguidelines.nih.gov/

[15] World Health Organization (2021). *Therapeutics and COVID-19: living guideline*. Geneva: World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/340374/WHO-2019-nCoV-therapeutics-2021.1-eng.pdf?sequence=1&isAllowed=y.

[16] Public Health Agency of Canada. (2020) *Ad-hoc COVID-19 Clinical Pharmacology Task Group: Statement on remdesivir.* Retrieved April 18, 2021, from https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/covid-19-clinical-pharmacology-task-group/statement-remdesivir.html

[17] National Institute of Health and Care Excellence. (2021) *Making GRADE the Irresistible Choice - Guidelines and Evidence summaries*. Retrieved April 28, 2021, from https://www.nice.org.uk/guidance/ng191

[18] Hegmann, K. T., Cowl, C. T., Harber, P., Hyman, M. H., Pacheco, K. A., Winters, T., ... & Tallman, M. N. (2021). *Coronavirus (COVID-19)*. 28(12), 1290-9.

[19] Centers for Disease Control and Prevention. (2021). Guidance for Wearing Masks. Retrieved April 29, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html

[20] Public Health Agency of Canada. (2020) Non-medical masks: About. Retrieved March 20, 2020, from https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks/about-non-medical-masks-face-coverings.html

[21] World Health Organization (2020). Mask use in the context of COVID-19.Retrieved March 20, 2020, from https://apps.who.int/iris/handle/10665/337199

[22] Lubrano, R., Bloise, S., Testa, A., Marcellino, A., Dilillo, A., Mallardo, S., ... & Ventriglia, F. (2021). Assessment of respiratory function in infants and young children wearing face masks during the COVID-19 pandemic. *JAMA Network Open*, *4*(3), e210414-e210414.

Appendix A:

Summary of the Types of Divergence

A total of four types of divergence were identified from our search. When recommendations diverge in strength, this means guidelines agree in the direction of the recommendation, but not in the confidence of their judgment. When recommendations diverge in direction, this means guidelines disagree on whether the intervention should be implemented. Recommendations could also diverge in the population the intervention is intended for or subgroup considerations of the intervention. For example, guideline developers can use different metrics when considering a specific intervention. This section will describe each step of the deductive content analysis in detail and how they were executed for our study.

Preparation Phase: The preparation phase entails selecting units of analysis for review and can include selecting between analyzing specific words in the texts or themes [1]. We defined 'recommendations' as our unit of analysis. For each recommendation, we reviewed their PICO question, evidence synthesis, methods, EtD criteria, and information about the panel. We reviewed the contents of the guideline, information on the eCOVID-19 recommendations catalogue, and any supplementary materials file to extract our data. We are not claiming these factors cause divergence. However, we are assessing possible patterns that can be explored more rigorously in future studies.

Organization phase: Per deductive content analysis, our second step is developing the categorization matrix and grouping our data into predetermined codes or themes [1]. We assessed themes in order of simple differences such as date of 'last

literature search' and 'Included studies'. Following this, we assessed more complex differences such as 'interpretation of the evidence' and 'judgment of EtD criteria'. Our matrix is unstructured, meaning we are not limiting our analysis to the predetermined themes. Clusters can also have multiple differences because numerous recommendations can be grouped in one cluster.

Reporting: Reporting of content analysis needs to be explicit, transparent, and consist of all relevant information extracted from the text [1]. Consequently, we have provided the complete step-by-step in the supplementary materials file. Information was extracted verbatim from each guideline included in our assessment. Furthermore, the excel sheet is structured by themes and comprises all the variables we assessed for each theme.

Appendix B: Table of Intervention Clusters and Type of Divergence Between Recommendations within Clusters

Table 1B: Type of Divergence and Intervention for Recommendation Clusters

Cluster	Intervention	Included Organizations[a]	Type of Divergence
Pharmacological and/or Clinical Interventions			
1	Anakinra	Australian Clinical Taskforce ACOR Asociación Española de Pediatría NIH	Differences between PICO elements of recommendation **(Different populations)**
2	Antibiotics	CMI WHO	Difference between strength of recommendations
3	Antipyretics strategies	NIH SSC	Difference between strength of recommendations
4	Bamlanivimab	Australian Clinical Taskforce ACOEM IDSA NIH	Difference between direction of recommendations
5	Baricitinib	Australian NIH	Differences between PICO elements of recommendation **(Difference in intervention considerations)**
6	Colchicine	Australian Clinical Taskforce EULAR NIH	Difference between strength of recommendation
7	Convalescent plasma	Australian Clinical Taskforce IDSA NIH SSC WHO	Difference between strength and direction of recommendation
8	Crystalloids vs. colloids	NIH SSC	Difference between strength of recommendation
9	Dexamethasone	Australian Clinical Taskforce IDSA NICE	Difference between strength of recommendations

		NIH PHAC NIH	
10	ECMO	Australian Clinical Taskforce NIH	Difference between direction of recommendations
11	Gelatin	SSC WHO	Difference between strength of recommendations
12	Heparin	Australian Clinical Taskforce CHEST Guideline NIH Management of Coagulopathy	Difference between strength of recommendations
13	Hydroxychloroquine	Australian ACOEM Brazilian Association IDSA NIH SSC Spanish Consensus Guideline WHO	Difference between strength of recommendations
14	Immunoglobulin	Australian SSC NIH	Difference between strength and direction of recommendations
15	Ivermectin	Australian IDSA NIH	Difference between strength and direction of recommendations
16	Nitric Oxide	NIH SSC	Difference between strength of recommendations
17	Norepinephrine	NIH SSC	Difference between strength of recommendations
18	PEEP Strategy	Australian NIH SSC WHO	Difference between strength of recommendations
19	Prone positioning in general	Australian Clinical Taskforce NIH SSC WHO	Difference between strength and direction of recommendations
20	Prone positioning in	Australian Clinical Taskforce	Difference between strength of recommendations

	intubated patients (time)	Indian NIH WHO	
21	Pulmonary vasodilator	NIH SSC	Difference between strength of recommendations
22	Remdesivir without mechanical ventilation	Australian Clinical Taskforce ACOEM IDSA NICE NIH PHAC SSC WHO	Differences between the direction of recommendations
23	Remdesivir with mechanical ventilation	IDSA NIH SSC	Difference between strength and direction of recommendations
24	Sarilumab	Australian Clinical Taskforce NIH	Difference between strength of recommendations
25	Serologic testing	CDC IDSA NIH WHO	Differences between the strength of recommendations
26	Tocilizumab	Australian Clinical Taskforce ACOEM Brazilian Association PHAC IDSA NICE NIH WHO	Differences between the strength and direction of the recommendations
27	Vitamin C	Australian clinical task force NIH	Difference between direction of recommendations
28	Vitamin D	Australian clinical task force NIH	Difference between direction of recommendations
29	Zinc	Australian clinical task force NIH	Difference between strength of recommendation
	Public Health/Non-Clinical Recommendations		
30	Alcohol percent for disinfecting Surfaces	PHAC US EPA WHO	Difference between PICO elements **(Different percentages)**

31	Antibody testing	AACC Australian Clinical Taskforce IDSA Latin America	Difference between PICO elements **(Different times)**
32	AstraZeneca vaccine	ATAGI EMA PHAC WHO	Difference between PICO elements **(Different recommended ages)**
33	Breastfeeding	Australian Clinical Taskforce CDC PHAC Rapid advice guidelines for management of children with COVID-19 WHO	Difference between PICO element **(Recommended vs. risk-based approach)**
34	Broader contact tracing	Australian CDC PHAC	Difference between choice of comparison measures **(Broader contact tracing vs. specific contact tracing)**
35	Cleaning surfaces	CDC PHO US EPA	Difference between PICO elements **(Daily once cleaning vs. more than daily)**
36	Cohorting classrooms during outbreak	AAP CDC WHO	Difference between PICO elements **(Difference in intervention considerations)**
37	Cohorting suspected patients in hospitals	CDC ISCCM WHO	Difference between recommendation directions
38	Delaying second dosage	CDC ECDC PHAC	Difference between recommendation directions
39	Discarding face mask after extended use	CDC WHO	Difference between PICO elements **(Difference in intervention considerations)**
40	Distancing between students and teachers in schools	AAP CDC England Education Ministry of health PHAC TAG (WHO and UNICEF)	Difference between PICO elements **(Difference in intervention considerations)**
41	Hand hygiene	Australian CDC	Difference between PICO elements **(% alcohol in disinfectants)**

		COVID-19 IPC Guidelines for South Africa PHAC	
42	Isolating COVID-19+ children from schools	CDC England Education French Pediatric Society	Difference between PICO elements **(Differences in intervention considerations)**
43	Masking for vaccinated people	CDC ECDC WHO	Difference between PICO elements **(Differences in intervention considerations)**
44	Masks in children	CDC NHS WHO	Difference between PICO elements **(Differences in population)**
45	Masks in children at school	AAP CDC Department of Education and skills England Education Ministry of health PHAC WHO	Difference between PICO elements **(Differences in population)**
46	Mode of Birth for COVID-19 mothers	Australian Clinical taskforce WHO	Difference between strength of recommendations
47	N95 masks for treating non-ventilated patients	CDC NIH SSC WHO	Difference between direction of recommendations
48	Patient Discharge	CDC ECDC WHO	Difference between PICO elements **(Differences in intervention consideration)**
49	Physical distancing for long-term care staff	PHAC WHO	Difference between PICO elements **(Differences in intervention consideration)**
50	Physical distancing in hospitals	CDC COVID-19 IPC Guidelines for South Africa PHAC	Difference between PICO elements **(Differences in intervention consideration)**
51	Physical distancing in schools	AAP CDC Department of Education & skills ECDC Ministry of health	Difference between PICO elements **(Differences in intervention consideration)**

		PHAC TAG	
52	PPE in adult staff in schools	CDC Department of Education & skills Ministry of health WHO	Difference between PICO elements **(Differences in intervention consideration)**
53	PPE in long term care staff	PHAC WHO	Difference between PICO elements **(Differences in intervention consideration)**
54	Quarantining close contacts	CDC PHAC WHO	Difference between PICO elements **(Differences in intervention consideration)**
55	Quarantining period after travel	CDC PHAC WHO	Difference between PICO elements **(Differences in intervention consideration)**
56	Quarantining for vaccinated travels	CDC PHAC	Difference between direction of recommendations
57	Reduced quarantining time for members of households	CDC PHAC	Difference between PICO elements **(Differences in intervention consideration)**
58	Reuse of masks	CDC ECDC WHO	Difference between PICO elements **(Differences in intervention consideration)**
59	Rooming COVID-19 positive mothers with newborns	Australian Clinical Taskforce WHO	Difference between strength of recommendations
60	School activity closures	CDC Ministry of health PHAC WHO	Difference between PICO elements **(Differences in intervention consideration)**
61	School closures	Australian - Department for Education CDC French Pediatric Society	Difference between PICO elements **(Differences in intervention consideration)**
62	Self-monitoring symptoms	CDC WHO	Difference between PICO elements **(Differences in intervention consideration)**

63	Social distancing for vaccinated populations	CDC ECDC WHO	Difference between PICO elements **(Differences in intervention consideration)**
64	Vaccinating Immunocompromised groups	Australian Vaccine Guidelines CDC PHAC WHO	Difference between PICO elements **(Risk-based approach vs. regular)**
65	Vaccinating lactating women	ACOG Australia CDC PHAC/NACI WHO	Difference between PICO elements **(Risk-based approach vs. regular)**
66	Vaccinating pregnant women	ACOG Australia CDC PHAC/NACI WHO	Difference between PICO elements **(Risk-based approach vs. regular)**

[a] Organizations with diverging recommendations for the intervention being clustered

Appendix C:

Organization abbreviations:

ACOEM	American College of Occupational and Environmental Medicine
Australia	Australian Clinical Taskforce
CDC	Centres for Disease Control and Prevention
ECDC	European Centres for Disease Control and Prevention
IDSA	Infectious Disease Society of America
NACI	National Advisory Committee on Immunization
NIH	National Institute of Health
PHAC	Public Health Agency of Canada
SSC	Surviving Sepsis Campaign
WHO	World Health Organization

This section provides a detailed qualitative analysis for all twelve recommendation clusters that were selected for review.

Therapeutic Interventions

The following therapeutic interventions were selected for comparative analysis of factors influencing divergence:

1. Convalescent plasma
2. Gelatin
3. Intravenous immunoglobulins
4. Ivermectin
5. Remdesivir
6. Zinc

Convalescent plasma:

Four organizations issued recommendations concerning the use of convalescent plasma to treat populations with COVID-19. The Australian and NIH guidelines published strong recommendations against the intervention, whereas the IDSA and SSC guidelines published conditional (or weak) recommendations against the intervention [2-5]. Differences in the date of publication (Theme 1) and the EtD criteria (Theme 4) were observed.

The most recent update of the SSC guideline was published in January 2021. Consequently, guideline developers excluded evidence from the RECOVERY trial, which was published after the update. Therefore, they primarily consider evidence from four other RCTs on the use of convalescent plasma. Because of the indirectness of these study results, the SSC downgraded their overall certainty of evidence to low [2].

The Australian guideline assessed a total of 15 outcomes for convalescent plasma. Overall, the Australian guidelines downgraded their certainty of evidence to moderate due to serious imprecision for critical outcomes like mortality and non-invasive ventilation. However, they rated the certainty as high for other outcomes such as progress to mechanical ventilation [2]. Deviating from the Australian guideline, the IDSA appraised the certainty of evidence as low for mortality due to imprecision. The IDSA guideline also appraised the certainty of evidence as low for its remaining critical outcomes [3]. Both guidelines applied findings from the RECOVERY trial and five additional studies. The IDSA excluded three studies that were included in the Australian guideline. However, this was not owing to the date of publication. Similarly, the Australian guideline excluded four studies used by the IDSA that were possible to include.

The NIH guideline does not apply the GRADE framework, contrasting the other guidelines. Instead, using their appraisal system, they defined the recommendation as strong because panellists used evidence from more than one clinical trial without significant limitations [5]. NIH predominantly used the findings from the RECOVERY trial to formulate their judgment. They also assessed the same studies and outcomes as the Australian and IDSA guidelines. However, the NIH used one additional study not considered in the IDSA or Australian guideline.

Gelatin:

Two organizations issued recommendations for the use of gelatin to resuscitate patients experiencing shock. The WHO guideline for clinical management issued a strong recommendation against gelatin, whereas the SSC issued a conditional (weak) recommendation against the intervention [2,6]. These guidelines did not complete formal EtD frameworks for this recommendation. Nevertheless, they purposefully assess specific criteria within the framework (Theme 4). Specifically, when balancing the desirable and undesirable effects of gelatin, WHO and SSC concur that the benefits of the comparator group, crystalloids, outweigh the benefits of the intervention. Moreover, both guidelines agree that gelatin is expensive, thus less cost-effective than its alternative option.

Intravenous Immunoglobulins:

Three guidelines issued recommendations regarding the use of immunoglobulins to treat patients with COVID-19. The SSC issued a conditional (weak) recommendation against the intervention while the Australian guideline issued a strong recommendation against the intervention [2,3]. The NIH did not make a recommendation for or against the intervention as a result of limited evidence [5]. We identified differences in the date of most recent literature search (Theme 1) and EtD criteria (Theme 4).

The NIH guideline was updated seven days before the Australian guideline on April 21, 2021. Interestingly, the NIH did not make any changes to their recommendation since the judgment for this intervention was last deliberated on July 17, 2020. The Australian guideline used results from a single RCT for their deliberation [3]. This RCT

was published before the most recent update of the NIH guideline. Therefore, we are uncertain why NIH excluded the study.

The SSC guideline did not specify many details for this recommendation in their update. However, in the initial guideline, they stated using evidence from efficacy data [2]. However, because of the very low certainty of evidence from a singular study, they issued a conditional recommendation against the intervention. They also excluded results from the RCT used by the Australian guidelines. These results were published before the guideline update.

Ivermectin:

Three organizations issued recommendations for the use of ivermectin to treat patients with COVID-19. The Australian guideline issued a strong recommendation against the intervention, and the IDSA guidelines issued a conditional recommendation against the intervention [3,4]. The NIH guideline did not issue a recommendation for or against the intervention as a result of limited evidence [5]. We identified differences in the interpretation of the results between these guidelines (Theme 3).

The Australian and IDSA guidelines used evidence from existing RCTs to formulate their recommendation. Both guidelines downgraded the certainty of evidence to very low as a result of very serious imprecision and risk of bias from inadequate randomization in the trials [3,4]. The included studies, however, do find a reduction in mortality and time to recovery for patients using remdesivir [3,4].

The Australian guideline similarly judged small net benefits or little differences between the alternatives [3]. However, as a result of the very low certainty of evidence and the common side effects associated with ivermectin, they do not recommend its use [3]. The NIH guideline also acknowledges the limitations of the current evidence on the use of ivermectin. However, due to the limited evidence, the panel did not make a judgment for or against the intervention [5].

Remdesivir:

Eight organizations issued recommendations concerning the use of remdesivir for treating adults with severe COVID-19 who are not under invasive mechanical ventilation. These guidelines include the Australian, PHAC, ACOEM, IDSA, NICE, NIH, SSC, and WHO [2-9]. Recommendations primarily diverged in the direction of the intervention with all guidelines except PHAC and WHO issuing conditional recommendations for the intervention. PHAC and WHO both issue conditional recommendations against the intervention [6,7]. We observed differences in the EtD framework for numerous criteria including, the balance of desirable and undesirable effects, certainty of the evidence, cost-effectiveness, and patient values and preferences (Theme 4). We also identified differences in the interpretation of the results (Theme 3) and additional factors (Theme 5).

The Australian and NICE guidelines formulate their recommendation using evidence from the SOLIDARITY trial [3,8]. Additionally, results from the ACTT-1 trial data, SIMPLE MODERATE, and Wang trial data were also used by most guidelines in the guideline [2-9].

The Australian guideline conducted an assessment comparing their methods with WHO's methods. According to their assessment, guideline developers state using the ICEMAN tool to evaluate the credibility of disease severity and input from experts such as clinicians, virologists, and immunologists to assist with their judgment [3]. Furthermore, the Australian guideline notes differences in statistical methods to compile and appraise the data [3]. WHO conducted a meta-analysis using odds ratio whereas the Australian guideline conducted pairwise comparisons using random-effect models to calculate risk ratio [3]. Lastly, differences between judgments of EtD criteria were identified for the benefits and harms, the certainty of the evidence, patient preferences, and values, resource implications, feasibility, equity, and others [3].

When comparing differences in the benefits of the intervention, WHO finds that evidence does not indicate any important effect on mortality, need for mechanical ventilation, time to clinical improvement, or other patient-important outcomes [6]. Conversely, Australian guideline panellists suppose that evidence from RCT indicates remdesivir has an acceptable safety profile compared to standard care [3]. The Australian guideline also states remdesivir may reduce the incidence of serious adverse events. WHO does not make the same conclusions as a result of inconclusive evidence. Overall, the Australian guideline judged certainty of evidence as moderate for critical outcomes, including death at 28 days, discharge from hospital, serious adverse events, time to recovery, and time to improve. WHO downgraded their certainty of evidence to low or very low across all outcomes because of factors like risk of bias in the studies and imprecision in the results [6]. The Australian guideline only found imprecision in the

results because of wide confidence intervals. The SSC guideline also downgraded the certainty of evidence to moderate due to serious imprecision for the outcomes of mortality and serious adverse events [2]. For time to clinical recovery, the SSC downgraded the quality of evidence to low due to the risk of bias [2].

We also observed differences between EtD criteria to contextualize recommendations for different settings. First, when comparing patient preferences and values, WHO judges most patients would not want to use remdesivir because of the uncertainty of evidence [6]. Conversely, the Australian guideline's consumer panel do not expect substantial variability in preference. Instead, they assume patients and clinicians would choose the intervention because the current evidence does not exclude the possibility of benefits [3]. The SSC guideline panel makes the same judgment as the Australian guideline. They state patients will place more importance on outcomes such as reduced mortality, adverse events and time to recovery. Lastly, when assessing the cost-effectiveness of the intervention, The Australian guideline state that opportunity costs are not concerning in the Australian setting when trying novel interventions, whereas WHO acknowledge opportunity costs may lead to inequities when conclusive evidence is not available [3,6].

The NICE guideline also applied results from RCT to evaluate the desirable and undesirable effects of the intervention. Like the Australian guideline, they conclude remdesivir has an acceptable safety profile and may reduce the incidence of serious adverse events [8]. The NICE guideline also reviewed evidence from an additional observational study but did not apply findings from the study to formulate their

recommendation [8]. The NICE guideline also appraised the certainty of evidence as moderate and downgraded due to serious imprecision only [8]. Additionally, NICE concludes that the direction of effect consistently favoured remdesivir [8]. Contrasting the Australian guideline, NICE agreed with WHO that substantial variability is expected for patients' values and preferences [6,8]. However, this was not enough to assert a recommendation against the intervention.

The IDSA guideline did not complete a formal EtD framework for their recommendation [4]. But, they did consider the benefits and harms of the intervention for a range of outcomes. These outcomes included mortality, progression to mechanical ventilation, and serious adverse events. Panellists used evidence from the ACTT-1 and SOLIDARITY trial to assess these outcomes [4]. The pooled analysis, however, fail to show a reduction in mortality. The IDSA primarily bases their judgment on the possible benefits of reducing time to improvement and shorter median recovery times. The certainty of the evidence was appraised as moderate, like Australia and NICE guidelines for these outcomes [2,4,9]. Lastly, the IDSA panel assessed harms associated with the intervention and found that patients receiving remdesivir do not experience greater serious adverse events [4]. The Australian and NICE guidelines also establish this.

The IDSA panel diverges from the remaining guidelines in the population the intervention is recommended for. The Australian and NICE guidelines do not recommend remdesivir for patients with severe COVID-19 on mechanical ventilation whereas the IDSA panel recommends it for patients on any supplemental oxygen, mechanical ventilation, and ECMO. The IDSA panel does not provide specific details regarding why

this population was selected [4]. They do state there is not sufficient evidence to recommend the use of remdesivir for populations, not on supplemental oxygen [4]. This was agreed by all guidelines.

The NIH makes the same judgment as NICE and Australia. Their rationale for a moderate (i.e., conditional) recommendation for the use of remdesivir in patients on supplementary oxygen is similar to IDSA where they found the intervention was associated with improved time to recovery [5]. Additionally, they found fewer patients using the intervention progressed to invasive mechanical ventilation or ECMO [5].

Lastly, the American College of Occupational and Environmental Medicine issued a conditional recommendation for the intervention. They excluded results from the SOLIDARITY trial because trial results were published after the guideline was issued. However, they did use evidence from other RCTs, including the ACTT-1, Goldman, and Wang trials [9]. Similar to other guidelines, their recommendation is founded on evidence of shortened ICU time (time to recovery) and possible improved survival [9]. Finally, although the American College did not complete a formal EtD table, they did appraise minimal adverse effects associated with the intervention. Lastly, they acknowledge the high costs of remdesivir, but this did not deter the panel from recommending the intervention [9].

Zinc:

Two organizations issued recommendations concerning the use of zinc to treat patients with COVID-19. The Australian guideline issued a strong recommendation against the intervention, and the NIH guideline did not issue a recommendation for or against the use of zinc [3,5]. However, the guideline did publish a moderate recommendation against prescribing larger doses [5]. We observed differences in the interpretation of results (Theme 3) between guidelines. Interestingly, both guidelines use data from the same RCTs. However, the NIH excluded one RCT used by the Australian guideline. Both guidelines agree there are serious limitations in the current studies assessing the use of zinc as a therapeutic intervention [3,5]. Correspondingly, the Australian guideline reduced the overall certainty of evidence due to study bias and imprecision. Although NIH does not use the same appraisal process, the panel also recognizes limitations in the small sample size [5]. Additionally, the panel recognizes the risk of study bias in one RCT without a placebo group.

Public Health Interventions

We also evaluated differences in the guideline development process between comparable public health interventions. The following recommendations were assessed:

1. Delaying the second dose of COVID-19 vaccines
2. Frequency of times to clean surfaces
3. The use of masks in children
4. Mode of birth for COVID-19 positive mothers
5. Rooming COVID-19 positive mothers and newborns together
6. Quarantining for vaccinated travellers

Delaying the second dose of COVID-19 vaccines:

Three organizations issued recommendations for delaying the second dose of COVID-19 vaccines that require two complete doses for maximum efficacy. These include the CDC, ECDC, and PHAC (NACI) [10-12]. ECDC and PHAC recommend actively delaying the second dose in response to vaccine shortages, whereas CDC does not recommend delaying the second dose unless it is necessary. The ECDC and CDC do not use formal judgments to make their recommendations, whereas PHAC made a strong recommendation for the intervention. Under circumstances where vaccines need to be delayed as a result of shortages, the CDC recommends delaying vaccines for a maximum of six weeks. PHAC recommends delaying vaccine dosages for up to four months. CDC calculated this time using the available evidence of vaccine efficacy. PHAC also used evidence of vaccine efficacy but recommended delaying the second dose up to four months.

We observed differences in contextualization factors (Theme 5) between guidelines. In their guideline, PHAC reports delaying the second dose, so more members

of the Canadian population are vaccinated in times of vaccine shortages [10]. Because of the high efficacy reported in clinical trials, PHAC concludes that delaying the second dose will not result in significant harms [10]. They also apply evidence from one observational study evaluating the duration of vaccine effectiveness after one dose. However, the data is not direct because it does not assess effectiveness in the population four months after vaccination [10]. The CDC also acknowledges the limitation on evidence for delaying vaccine doses. This is the primary reason they do not recommend delaying doses beyond 6-weeks [11]. ECDC also recommends delaying the second dose in areas facing supply shortages. However, their concerns rely primarily on mitigating the threat of variants [12]. Although this was not a primary cause of formulating the recommendation for PHAC, panellists do speculate that delaying the second dose might reduce variants as a consequence of less community transmission [10]. ECDC does not state how long second doses can be delayed [12].

Frequency of cleaning surfaces to reduce transmission of COVID-19:

Three organizations issued recommendations for cleaning surfaces as an infection prevention and control measure to reduce COVID-19 transmission. These organizations include the CDC, PHAC, and WHO [13-15]. All guidelines recommend routinely cleaning surfaces but diverge on how frequently surfaces should be cleaned. We identified differences in the date of publication (Theme 1) and the use of incomplete evidence (Theme 5) between guidelines.

The CDC recommends cleaning all surfaces once daily [13]. PHAC recommends cleaning surfaces routinely and cleaning more frequently touched surfaces more than once daily and whenever visibly dirty [14]. WHO issues the same recommendation as PHAC [15]. The CDC is the only organization that provides research evidence for its recommendation. Specifically, guideline developers assessed the available evidence from quantitative microbial risk assessment (QMRA) studies and determined a low risk of transmission from physical contact with surfaces. PHAC does not provide evidence for their recommendation, so it is unknown whether they applied a pragmatic approach or used formal evidence to determine the frequency surfaces should be cleaned [14]. Finally, WHO formulated its recommendation in June 2020, whereas CDC updated its guideline in April 2021.

Masks in children:

Three organizations issued recommendations concerning the use of masks in children. All guidelines favour the intervention of masking in children but diverge on the minimum recommended age of children who should mask. CDC and PHAC recommend masking for children above the age of two, whereas WHO recommends masking for children above the age of five [16-18]. We identified differences in the included evidence (Theme 2), interpretation of the evidence (Theme 4), and the use of unclear evidence (Theme 5) between guidelines.

CDC conducted a scientific brief synthesizing existing research regarding the use of masks for preventing infectious diseases. In their brief, they used a single cohort study

evaluating oxygenation levels in children wearing masks [16]. The results from the study indicate no significant differences in oxygenation level between children younger than 24 months compared to children older than 24 months [16]. Therefore, it is unclear how the age for children who are required to mask was determined. Nonetheless, the CDC does consider other criteria, including discomfort and proper mask.

Like the CDC, WHO primarily considered factors such as discomfort and proper use of masks to determine its age cut off for their recommendation [16,17]. WHO indicates that children below the age of five should not wear masks, and children between the ages of five and eleven should wear masks after completing a risk assessment. The risk assessment includes evaluating the level of transmission in the community, the child's capacity to comply with appropriate mask use, and the availability of appropriate supervision for children masking [17]. WHO applied evidence from three studies to formulate their recommendation. Two of these studies were RCTs evaluating the use of facemasks and hand hygiene to prevent influenza transmission. One was a cross-sectional study evaluating the feasibility of wearing masks in elementary school children during influenza season [17].PHAC also does not recommend masking for children below the age of two. For children between the ages of two and five, PHAC recommends masking if they are supervised, able to tolerate it, and can adequately wear and take off their mask [18]. PHAC does not provide any direct or indirect evidence for their recommendation.

Mode of birth for Pregnant People with COVID-19:

We identified two guidelines with recommendations concerning the mode of birth for pregnant people who have COVID-19. The Australian and WHO clinical guidelines agree that the mode of birth should not change because of COVID-19 status [3,6]. However, WHO issued a strong recommendation, whereas the Australian guideline issued a conditional recommendation. We observed differences in the EtD criteria (Theme 4) and the use of unclear evidence (Theme 5) between guidelines.

WHO recommends cesarian birth to be performed only when medically justified. The panel concludes that the risk of vertical transmission through blood is minimal [6]. Similarly, WHO does not find any evidence of vertical transmission from delaying cord clamping [6]. Moreover, WHO assessed the advantages of vaginal birth, including the known benefits of delaying clamping and compares it to the unproven or unknown risk of vertical transmission [6]. The Australian guideline applied evidence from one systematic review across 49 studies comparing vertical transmission and newborn infection rates between vaginal birth and caesarian birth. In their EtD judgment, their consumer panel also states that most individuals in the population would choose the intervention [3]. However, the certainty of evidence was downgraded to very low [3].

Rooming mothers with COVID-19 and their newborns:

We identified two guidelines with recommendations concerning rooming mothers with COVID-19 with their newborns after delivery. The WHO and Australian clinical guidelines favour the intervention of rooming but diverge in the strength of the

recommendations [3,6]. WHO issues a strong recommendation, whereas the Australian guideline issues a conditional recommendation for the intervention. Differences were observed in the included studies (Theme 2) and EtD criteria (Theme 4) between guidelines.

Both guidelines compare the desirable and undesirable effects of the intervention. However, WHO does not complete a formal EtD table. The Australian guideline also downgraded the certainty of evidence to low, whereas WHO does not conduct a formal assessment evaluating certainty [3,6]. WHO also deliberates the benefits of rooming. Specifically, WHO states rooming provides new mothers with an opportunity to continue breastfeeding and initiate and practice skin-to-skin contact [6]. Furthermore, WHO concludes the risk of infection is low from mother to baby but there are significant known consequences of separating the newborn from their mother [6].

The Australian guideline also considers the desirable and undesirable effects associated with rooming mothers with COVID-19 with their newborns. Like WHO, the Australian guideline judges substantial net benefits of the recommendation [3]. These include allowing the mother to bond with their newborn, increasing the probability of exclusive breastfeeding at discharge, and increasing the duration of breastfeeding [3]. The Australian consumer panel also judged there would be no substantial variability in patients' preferences and values. The panel states they anticipate most women would want to room with their newborns because there is no evidence to suggest there are risks associated with the practice [3]. Nonetheless, the Australian guideline acknowledges the limited evidence available on separating mothers from their newborns to prevent COVID-

19 transmission. Specifically, the Australian guideline downgrades the certainty of evidence to 'very low' due to reliance on case reports and case series comparing infection rates between mothers rooming-in and not rooming-in [3]. It is likely the recommendation is conditional because of the low quality of available evidence.

Quarantining practices for vaccinated travels:

Three organizations published recommendations concerning quarantining practices for vaccinated populations who are returning from international travel. These organizations consisted of the CDC, PHAC, and WHO [19-21]. PHAC and WHO concur that vaccinated populations should continue to quarantine for 14 days after returning from international travel. Conversely, CDC issued interim recommendations stating vaccinated populations can officially refrain from self-quarantining. We observed differences in the included evidence (Theme 2) and other contextualization factors (Theme 5) between guidelines.

WHO published an interim position paper concerning their recommendation. In their paper, WHO discusses the scientific unknowns regarding vaccination. Unknown outcomes included vaccine effectiveness in the population, duration of protection, and efficacy in limiting the transmission of variants [19]. PHAC uses WHO's evidence and considers our limited understanding of viral transmission and duration of immunity to deliberate their judgment [20].

CDC also recognizes there is limited evidence on how long vaccine protection lasts and on the extent of vaccine protection against SARS-CoV-2 variants [21]. However, CDC also included contextual factors such as the American population’s

attitudes and preferences for getting vaccinated in their judgement [21]. In their scientific brief, CDC shares result from a survey completed by members of the US population [21]. The survey found that two-thirds of Americans were at least somewhat likely to receive a vaccine, which was insufficient for CDC [21]. Furthermore, when asked why they would get vaccinated, surveyors shared a return to normal life as an element influencing their decision. Therefore, CDC partially issues the recommendation to incentivize the US population to get vaccinated and consider emerging evidence on vaccine effectiveness [21].

Appendix References

[1] Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. J*ournal of advanced nursing,* 62(1), 107-115.

[2] Alhazzani, W., Evans, L., Alshamsi, F., Møller, M. H., Ostermann, M., Prescott, H. C., Arabi, Y. M., Loeb, M., Ng Gong, M., Fan, E., Oczkowski, S., Levy, M. M., Derde, L., Dzierba, A., Du, B., MacHado, F., Wunsch, H., Crowther, M., Cecconi, M., … Rhodes, A. (2021). Surviving Sepsis Campaign Guidelines on the Management of Adults with Coronavirus Disease 2019 (COVID-19) in the ICU: First Update. Critical Care Medicine, E219–E234. https://doi.org/10.1097/CCM.0000000000004899

[3] The Australian Clinical Taskforce. *National COVID-19 Clinical Evidence Taskforce.* (2021). Retrieved April 24, 2021, from https://covid19evidence.net.au/

[4] Bhimraj, A., Morgan, R. L., Shumaker, A. H., Lavergne, V., Baden, L., Cheng, V. C., Edwards, K. M., Gandhi, R., Muller, W. J., O'Horo, J. C., Shoham, S., Murad, M. H., Mustafa, R. A., Sultan, S., & Falck-Ytter, Y. (2020, November 4). *IDSA Guidelines on the Treatment and Management of Patients with COVID-19*. 4/11/2020. https://www.idsociety.org/practice-guideline/covid-19-guideline-treatment-and-management/

[5] COVID-19 Treatment Guidelines Panel. *Coronavirus Disease 2019 (COVID-19) Treatment Guidelines.* National Institutes of Health. from https://www.covid19treatmentguidelines.nih.gov/

[6] World Health Organization (2021). *Therapeutics and COVID-19: living guideline*. Geneva: World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/340374/WHO-2019-nCoV-therapeutics-2021.1-eng.pdf?sequence=1&isAllowed=y.

[7] Public Health Agency of Canada. (2020) *Ad-hoc COVID-19 Clinical Pharmacology Task Group: Statement on remdesivir.* Retrieved April 18, 2021, from https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/covid-19-clinical-pharmacology-task-group/statement-remdesivir.html

[8] National Institute of Health and Care Excellence. (2021) *Making GRADE the Irresistible Choice - Guidelines and Evidence summaries*. Retrieved April 28, 2021, from https://www.nice.org.uk/guidance/ng191

[9] Hegmann, K. T., Cowl, C. T., Harber, P., Hyman, M. H., Pacheco, K. A., Winters, T., ... & Tallman, M. N. (2021). *Coronavirus (COVID-19)*. 28(12), 1290-9.

[10] Kriss JL, Reynolds LE, Wang A, et al. COVID-19 Vaccine Second-Dose Completion and Interval Between First and Second Doses Among Vaccinated Persons —

United States, December 14, 2020−February 14, 2021. MMWR Morb Mortal Wkly Rep 2021;70:389–395. DOI: http://dx.doi.org/10.15585/mmwr.mm7011e2

[11] Government of Canada. Extended dose intervals for COVID-19 vaccines to optimize early vaccine rollout and population protection in Canada in the context of limited vaccine supply. Retrieved April 20, 2021, from https://www.canada.ca/en/public-health/services/immunization/national-advisory-committee-on-immunization-naci/extended-dose-intervals-covid-19-vaccines-early-rollout-population-protection.html

[12] European Centre for Disease Prevention and Control. (15 Feb 2021) Risk assessment: SARS-CoV-2 - increased circulation of variants of concern and vaccine rollout in the EU/EEA, 14th update. Accessed April 20, 2021, from https://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-variants-vaccine-fourteenth-update-fehttps://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-variants-vaccine-fourteenth-update-february-2021bruary-2021

[13] Centre for Disease Prevention and Control. (5 April 2021). *Cleaning Your Facility.* Accessed April 20, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/community/disinfecting-building-facility.html

[14] Government of Canada. (23 September 2020). *COVID-19: Cleaning and disinfecting.* Accessed April 20, 2021, from https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks/cleaning-disinfecting.html

[15] World Health Organization (15 May 2020). *Cleaning and Disinfection of environmental surfaces in the context of COVID-19.* Accessed April 20, 2021, from https://www.who.int/publications/i/item/cleaning-and-disinfection-of-environmental-surfaces-inthe-context-of-covid-19

[16] Centers for Disease Control and Prevention. (2021). Guidance for Wearing Masks. Retrieved April 29, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html

[17] World Health Organization (2020). Mask use in the context of COVID-19.Retrieved March 20, 2020, from https://apps.who.int/iris/handle/10665/337199

[18] Public Health Agency of Canada. (2020) Non-medical masks: About. Retrieved March 20, 2020, from https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks/about-non-medical-masks-face-coverings.html

[19] World Health Organization (5 February 2021). Interim position paper: considerations regarding proof of COVID-19 vaccination for international travellers. Accessed April 28 2021 from https://www.who.int/news-room/articles-detail/interim-position-paper-considerations-regarding-proof-of-covid-19-vaccination-for-intern

[20] Government of Canada. (2021). COVID-19 vaccinated travellers entering Canada. Accessed April 28, 2021 from https://travel.gc.ca/travel-covid/travel-restrictions/covid-vaccinated-travellers-entering-canada

[21] Centre for Disease Control and Prevention (2021) Guidance for Fully Vaccinated People. Accessed April 28, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/vaccines/fully-vaccinated-guidance.html

CHAPTER THREE: IMPLICATIONS AND CONCLUSION

In this section, we provide possible considerations for users and developers of the COVID-19 recommendations catalogue to review when selecting the most appropriate recommendation when multiple possibilities exist. The purpose of this section is to discuss the implications of our research for the development of the recommendations catalogue and public health in general. However, these are only considerations that may assist with more efficient decision-making. The final decision should be at the discretion of the developers and users.

3.1 Considerations for developers of the COVID19 Recommendations and Gateway to Contextualization Catalogue

Before comparing recommendations with diverging judgments, it is beneficial to appraise whether recommendations truly answer the same PICO question. Suppose recommendations are comparable but have differences in sub-group considerations that are resulting in different formal judgments. In that case, developers may not need to conduct a further evaluation as the recommendations are inherently distinct from each other. Otherwise, we can evaluate two types of overarching differences in the guideline development process that may have varying implications for divergence. First, diverging recommendations can apply different methods, such as incorporating different studies or having distinctive interpretations of the same evidence. Second, diverging

recommendations can consider different contextualization factors such as specific judgments of their EtD criteria and setting.

Guideline developers may need to apply different approaches to address these variances in the catalogue. For example, developers may consider archiving older recommendations that were formulated using pragmatic models or incomplete evidence if comparable recommendations for the same intervention formulated using higher quality evidence and a more rigorous process exists. Developers may also consider conducting a credibility assessment evaluating different dynamics between comparable recommendations. If diverging recommendations have different judgments in EtD criteria or other contextualization factors, catalogue developers may keep their current system of reporting each criterion for users to appraise themselves. This will provide users with a comprehensive overview of the variation that can exist in the judgment of comparable recommendations, allowing them to review and select the recommendation most appropriate for their setting or concern.

3.2 Considerations for Users of the COVID19 Recommendations and Gateway to Contextualization Catalogue

Users should also compare recommendations with formal judgments to ensure they are genuinely duplicate. If there are differences in subgroup considerations that have led to differences in the formal judgments of the recommendations, then users should select the most appropriate recommendation for their specific question or concern. This can also be considered if diverging recommendations assess different outcomes. In these

circumstances, users should consider reviewing the recommendation formulated by evaluating their specific outcome of interest.

If diverging recommendations were published during different periods, users might want to appraise the relevance of each recommendation at the time of assessment. This includes reviewing if older recommendations exclude any new evidence that may have led to a different judgment. Users may also want to evaluate factors that can contribute to disparities in the included evidence by evaluating pragmatic factors such as the last date of literature search for any evidence synthesis used to develop recommendations or the inclusion criteria applied for literature searches. Moreover, users may also consider applying a credibility assessment. For example, the catalogue provides AGREE-II scores alongside each recommendation. The AGREE-II reporting checklist is a list of criteria used to assess the methodological rigour of the guideline development process. Accordingly, users may wish to compare scores between diverging recommendations published on the recommendation catalogue.

If differences are attributed to variation in the EtD criteria or contextualization factors, then users may review each criterion to make an informed decision. For example, if resource constraints and cost-effectiveness are not pressing concerns in their setting, and this prompted a conditional recommendation instead of a strong recommendation, users may consider implementing the intervention. Lastly, if organizations formulate comparable recommendations aimed at different settings, users may want to select the recommendation that complies with the regulations in their region.

3.3 Discussion

This study explored differences in the formal judgments of comparable recommendations issued on the eCOVID-19 Recommendations Catalogue. From our findings, we were able to identify 223 recommendations in divergence with at least one recommendation issued by a different guideline addressing the same intervention or health scope. Therapeutic interventions in our sample diverged in formal judgment more often than PICO element. Conversely, public health interventions in our sample diverged in population or subgroup considerations of the intervention more often than formal judgment. However, public health recommendations were less likely to use formal ratings than therapeutic interventions, possibly explaining these differences.

We applied existing definitions of formal judgments using the GRADE framework to differentiate between our recommendations. Consequently, we did not assume strength or direction for recommendations that did not provide a formal judgment. Nonetheless, differences in population or subgroup considerations may be interpreted as differences in the direction of recommendations in many of our examples. For instance, we can theoretically assume divergence in direction when guidelines recommend masking for different age groups or when they recommend different frequencies of time to clean public surfaces. Because we did not implicitly suppose formal judgment, we interpreted these types of recommendations as divergence in population and subgroup consideration of the intervention, respectively. Therefore, the magnitude of recommendations with divergence in direction may be higher than what is suggested from our study. Nonetheless, the findings from our research provide a vital baseline to describe the

frequency and types of divergence that can exist between comparable recommendations for the management of the novel coronavirus. Finally, we postulate a possible thematic framework for guideline users and developers to reflect when evaluating differences in the guideline development methods to facilitate their decision-making.

ADDITIONAL TABLES AND FIGURES:

Table One: Guideline Characteristics Extraction Table

Characteristic	Description of Characteristic
Cluster Number	A set of two or more recommendations with similar PICO elements but containing varying judgment in their conclusion
Guideline Title	Title of the guideline individual recommendation in the cluster was extracted from
Guideline Source	Organization or institution publishing guideline containing diverging recommendations
Date of publication or last literature review	Date of most updated version of the guideline at the time study was completed or date of the latest literature search
PICO elements	The population, intervention, comparison, and outcome groups for each recommendation as defined by Cochrane
Overarching Healthscope	Appraising if cluster intervention is therapeutic or public health
Intervention subtype	Appraising if cluster intervention is drug therapeutic, other clinical, IPC-measure, school-measure, or vaccine-measure
Recommendation	Complete recommendation extracted verbatim from the guideline
Diverging variable	Divergence in the recommendation strength, direction, and/or subgroup considerations (i.e., PICO element)

Table Two: Categorization Matrix of Predetermined Themes for Content Analysis

Question	Deductive themes for Appraising Differences in Methods				
What factors influence divergence in recommendations?	Date of publication	Included studies and assessed outcomes	Interpretation of the results	EtD Criteria	Additional factors
Total number (n) of clusters identified	3	3	3	6	5

Three: Detailed Frequencies of Divergence Within Clusters

3A) Divergence in Formal Judgment

	Divergence of Formal Judgment	
Types of Divergence	**Frequency (n)[a]**	**Percentage (%)[b]**
Strength	19	55.9
Direction	9	26.5
Divergence in both	6	17.6
Total	**34**	**100**

3B) Divergence in PICO Element

	Divergence in PICO element	
Types of Divergence	**Frequency (n)**	**Percentage (%)**
Population	6	18.8
Intervention	26	81.3
Total	**32**	**100**

3C) Total Divergence Across Clusters

	Overall Divergence	
Types of Divergence	**Frequency (n)**	**Percentage (%)**
Strength	19	28.8
Direction	9	13.6
Divergence in both	6	9.1
Population	6	9.1
Intervention	26	39.4
Total	**66**	**100**

[a] Frequency of clusters

[b] Percentage of clusters

Table Four: Chi-Square Table comparing the type of divergence between public health and therapeutic interventions

Table 4A) Contingency Table

Health scope	Type of divergence		
	Formal judgment (n)	PICO element (n)	Total (n)
Therapeutic intervention	27	2	29
Public Health intervention	7	30	37
Total (n)	34	32	66

Table 4B) Chi-Square

	Value	Df	Exact Sig. (1 sided)	Exact Sig. (2 sided)
Pearson chi-square	35.821[a]	1	<0.001	<0.001
Fisher's exact test		1	<0.001	<0.001
N of Valid cases	66			

[a] Zero cells (0.0%) have an expected count of less than 5. The minimum expected count is 13.78.

Figure One: Bar Graph of the Type of Divergence Identified within public health and therapeutic clusters