BMJ Global Health

The use of network meta-analysis in updating WHO living maternal and perinatal health recommendations

Myfanwy J Williams [1], Joshua P Vogel [2], Ioannis D Gallos [3], Jenny A Ramson [2], Doris Chou [3], Olufemi T Oladapo [3]

To cite: Williams MJ, Vogel JP, Gallos ID, *et al*. The use of network meta-analysis in updating WHO living maternal and perinatal health recommendations. *BMJ Glob Health* 2023;**8**:e013109. doi:10.1136/bmjgh-2023-013109

Handling editor Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgh-2023-013109).

Received 12 June 2023
Accepted 21 October 2023

[1]Independent, West Kirby, UK
[2]Maternal, Child and Adolescent Health Program, Burnet Institute, Melbourne, Victoria, Australia
[3]UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Sexual and Reproductive Health and Research, World Health Organization, Geneva, Switzerland

Correspondence to
Dr Myfanwy J Williams; williams.myfanwy@icloud.com

ABSTRACT

Drawing on two recent examples of WHO living guidelines in maternal and perinatal health, this paper elucidates a pragmatic, stepwise approach to using network meta-analysis (NMA) in guideline development in the presence of multiple treatment options. NMA has important advantages. These include the ability to compare multiple interventions in a single coherent analysis, provide direct estimates of the relative effects of all available interventions, infer indirect effect estimates for interventions not directly compared and generate rankings of the available treatment options. It can be difficult to harness these advantages in the face of a lack of current guidance on using NMA evidence in guideline development, with several challenges emerging. Challenges include the choice of conceptual approach, the volume and complexity of the evidence, the contribution of treatment rankings, and the fact that the preferable treatment is not always obvious. This paper describes a layered approach to resolving these challenges, which supports systematic guideline decision-making and development of trustworthy clinical guidelines when multiple treatment options are available.

SUMMARY BOX

⇒ While network meta-analysis (NMA) can provide streamlined and methodologically coherent synthesis of multiple treatments for guideline development, incorporating evidence from NMA into guideline Evidence-to-Decision (EtD) frameworks is not always straightforward.

⇒ NMA can generate a large volume of complex evidence that is both conceptually challenging and not easily presented using standard summary approaches, while some appealing features of NMA (such as treatment rankings) are potentially misleading for guideline panel members.

⇒ We elucidate how, for two sets of WHO guideline updates using effectiveness evidence from two NMAs, we adopted a layered approach to conceptualising and communicating the effectiveness evidence to guideline panel members.

⇒ We describe an approach that other guideline technical teams may adopt when preparing the 'effects of interventions' domain of EtD frameworks using NMA, which can facilitate interpretation, aid decision-making and support development of trustworthy recommendations.

INTRODUCTION

Network meta-analysis (NMA) is a technique used in systematic reviews to compare multiple treatments for a single condition.[1 2] It can also produce rankings of treatments for different outcomes.[3] In recent years, NMA has provided effectiveness evidence for the development of several WHO maternal and perinatal health recommendations.[4 5] We describe the process of using NMA for developing recommendations in the context of Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence-to-Decision (EtD) frameworks.[6 7]

While NMA has advantages over conventional pairwise meta-analysis for guideline development, several challenges emerged. These concerned the choice of conceptual approach, volume and complexity of the evidence, the contribution of treatment rankings, and the fact that the preferable treatment is not always obvious. We sought to achieve clarity while maintaining transparency, by adopting a pragmatic, stepwise approach. For two sets of recommendations, we provided guideline decision-makers with an accessible package of information designed to aid interpretation of the evidence.

WHO LIVING GUIDELINE DEVELOPMENT USING GRADE ETD FRAMEWORK

WHO guidelines are typically intended for global use, and are developed to rigorous methodological standards.[8] The scientific evidence supporting a WHO recommendation is synthesised using the GRADE approach, including the use of structured EtD frameworks.[6 7] These inform Guideline Development Group (GDG) deliberations and allow systematic and transparent use

Box 1 Uterotonics for preventing postpartum haemorrhage: 2018 recommendation update

Globally, nearly a quarter of maternal deaths are associated with postpartum haemorrhage (PPH), while uterine atony is the most common cause of PPH. Uterotonic agents work by increasing contractility of the uterus—when administered to all women prophylactically after birth, they can reduce postpartum blood loss. WHO published updated recommendations on the use of uterotonics for preventing PPH in 2018, following publication of a major new trial which found that heat-stable carbetocin was non-inferior to oxytocin for the prevention of PPH.[4] These recommendations used effectiveness evidence from a Cochrane systematic review and network meta-analysis (NMA) that included 196 trials involving 134 414 women.[11] This NMA included seven different uterotonic agents (oxytocin, misoprostol, carbetocin, ergometrine, injectable prostaglandins, eg, carboprost, syntometrine (oxytocin plus ergometrine), oxytocin plus misoprostol), as well as placebo/no treatment. The updated recommendations considered 18 outcomes that the Guideline Development Group agreed were critical and important for global guideline decision-making. These included outcomes prioritised when this guideline was previously updated in 2012 (which were identified through consultation with international stakeholders), plus three additional outcomes selected to reflect a relevant core outcome set published in the intervening period and to ensure that the final recommendations would be woman-centred.

Box 2 Tocolytics for delaying preterm birth: 2022 recommendation update

Preterm birth (before 37 completed weeks of pregnancy) is the single largest cause of neonatal death worldwide. The earlier babies are born, the greater the risk of respiratory, infectious, metabolic and neurological morbidities. Tocolytic drugs can inhibit or arrest contractions of the uterus, and thus can prolong pregnancy. This allows more time for in-utero fetal maturation, administration of antenatal corticosteroids and other medications that can improve preterm newborn outcomes, and also provide time for transferring a woman to a higher level of care. WHO published updated tocolytic recommendations in 2022 in the context of new, important evidence on the use of antenatal corticosteroids, whose effects are closely linked to the use of tocolysis.[5] For the recommendation update, evidence on the effectiveness and safety of tocolytics was provided by a new Cochrane systematic review and network meta-analysis.[12] This review identified 122 individually randomised trials, involving 13 697 women. The trials included comparisons of six different classes of tocolytic drugs (betamimetics, cyclo-oxygenase inhibitors, calcium channel blockers, magnesium sulphate, oxytocin receptor antagonists and nitric oxide donors), combinations of tocolytics, and placebo or no treatment with a tocolytic. The recommendation update considered 31 outcomes that the Guideline Development Group agreed were critical and important for global guideline decision-making, including those prioritised in the previous iteration of this recommendation (following consultation with international stakeholders), plus two additional outcomes to ensure that the final recommendations would be woman-centred.

of available evidence to formulate recommendations.[7] NMA can provide evidence that informs the EtD domain 'effects of interventions' (which includes the size of the desirable effects, size of the undesirable effects, certainty of the evidence of effects, and balance of desirable and undesirable effects). Other EtD domains (such as acceptability and feasibility) are informed by other processes.

Since 2017, WHO's Department of Sexual and Reproductive Health and Research has adopted a 'living guideline' approach to updating maternal and perinatal health recommendations. Individual recommendations rather than whole guidelines are prioritised for updating by an independent Executive Guideline Steering Group, on the basis of emerging evidence.[9] WHO has released more than 40 updated recommendations using this approach, including 11 informed by commissioned Cochrane reviews using NMA: ten 2018 recommendations on uterotonics for preventing postpartum haemorrhage (see box 1); and one 2022 recommendation on tocolytic therapy for improving preterm birth outcomes (see box 2).[4 5]

ADVANTAGES OF EVIDENCE SYNTHESIS USING NMA

NMA has several advantages over conventional pairwise meta-analysis. When sufficient trial data are available, it can provide direct estimates of the relative effects of all available interventions for a common set of outcomes. Indirect effect estimates can also be obtained when the relative effectiveness of two interventions is inferred through a common comparator. Network effect estimates combine the entirety of the available direct and indirect evidence connecting two interventions to yield an estimate that draws on all connected sources of evidence.[1 2] NMA also enables comparison of interventions that have not been directly compared, such as newer agents with placebo in the absence of placebo-controlled trials.[10]

NMA can also generate rankings of available treatments.[3] The Cochrane reviews that underpinned the 2018[11] and 2022[12] updates expressed rankings as the Surface area Under the Cumulative RAnking line (SUCRA). For each outcome, this indicates the cumulative probability of being the best agent, second best and so forth. SUCRAs are attractive because they are the only graphical option that allows simultaneous comparison of all treatments. Furthermore, they simplify the information about the relative effect of each treatment into a single number.[3] The closer the SUCRA value is to 100%, the more likely it is that the treatment is in the top rank or one of the top ranks.[13] For instance, figure 1 includes three interventions with high SUCRAs of 72%–76% (combinations of tocolytics, calcium channel blockers and nitric oxide donors).

ADVANTAGES OF USING NMA EVIDENCE IN GUIDELINE DEVELOPMENT

GRADE EtD frameworks were designed to integrate evidence from a standard pairwise review.[6 7] We have previously encountered indications with multiple treatment options, and hence multiple pairwise systematic

Effects of tocolytics for preventing preterm birth: Delay in birth by 48 hours

Source: Gallos ID, Wilson A, Hodgetts-Morton VA, Marson E, Markland A, Papadopoulou A, et al. Tocolytics for delaying preterm birth: a network meta-analysis. Cochrane Database 2021, (4), [CD014978]. (https://doi.org/10.1002/14651858.CD014978, accessed 1 July 2022).

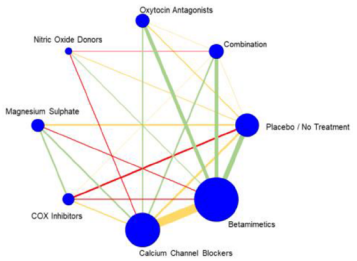
Patient or population: Women with signs and symptoms of preterm labour

Settings: Hospital setting

Intervention: Betamimetics, calcium channel blockers, COX inhibitors, magnesium sulfate, nitric oxide donors, oxytocin receptor antagonists, combinations of tocolytics

Comparison: Placebo or no treatment

Network diagram: The nodes represent an intervention and their size is proportional to the number of trials comparing this intervention to any other in the network. The lines connecting each pair of interventions represent a direct comparison and are drawn proportional to the number of trials making each direct comparison. Numbers on the lines represent the number of trials and participants for each comparison. The colour of the line is green for high-certainty evidence; light green for moderate-certainty evidence; orange for low-certainty evidence and red for very low-certainty evidence. Multi-arm trials contribute to more than one comparison.



	Direct Evidence		Indirect Evidence		Network Evidence		Anticipated absolute effects for network estimate[1]			SUCRA
	RR (95% CI)	Certainty	RR (95% CI)	Certainty	RR (95% CI)	Certainty	Risk with placebo or no treatment	Risk with tocolytic agent	Risk difference with tocolytic agent	
Betamimetics	1.27 (1.11, 1.45)	⊕⊕⊕⊖ MODERATE[a]	1.04 (0.96, 1.12)	⊕⊕⊖⊖ LOW[b]	1.12 (1.05, 1.20)	⊕⊕⊖⊖ LOW[c]	645 per 1,000	722 per 1,000	77 more per 1,000 (from 32 to 129 more)	42%
COX inhibitors	2.02 (0.81, 5.08	⊕⊖⊖⊖ VERY LOW[d]	1.10 (0.98, 1.23)	⊕⊖⊖⊖ VERY LOW[e]	1.11 (1.01, 1.23)	⊕⊕⊖⊖ LOW[f]	645 per 1,000	716 per 1,000	71 more per 1,000 (from 6 to 148 more)	42%
Calcium channel blockers	1.87 (1.06, 3.28)	⊕⊕⊖⊖ LOW[g]	1.17 (1.08, 1.26)	⊕⊕⊖⊖ LOW[b]	1.16 (1.07, 1.24)	⊕⊕⊖⊖ LOW[h]	645 per 1,000	748 per 1,000	103 per 1,000 (from 45 to 155 more)	72%
Magnesium sulfate	1.06 (0.88, 1.29)	⊕⊕⊖⊖ LOW[i]	1.14 (1.02, 1.28)	⊕⊖⊖⊖ VERY LOW[e]	1.12 (1.02, 1.23)	⊕⊕⊕⊖ MODERATE[k]	645 per 1,000	722 per 1,000	77 more per 1,000 (from 13 to 148 more)	44%
Oxytocin receptor antagonists	1.07 (0.91, 1.27)	⊕⊕⊖⊖ LOW[l]	1.17 (1.06, 1.29)	⊕⊕⊕⊖ MODERATE[m]	1.13 (1.05, 1.22)	⊕⊕⊕⊖ MODERATE[n]	645 per 1,000	729 per 1,000	84 more per 1,000 (from 32 to 142 more)	50%
Nitric oxide donors	1.18 (0.76, 1.84	⊕⊕⊖⊖ LOW[o]	1.20 (1.06, 1.36)	⊕⊕⊕⊖ MODERATE[m]	1.17 (1.05, 1.31)	⊕⊕⊕⊖ MODERATE[n]	645 per 1,000	755 per 1,000	110 per 1,000 (from 32 to 200 more)	74%
Combinations of tocolytics	1.05 (0.84, 1.31)	⊕⊕⊖⊖ LOW[p]	1.18 (1.08, 1.30)	⊕⊕⊕⊖ MODERATE[m]	1.17 (1.07, 1.27)	⊕⊕⊕⊖ MODERATE[n]	645 per 1,000	755 per 1,000	110 per 1,000 (from 45 to 174 more)	76%

1. Throughout the summary of findings tables, the three columns under 'Anticipated absolute effects for network estimate' list the anticipated number of women per 1000 who would experience the outcome if given placebo or no treatment; the anticipated number of women per 1000 who would experience the outcome if given the tocolytic agent; and the difference between these two (and confidence interval).

Footnotes

[a]Direct evidence downgraded -1 due to multiple limitations in study design.
[b]Indirect evidence downgraded -2 due to multiple limitations in study design and suspected publication bias.
[c]Network evidence downgraded -2 due to moderate certainty direct evidence further downgraded -1 because of incoherence.
[d]Direct evidence downgraded -3 due to multiple limitations in study design, severe unexplained statistical heterogeneity, and very serious imprecision.
[e]Indirect evidence downgraded -3 due to multiple limitations in study design, and very serious imprecision.
[f]Network evidence downgraded -2 due to very low certainty direct and indirect evidence upgraded +1 since the network estimate is precise.
[g]Direct evidence downgraded -2 due to multiple limitations in study design and severe unexplained statistical heterogeneity.
[h]Network evidence downgraded -2 due to low certainty direct and indirect evidence.
[i]Direct evidence downgraded -2 due to multiple limitations in study design and serious imprecision.
[k]Network evidence downgraded -1 due to low certainty direct evidence upgraded +1 since the network estimate is precise.
[l]Direct evidence downgraded -2 due to severe unexplained statistical heterogeneity and serious imprecision.
[m]Indirect evidence downgraded -1 due to multiple limitations in study design.
[n]Network evidence from moderate certainty indirect evidence.
[o]Direct evidence downgraded -2 due to very serious imprecision.
[p]Direct evidence downgraded -3 due to multiple limitations in study design and very serious imprecision.

Cumulative rankograms comparing each of the tocolytic drugs for **delay in birth by 48 hours**. Ranking indicates the cumulative probability of being the best agent, the second best, the third best, etc. The x axis shows the relative ranking and the y-axis the cumulative probability of each ranking. We estimate the SUrface underneath this Cumulative RAnking line (SUCRA); the larger the SUCRA the higher its rank among all available agents.

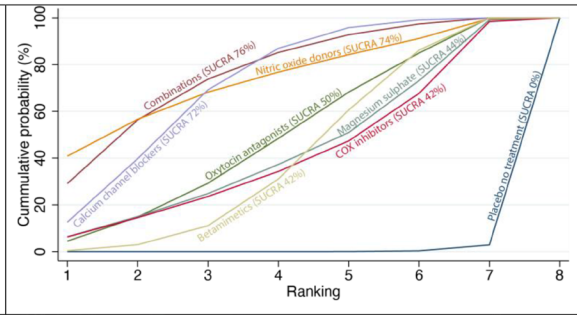


Figure 1 Example of NMA summary of findings table from Evidence-to-Decision framework on tocolytics for delaying preterm birth.[32]

reviews requiring an EtD framework per comparison, as in the previous iterations of the uterotonic and tocolytics recommendations.[14 15] This was cumbersome and difficult for GDG members to interpret. It was also challenging to assess the impact of methodological differences between reviews (eg, eligibility criteria, reported outcomes and handling of subgroup analyses). As searches were conducted at different times (sometimes years apart) and review methods have improved over time, methodological approaches may have differed. Also, some treatments appeared in more than one review, causing confusion as to which was more 'correct'. While not insurmountable, such issues risk introducing subtle but potentially important sources of bias into guideline decision-making.[16]

Generating one EtD framework including effectiveness evidence from only one NMA of multiple treatments resolves several of these difficulties, as it is a single evidence base, assembled with a standardised methodology. This is easier for GDG members to understand and uses optimal analytical methods.[10 16] Also, the use of evidence from NMA is particularly efficient because each guideline update requires only one (though often large) systematic review.

CHALLENGES OF USING EVIDENCE FROM NMA IN ETD FRAMEWORKS

Using evidence from NMA brings new challenges to guideline development. NMA does not currently appear in the WHO handbook for guideline development,[8] although there are a number of case studies available, notably the WHO living guideline on drug treatments for COVID-19.[17] Similarly, there is no official GRADE guidance on incorporating evidence from NMA into EtD frameworks, or on producing NMA evidence tables, although there are references available.[18 19]

For guideline methodologists, the absence of established conventions on these issues presents five significant challenges.

The usefulness of NMA evidence rests on conceptual decisions about the choice of question and reference agent

Guideline developers must decide the most logical and useful conceptual approach to structuring the evidence for decision-makers and end users, including decisions about the PICO (population, intervention, comparison, outcomes) question(s) and relatedly the choice of reference agent. Ideally, the PICO question(s) for the recommendation aligns with the systematic review PICO question(s). NMA can make it easier to address multiple treatment comparison or even multiple PICOs, however the best question strategy may be hard to determine when a single guideline must speak to different healthcare contexts. For example, the 2018 WHO uterotonics recommendations aimed to address high-income, medium-income and low-income countries, across which uterotonics availability varies from seven in higher-resource settings to one or two in limited-resource settings. Even though we had NMA evidence and would ultimately need to compare all options with one another, in the absence of widespread access to all treatments, the appropriate conceptual starting point was not a multiple treatment comparison.[4]

NMA generates evidence on all possible pairwise comparisons of the included interventions, which are potentially numerous. To interpret NMA results, a 'reference' intervention against which all other interventions are compared is chosen. The choice of reference is often placebo or no treatment, or the most commonly used comparator treatment, that is, the best-connected intervention in the network.[1] While it has an important bearing on the validity and utility of the resulting recommendation, the best choice of reference agent may not always be obvious. For the uterotonics guideline, the currently recommended treatment and best choice of reference agent was oxytocin, however placebo or no treatment was also a relevant comparator in contexts where oxytocin was not available.

A large volume of streamlined evidence is still a large volume of evidence

Effectiveness evidence generated by NMA is more coherent and streamlined than referring to multiple (potentially conflicting) systematic reviews. It also leverages indirect evidence, which can improve effect estimates. However, the GDG may still need to consider a large volume of evidence. The uterotonics question included evidence on 7 interventions plus placebo or no uterotonic treatment, and 18 outcomes (see box 1); the tocolytics question included 7 interventions plus placebo or no tocolytic treatment, and 31 outcomes (see box 2). We generated a summary of findings table for each outcome, presenting effect estimates for all interventions versus the reference. This is a lot of information for the panel to assimilate, even when synthesised and streamlined in this way.

NMA findings can be difficult for guideline panels to interpret

With multiple interventions and many outcomes comes complexity. It can be hard to discern which, if any, intervention is clinically superior, especially when results vary across outcomes.[20] Panel members found it challenging to understand how much weight to place on low or very low certainty evidence, and how to make decisions where findings for different outcomes appeared contradictory. For example, in figure 2 several tocolytic interventions were likely more effective at delaying birth by 7 days compared with placebo, while there was probably no difference between the intervention and placebo groups in the mean time between therapy and birth. These differences occur because the two outcomes included data from different trials. Although these issues may be familiar to GDG members considering evidence from pairwise systematic reviews, their impact is multiplied in line with the complexity of NMA.

Treatment rankings are problematic and potentially misleading

For the 2022 tocolytic guideline, we presented the treatment rankings. Rankings are produced on a per-outcome basis, and thus one drug can have different rankings for different outcomes. For instance, the SUCRA for the tocolytic intervention nitric oxide donors ranged from 1% to 100% (see table 1).

Methodologists advise caution in interpreting rankings.[2 13] A higher ranking cannot be relied on to consistently identify better treatments. Differences in rank might not be clinically significant, while the relative importance of an outcome may vary. Treatments can have a higher rank without evidence that they have a better effect,[2] and

Key:	High certainty (benefit)	Moderate certainty (probable benefit)	Low certainty (possible benefit)	High certainty (harm)	Moderate certainty (probable harm)	Low certainty (possible harm)	High certainty (no difference)	Moderate certainty (probable no difference)	Low certainty (possible no difference)	Very low certainty (uncertain)

Outcome	*Placebo/ no tocolytic (reference)*	Betamimetics vs placebo/no tocolytic	COX inhibitors vs placebo/no tocolytic	Calcium channel blockers vs placebo/no tocolytic	Magnesium sulfate vs placebo/no tocolytic	Oxytocin receptor agonists vs placebo/no tocolytic	Nitric oxide donors vs placebo/no tocolytic	Combinations of tocolytics vs placebo/no tocolytic
Pregnancy prolongation								
Delay in birth 48 hours	*645 per 1,000*	**Possibly increased** 77 more (32 more to 129 more)	**Possibly increased** 71 more (6 more to 148 more)	**Possibly increased** 103 more (45 more to 155 more)	**Probably increased** 77 more (13 more to 148 more)	**Probably increased** 84 more (32 more to 142 more)	**Probably increased** 110 more (32 more to 200 more)	**Probably increased** 110 more (45 more to 174 more)
Delay in birth 7 days	*742 per 1,000*	**Possibly increased** 104 more (22 more to 186 more)	**Probably no difference** 30 more (89 fewer to 178 more)	**Probably increased** 111 more (30 more to 200 more)	**Uncertain**	**Increased** 134 more (52 more to 223 more)	**Probably increased** 134 more (15 more to 275 more)	**Probably Increased** 141 more (37 more to 252 more)
Birth <28 weeks[1]	*158 per 1,000*	**Not estimable**	**Not estimable**	**Not estimable**	**Uncertain**	**Probably increased** 333 more (3 more to 1000 more)	**Possibly no difference** 79 fewer (122 fewer to 14 more)	**Not estimable**
Birth <32 weeks	*476 per 1,000*	**Possibly no difference** 67 fewer (129 fewer to 5 more)	**Not estimable**	**Possibly no difference** 5 fewer (138 fewer to 186 more)	**Uncertain**	**Possibly no difference** 67 more (90 fewer to 295 more)	**Uncertain**	**Uncertain**
Birth <34 weeks	*313 per 1,000*	**Uncertain**	**Uncertain**	**Uncertain**	**Uncertain**	**Uncertain**	**Possibly no difference** 44 fewer (128 fewer to 85 more)	**Uncertain**
Birth <37 weeks	*593 per 1,000*	**Possibly no difference** 17 fewer (97 fewer to 74 more)	**Uncertain**	**Possibly no difference** 51 fewer (126 fewer to 40 more)	**Uncertain**	**Probably no difference** 57 more (63 fewer to 206 more)	**Possibly no difference** 131 fewer (243 fewer to 0 more)	**Uncertain**
Mean time from tocolytic therapy to birth (days)	*20 days*	**Probably no difference** 1 day more (3 fewer to 5 more)	**Possibly no difference** 3 days more (4 fewer to 11 more)	**Increased** 5 days more (0.13 more to 9 more)	**Uncertain**	**Possibly increased** 10 days more (2 more to 17 more)	**Probably no difference** 7 days more (0.44 fewer to 15 more)	**Uncertain**
Gestational age at birth	*35.2 weeks*	**Probably no difference** MD 0.2 weeks fewer (0.7 fewer to 0.2 more)	**Uncertain**	**Probably no difference** MD 0.2 weeks more (0.3 fewer to 0.7 more	**Uncertain**	**Possibly no difference** MD 0.1 weeks fewer (0.7 fewer to 0.6 more)	**Possibly increased** 1.4 weeks more (0.4 more to 2.3 more)	**Uncertain**

1. Direct evidence only
2. This outcome was defined as time from trial entry (i.e. time when a woman was randomised) until time of birth.

Figure 2 Excerpt from summary table of anticipated absolute effects of tocolytics versus reference (placebo/no treament) for delaying preterm birth.[32]

rankings based on low or very low certainty evidence are not conclusive.[13] While effect estimates usually include a conventional level of significance such as a p value or CI, SUCRA rankings do not (although other approaches, such as the use of median ranking, can).

NMA is not a panacea for GDG decision-making

While NMA has advantages, it is not a panacea for all the challenges the GDG faces when navigating effectiveness evidence. NMA does not automatically provide 'the answer' as to the clinically superior intervention. Given the significant resources invested in conducting NMA, and the promise inherent in using up-to-date methods, GDG members may find this disappointing or frustrating.

NMA provides a coherent picture of the available evidence, but the evidence may be incomplete or 'patchy'. The indirect evidence may help address any gaps but cannot resolve them all. Also, assessors may have low confidence in the evidence that is available—for example, the tocolytics NMA identified 122 trials of 13 697 women, however, most outcomes had low or very low certainty evidence.

PRAGMATIC SOLUTIONS

We adopted a pragmatic approach to addressing these inter-related challenges. Where necessary we broke down the guideline decision-making process into multiple stages. We had two guiding principles: to make all the effectiveness evidence available in a structured manner so that any detail was accessible and readily retrievable; to facilitate interpretation and interrogation of the evidence by providing multiple navigable layers, from simple to complex. We aimed to produce materials that maximised clarity by using carefully considered layout, graphical elements and consistent colour-coding. We describe our solutions in the following sections.

Starting point: conceptual approach

In the example described earlier (section Challenges of using evidence from NMA in EtD frameworks)

Table 1 Example of different treatment rankings for a single intervention.[32]

Outcome	Nitric oxide donors vs placebo/no tocolytic SUCRA treatment ranking (Surface area Under the Cumulative RAnking line)
Pregnancy prolongation outcomes	
Delay in birth 48 hours	74%
Delay in birth 7 days	76%
Birth<32 weeks	67%
Birth<34 weeks	60%
Birth<37 weeks	94%
Mean time from tocolytic therapy to birth (days)	80%
Gestational age at birth	98%
Maternal outcomes	
Serious adverse effects	56%
Cessation of treatment due to side effects	40%
Palpitations	80%
Headache	1%
Nausea or vomiting	57%
Tachycardia	84%
Cardiac arrhythmias	–
Hypotension	32%
Pulmonary oedema	60%
Dyspnoea	81%
Maternal infection	–
Fetal and neonatal outcomes	
Perinatal death	83%
Fetal death	72%
Neonatal death before 28 days	69%
Respiratory morbidity	85%
Neurodevelopmental morbidity	74%
Gastrointestinal morbidity	47%
Neonatal infection	—
Mean birth weight	100%
Birth weight<2000 g	—
Birth weight<2500 g	87%

concerning the best approach to take to the 2018 uterotonics guideline, the GDG needed to develop recommendations that addressed the variable availability of uterotonic agents across high-resource, medium-resource and low-resource settings. Given this variation, in order to be useful to policy-makers and clinicians in all contexts it was crucial that the guideline first establish to what extent each agent was better or worse than placebo or no treatment. To support the GDG in tackling this issue thoroughly and systematically, although all effectiveness evidence was drawn from a single NMA, we divided the decision-making process into two sets of PICO questions. In a first phase of GDG meetings, we asked whether each intervention improved outcomes, developing individual EtD frameworks for each intervention versus placebo or no treatment. These EtDs are available in a series of web annexes that accompany the published guideline.[21–26] In addition to supporting the GDG in their decision-making, this deconstructed approach had the additional benefit of providing a clear evidence base for policy-makers in diverse global contexts where only one or two treatments are available. Second, speaking to settings where all interventions are available, we prepared a single EtD that compared all interventions to a reference (oxytocin), which is available in a further web annex[27] and is also included in online supplemental appendix A. While this entailed multiple GDG meetings, it meant we could be confident in the completeness and relevance of the resultant recommendations.

Oxytocin was chosen as the reference because it was the current standard of care, and the most frequently investigated uterotonic across all outcomes. For the 2022 tocolytics update (EtD included in online supplemental appendix B), placebo or no tocolytic treatment was the chosen reference because there was no standard tocolytic treatment, tocolytics were not recommended for women at risk of imminent preterm birth, and in the NMA, placebo/no tocolytic treatment was the best-connected node across most outcomes.

Evidence foundation: NMA summary of findings tables

Appraisal of the evidence on effectiveness and safety of interventions is usually captured in GRADE evidence profiles (see box 3), and as such these tables provide the basis for all GDG decision-making on the undesirable and undesirable effects and certainty of the evidence. Evidence profiles have a standard format that is familiar to guideline decision-makers, being automatically generated using GRADEpro software.[28] GRADEpro does not currently support NMA results (although one pilot project has trialled a new (not publicly released) GRADEpro

Box 3 GRADE certainty assessments

A key step in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence-to-Decision process involves rating the 'certainty' of effect estimates for each outcome from high to very low certainty. Certainty captures how confident assessors are that the evidence describes the true intervention effect. To determine the certainty level of evidence on the effectiveness and safety of interventions, evidence for each comparison and outcome is assessed against predefined criteria (study design, risk of bias, inconsistency, indirectness, imprecision and publication bias). Assessments are summarised for guideline decision-makers in 'evidence profiles' (detailed 'summary of findings' tables that include explicit judgements for each GRADE criterion so that panel members can interrogate and reach agreement on these assessments).[40]
A modified method for assessing the certainty of effect estimates generated by network meta-analysis has been described by the GRADE Working Group.[29 30]

module to support multiple intervention comparison[19]). Therefore we undertook the process manually, using Excel and Word.[29 30] Since we undertook this process, further detailed guidance has been published providing practical strategies for GRADE-assessing NMA effect estimates, that seeks to reduce the significant workload involved. This approach may be especially beneficial for developers of guidelines based on NMA with a large number of interventions.[31] As NMA generates three relevant effect estimates for each outcome—direct, indirect and network estimates—and the certainty of evidence may differ for each, it would be hard to include a complete breakdown of all GRADE assessments in the standard 'evidence profile' format. Our results tables therefore more closely resembled adapted 'summary of findings' tables.

While there is no established format for NMA evidence tables, guidance is available and there are examples in recent literature. Yepes-Nuñez *et al* have explored the optimal presentation of NMA results in summary of findings tables.[18] Taking as their starting point published expert guidance on the aspects of NMA that should be included (relative and absolute effects; GRADE certainty; rank probabilities; NMA geometry), the authors develop a template intended to facilitate understanding and enhance decision-making. While the paper does not constitute official GRADE guidance, we adopted many aspects of the suggested approach.

For our two sets of WHO recommendations, we included one NMA summary of findings table for each outcome, detailing effectiveness evidence for all interventions versus a reference comparator (see boxes 1 and 2, and example figure 1). All effect estimates are detailed alongside their GRADE, anticipated absolute effects for the 'headline' estimate (as natural frequency) and the ranking (as SUCRA). The PICO, network diagram and SUCRA graph are also provided. The network estimate is usually the 'headline' result, although this is not always the case (eg, where no indirect evidence is available). Our summary of findings tables differed slightly from the table described by Yepes-Nuñez *et al*.[18] We retained effect estimates and GRADE assessments for the indirect and direct evidence, alongside the network estimate, consistent with previous GRADE Working Group advice.[29 30] In line with the source Cochrane review, we used SUCRA rather than median rank to express treatment rankings, and so included the SUCRA graph. We also included explanations of network diagrams and SUCRA graphs. Like Yepes-Nuñez *et al*, we listed the interventions in the same order (and not by order of rank), reflecting our caution about the reliability of rankings, a challenge discussed in more detail in later sections.

Summary of findings: collating all outcomes

The 2018 and 2022 EtD frameworks included 18 and 31 detailed summary of findings tables, respectively (see boxes 1 and 2, and figure 1). To facilitate comprehension of this large body of evidence, we included a collated summary of intervention effects in the main EtD framework. The full summary of findings tables were included as an appendix (later published in web annexes to the main guideline, available as online supplemental appendices A and B).[27 32]

The collated summaries comprised colour-coded tables. Figure 3 shows the summary from the 2018 uterotonics update. GDG assessments of the balance of the desirable and undesirable effects must factor in the magnitude of the effect; we modified the design for the 2022 tocolytics update (see figure 2), including anticipated absolute effects (natural frequency per 1000 women, with 95% CI) for each intervention versus the reference. The colours signal benefit (green), harm (orange/red) or no difference (grey). The shades (darker to lighter) reflect the certainty of the evidence, with a neutral yellow signalling very low certainty. The narrative interpretation of each result is included for clarity, and to increase accessibility for colour-blind readers, using language that reflects guidance published by the GRADE Working Group.[33]

These summaries brought together evidence on the size of the desirable and undesirable effects and the certainty of the evidence in a highly digestible format, enabling the panel to begin to make assessments of the balance of effects for all interventions. Although differing in some details, developers of other WHO guidelines have produced similar tabular collations of NMA results, and the authors have observed that this approach is 'optimal'.[17 34] This adaptable format provides an accessible overview and interpretative aid for guideline decision-makers.

Summary of findings: communicating treatment rankings

For the 2022 tocolytics recommendation update, we included treatment rankings in the summary of findings tables (figure 1) and provided a collated array in an appendix (figure 4). As noted earlier, the interpretation of rankings is fraught with difficulty. As indicated by the colour-coding and narrative interpretation in figure 4, the certainty of the evidence varied widely, and we had little confidence in some ostensibly 'high ranking' treatments for some outcomes (signalled by the yellow boxes). Recent GRADE guidance acknowledges these challenges, and offers an approach to drawing conclusions from NMA that takes into account primarily the effect estimates and certainty of the evidence, and secondarily the rankings, however we found this approach overly cumbersome to implement given the number of outcomes involved.[20 35] The collated summary of SUCRA rankings provided another prism through which panel members could assess the limitations of the available evidence.

Making judgements

Guideline panel members study the EtD before the group meets. The highly visual collated summary supported identification of treatments that signalled potentially important benefits across priority outcomes, and red flags (eg, in side effect profiles). During meetings, the GDG

Desirable outcomes	Oxytocin (absolute risk)	Carbetocin	Misoprostol	Injectable prostaglandins	Ergometrine	Oxytocin plus ergometrine	Misoprostol plus oxytocin
Maternal death	1 per 1000	Probably similar effect	Probably similar effect	Don't know	Don't know	Don't know	Don't know
PPH ≥ 1000 ml	37 per 1000	Uncertain	Inferior	Uncertain	Possibly similar effect	Similar effect	Similar effect
Blood transfusion	22 per 1000	Probably similar effect	Probably similar effect	Uncertain	Possibly similar effect	Possibly similar effect	Probably superior
ICU admissions	2 per 1000	Probably similar effect	Probably similar effect	Don't know	Uncertain	Uncertain	Uncertain
PPH ≥ 500 ml	145 per 1000	Probably superior	Possibly similar effect	Possibly similar effect	Possibly similar effect	Probably superior	Possibly superior
Additional uterotonics	135 per 1000	Possibly superior	Possibly similar effect	Possibly superior	Uncertain	Possibly superior	Probably superior
Blood loss	301.5 mL (98-1299 mL)	Possibly superior	Uncertain	Uncertain	Possibly similar effect	Uncertain	Probably superior
Change in haemoglobin	11.37 g/L (2.30-27.9 g/L)	Possibly superior	Uncertain	Uncertain	Possibly similar effect	Possibly similar effect	Possibly superior
Breastfeeding	849 per 1000	Uncertain	Don't know	Don't know	Don't know	Similar effect	Don't know

ICU: intensive care unit; PPH: postpartum haemorrhage
Superior or inferior or similar effect– high certainty evidence of different effect or no effect
Probably superior or probably inferior or probably similar effect– moderate certainty evidence of different effect or no effect
Possibly superior or possibly inferior or possibly similar effect - low certainty evidence of different effect or no effect
Uncertain – very low certainty evidence (regardless of effect)
Don't know – outcome not reported / not estimable

Figure 3 Summary table of anticipated treatment effects (beneficial outcomes) of uterotonic agents versus reference agent (oxytocin) for preventing postpartum haemorrhage.[27]

used the collated summary to identify a shortlist of potential candidate treatments based on their benefit/risk profiles. The panel identified the most promising three options based on key outcomes, and ruled out others that were clearly unhelpful due to having the fewest benefits and/or worst side effects. When discussing the safety and effectiveness of the shortlisted options, the panel considered whether the differences between them were clinically meaningful. This discussion focused on the magnitude of the effect and certainty of the evidence, as well as outcome importance, and was less dependent on SUCRA rankings.

In the EtD frameworks that compared multiple treatments, we modified judgement tables for all EtD domains to include rows summarising judgements for all treatments (online supplemental appendices A and B).[27 32] Final judgements depended on further discussion of all EtD domains (and not just effectiveness). For example, while the tocolytics recommendation highlights nifedipine, the accompanying remarks made by the GDG note that oxytocin receptor agonists and nitric oxide donors can be effective in prolonging pregnancy, but are not available in many countries and can be more costly.[5]

STRENGTHS AND LIMITATIONS OF OUR APPROACH

This stepwise, layered approach enabled us to organise a productive guideline decision-making process. Both GDGs were able to navigate complex effectiveness evidence in readiness for consideration of other EtD domains. Presenting top layers of the most salient, simplified evidence did not obscure detail, but rather enabled precise and transparent signalling of areas of uncertainty or inconsistency. Ultimately, this meant that GDG discussions could relatively systematically move towards

Key:	High certainty (benefit)	Moderate certainty (probable benefit)	Low certainty (possible benefit)	High certainty (harm)	Moderate certainty (probable harm)	Low certainty (possible harm)	High certainty (no difference)	Moderate certainty (probable no difference)	Low certainty (possible no difference)	Very low certainty (uncertain)

Outcome	*Placebo/ no tocolytic (reference)*	Betamimetics vs placebo/no tocolytic	COX inhibitors vs placebo/no tocolytic	Calcium channel blockers vs placebo/no tocolytic	Magnesium sulfate vs placebo/no tocolytic	Oxytocin receptor agonists vs placebo/no tocolytic	Nitric oxide donors vs placebo/no tocolytic	Combinations of tocolytics vs placebo/no tocolytic
Pregnancy prolongation								
Delay in birth 48 hours	*0%*	**Possibly increased** 42%	**Possibly increased** 42%	**Possibly increased** 72%	**Probably increased** 44%	**Probably increased** 50%	**Probably increased** 74%	**Probably increased** 76%
Delay in birth 7 days	*16%*	**Possibly increased** 55%	**Probably no difference** 30%	**Probably increased** 61%	**Uncertain** 6%	**Increased** 78%	**Probably increased** 76%	**Probably increased** 79%
Birth <32 weeks	*52%*	**Possibly no difference** 79%	**Not estimable** —	**Possibly no difference** 56%	**Uncertain** 22%	**Possibly no difference** 28%	**Uncertain** 67%	**Uncertain** 46%
Birth <34 weeks	*35%*	**Uncertain** 18%	**Uncertain** 79%	**Uncertain** 64%	**Uncertain** 47%	**Uncertain** 17%	**Possibly no difference** 60%	**Uncertain** 79%
Birth <37 weeks	*38%*	**Possibly no difference** 47%	**Uncertain** 33%	**Possibly no difference** 70%	**Uncertain** 26%	**Probably no difference** 15%	**Possibly no difference** 94%	**Uncertain** 77%
Mean time from tocolytic therapy to birth (days)	*20%*	**Probably no difference** 28%	**Possibly no difference** 52%	**Increased** 67%	**Uncertain** 25%	**Possibly increased** 92%	**Probably no difference** 80%	**Uncertain** 37%
Gestational age at birth	*32%*	**Probably no difference** 14%	**Uncertain** 82%	**Probably no difference** 64%	**Uncertain** 24%	**Possibly no difference** 31%	**Possibly increased** 98%	**Uncertain** 55%

Figure 4 Excerpt from summary table of treatment rankings from Evidence-to-Decision framework on tocolytics for delaying preterm birth.[32]

conclusions about the balance of effects of the available interventions.

We have described our approach to handling evidence on the effect of interventions from NMA, but we have not attempted here to address the challenges posed to other EtD domains by the availability of multiple interventions. Other authors have recently highlighted that NMA does not address EtD domains beyond those concerned with effectiveness evidence, and have described the ongoing development of a solution to this wider issue.[19]

One further limitation in our approach may be the considerable resources involved in preparing the NMA (for author teams) and the EtDs (for guideline technical teams). When commissioning a systematic review, guideline development teams must weigh up whether this investment is likely to pay off, that is, whether NMA is warranted as opposed to relying on pairwise meta-analysis. For both the examples presented, the resource investment in EtD preparation improved usability and efficiency for the GDG.

Recently, several WHO guidelines have been developed using WHO-INTEGRATE, a modified version of the GRADE EtD framework that places greater emphasis on WHO's distinctive norms and values.[36] The process described in this paper will be informative for appraisal of effectiveness evidence from NMA for clinical guideline development using the standard GRADE approach or such adaptations.

POSSIBLE IMPROVEMENTS FOR FUTURE GUIDELINE UPDATES

There may be ways to improve this approach, that modify or adapt the components we have described (summary of findings tables, collated summary of findings and treatment rankings), or add additional components into the package. For instance, a modified summary of findings table could incorporate alternative ranking methods or colour-coded narrative interpretation of results. It would also be helpful if software were available for creating NMA GRADE summary of findings tables, as doing this manually is time-consuming and risks copy-paste errors.

While it is no longer a novel method of meta-analysis, NMA remains at the cutting edge of evidence synthesis for clinical guideline development. Efforts are ongoing to improve the ability of clinicians and decision-makers to fully and accurately use NMA findings. Guideline technical teams could explore the potential of novel visual approaches to augment evidence summaries and enhance interpretation.[37–39]

CONCLUSIONS

Guideline development involves difficult decisions about the best conceptual starting points and ultimate recommendations. NMA offers guideline developers many advantages over standard meta-analysis when multiple treatments are available. However, NMA may not resolve all difficulties, and can create distinctive challenges. These challenges are not insurmountable, and we have provided some solutions that are characterised by a stepwise approach to conceptualising, presenting and interpreting effectiveness evidence. Although it involved significant preparation by the technical team, this enabled guideline panel members to develop trustworthy recommendations. Developers of future clinical guidelines may build on this approach, potentially

incorporating innovations such as novel depictions of NMA results or using more advanced software aids. We anticipate evolving consensus in this area and look forward to further advances.

Contributors This article was conceived by OTO and JPV. MJW drafted and revised the paper, with supervision by JPV and incorporating feedback from OTO and IDG. JAR helped to edit the final version of the paper. JPV is guarantor. OTO, JPV, IDG and MJW developed the 2018 uterotonics EtD frameworks discussed. OTO, JPV, IDG, JAR and DC developed the 2022 tocolytics EtD framework discussed. All authors critically reviewed the manuscript and approved the final version. The corresponding author (MJW) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding The authors received no specific funding to develop this article. Both the tocolytics and uterotonics recommendations were developed with financial support from USAID and the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), a co-sponsored programme executed by the World Health Organization. The donors did not participate in any decision related to the recommendation development process, including the composition of research questions, membership of the recommendation development groups, conducting and interpretation of systematic reviews, or formulation of the recommendations. The views of the funding bodies did not influence the content of the recommendations.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study. Not applicable.

ORCID iDs
Myfanwy J Williams http://orcid.org/0000-0001-8201-290X
Joshua P Vogel http://orcid.org/0000-0002-3214-7096
Ioannis D Gallos http://orcid.org/0000-0002-2766-358X
Jenny A Ramson http://orcid.org/0000-0002-4091-6529
Doris Chou http://orcid.org/0000-0003-0250-4010
Olufemi T Oladapo http://orcid.org/0000-0002-3371-5892

REFERENCES

1 Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for Clinicians. *Intern Emerg Med* 2017;12:103–11.
2 Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed* 2019;104:F8–12.
3 Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and Tutorial. *J Clin Epidemiol* 2011;64:163–71.
4 WHO recommendations: uterotonics for the prevention of postpartum haemorrhage. Geneva: World Health Organization; 2018. Available: https://www.who.int/publications/i/item/9789241550420 [Accessed 16 Nov 2023].
5 WHO recommendation on tocolytic therapy for improving preterm birth outcomes. Geneva: World Health Organization; 2022. Available: https://apps.who.int/iris/handle/10665/363128 [Accessed 16 Nov 2023].
6 Alonso-Coello P, Oxman AD, Moberg J, *et al*. GRADE evidence to decision (Etd) frameworks: a systematic and transparent approach to making well informed Healthcare choices. 1: introduction. *BMJ* 2016;353:i2089.
7 Alonso-Coello P, Oxman AD, Moberg J, *et al*. GRADE evidence to decision (Etd) frameworks: a systematic and transparent approach to making well informed Healthcare choices. 2: clinical practice guidelines. *BMJ* 2016;353:i2089.
8 World Health Organization. WHO Handbook for guideline development. 2nd edn. Geneva: World Health Organiztion, 2014.
9 Vogel JP, Dowswell T, Lewin S, *et al*. "Developing and applying a "living guidelines" approach to WHO recommendations on maternal and perinatal health". *BMJ Glob Health* 2019;4:e001683.
10 Kanters S, Ford N, Druyts E, *et al*. Use of network meta-analysis in clinical guidelines. *Bull World Health Organ* 2016;94:782–4.
11 Gallos ID, Papadopoulou A, Man R, *et al*. Uterotonic agents for preventing postpartum haemorrhage: A network meta-analysis. *Cochrane Database Syst Rev* 2018;12:CD011689.
12 Wilson A, Hodgetts-Morton VA, Marson EJ, *et al*. Tocolytics for delaying Preterm birth: a network meta-analysis (0924). *Cochrane Database Syst Rev* 2022;8:CD014978.
13 Mbuagbaw L, Rochwerg B, Jaeschke R, *et al*. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev* 2017;6:79.
14 WHO recommendations for the prevention and treatment of postpartum haemorrhage: evidence base. Geneva: World Health Organization; 2012. Available: https://www.who.int/publications/i/item/9789241548502 [Accessed 21 Apr 2023].
15 WHO recommendations on interventions to improve preterm birth outcomes: Evidence base. Geneva: World Health Organization; 2015. Available: https://www.who.int/publications/i/item/9789241508988 [Accessed 21 Apr 2023].
16 Vogel JP, Williams M, Gallos I, *et al*. WHO recommendations on Uterotonics for postpartum haemorrhage prevention: what works, and which one *BMJ Glob Health* 2019;4:e001466.
17 Siemieniuk RA, Bartoszko JJ, Zeraatkar D, *et al*. Drug treatments for COVID-19: living systematic review and network meta-analysis. *BMJ* 2020;370:m2980.
18 Yepes-Nuñez JJ, Li S-A, Guyatt G, *et al*. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol* 2019;115:1–13.
19 Piggott T, Brozek J, Nowak A, *et al*. Using GRADE evidence to decision frameworks to choose from multiple interventions. *Journal of Clinical Epidemiology* 2021;130:117–24.
20 Brignardello-Petersen R, Florez ID, Izcovich A, *et al*. GRADE approach to drawing conclusions from a network meta-analysis using a minimally Contextualised framework. *BMJ* 2020;371:m3900.
21 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 1: Oxytocin versus placebo or no treatment. Geneva: World Health Organization; 2018. Available: https://apps.who.int/iris/handle/10665/277277 [Accessed 13 Dec 2022].
22 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 2: Carbetocin versus placebo or no treatment. Geneva: World Health Organization; 2018. Available: https://apps.who.int/iris/handle/10665/277278 [Accessed 13 Dec 2022].
23 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 3: Misoprostol versus placebo or no treatment. Geneva: World Health Organization; 2018. Available: https://apps.who.int/iris/handle/10665/277279 [Accessed 13 Dec 2022].
24 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 4: Ergometrine / Methylergometrine versus placebo or no treatment. Geneva: World Health Organization; 2018. Available: https://apps.who.int/iris/handle/10665/372388 [Accessed 13 Dec 2022].
25 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 5: Oxytocin and Ergometrine versus placebo or no treatment. Geneva: World Health Organization, . 2018Available: https://apps.who.int/iris/handle/10665/277280 [Accessed 13 Dec 2022].
26 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 6: Injectable prostaglandins versus placebo or no treatment. Geneva: World Health Organization;

2018. Available: https://apps.who.int/iris/handle/10665/277281 [Accessed 13 Dec 2023].
27 WHO recommendations: Uterotonics for the prevention of postpartum haemorrhage. Web annex 7: Choice of uterotonic agents. Geneva: World Health Organization; 2018. Available: https://apps.who.int/iris/handle/10665/277283 [Accessed 21 Apr 2023].
28 GRADEpro GDT. *GRADEpro Guideline Development Tool [Software]*. McMaster University and Evidence Prime, 2022.
29 Puhan MA, Schünemann HJ, Murad MH, *et al*. A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630.
30 Brignardello-Petersen R, Bonner A, Alexander PE, *et al*. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol* 2018;93:36–44.
31 Izcovich A, Chu DK, Mustafa RA, *et al*. A guide and pragmatic considerations for applying GRADE to network meta-analysis. *BMJ* 2023;381:e074495.
32 WHO recommendation on tocolytic therapy for improving preterm birth outcomes. Web annex: Evidence-to-decision frameworks. Geneva: World Health Organization; 2022. Available: https://apps.who.int/iris/bitstream/handle/10665/363130/9789240057241-eng.pdf [Accessed 13 Dec 2022].
33 Santesso N, Glenton C, Dahm P, *et al*. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol* 2020;119:126–35.
34 Chu DK, Brignardello-Petersen R, Guyatt GH, *et al*. Method's corner: Allergist's guide to network meta-analysis. *Pediatr Allergy Immunol* 2022;33:e13609.
35 Brignardello-Petersen R, Izcovich A, Rochwerg B, *et al*. GRADE approach to drawing conclusions from a network meta-analysis using a partially Contextualised framework. *BMJ* 2020;371:m3907.
36 Rehfuess EA, Stratil JM, Scheel IB, *et al*. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. *BMJ Glob Health* 2019;4(Suppl 1):e000844.
37 Seide SE, Jensen K, Kieser M. Utilizing radar graphs in the visualization of simulation and estimation results in network meta-analysis. *Res Synth Methods* 2021;12:96–105.
38 Daly CH, Mbuagbaw L, Thabane L, *et al*. Spie charts for Quantifying treatment effectiveness and safety in multiple outcome network meta-analysis: a proof-of-concept study. *BMC Med Res Methodol* 2020;20:266.
39 Ostinelli EG, Efthimiou O, Naci H, *et al*. Vitruvian plot: a Visualisation tool for multiple outcomes in network meta-analysis. *Evid Based Ment Health* 2022;25:e65–70.
40 Guyatt G, Oxman AD, Akl EA, *et al*. GRADE guidelines: 1. introduction - GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.