World Health Organization

Report on

the Third Meeting of the WHO Technical Advisory Group on Gestational Weight Gain

12-14 March 2024

World Health Organization

Report on

The Third Meeting of the WHO Technical Advisory Group on Gestational Weight Gain

12-14 March 2024

CONTENTS

ACRONYMS 4
1. INTRODUCTION 5
2. SUMMARY OF DAY 1 PRESENTATIONS AND DISCUSSIONS 5
Session 1: Overview of the GWG Project 6
Session 2: Systematic literature review on association of GWG with maternal and infant outcomes 7
Session 3: Working Group 1 – Individual-level eligibility criteria (updates) 13
Session 4: Working Group 2 – Identification and harmonization of databases (updates) 18
Session 5: Data acquisition (updates) 22
Session 6: Working Group 3 – Methods for the development of GWG Standards (updates) 25
Session 7: WHO Guidelines for Declaration of Interests 29
3. SUMMARY OF DAY 2 PRESENTATIONS AND DISCUSSIONS 30
Session 1: Guide for the Working Group Discussions 31
Session 2: "Considerations around separate charts for subgroups with different weight gain patterns" 31
Session 3: Working Group 1 presentation 34
Session 4: Working Group 3 presentation 36
Session 5: *Modus operandi* of working groups – way forward 37
4. SUMMARY OF DAY 3 PRESENTATIONS AND DISCUSSIONS 38
Session 1: How to call the GWG centile curves, which will be developed based on "semi-prescriptive" sample? 38
Session 2: Use of GWG standards in WHO recommendations 40
Next steps 43
Closing remarks 44
ANNEX I. Meeting participants 45
ANNEX II. Meeting agenda 47
ANNEX III. Working groups guide, points for discussion (Day 2) 50
ANNEX IV. 2024–2025 work plan (Phase 1) 52

ACRONYMS

AHS	Food and Nutrition Actions in Health Systems
ANC	Antenatal care
APGAR	Appearance, pulse, grimace, activity and respiration
BMI	Body mass index
C-section	Caesarean section
COI	Conflict of interest
GDM	Gestational diabetes mellitus
GWG	Gestational weight gain
HIV	Human immunodeficiency virus
IOM	Institute of Medicine
IPD	Individual participant data
kg	Kilogram
LGA	Large-for-gestational age
LMP	Last menstrual period
MAH	Maternal Health Unit
MCA	Department Maternal, Newborn, Child and Adolescent Health and Ageing
MNF	Monitoring Nutrition and Food Safety Events Unit
MPH	Maternal and Perinatal Health Unit
NFS	Department of Nutrition and Food Safety
NICHD	National Institute of Child Health and Human Development
OR	Odds ratio
PI	Principal investigator
PPWR	Postpartum weight retention
PRISMA	Prospective, Longitudinal Study of Maternal and Newborn Health of the Pregnancy Risk Stratification Innovation and Measurement Alliance
QNS	Quality norms and standards
RR	Relative risk
SES	Socioeconomic status
SGA	Small-for-gestational age
SRH	Sexual and Reproductive Health and Research
SSD	Standardized Site Difference
TAG-GWG	Technical Advisory Group on Gestational Weight Gain
WHO	World Health Organization

1. INTRODUCTION

Pregnancy is a unique period in the life cycle for implementing interventions to optimize longer-term maternal and child health. Pregnant women's frequent contact with the health care system and their strong motivation to optimize their health make pregnancy a critical period where lifestyle factors, such as weight management, can influence longer-term outcomes for women and their children.

There is a lack of evidence-based public health tools for monitoring gestational weight gain (GWG) that apply to women of all body mass index (BMI) categories and geographic locations. The GWG recommendations used worldwide and endorsed by the World Health Organization (WHO) in the most recent antenatal care (ANC) guidance[1] are based on Institute of Medicine (IOM) guidelines, which were developed primarily using observational studies from high-income countries.

To address this critical gap, WHO is undertaking a normative process to develop global standards for GWG and optimal GWG ranges[2] to reduce the risk of adverse maternal and infant outcomes. This initiative is led by the WHO Departments of Nutrition and Food Safety (NFS) and Sexual and Reproductive Health and Research (SRH), with contributions from the Department of Maternal, Newborn, Child and Adolescent Health and Ageing (MCA), in partnership with a research group comprised of scientists from the Federal University of Rio de Janeiro (Brazil), Cornell University (United States), and the University of British Columbia (Canada). The GWG Steering Committee includes some members of these external institutions and the WHO Secretariat (see **Annex I** for details).

In 2023, WHO established a multidisciplinary Technical Advisory Group on Gestational Weight Gain (TAG-GWG) to advise on the process of developing GWG standards and optimal ranges (see **Annex I**).[3] TAG-GWG advisers provide advice on the development of the detailed research protocol, eligibility criteria for determining a sample that is as prescriptive as data and evidence allow, and methods and approaches for developing the global GWG standards and optimal ranges.

The Third Meeting of the WHO TAG-GWG was convened at WHO Headquarters in Geneva from 12–14 March 2024. The objectives of the meeting were to 1) provide an opportunity for the three TAG-GWG working groups to present updates on their work and 2) to discuss and provide recommendations on key questions proposed by the GWG Steering Committee. This report provides a summary of discussions and recommendations emanating from this meeting. A list of participants and meeting agenda are available in **Annexes I** and **II**.

2. SUMMARY OF DAY 1 PRESENTATIONS AND DISCUSSIONS

Elaine Borghi, WHO/NFS, opened the meeting and thanked the TAG-GWG working groups for their work since the previous meeting. She introduced the meeting chairs, Helena Teede and Suzanne Phelan. All TAG-GWG advisers were introduced and asked to declare any conflicts of interest (COI); none were declared.

Helena Teede presented the meeting agenda (see **Annex II**) and the two meeting objectives.

[1] WHO. 2016. WHO recommendations on antenatal care for a positive pregnancy experience. Available at: https://www.who.int/publications/i/item/9789241549912

[2] For more information about the process, see: https://www.who.int/teams/nutrition-and-food-safety/development-of-global-gestational-weight-gain-standards

[3] For more information on the Technical Advisory Group on Gestational Weight Gain, see: https://www.who.int/groups/technical-advisory-group-on-gestational-weight-gain-(tag-gwg)

Session 1: Overview of the GWG project

Gilberto Kac presented an overview of the rationale, objectives and methods for the development of GWG standards.

There is robust evidence from individual studies, literature reviews and meta-analyses[4] that both insufficient and excessive weight gain during pregnancy are strongly associated with adverse maternal and child health outcomes.[5] Monitoring weight gain during pregnancy and intervening to support healthy weight and weight gain through counselling or other behavioural interventions has the potential to reduce the risk of adverse outcomes, including gestational diabetes mellitus (GDM), emergency caesarean delivery (C-section), macrosomia/large-for-gestational age (LGA) and excess postpartum weight retention (PPWR).[6]

Several existing charts, standards and curves used to monitor GWG have informed the methods and objectives of a new GWG standard. The terminology used to describe the intended outputs of the project was presented to ensure language consistency. A 'standard' refers to the distribution of weight gain based on an ideal/low-risk population, reflecting aspirational weight gain (i.e., prescriptive). In contrast, a 'reference' refers to the distribution of weight gain based on a reference/unselected population (i.e., descriptive). There may be limitations in developing a GWG 'standard' (e.g., due to issues with data availability), so it is unclear how prescriptive the project output will be.

Previous research has identified maternal weight as a concern worldwide, and most countries anchor their GWG target on pre-pregnancy BMI and use the IOM 1990/2009 guidelines as a starting point to inform their national policies.[7] However, several issues with the global adoption of the 2009 IOM guidelines were noted, including that they: 1) were defined for women in the United States; 2) were based on data from high-income countries; 3) lack good quality evidence on several relevant outcomes (e.g., pre-eclampsia); and 4) were never properly tested in low-and-middle income countries. This project addresses the gap in globally relevant GWG standards and related tools and recommendations. The project has been conceptualized in two phases: 1) to develop global GWG standards and 2) to identify ranges on charts associated with the lowest risks of adverse maternal and infant outcomes to inform recommendations on optimal GWG ranges.

The creation of GWG standards would be preferable as it allows for the monitoring of weight gain throughout pregnancy and facilitates intervention as needed. Further, it was noted that GWG standards would allow for the calculation of z-scores, facilitate comparison between individuals and be a useful tool for ANC providers. However, even if a standard is prescriptive, the optimal ranges associated with a lower risk of adverse outcomes must also be defined (i.e., is it 'safe' for women/children to be at a given percentile at a certain gestational age?). In the INTERGROWTH-21st GWG standard, half of the GWG z-scores were associated with an increased risk of excess interpregnancy weight gain. However, no thresholds have been developed from

[4] See: (1) Viswanathan, M., Siega-Riz, A. M., Moos, M. K., Deierlein, A., Mumford, S., Knaack, J., Thieda, P., Lux, L. J., & Lohr, K. N. (2008). Outcomes of maternal weight gain. Evidence report/technology assessment, (168), 1–223; (2) Ohadike, C. O., Cheikh-Ismail, L., Ohuma, E. O., Giuliani, F., Bishop, D., Kac, G., Puglia, F., Maia-Schlüssel, M., Kennedy, S. H., Villar, J., & Hirst, J. E. (2016). Systematic Review of the Methodological Quality of Studies Aimed at Creating Gestational Weight Gain Charts. *Advances in nutrition (Bethesda, Md.)*, *7*(2), 313–322. https://doi.org/10.3945/an.115.010413; and (3) Moore Simas, T. A., Waring, M. E., Sullivan, G. M., Liao, X., Rosal, M. C., Hardy, J. R., & Berry, R. E., Jr (2013). Institute of medicine 2009 gestational weight gain guideline knowledge: survey of obstetrics/gynecology and family medicine residents of the United States. *Birth (Berkeley, Calif.)*, *40*(4), 237–246. https://doi.org/10.1111/birt.12061

[5] Insufficient GWG is associated with fetal growth restriction, low birth weight/small gestational age (SGA), preterm birth, problems with initiating or maintain breastfeeding and neonatal mortality. Excessive GWG is associated with macrosomia/large for gestational age (LGA), postpartum weight retention (PPWR), childhood obesity, hypertensive disorders of pregnancy, gestational diabetes and neonatal mortality

[6] Cantor, A. G., Jungbauer, R. M., McDonagh, M., Blazina, I., Marshall, N. E., Weeks, C., Fu, R., LeBlanc, E. S., & Chou, R. (2021). Counseling and Behavioral Interventions for Healthy Weight and Weight Gain in Pregnancy: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*, *325*(20), 2094–2109. https://doi.org/10.1001/jama.2021.4230

[7] Scott, C., Andersen, C. T., Valdez, N., Mardones, F., Nohr, E. A., Poston, L., Loetscher, K. C., & Abrams, B. (2014). No global consensus: a cross-sectional survey of maternal weight policies. *BMC pregnancy and childbirth*, *14*, 167. https://doi.org/10.1186/1471-2393-14-167

the INTERGROWTH-21st tool so far.[8] Identifying cut-offs is essential in creating a complete, globally relevant GWG monitoring and counselling tool.

Key elements of the project data sources and methods were presented. The project is limited to using secondary data only to develop the underlying sample for analysis. These data are sourced through a variety of channels, including 1) a global call for data, 2) literature reviews and 3) contact with principal investigators (PIs) of relevant studies (see Day 1, Session 5 for more detail). Eligibility criteria for studies/datasets to be accepted as part of the sample have been defined.

There are ongoing discussions within the TAG-GWG working group 1 to define individual-level eligibility criteria (i.e., the criteria for an individual to be part of the underlying sample to construct the standard). Criteria should include GWG determinants and outcomes, as the GWG standard should be based only on individuals with optimal conditions during pregnancy, delivery and the postpartum period. Working group 1 compiled a comprehensive list of GWG determinants and outcomes based on the 2009 IOM guidelines and more recent references[9] (see Day 1, Session 3 for more detail on this work).

The GWG standard will be stratified by pre-pregnancy BMI because it remains the strongest effect modifier of the association between GWG and outcomes. However, there will be discussions regarding the potential for separate charts or thresholds/cut-offs for other characteristics (see Day 2, Session 2 for more detail).

The TAG-GWG working group 2 will evaluate the possibility of using self-reported weight to calculate pre-pregnancy BMI and GWG (see Day 1, Session 4 for more detail). Working group 3 is developing plans to: 1) identify data entry errors and statistical outliers of weight, weight gain and other relevant variables in the underlying sample; 2) identify and evaluate the heterogeneity of weight and weight gain according to gestational age when dealing with pooled data from multiple sources; and 3) define the best model to fit GWG standard (see Day 1 Session 6 for more detail).

Optimal GWG ranges and thresholds for counselling will be defined in phase 2 of the project. The defined optimal ranges and thresholds must consider a range of short- and long-term adverse maternal and child health outcomes linked with excess or inadequate weight gain. A systematic review may be conducted to identify methods to define cut-offs. The developed GWG standard with optimal ranges and cut-offs can be used as a final project output to facilitate GWG monitoring during ANC. In the future, a knowledge translation working group will discuss strategies for disseminating the standard.

Discussion and questions:

No questions or comments were raised following the presentation.

Session 2: Systematic literature review on association of GWG with maternal and infant outcomes

Aya Mousa and Rebecca Goldstein from Monash University presented the initial results from a systematic literature review and meta-analysis on the association of GWG with maternal and infant outcomes.

[8] Hutcheon, J. A., Chapinal, N., Bodnar, L. M., & Lee, L. (2017). The INTERGROWTH-21st gestational weight gain standard and interpregnancy weight increase: A population-based study of successive pregnancies. Obesity (Silver Spring, Md.), 25(6), 1122–1127. https://doi.org/10.1002/oby.21858

[9] See reports on the first and second meetings of the WHO TAG-GWG at: https://www.who.int/publications/m/item/report-on-the-first-meeting-of-the-who-technical-advisory-group-on-gestational-weight-gain-(tag-gwg) and https://www.who.int/publications/m/item/report-on-the-second-meeting-of-the-who-technical-advisory-group-on-gestational-weight-gain-(tag-gwg)

The group of researchers contributing to the systematic review is a multidisciplinary team from both Monash University and other institutions (including Helena Teede, a TAG-GWG adviser).

The research team identified existing individual participant data (IPD) meta-analyses**.** However, many of the analyses are specific to certain geographies, use randomized control trial data or report data in different ways (e.g., timing of measurements and self-reported or imputed weight). These separate analyses have not yet been combined for a 'global' overview of the impact of GWG on adverse outcomes.

The research question underpinning the systematic review was: What is the impact of GWG on immediate and long-term adverse maternal and infant outcomes? The study was registered and the literature search began in November 2023. The review team specifically searched for observational studies that: 1) included a sample size >300 women; 2) were published in any language (118 of the identified studies were not in English); 3) were conducted from 2009 onwards; and 4) were available in full text. In addition, reviewers considered systematic reviews (42 were identified).

The eligibility criteria for the systematic review were presented by population, intervention/exposure and outcomes. The population inclusion criteria were: 1) adult women; and 2) singleton pregnancies. Interventions were excluded. However, criteria for exposure considered how studies categorized GWG and BMI. Studies were required to report total GWG in categorical form as either below, above, or within the reported guideline/recommendation specified in each study. Further, total GWG must have been stratified by one or more BMI categories (for any BMI category system). Criteria for timing required that BMI be either self-reported, measured at preconception or measured before 14 weeks of gestation. Similarly, initial gestational weight must have been either self-reported, measured at preconception or measured before 14 weeks of gestation. The last gestational weight was also required to be measured at the last visit or at time of delivery. All maternal and neonatal outcomes were considered.

A visual was presented of the ideal way that reviewers would want studies to have reported their data, stratified by BMI and GWG with odds ratio (OR), relative risk (RR) or frequency within each group for each particular outcome (see **Figure 1**).

Figure 1: Example ideal report for data, stratified by BMI and GWG

BMI / GWG Stratification	Insufficient GWG (Below Recommendations)	Adequate GWG (Within Recommendations)	Excessive GWG (Above Recommendations)
Underweight (e.g. BMI <18.5 kg/m²)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)
Normal weight (e.g. BMI 18.5-24.9 kg/m²)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)
Overweight (e.g. BMI 25-29.9 kg/m²)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)
Obese (e.g. BMI ≥30 kg/m²)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)	OR (95%CI) RR (95% CI) Frequency, n (%)

More than 20,000 studies were identified in the literature search. After removing duplicates and studies screened out based on the title and abstract, a total of 2,163 studies were eligible for full text review. The vast majority of these (n=2,122) were excluded. Some of the most common reasons for exclusions included: 1) wrong study design; 2) the BMI and GWG stratification were not presented as required; 3) missing or

incomplete data on population age (i.e., it was not possible to determine if adolescents were included in the sample) or inclusion of adolescents in the sample; and 4) GWG was used as an outcome or effect modifier. A total of 119 of the studies were identified as lacking data on population age. The PIs of these 119 studies were contacted for clarification on age of the study sample. Only 17 responses have been received thus far, seven of which have resulted in studies being considered eligible. In total, 49 eligible studies were identified for inclusion in the review (42 from the literature search and seven from contacting study authors).

A snapshot of data from the more than 2,000 studies reviewed was presented to illustrate the availability of some of the determinant and outcome variables considered by working group 1. The most frequently available variables were maternal risk factors and behaviours, including variables related to parity, education, smoking/tobacco use and diet/micronutrients. The most frequently occurring variables related to the environment included area of residence, disasters, heat intolerance, food, toxins and socioeconomic status (SES). The most frequently occurring psychosocial variable was stress. While some of the variables may be available in identified studies, it is unclear if these variables could be harmonized across studies.

The characteristics of the 42 eligible studies were presented (the seven studies added following clarification on population age were not presented). The overall sample from all 42 studies is 1.4 million, with data sourced largely from retrospective studies, maternity clinics or large datasets. The geographic distribution of the studies and sample population was also presented,[10] with the greatest number of studies from China (n=14), but the largest proportion of the sample from studies in North America (48%) and Europe (33%). Large datasets of >50,000 women were identified in China, North America, Germany, Japan and Sweden, while most low-and-middle income countries (e.g., Thailand, the Islamic Republic of Iran, Brazil and Türkiye) had studies with samples <500.

The conceptual framework used to identify inclusion/exclusion criteria by working group 1 (see the Report on the Second Meeting of the WHO TAG-GWG[11]) was presented to illustrate variables within the framework that were either poorly captured, had varying definitions or had weak clinical linkages in the 42 studies identified. Poorly captured variables included media, policies, stress and attitude towards weight gain. Depression and stress had varying definitions, and altitude, environmental toxicants and natural and humane-made disasters had weak clinical linkages. Therefore, the variables used in the review were limited by availability and measurement. The variables that were reasonably available were presented (e.g., age, race/ethnicity, SES, education in years, smoking status, alcohol use, area of residence, employment status, income, pre-pregnancy BMI, pre-pregnancy weight, diabetes, hypertension, etc.). However, even race/ethnicity was only presented in 33% of the studies (and was less likely to be presented in homogeneous populations), education was presented in only 50% of studies and smoking in 43%. The review team tagged variables related to dietary intake, physical activity and sedentary time during preconception and pregnancy in case they need to be used in future analyses.

Variables not assessed due to a weak link to GWG, or that are difficult to measure/unlikely to be available, included hormonal milieu, genetic characteristics, social support, food insecurity, heat influences and unintended pregnancies.

Most (76%) of the studies used the standard WHO BMI categorization while others used Chinese or metropolitan life insurance table categories. Using aggregate data, it is difficult to compare different BMI

[10] The number of eligible studies identified and the proportion of the total sample was s as follows: China (14 studies, 13% of the sample); North America (10 studies, 48% of the sample); Europe (7 studies, 22% of the sample); Other Asian countries (6 studies, 5% of the sample); South America (3 studies, 2% of the sample); and the Middle East (2 studies, 0.1% of the sample).

[11] Available at: https://www.who.int/publications/m/item/report-on-the-second-meeting-of-the-who-technical-advisory-group-on-gestational-weight-gain-(tag-gwg)

categorizations. For GWG categorization, 86% used IOM guidelines while 14% used region-specific (e.g., China and Thailand) or study-specific guidelines.

Selected maternal and fetal outcomes for the review were presented (see **Table 1**). The review team found that several outcomes were defined differently across studies, complicating efforts for a meta-analysis.

Table 1: Outcomes considered in the review

Immediate maternal outcomes	**Immediate fetal outcomes**
• Pre-eclampsia, eclampsia • GDM • Hypertension in pregnancy • Miscarriage • Premature rupture of membranes • C-section (emergency, elective, uncertain) • Induction of labour • Instrumental delivery • Postpartum haemorrhage • Perineal tears • Episiotomy • Maternal mortality	• Preterm birth • Stillbirth / fetal demise • Composite birth trauma • Gender • Birthweight • Small-for-gestational age (SGA) • LGA • Intrauterine growth restriction • Low birthweight, Macrosomia • APGAR score • Neonatal hypoglycaemia • Respiratory distress • Jaundice requiring phototherapy • Special care nursery admission • Neonatal intensive care unit admission
Long-term maternal outcomes	**Long-term offspring outcomes**
• Breastfeeding • Postpartum depression and anxiety • Hypertension • Diabetes • Cardiovascular disease • Obstructive sleep apnea • Overweight and obesity • All-cause mortality	• Tagged

There were several issues presented regarding study presentation of frequencies or ORs stratified by BMI and GWG categories. The review team decided that the best way to look at the studies was by subgroup. Subgroups were defined as follows: 1) studies using WHO BMI categories versus Asian BMI categories; 2) studies grouping women with overweight and obesity together; 3) analysis of studies by obesity group class I, II and III; and 4) low-and-middle income countries versus high-income countries.

Initial results were presented for some of the subgroup categories. For studies using the WHO BMI categories, a summary of the pooled ORs for the association between GWG below and above guidelines with adverse outcomes were presented. For women who had GWG below guidelines for their BMI, it corresponded to a decreased risk for LGA, a higher risk for SGA and a lower risk of C-section. For women who had GWG above guidelines for their BMI, it corresponded to an increased risk for LGA, a lower risk for SGA and a higher risk of C-section. The same analysis will also be conducted for Asian BMI categories and other subgroups/categories at a future time.

These results were then compared with a 2017 systematic review[12] [13] that investigated maternal and infant risks associated with weight gain below or above the 2009 IOM recommendations. The review investigated size outcomes only (e.g., SGA, preterm birth, LGA, macrosomia, C-section and GDM) and included only 23 studies (a total of 1.3 million women, with studies from only the United States, Europe and Asia). At the onset of pregnancy for the women in this sample, nearly 38% of the women had overweight or obesity, and at the end of pregnancy nearly 50% of women had weight gain above guidelines. The summary of the pooled ORs for the association between GWG below guidelines with adverse outcome showed that women with lower weight gain were more likely to develop SGA and preterm birth and women with excess weight gain were more like to have LGA and macrosomia. Key differences between the 2017 systematic review and the current review were summarized (see **Table 2**).

Table 2: Comparison between 2017 and current literature reviews

Search criteria	2017 review	2024 review
Period (in years)	1999–2017	2009–2023
Population	>500	>300
Age limit	Large database studies, accept if does not specify age <18	Limited to age >18 (contacted >110 study authors to clarify age)
Final pregnancy weight	Can be self-reported	Must be measured
GWG guidelines	IOM	Any
Outcomes	3 maternal and 3 neonatal	Multiple maternal and neonatal
Language	English only	Non-English and English
Data	Must provide OR with ref range as normal GWG within each BMI group	Any OR any frequency
Data re-analysis	Yes, for 13 studies	No
No. of studies	23	49

A list of challenges with the current review were summarized as follows: 1) limiting the age to women >18 years of age (100+ studies do not include age information); 2) different categorization of BMI and GWG; and 3) different definitions of outcomes. As a result, there might be a limited number of studies appropriate for meta-analysis. It was noted that an IPD would avoid many of these challenges.

Next steps for the review team are to continue data cleaning and analysis for the other outcomes, extract data for the additional seven studies and do a subgroup analysis.

Discussion and questions:

There was a question on why gender was listed as a fetal outcome of interest for the review (see **Table 1**).

- The presenters noted that the sex of a child is required to calculate SGA and LGA. Gender was noted rather than sex of the child; however, this is not an outcome itself but a variable required to calculate certain outcomes.

There was a comment regarding the majority of the sample being sourced from studies in North America and

[12] Goldstein, R. F., Abell, S. K., Ranasinha, S., Misso, M., Boyle, J. A., Black, M. H., Li, N., Hu, G., Corrado, F., Rode, L., Kim, Y. J., Haugen, M., Song, W. O., Kim, M. H., Bogaerts, A., Devlieger, R., Chung, J. H., & Teede, H. J. (2017). Association of Gestational Weight Gain With Maternal and Infant Outcomes: A Systematic Review and Meta-analysis. JAMA, 317(21), 2207–2225. https://doi.org/10.1001/jama.2017.3635

[13] Goldstein, R. F., Abell, S. K., Ranasinha, S., Misso, M. L., Boyle, J. A., Harrison, C. L., Black, M. H., Li, N., Hu, G., Corrado, F., Hegaard, H., Kim, Y. J., Haugen, M., Song, W. O., Kim, M. H., Bogaerts, A., Devlieger, R., Chung, J. H., & Teede, H. J. (2018). Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC medicine, 16(1), 153. https://doi.org/10.1186/s12916-018-1128-1

Europe (nearly 80%), while low-and-middle income countries had studies with small sample sizes. There was also a question about whether the review had helped improve the global perspective.

- The review team collected 118 studies in non-English languages and tried to gather data from as many contexts as possible. However, many studies from underrepresented contexts were excluded because they did not specify age. Where there is IPD data this will not be an issue and may allow for more representation from low-and-middle income countries.
- An adviser noted that it would be helpful to understand why studies from certain contexts (e.g., Africa) were excluded. This can help the development of the global GWG standards and encourage and inform more research in these contexts.
- Another adviser noted that a meta-analysis can account for small sample size. There are quite a few studies of GWG with a sample size >500 identified from Africa in previous searches. It was suggested that some studies from these contexts may have been excluded from this systematic review because they were trials. There may be a way to include studies and account for the effect of the trial or intervention so that they can contribute to this work.

There was a comment on a difference between the eligibility criteria in this systematic review and those in the GWG standards project. Criteria for the global GWG standard requires repeated measurements of weight gain, rather than total weight gain.

There was a request for clarification on what was meant by 'wrong study design' as a rationale for study exclusion.

- The review team responded that this is a term describing a variety of publication types that were ineligible, including editorials, letters or reviews (i.e., not an observational dataset).

There was a question on why PPWR was not considered in the list of maternal outcomes for the systematic review.

- The reviewers noted that PPWR was meant to be included in the list of outcomes (as a long-term outcome) but was erroneously left out from the list presented. It was an outcome considered and tagged in the analysis. However, the reviewers did not believe that any of the eligible 42 studies included PPWR.

There was a comment on studies that may have reported risk ratios and thus were not included in the systematic review. Mathematically, RR can be converted into ORs, so should these have been restricted?

- The review team will come back to the TAG-GWG with further detail on this at a later date.

There was a request for clarification on the charts used by the team to calculate SGA and LGA.

- If the raw data on birthweight and sex were available, they used country-specific GWG charts to calculate SGA or LGA. Most of the studies for this review, however, defined SGA using their own definitions (noted in the publication).

There was a comment on the exclusion of studies that did not report GWG stratified by pre-pregnancy BMI and whether it was possible to request the raw data from study authors, rather than automatically excluding them.

- It was noted that this will be part of the IPD being considered by the global GWG standard project.
- The systematic review presented here is not a funded part of the GWG standard project. There was not sufficient time in this systematic review to go back to study authors.

There was a request for clarification on the rationale for this systematic review and its relevance to the GWG

standard project.

- The GWG Steering Committee clarified that Monash University had already conducted a systematic review in 2017 that could support the work of working group 1 of the GWG project and offered to update it. The updated version includes some aspects related to the GWG standard with the intention of helping to validate and inform project findings and decisions on eligibility criteria.
- The systematic review is providing valuable insight into what data may (or may not) be available for the GWG standard. It also highlights the deficiencies of the classic approach to systematic reviews and meta-analyses and illustrates there has not yet been a good attempt at global representation in a systemic review on this topic.
- The GWG Steering Committee thanked the review team for their extensive efforts and for enlarging the scope of their review to benefit the GWG standard project and accommodating the project's short timeline.

There was a comment that it is encouraging to see that the systematic review confirms that both too little and too much GWG by various criteria is associated with adverse outcomes. This is a good validation that the GWG project is on right track.

There was a concern regarding use of race/ethnicity data as a maternal factor, given that there is extensive evidence showing that race/ethnicity is a social determinant, and many adverse pregnancy outcomes are actually related to inequity of care or access to care. Further, many studies use self-reported race/ethnicity, which is often inaccurate. The group was encouraged to be careful not to ascribe any differences in outcomes to self-declared variables such as race/ethnicity and in how any related data are presented.

- An adviser noted the difficulty in understanding how these variables are defined and collected across studies.
- It was also noted that there are significant metabolic differences across the world. For example, GDM can occur at a mean BMI of 21-22 in Asian women, at a mean BMI of 23-24 in women from South Asia and at a BMI of 32 in Caucasian women.[14] It is important not to ignore metabolic differences across regions that impact GWG and adverse outcomes.

Session 3: Working group 1 – Individual-level eligibility criteria (updates)

Suzanne Phelan presented the goal, expected outputs and process for working group 1 as well as progress towards outputs and proposed recommendations.

The goal of working group 1 is to: propose individual-level eligibility criteria to be considered when creating the GWG standards, considering both GWG determinants and outcomes. To achieve this goal, the working group is focusing on two key outputs: 1) an ideal initial list of recommended individual-level criteria to define the sample for developing the GWG standards; and 2) a revised list of recommended individual-level criteria that considers the availability of the data in eligible studies. The three processes designed to achieve these outputs were presented.

The first process was to conduct a literature review to identify relevant GWG conceptual models and a comprehensive list of determinants and outcomes. The working group reviewed relevant literature, available

[14] Wan, C.S., Abell, S., Aroni, R., Nankervis, A., Boyle, J. and Teede, H., 2019. Ethnic differences in prevalence, risk factors, and perinatal outcomes of gestational diabetes mellitus: a comparison between immigrant ethnic Chinese women and Australian-born Caucasian women in Australia. *Journal of diabetes*, *11*(10), pp.809-817. DOI: https://doi.org/10.1111/1753-0407.12909

conceptual frameworks and individual eligibility criteria from studies conducted to develop GWG curves published in the last 10 years.[15] Based on their review, the working group decided to use the 2009 IOM conceptual framework as a basis, taking into consideration the latest literature and adding determinants and outcomes. Summaries of the determinants and potential outcomes associated with GWG identified by the working group are presented in **Tables 3** and **4.**

Table 3: Summary of identified determinants of GWG

Systems-level	Individual-level
• Macrosystem: Sociocultural norms, economic, political and media influences • Microsystem: Access to health services, clean water, healthy food and opportunities for physical activity • Geographic: Altitude, natural and human-made disasters and exposure to toxins	• Sociodemographic: Pre-pregnancy BMI, age, race/ethnicity, SES, food insecurity, employment and interpregnancy interval • Behavioural: Sleep, sedentary behaviour, dietary intake/micronutrient status, smoking, alcohol, drugs and unintended pregnancy • Psychosocial: stress/depression/social support/attitudes, eating disorders, perceived discrimination and acculturation • Biological: Genetics/epigenetics, pre-existing comorbidities (hypertension and diabetes), thyroid, autoimmune diseases, human immunodeficiency virus (HIV), malaria, hyperemesis, irregular menstrual cycle and child sex

Table 4: Summary of potential outcomes associated with GWG, to be considered as exclusionary criteria

Maternal outcomes	Child outcomes
• **Prenatal/delivery** o Pregnancy-induced hypertension or pre-eclampsia o Spontaneous loss o Mortality o Emergency caesarean delivery • **Postpartum** o Excessive postpartum weight retention or interpregnancy weight change o Lactation problems o Depression, stress and anxiety o *Longer-term:* Type 2 diabetes or any other metabolic disorders, cardiovascular disorders, metabolic syndrome	• **Neonatal** o Poor appearance, pulse, grimace, activity, and respiration (APGAR) score o Admission to specialty care/intensive care o Birth defects o Preterm birth o Perinatal death (including stillbirth) o Neonatal body composition, SGA, LGA • **Childhood** o Excessive weight gain o Overweight/obesity o Cancer o Asthma, allergies o Neurodevelopment

The working group also considered the impact of including PPWR as an exclusion criterion by reviewing and presenting evidence related to the impact of PPWR on GWG percentiles. Data from the nuMoM2b (2010-2013) cohort in the United States found that excluding people with a >5 kg PPWR shifted the GWG percentiles.[16] However, data from a Swedish cohort (2008-2014) excluding people with >5 kg interpregnancy weight gain did *not* shift GWG percentiles. Given the conflicting evidence, and to ensure that the exclusion criteria result in a population that is globally representative, the working group concluded that eligibility

[15] Hutcheon et al., 2013; Xu et al., 2014; Hutcheon et al., 2015; Cheikh Ismail et al., 2016; Johansson et al., 2016; Santos et al., 2018; Huang et al., 2020; Kac Carrilho et al., 2021; Thiruvengadam et al., 2022

[16] Carrilho TRB, Bodnar LM, Johansson K, Kac G, Hutcheon JA. The impact of cohort inclusion/exclusion criteria on pregnancy weight gain chart percentiles. Journal of Epidemiology and Community Health. 2024 submitted.

criteria should be selected considering population characteristics, the different measurements used across studies, whether it is evidence-based and the generalizability of findings.

The second process was to identify factors to assist in narrowing down the identified determinants and outcomes (e.g., inclusion and exclusion criteria). Key considerations for variable selection were as follows: 1) criteria must be rooted in evidence-based and conceptual models; 2) generalizability to the broader population; 3) clinical significance or severity of an outcome; and 4) actual availability in datasets. Based on these identified factors, the working group adopted a pragmatic approach for identifying inclusion and exclusion criteria that balanced theoretical consideration with practical issues of data availability.

The list of identified variables (see **Tables 3** and **4**) for inclusion and exclusion were grouped into three categories as follows: 1) available (e.g., an existing study-level criteria); 2) likely available in most studies; and 3) likely not measured or available in most studies (due to the complexity/challenges in measurement). Within these three categories, the working group further categorized the variables into priority groups:

1) Mandatory: Study-level criteria, commonly measured, soundly rooted in literature or strong face validity.
2) Conditional: The final inclusion or exclusion of determinants/outcomes will be determined using a data-driven approach to assess whether including or excluding individuals with that determinant or outcome would meaningfully change the weight gain chart's percentiles (considering and examining selection bias).
3) Drop/pending: Limitations in data availability, evidence, generalizability and/or clinical significance.

The working group voted to achieve consensus on variables that are mandatory, conditional (if their inclusion/exclusion meaningfully change the percentiles) and that will be dropped/are pending based on data availability. The recommended inclusion and exclusion criteria for the determinants and outcomes are presented in **Tables 5** and **6.**

Table 5: Recommended determinant variables, by priority categorization

Mandatory inclusion criteria	Conditional inclusion/exclusion criteria	Drop/Pending criteria
• Maternal age >18 years • Singleton pregnancy • All pre-pregnancy BMI categories	• Altitude >2500 metres • Smoking, alcohol and drugs • Pre-existing comorbidities (hypertension and diabetes), thyroid, autoimmune diseases, HIV, malaria, hyperemesis and depression	• Stress/anxiety

Table 6: Recommended conditional exclusion criteria for GWG adverse outcomes

Maternal outcomes	Child outcomes
• GDM, pregnancy-induced hypertension or pre-eclampsia • Spontaneous loss • Delivery complications (e.g., C-section delivery) • Mortality • Excessive PPWR or interpregnancy weight change • Type 2 diabetes, cardiovascular disease, metabolic syndrome, depression	• Preterm birth • SGA/LGA • LBW/macrosomia • Lower APGAR scores • Perinatal death (including stillbirth) • Excessive weight gain during infancy or childhood

Additional determinant[17] and outcomes variables[18] that are pending based on data availability were also presented.

Finally, the third process was to operationalize/define the identified determinants and outcomes. The working group presented initial definitions for the list of identified determinants and conditional maternal and child outcomes/exclusions. Operationalization of maternal height and maternal age, and a decision on whether or not to classify birthweight for gestational age, will not take place until the availability of these variables can be assessed.

The presentation concluded with a list of next steps for working group 1, including: the creation of a plan for conditional variables and the data-driven approach (e.g., how is a 'meaningful' effect defined?), integration of the preliminary results from the systematic review and the continued operationalization/definition of the determinant and outcome variables.

Discussion and questions:

There was a concern raised regarding the use of C-section delivery as an exclusion criterion as it may decrease the sample of women from certain contexts or backgrounds. There are large differences in obstetric practices between geographic contexts, and an increased likelihood of C-section delivery in women in higher BMI categories.

- It was acknowledged by several advisers that an additional challenge of using C-section delivery as an exclusion criterion is the lack of data on the rationale for C-section (e.g., maternal request, emergency delivery, etc.). In many countries (e.g., Bangladesh), the prevalence of C-section is above 60%, with a large proportion based on maternal request. The exclusion of C-section delivery may introduce bias.
- The working group responded that the inclusion/exclusion of C-section delivery will be tested to determine if it shifts the percentile before making a final decision on its use as a criterion.

There was a comment regarding the exclusion of women under 18 years of age in the sample. Adolescent pregnancy is a concern worldwide and adolescents represent a significant proportion of the female population in many countries.

- The working group responded that while having a standard for adolescents is important, it is not the plan for this project.
- It was also noted that there are additional complications with the data required to create a standard for adolescents compared to adults (e.g., changes in maternal height throughout pregnancy while the adolescent is still growing, timing of pregnancy after first period, etc.).

There was a concern regarding the use of variables such as maternal education, income and SES, which are often measured differently across contexts and difficult to standardize.

- The working group responded that only raw, continuous data for these variables has been requested (i.e., not already categorized variables that would be difficult to standardize).

There was a question regarding the use of the APGAR score as an outcome variable, given that there is a

[17] Pending determinant variables included: sociocultural norms, economic, political and media; access to health services, clean water, healthy food and opportunities for physical activity; natural and manmade disasters and exposure to toxins; sleep, sedentary behavior, dietary intake/micronutrient status and unintended pregnancy; attitudes, eating disorders, family violence and acculturation; and genetics/epigenetics and irregular menstrual cycle.

[18] Pending child-related outcome variables included: lactation; admission to specialty care/intensive care; child asthma, allergy and neurodevelopment; shoulder dystocia; NICU admission; neonatal hypoglycemia; neonatal body composition; cancer; and stress and anxiety.

community of clinicians who do not support its use as an assessment tool.

- An adviser responded that while some clinicians may not support its use currently, it has been used clinically for a very long time and should thus be considered as an outcome.

There was a question on whether the conditional variables identified will be used as exclusion criteria or if they will be used for stratification factors.

- The working group responded that adverse outcomes are exclusion criteria. Most of the determinants will also be part of the exclusion criteria (e.g., pre-existing diabetes), but the working group is also considering using some of the determinants as potential stratifying variables (e.g., parity, SES, BMI, etc.).

There was a question regarding the rationale for assessing whether some determinants shift the curve when included/excluded even if the literature already clearly indicates a determinant is linked to adverse outcomes (e.g., smoking).

- The working group responded that while it is clear that smoking is a determinant of weight gain, and LGA and SGA are related to weight gain, it is not clear whether excluding a variable such as smoking from the curve will have a practical impact on the distribution of weight gain. Results from the nuMoM2b (2010-2013) study were shared to illustrate this point. The nuMoM2b researchers looked at several factors in addition to PPWR. The sample began with 3,000 women and the eventual prescriptive cohort had a sample of only 100 women. Even after losing 95% of the sample after applying all possible criteria, the distribution of weight gain in the prescriptive cohort was the same as in the descriptive cohort.
- Working group 1 will test each variable to determine if there are important variables to consider and/or ones that are not strongly associated with weight gain or do not meaningfully change the weight gain distribution. This process will help ensure that the working group can use as much data as possible to construct the standards.
- It was also noted that for variables such as smoking, WHO must consider the message it sends to include smokers in the sample, regardless of its effect on the distribution. Smoking is not a recommended practice for any population.
- An additional point made was the potential that there might not be information on all covariates in all datasets. With testing, the group can focus on variables that effect distribution of the curve and only these have to be in all datasets. Therefore, datasets will not need to be excluded for lack of certain covariate detail.
- PPWR was singled out as a concern due to limited availability in the data. PPWR is often not measured/available in the data, but there is a high chance it may shift the distribution (to be confirmed through testing).

A comment was made regarding the potential to calculate interpregnancy weight gain as a proxy of PPWR, by looking at the difference in pre-pregnancy BMI between a first and second pregnancy.

- The working group responded that it would be happy to use either PPWR or interpregnancy weight change as these are often unavailable in datasets. A sensitivity analysis could also be done to determine if there are any differences between the indicators.

There was a question regarding the definition of a meaningful shift in the curve and whether variables should be excluded one-by-one to assess the presence or absence of a shift or as a group.

- The working group responded that these questions will be addressed during the working group 1 discussion scheduled for Day 2 (see Day 2, Session 3).

There was a question on whether the analysis would investigate deviation from a pooled mean or differences between regions.

- The working group responded that regional differences will be investigated in the homogeneous datasets (i.e., after the heterogeneity assessment is complete). The analysis flow between working groups will need to be considered.
- An adviser noted their preference for analysing deviation from a pooled mean, as investigating differences between regions may introduce bias.

There was a comment on the difference between what is possible statistically and what is sensible clinically. While it may make statistical sense to create different charts for different subgroups, this may create significant challenges for implementation. Further, it is not advisable from a clinical perspective to consider separate charts for pathological conditions.

- It was noted that there is a specific session on Day 2 to discuss the implications of separate charts for different subgroups (see Day 2, Session 2).

There was a concern regarding use of variables such as education, SES as inclusion or exclusion criteria.

- The working group responded that it is aware that not all identified determinants can be used as exclusion criteria (e.g., parity, education, SES, etc.). These variables will be tested separately to see whether or not they shift the curves. A decision will be made on how to consider these variables once those results are available.

There was a question regarding the availability of the identified determinant and outcome variables across available datasets, and whether the variables of interest are largely unavailable in reality.

- It was noted that the working group is aware of and considering this issue. Also, the results from the systematic review will help inform what is widely or rarely reported in datasets (see Day 1, Session 2).

There was a comment on the large number of potential analyses to be conducted. A clear understanding of the amount of work, the workflow and the capacity required to conduct these analyses is needed.

Session 4: Working group 2 – Identification and harmonization of databases (updates)

Nandita Perumal reviewed working group 2 objectives and presented progress towards the objectives since the second TAG-GWG meeting.

Working group 2 objectives are as follows: 1) propose additional search strategies for identifying eligible studies for the GWG project; 2) propose approaches for adjustments that might be required in specific variables (e.g., hierarchy, conversions, plausible ranges, etc.); and 3) provide feedback on harmonization strategies.

A recap of recommendations from the second TAG-GWG meeting was presented. See the report on the second meeting of the TAG-GWG for more detail.[19]

Working group 2 has met twice since the second TAG-GWG meeting in order to: 1) provide recommendations on the methodological approaches to define the hierarchy for the initial weight; 2) discuss methodological approaches for the agreement analysis between measured and self-reported pre-pregnancy/first trimester

[19] Available at: https://www.who.int/publications/m/item/report-on-the-second-meeting-of-the-who-technical-advisory-group-on-gestational-weight-gain-(tag-gwg)

weight and potential corrections for both; and 3) define the hierarchy of methods to calculate gestational age. Progress on each of these topics was summarized.

For recommendations on methodological approaches to define the hierarchy for initial weight,[20] the working group defined the eligibility criteria for study inclusion in the GWG standard. For the pre-pregnancy (initial) weight criteria, the 'ideal' data is a weight measured before or close to (a maximum of three months before) conception, while 'acceptable' data could be either weight measured in the first trimester (up to 13+0 weeks) or self-reported pre-pregnancy weight. These pre-pregnancy weight criteria allow for a more comprehensive/ larger sample of eligible studies. However, problems with using either measured first trimester weight or self-reported pre-pregnancy weight include: 1) there is evidence from studies that maternal education, SES and other background characteristics can influence the quality of the reported weight; and 2) comparing the self-reported pre-pregnancy weight with the weight measured at any time point in the first trimester makes it difficult to differentiate between error and actual weight gained during the first trimester.

To support discussions on the assessment of pre-pregnancy weight, the GWG Steering Committee provided working group 2 with an overview of two relevant studies. A study from Brazil[21] investigated the agreement between self-reported pre-pregnancy weight and measured first trimester weight at various gestational age intervals using both administrative datasets and research studies. The analysis found substantial agreement between self-reported pre-pregnancy weight and weight measured during the first trimester. However, the level of agreement between the two measures decreased slightly when using weight measurements at a later gestational age, potentially due to an influence of GWG in the measured weights. Self-reported pre-pregnancy weight is more readily available but may underestimate actual pre-pregnancy weight and may be affected by characteristics of the measurement procedure. In the second study carried out in the United States and Mexico,[22] authors developed a multiple regression model to predict pre-pregnancy weight from the first trimester measured weight, gestational age, maternal height, maternal age at conception and parity. The analysis found that pre-pregnancy weight predictions from measured weight during pregnancy and other maternal information can be useful especially when self-reported pre-pregnancy weight may not be accurate or is unavailable. However, the study had limited data on actual pre-pregnancy weight and used self-reported last menstrual period (LMP) date for the gestational age calculation, which could have biased the model. Overall, the second study is a successful proof of concept, but is not fully generalizable to the United States and other international contexts because the data used contained too few women with either high or low BMI.

Key discussions points among working group 2 members following presentation of these two studies included:

- A concern that there may be a lot of weight gain in the first trimester, which should be considered as part of the total weight gain during pregnancy.
- The degree of agreement across different studies and contexts was approximately 2 kilograms (kg); but maternal education, wealth or access to scales might be an important factor in the degree of agreement.

[20] Pre-pregnancy weight is needed to calculate both BMI and the cumulative weight gain during pregnancy. The gold standard to determine pre-pregnancy weight is the weight at the conception, however, this is not available in most settings. Therefore, most studies have self-reported pre-pregnancy weight or a measured first trimester weight.

[21] Rangel Bousquet Carrilho, T., M Rasmussen, K., Rodrigues Farias, D., Freitas Costa, N. C., Araújo Batalha, M., E Reichenheim, M., O Ohuma, E., Hutcheon, J. A., Kac, G., & Brazilian Maternal and Child Nutrition Consortium (2020). Agreement between self-reported pre-pregnancy weight and measured first-trimester weight in Brazilian women. *BMC pregnancy and childbirth*, *20*(1), 734. https://doi.org/10.1186/s12884-020-03354-4

[22] Thomas, D. M., Oken, E., Rifas-Shiman, S. L., Téllez-Rojo, M., Just, A., Svensson, K., Deierlein, A. L., Chandler-Laney, P. C., Miller, R. C., McNamara, C., Phelan, S., Yoshitani, S., Butte, N. F., & Redman, L. M. (2019). Do Women Know Their Prepregnancy Weight?. Obesity (Silver Spring, Md.), 27(7), 1161–1167. https://doi.org/10.1002/oby.22502

- The presented studies focused on pre-pregnancy weight, but it should be noted that the measurement quality of final pregnancy weight is also an issue. Conceptually, weight gain is from pre-pregnancy to delivery, so the last weight measurement must be as close to delivery as possible; otherwise, the working group would need to rely on weight gain rate.
- The study eligibility criteria (i.e., studies must have initial weight in addition to at least two weight measures in different trimesters) may help mitigate this issue of not having the last weight measure as close as possible to delivery.

Three questions were posed to working group 2 members:

1. When performing an agreement analysis between self-reported pre-pregnancy weight and measured first trimester weight, what analysis steps and variables should be considered? And what are the criteria from which to consider the agreement good/acceptable?
2. If the agreement between self-reported and measured weight is low or lower than expected, what should be the next step? (e.g., use only measured weights; conduct corrections for pre-pregnancy/first trimester weight, etc.)
3. In the future, depending on the results of the first two questions, what will be the hierarchy to define the initial weight for women with both self-reported pre-pregnancy weight and measured first trimester weight?

Working group 2 agreed to investigate the agreement between measured pre-pregnancy weight and self-reported weight before deciding whether corrections would be needed (notably, there is no currently gold standard for performing such corrections).

For the first question, working group 2 recommended three agreement analyses as follows: 1) measured pre-pregnancy weight (3 months prior to conception) versus measured first trimester weight (in intervals up to 6, 8, 9, 13 weeks);[23] 2) self-reported weight versus measured first trimester weight (in intervals up to 6, 8, 9, 13 weeks);[24] and 3) comparison of women with weights at 6 and 3 months prior to conception.[25] The working group recommended to perform the analysis between self-reported pre-pregnancy weight and measured first trimester weight using Bland & Altman plots and to assess the potential for bias in BMI classification when using self-reported and measured weights. Regarding the criteria to consider a good/ acceptable agreement, the working group recommended that 'good/acceptable' agreement is <±2 (kg) between the two weight measurements and no systematic differences according to pre-pregnancy BMI category and other variables such as income, education and ethnicity (pending the availability of these variables). It was also discussed whether to use a more stringent criteria of ±1 kg differences as 'good/acceptable' agreement and to look at other thresholds based on the analysis between measured pre-pregnancy versus measured first trimester weight (e.g., a data-driven approach).

For the second question regarding possible low agreement between self-reported and measured weight, the working group recommended examining the results of the agreement analyses before making any decisions. This would mean exploring the implications of exclusions and looking at the proportion of different types of weights (i.e., self-reported and measured) to establish the hierarchy that should be considered for BMI and GWG calculation.

The working group also discussed the hierarchy of methods to calculate gestational age. The GWG Steering Committee had proposed to prioritize the use of gestational age estimated by ultrasound performed before

[23] Applied to studies with measured pre-pregnancy weight (e.g., the Women's First trial and Early Nutrition and Immune Development (ENID)).
[24] Understanding the timing of when self-reported weight was obtained is important and needs to be considered in the analysis whenever possible.
[25] Weights measured at 6 months correlate with an intention to become pregnant, which is most relevant in high-income countries.

24 weeks, and to use LMP date confirmed by ultrasound performed before 24 weeks as a second-tier consideration. Working group 2 members agreed with these proposals; however, not all working group members were present for quorum.

Pending issues for working group 2 include: 1) all working group members must express their agreement/disagreement on the hierarchy of gestational age proposed by the GWG Steering Committee (all members must be present for quorum); and 2) discussion is needed to define the lower/upper limits for gestational age at visits and at delivery in the data harmonization process.

Discussion and questions:

There were several questions and comments regarding the identified criteria to assess 'good/acceptable' agreement (<±2 kg between the two weight measurements), and whether or not BMI categories should be considered in the agreement criteria.

- The working group noted that ± 2 kg threshold was based on the two high quality studies that the working group reviewed. The group discussed several different thresholds and considered whether women in different BMI categories would require different thresholds. However, the initial consensus was that ± 2 kg difference would be appropriate for almost all BMI categories.
- An adviser suggested two potential alternative thresholds for consideration: 1) using the confidence interval from the Brazilian study (rather than the mean or median) would suggest a threshold of ± 3 kg, and 2) whether ± 2 kg or ± 3 kg is used, an additional condition that the BMI class must not change. However, a concern was raised regarding the use of the upper-bound as a threshold, as it is dependent on variability.
- A concern was raised regarding weight measured late in the first trimester or for obese women, noting that a 2 (or 3) kg gain is >50% of the weight gain recommended for that trimester.
- The working group acknowledged that a higher threshold is more challenging for higher BMI categories. Therefore, there was general agreement within working group 2 to keep the threshold as conservative as possible, while remaining inclusive. It was noted that understanding if there are studies where the majority of the sample had weight measured early in the first trimester versus late, and if there are similarities or differences in the results between these, would be helpful in these discussions.
- An adviser noted that different cut-offs (2 or 3 kg) can be tested to see how many women experience these levels of difference. Regarding BMI categories, the adviser noted that a 1 or 2 kg gain will only change the BMI category for women who are already on the borderline between categories.
- There was support expressed for the use of Bland & Altman plots to assess agreement, as more hypothesis-driven approaches (e.g., Kappa statistics) are not ideal for agreement analyses.

There was a comment on the accuracy of pre-pregnancy weight and how this might be context specific. The accuracy of these data may be different in different settings (e.g., where access to scales is limited).

There was a question about what can be done when a dataset does not have self-reported pre-pregnancy weight (i.e., only has weight measured in the first trimester).

- An adviser noted that a Harvard project addressed this by constructing a hierarchy based on the results of an agreement analysis. First, measured first trimester weight (up to 8 weeks) was prioritized. If that was not available, self-reported weight was used. Where self-reported data were not available,

weight was extracted from medical records. There is also a paper from the Harvard team that imputed weight at 9 weeks based on other measurements. This method could also be an alternative to impute an early pregnancy weight, but it was suggested to wait for the results of the agreement first. It was noted that the prediction models could be used in this analysis as well.

There was a comment on the proposed hierarchy for gestational age.

- There was a suggestion to include a third tier in the hierarchy to include gestational age estimated between 14 and 24 weeks.
- The working group agreed that gestational age assessed by ultrasound in the first 14 weeks makes sense and subsequently up to 24 weeks where data is available.

A comment was made regarding the potential to use machine learning and other technologies to allow for personalized assessment. In some contexts, tools have been developed that consider the type and timing of weight measurements, a woman's BMI and region of birth and produce personalized recommendations

- The working group acknowledged that it would be useful to see what might be possible with machine learning.
- A question was raised about the machine learning algorithm used to correct pre-pregnancy weight where this has been applied. The adviser noted that pre-pregnancy weight correction was not calculated using machine learning.
- There are papers forthcoming on the development and implementation of machine learning and this personalized approach.

There was a comment noting that the working groups plan to first assess agreement between measured pre-pregnancy weight and measured first trimester weight was ideal. The results of this analysis can help inform or be applied to the use of self-reported weight.

There was another question that there was not time to respond to during the session on the method for assessing agreement and how it will account for factors such as income and ethnicity. How do these variables influence the agreement between measured and pre-pregnancy weight? This will be discussed in detail in the next working group 2 meeting.

There was a comment that decisions should be made based on clinical and conceptual priorities and not simply based on data availability. Decisions should also be well documented and published.

Session 5: Data acquisition (updates)

Giovanna Gatica-Domínguez, WHO/NFS, provided an update on the first global call for GWG data and studies identified in a literature review.

The global call for data was launched on 31 July 2023 and closed on 29 February 2024. A total of 117 submissions were received. Of these, 64 have been deemed ineligible, six are not possible to assess for eligibility, and eight may be eligible pending further investigation. In total, 40 submissions (34% of the total) have been deemed eligible.

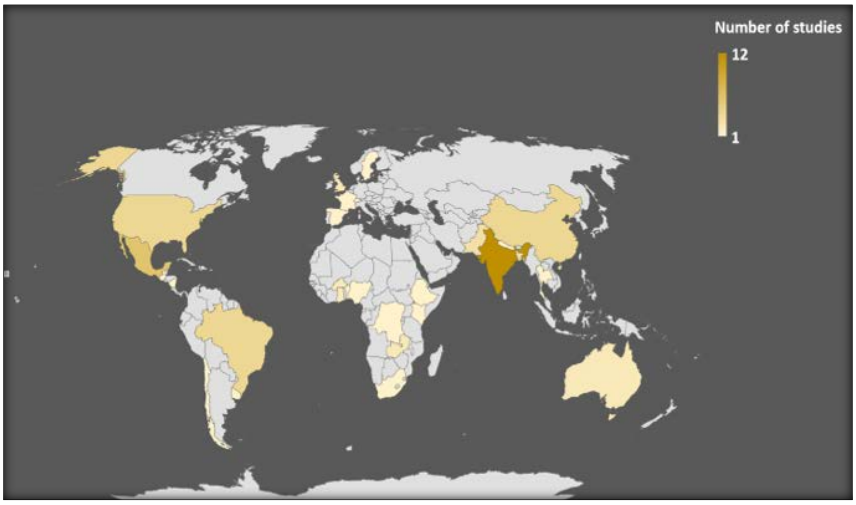
In parallel to the global call for data, a literature review identified 121 studies that may be eligible for consideration. These studies were separated into five priority categories:

- Priority 0 studies were those with a sample size >1,000 and/or were multicentric and deemed likely eligible after review. A total of 18 Priority 0 studies identified. The PIs from these 18 studies were invited to submit data for consideration; however, nine PIs did not reply to the invitation.
- Priority 1 studies were those with a sample size >1,000 and/or were multicentric and for which eligibility was uncertain after review. A total of 42 Priority 1 studies were identified and their PIs were invited to submit data. However, 27 PIs did not reply to the invitation.
- Priority 2 studies were those with a sample size <1,000. A total of 23 Priority 2 studies were identified. All PIs from these studies were invited to submit data; however, 15 did not reply.
- Priority 3 studies were those identified in Synapse. A total of 25 Priority 3 studies were identified. Only four of the PIs of these studies did not reply to the invitation to submit data.
- Priority 4 studies were those for which the PIs had already applied to the global call showing their interest in sharing data if their studies were eligible (a total of 13 studies).

In total, of the 121 studies identified through the literature review, 55 (45% of the total) did not reply to invitations to submit data for consideration. After invitation, a total of 40 submissions were received. Of these, 18 were deemed ineligible, one submission was a duplicate, two were not possible to assess, two are pending assessment and five may be eligible pending further investigation. In total, 12 submissions (30%) from the literature review and invitations have been deemed eligible.

From both the global call for data and the literature review invitations, a total of 52 eligible studies have been identified from 31 countries (representing 15.7% of the 198 countries across the six WHO regions). Several countries have more than one eligible study. The coverage of eligible studies was presented by both geography (WHO region) and country-income classification.[26]

Figure 3: Map of countries in which eligible studies have been identified, and the number of studies identified in each



A Data Sharing Agreement template, which is compatible with WHO's privileges and immunities, was shared with PIs of identified eligible studies in February 2024. Thus far, Data Sharing Agreements for nine studies

[26] The number of eligible studies identified and population coverage of by WHO Region is as follows: Africa (10 countries, 21.7% coverage); the Americas (7 countries, 20.0% coverage); Eastern Mediterranean (2 countries, 8.7% coverage); Europe (5 countries, 9.4% coverage); South-East Asia (4 countries, 36.4% coverage); and Western Pacific (3 countries, 11.1% coverage). The number of eligible studies identified and population coverage by country-income classification is as follows: low-income countries (4 countries, 15.4% coverage); lower-middle-income (11 countries, 20.4% coverage); upper-middle-income (6 countries, 11.3% coverage); and high-income (10 countries, 16.1% coverage).

have already been signed by WHO and the data provider,[27] eight Data Sharing Agreements are currently under review by legal offices[28] and several additional countries are still in the process of reviewing the agreement documents.[29] Links to begin data sharing have been sent to data providers for whom Data Sharing Agreements have already signed by both parties (i.e., WHO and the PI or PI's institution).

Discussion and questions:

Regarding the review of study eligibility, there were questions about whether an analysis had been conducted to identify the main reasons studies have been deemed ineligible, and if the current standards for eligibility are resulting in the exclusion of data from certain population groups (e.g., low-income countries).

- Reasons for data ineligibility were well documented during the review process. An assessment on the main reasons for ineligibility was conducted. One of the primary reasons for exclusion is that the method used to estimate gestational age was not aligned with WHO guidelines (e.g., a lack of ultrasound data confirming gestational age). Other common reasons included that data were exclusive to adolescents, were for those who became pregnant with technology assistance or were representative only of women with pre-eclampsia or other adverse conditions.
- The protocol for determining eligibility will be published.
- Efforts have been made to be as inclusive as possible when assessing data eligibility while adhering to standards of care in line with current global guidance.[30] While it was noted that assessing gestational age with an ultrasound may be challenging in certain contexts, the WHO recommendation on how to estimate gestational age must be followed to ensure the highest quality data is used to develop new WHO standards on GWG.[31]
- Studies that are currently considered eligible may be excluded in later stages if the data quality is found to be below defined standards.
- If there are contexts missing from the sample after data eligibility has been determined, alternatives or adjustments to the review process may need to be considered. However, currently, strict criteria are required to ensure the data used to construct the curves are of the highest quality.

There was a question regarding data acquisition from institutions that have already built and quality assured relevant datasets.

- There are ongoing discussions with several institutions, including the Bill & Melinda Gates Foundation (who maintain the Prospective, Longitudinal Study of Maternal and Newborn Health of the Pregnancy Risk Stratification Innovation and Measurement Alliance (PRISMA) study and Knowledge Integration[32] studies in Synapse Open Source Collaboration Platform), to finalize data sharing agreements.

There was a request for an update on the status of data acquisition and review from fetal growth projects, such as the National Institute of Child Health and Human Development (NICHD), INTERGROWTH and WHO fetal growth charts.

- The NICHD data was submitted and discussions on the data sharing agreements are ongoing.

[27] Spain, Mexico, Brazil (2 studies), Singapore, Nigeria, Ethiopia, Iran and China.
[28] Mexico, United States of America (3 studies), Australia, India, Sweden and France.
[29] Switzerland, Chile, Brazil, United Kingdom, India, United States of America and Uruguay. A data sharing agreement between Bill and Melinda Gates and WHO to access PRISMA data is still under review.
[30] WHO. 2016. WHO recommendations on antenatal care for a positive pregnancy experience. Geneva: World Health Organization. Available at: https://www.who.int/publications/i/item/9789241549912
[31] WHO. 2022. WHO antenatal care recommendations for a positive pregnancy experience. Maternal and fetal assessment update: imaging ultrasound before 24 weeks of pregnancy. Geneva: World Health Organization. Available at: https://www.who.int/publications/i/item/9789240046009
[32] https://www.kiglobalhealth.org/

- A follow-up communication has already been sent regarding the INTERGROWTH data and the WHO Secretariat will follow-up on the WHO fetal growth data.

There was a question about whether data from the control arms of a randomized controlled trial may be considered.

- Several submissions of data from control arms have already been received and are included in the list of 52 eligible studies.
- Any interventions delivered to a control arm must adhere to WHO standards of care. If any interventions were provided to the control arm that are not in line with WHO standard care recommendations, these data will be considered ineligible.

There was a question on whether the sample size from the 52 eligible studies, and the sample size distribution across contexts, had been calculated. Distribution of country representation can vary depending on study sample size, and often the largest studies come from high-income countries.

- An analysis on the sample size and its distribution across countries and contexts has not yet been conducted, but these data will be made available to share at future meetings.
- It was noted that there are lower income countries or underrepresented contexts with multiple eligible studies (such as India) or eligible national-level data (Uruguay).

A question was asked about the inclusion of medical records data in the 52 eligible studies.

- There are eligible data from a few countries that are medical records (e.g., Sweden and Switzerland).
- There are other countries with individual-level records that cannot be shared outside the country (e.g., Ireland). It is not feasible to use these data to construct the curves, however, these data may be considered during the validation stage.

There were several questions related to new data submissions after the close of the global call.

- Any submissions received through the global call after the 29 February deadline will still be considered.
- Additional data can still be submitted for consideration using a separate link, currently used with PIs invited to share data. It is expected that datasets will be received up until September.
- It was advised to finalize all submissions a minimum of six to eight weeks prior to the September deadline to allow discussions of the Data Sharing Agreement to be finalized, as well as review and harmonization of the data.
- A spreadsheet of the list of studies and their PIs who were contacted will be shared with the TAG-GWG. TAG members were requested to contact any of the listed PIs that are their networks that did not respond to the invitation of the GWG Steering Committee or to add studies and contact information for their PIs that are not currently captured in the list.

Session 6: Working group 3 – Methods for the development of GWG Standards (updates)

Eric Ohuma reviewed working group 3 objectives and expected outputs and provided an update on the progress towards outputs and proposed recommendations.

The objective of working group 3 is to: Propose methodological approaches to develop GWG standards. Key intermediate outputs related to this objective include:

1. Data analysis flow for data exclusions and heterogeneity

2. Method selection to flag outliers, considering cross-sectional and/or longitudinal approaches
3. Method selection to assess heterogeneity after pooling data from multiple sources, including the levels to be considered in this assessment (e.g., study, country, region)
4. Method selection to calculate the minimum sample size to model the GWG centiles
5. Selection of statistical approaches to develop GWG standards, including diagnostic and validation procedures to be adopted for each model.

Following in-depth discussions, a set of recommendations for outputs 1 and 3 were presented to the TAG for feedback and endorsement. Regarding the first output on the analysis flow for data exclusions and the heterogeneity analysis, the working group recommended the following:

1. Identify outliers in the following variables: maternal weight, height, weight gain,[33] pre-pregnancy BMI and gestational age
2. Identify gross data entry errors (e.g., 9999, 0) for exclusion in the individual datasets before combination (no other data to be excluded at this stage)
3. Apply other methods for cross-sectional and longitudinal outliers in the individual datasets before any other step
4. Evaluate the heterogeneity of weight gain and other key variables/distribution parameters; then possibly identify and remove outliers in the key variables in the complete dataset.

Regarding the methods to flag outliers, the working group provided separate recommendations for cross-sectional and longitudinal variables. For cross-sectional variables, the recommendations were as follows:

1. Use z-scores of external references, where available (e.g., for pre-pregnancy BMI and height)
2. For gestational age it is recommended to use the biological limit
3. For variables without an external reference (e.g., weight throughout pregnancy and GWG), different statistical approaches will be used (modified jackknife residuals, single model outlier measurement). For each statistical approach, the linear regressions will include a non-linear term for gestational age. The results of each approach will be reviewed prior to final decisions.

For longitudinal variables, the working group recommended the following:

1. Identification of longitudinal outliers (e.g., weight and GWG) using approaches such as mixed-effects models with the extraction of residuals
2. Modification of models to account for the non-linearity of gestational age
3. Addition of other variables or using nested models after initial explorations
4. Testing of other types of residuals (besides the studentized residuals)
5. In those models, individuals with only one measurement will still be considered.

Recommendations for the outliers' assessment included:

1. Only measurements flagged as outliers will be removed, not the whole set of measurements for that individual, in case of multiple measurements
2. Methods that assume that one of the measurements is correct will not be used.

Regarding the second output on the methods to assess heterogeneity after pooling data from multiple sources, the working group discussed how best to prioritize the different methods to assess heterogeneity. These methodological considerations included: 1) reproducibility; 2) flexibility across different variables (longitudinal versus cross-sectional); 3) stringency/liniency of the methods (distribution without following very extreme values); and 4) simplicity. In addition, the working group discussed methods to assess

[33] Weight gain will be calculated after cleaning weights in each dataset, followed by identification of extreme values.

comparability, what gestational age intervals should be used to assess comparability, what the comparison unit should be (e.g., dataset, country, region) and the parameters/characteristics (e.g., means, medians, etc.) to consider in this step and their thresholds. Recommendations for output 2 were as follows:

1. Use the Standardized Site Differences (SSD) (similar to a z-score) and the variance component analysis for assessing heterogeneity in the pooled dataset
2. Consider weight and weight gain by gestational age as the main variables
3. Verify the need to adjust by gestational age (e.g., due to differences in the timing of the measurements in each dataset)
4. Consider strategies to mitigate the sample size contribution to heterogeneity, such as adopting wider gestational age intervals
5. Examine heterogeneity across datasets, countries and regions.

A list of pending decisions relevant to these two outputs were presented, with a dedicated session to discuss them on Day 2 (see Day 2, Session 4):

- Discuss whether it is appropriate to assess outliers again in the pooled dataset after heterogeneity assessment.
- Decide the minimum sample size and number of measurements per individual after data cleaning and removal of outliers.
- Decide how to account for the non-linearity of gestational age and how this will be incorporated into the models (e.g., test splines, fractional polynomials, or transformations of the variable?)
- Discuss the gestational age intervals to be used for assessing comparability (e.g., 14-18, 19-24 weeks, etc.).
- Discuss the parameters/characteristics (e.g., means, medians, etc.) to be used, including their thresholds.

Additional discussion points and questions identified by members of working group 3 for future consideration included:

- The possibility of Latent Profile analysis for GWG trajectory heterogeneity – not considered by the majority of the group as possible
- One group member expressed his concerns about adopting methods that rely on the variance of the measurements
- One group member also mentioned that the cut-offs used to determine heterogeneity across the various sites/regions should be linked to functional consequence or observed alterations in GWG in intervention trials, instead of being based exclusively or primarily on statistical reasoning
- The working group agreed that any decisions should be done a priori, instead of after evaluation of the available data

Discussion on outputs 3 and 4 will take place in future working group 3 meetings.

Discussion and questions:

A question was raised on what will be done if heterogeneity is found across datasets.

- There is a planned working group discussion to address this on Day 2 (see Day 2, Session 4).

There was a request for clarification on whether outlier assessment will be conducted before or after the heterogeneity assessment.

- The working group clarified that there are three processes: 1) first, only gross data entry errors will be excluded; 2) in each dataset, the selected models to flag outliers will be applied; and 3) when the cleaned datasets are pooled, then they will proceed to the evaluation of heterogeneity of weight gain and other variables. The working group still needs to discuss the need to identify and remove outliers again in the complete dataset after heterogeneity assessment.

There was a question on whether the working group had discussed how weight loss during pregnancy may be considered.

- The working group did not specifically discuss considerations for assessing weight loss. However, weight loss will be assessed based on the extent to which it deviated from normal patterns. Any implausible weight loss cases will likely be identified using the longitudinal methods selected to flag outliers.

There was a question regarding the use of SSD and what cut-off would be used to assess heterogeneity in the pooled dataset.

- The working group has not discussed this yet.
- A 0.5 cut-off has been used in other research; however, reflection is needed on what a threshold of 0.5 standard deviations (SD) means clinically (i.e., what are the clinical implications of this threshold?). In addition, the 0.5 SD has been used for length-related variability, while the variability for weight is much higher. Consideration is needed to understand if use of the 0.5 SD would be too conservative for weight. If the GWG Standard is viewed as a screening tool, then it is important to focus on the tail ends of the distributions. The difference between the 3rd and the 10th percentile, for example, is approximately 0.32 SD. In INTERGROWTH-21st, it was proposed that 0.3 SD would make a difference from someone in the 3rd percentile and the 10th percentile. INTERGROWTH-21st used 0.5 SD and considered 0.32 SD to be a stricter criterion for screening. A co-variance analysis showed that intercountry variability for GWG was approximately 0.96 SD, while for child growth (i.e., length) it was 0.35 SD. Thus, there may be three times as much variability for GWG as for length. Discussion is needed to determine then if even a 0.5 SD is too conservative for changes in weight.

More work is needed to determine what change is clinically significant in terms of outcomes.

One group member suggested that the primary concern in these discussions should be the clinical or public health application of the GWG standards. Regardless of the SD used, there a risk of stigma in classifying a woman with inadequate GWG when there are limited interventions available to remedy the GWG during the short pregnancy period.

- The GWG Steering Committee noted the concern and that the working group will focus not only on a global approach but also on context-specific recommendations so that the standard is applicable to every woman in the world.

There was a suggestion to outline a concrete workflow for the TAG considering all remaining tasks to determine whether existing and planned capacity are sufficient.

- One adviser acknowledged the overlap between the outputs across working groups. The working groups are working separately but their decisions overlap. The sequencing of the working group decisions and outputs needs to be determined.
- The GWG Steering Committee stressed that there is enough capacity within the current team and plans for additional support for specific tasks. There is already a call out for a full-time consultancy to support data harmonization.

There was a request for clarification on the timeline for analysis.

- Timelines for analysis will be revisited during the next step discussion on Day 3.

Session 7: WHO Guidelines for Declaration of Interests

Monica Flores-Urrutia, WHO/NFS, revisited the WHO Guidelines for Declarations of Interest and necessary measures to avoid conflicts of interest (COI). See the Report on the First Meeting of the TAG-GWG for more detail.[34]

WHO expects TAG members to represent their personal views only and not seek or accept instructions from any entity or authority (i.e., institutions). The authorship for key publications stemming from the TAG-GWG will be decided by the WHO Secretariat and all publications will be noted as 'for the Technical Advisory Group on Gestational Weight Gain'.

The most relevant items of the *code of conduct* expected from the TAG-GWG were reiterated: disclose all relevant interests and biases, report any material changes to disclosed interest, respect the confidential nature of meetings and decisions and not make any public statements of the work without consent by WHO, not to engage in activities that may bring harm to WHO, to represent views in a personal and individual capacity and actively and fully participate in discussions and deliberations.

Aspects related to the COI were also discussed. First, it was emphasized that TAG-GWG members should update their Declaration of Conflict of Interest before every meeting and that they are responsible for continuously disclosing any changes in their status of COI. Different types of potential COIs were described (i.e., direct, of others, bias, unfair of competitive advantage and tobacco) for member consideration.[35]

Second, TAG members were informed about measures that would be taken in the event of a COI, such as:

- Conditional participation (minor COI): an expert continues in the TAG-GWG, but the COI is disclosed at the start of the meeting and in the meeting report.
- Partial exclusion (moderate COI): expert involvement is limited by a) excluding the expert from that portion of the meeting of work where the conflict was identified, and b) excluding the expert from participating in the decision-making process.
- Total exclusion (significant COI): expert is excluded from the meeting or work altogether where the nature of the COI is too significant vis-à-vis the overall objective or were limiting the expert's involvement to only a portion of the meeting or work is not feasible.

Regarding publications, it was stressed that TAG-GWG members should not express opinions on behalf of WHO. If a TAG-GWG member is "obliged", however, to include a disclaimer that they have a relationship with WHO, the disclaimer should state that the publication does not reflect the views of WHO or any TAG group to which the member belongs. TAG members should not use any materials or ideas from the TAG-GWG in any publication without prior disclosure to, and agreement from, WHO.

Finally, the TAG-GWG working group modality was reviewed: (1) the quorum shall normally be two thirds of the members (TAG-GWG meeting and working groups); and (2) TAG-GWG membership will be terminated in the event of any of the following situations:

[34] Available at: https://www.who.int/publications/m/item/report-on-the-first-meeting-of-the-who-technical-advisory-group-on-gestational-weight-gain-(tag-gwg)

[35] The threshold for what is considered a significant financial interest is US$ 5,000.

- The expert fails to attend two consecutive meetings
- There is a change in the expert's affiliation that results in a COI
- There is a lack of professionalism (breach of confidentiality)

Discussion and questions

There was a comment regarding a perceived COI for an editorial written by an adviser. The adviser disclosed their role in the TAG-GWG when writing the editorial and the journal used its own language to disclose this, which implied that the adviser had a relationship with WHO.

- It was suggested that TAG-GWG members used a standard disclosure statement that their publication is 'not reflective of the position of WHO'. The creation of a standard, generic statement to use for any publications related to GWG will avoid any perception of COI and to prevent journals from misrepresenting the TAG-GWG membership as a relationship to WHO/that the content is attributable to WHO.
- It is difficult to qualify when and how work with the TAG-GWG may influence publications written by TAG-GWG members. Membership to the TAG-GWG is also publicly available. It was suggested to take a conservative stance and default to disclosing membership to the TAG-GWG and adding the standard disclaimer when writing publications related to GWG.
- TAG-GWG members were advised to contact the GWG Steering Committee with any questions or doubts on situations where such a disclosure is required. Specific questions can also be relayed through the GWG Steering Committee to the WHO Office of Compliance for further clarification and input. It was noted that there will be a quick turnaround time for a response from WHO.
- TAG-GWG members must not publish ideas or content taken from the TAG-GWG.

3. SUMMARY OF DAY 2 PRESENTATIONS AND DISCUSSIONS

Day 2 sessions were chaired by Suzanne Phelan. A summary of presentations and key discussion points from Day 1 was provided.

A request was made for clarification on the differentiation between phase 1 (i.e., the development of the standard) and phase 2 (i.e., the validation of the standard).

- The GWG Steering Committee clarified that a two-phase approach is needed in part due to potential issues with availability of outcome data in the underlying sample. In phase 1, the GWG standards will be created using an underlying sample that is free from adverse outcomes. In phase 2, outcome data will be used to validate the GWG standards and define thresholds that may put women at risk of adverse pregnancy outcomes in different contexts. The thresholds will be defined using different study data than will be used to generate the standard.
- It was noted that it is now timely to start planning for phase 2.
- The GWG Steering Committee reported that a new working group (working group 4) will be formed later in 2024 to begin conceptualizing the methods for phase 2.

A question was raised regarding the timing of publication of the project output/product. Will the GWG standards developed in phase 1 (without optimal ranges) be published first, and a revised/amended version

with optimal ranges published later after the completion of phase 2?

- The GWG Steering Committee decided that only one product will be released which includes both the GWG standards and the optimal ranges (i.e., the products from phase 1 and phase 2 together).

Session 1: Guide for the working group discussions

The TAG-GWG members were split into two groups for discussions related to working groups 1 and 3. Working group 2 members were asked to join either group. Discussion points for each group were shared in advance and each group was given one hour for discussion (see **Annex III**). Results from each working group discussion are summarized in Day 2 Sessions 3 and 4.

Session 2: Considerations around separate charts for subgroups with different weight gain patterns

Jennifer Hutcheon reviewed the scientific rationale for creating subgroup-specific charts and presented practical considerations for their production and implementation.

There is precedent for subgroup-specific charts: there are separate child growth charts for boys and girls and separate charts for pregnancy weight gain by different pre-pregnancy BMI categories. One key consideration for the development of a global GWG standard is whether separate charts will be needed for characteristics such as race/ethnicity and/or different world regions.

The scientific rationale for separate subgroup charts is effect measure modification (i.e., interaction). Effect modification in this context means that the association between weight gain and health risks may be different across different population subgroups (i.e., the optimal range of pregnancy weight gain differs between subgroups). When the same amount of weight gain leads to different risks of adverse health outcomes in different subgroups, separate charts may be recommended. Effect modification by pre-pregnancy BMI is already well established and is the reason why there are separate charts by BMI category.

However, separate curves/charts may not be required in all cases where patterns/distributions of weight gain differ between subgroups. Separate charts should only be considered if the optimal ranges of pregnancy weight gain are meaningfully different between subgroups. In other words, the key consideration is the association between weight gain and outcomes, not the distribution of weight gain. To illustrate this point, results were presented from a recent analysis that used Stockholm-Gotland perinatal cohort data[36] to determine whether separate GWG charts are needed for different obesity classes. The analysis found that while the distribution of pregnancy weight gain differs between individuals with class 1 and class 2 obesity, the patterns of risk and optimal ranges were similar between the two classes. Therefore, the example presented suggests that separate GWG curves/charts for class 1 versus class 2 obesity may not be warranted.

Practical considerations for the implementation of separate subgroup charts were also presented. Tools developed for care providers are most effective if they are simple (e.g., a one-page tool). Separate charts may increase the complexity of a tool and thus be more challenging to implement. To illustrate this point, the experience implementing subgroup-specific fetal growth charts in the United States was presented: In 2015, the NICHD published race-specific fetal growth charts. However, by 2020, the Society for Maternal-Fetal

[36] Johansson, K., Bodnar, L. M., Stephansson, O., Abrams, B., & Hutcheon, J. A. (2024). Safety of low weight gain or weight loss in pregnancies with class 1, 2, and 3 obesity: a population-based cohort study. Lancet (London, England), 403(10435), 1472–1481. https://doi.org/10.1016/S0140-6736(24)00255-1

Medicine[37] reported that did not endorse the use of separate, race-specific charts and instead recommended using a single population-based chart. As a result, the NICHD group re-published their fetal growth chart as a single chart applicable to all race groups in 2022.[38] When determining the need for separate subgroup curves/charts, the differences in optimal ranges should be clinically important enough to outweigh challenges to implementation of multiple charts.

An alternative approach to addressing effect modification was also presented. Rather than separate charts for each subgroup, a single chart with different thresholds to classify risks for different groups is possible. For example, the WHO Fetal Growth Chart Group published a single standard but recommended that for the purposes of clinical testing, different cut-offs (i.e., percentile thresholds) for different subgroups may be warranted.[39] This is also an approach endorsed by the Federation of Obstetricians and Gynaecologists in their position statement on which fetal growth charts to use, noting that international standards for growth may be used with locally appropriate thresholds for risk interpretation.[40] There is therefore precedent for having a single global curve/chart with subgroup-specific percentile thresholds/cut-offs to define risk where there are differences between optimal ranges between subgroups.

Finally, a brief pro and con list for single versus separate charts was presented: Single charts both simplify implementation of the tool and facilitate comparisons between region, whereas separate charts allow for more accurate classification of risk in a given population and potentially facilitate better local uptake of the tool.

Discussions and questions:

There was some discussion related to the impact of different BMI categorizations applied to women in the underlying sample used to create the GWG standard. Some world regions have separate BMI categorizations (e.g., Asia) and applying the IOM BMI categories to women in these regions may result in a misclassification of the adequacy of their GWG and risk for adverse outcomes. How will the impact of any differences in BMI categorization on outcomes and clinical impact be considered when determining whether a single or multiple curves/charts is required?

- The presenter responded that the impact of different BMI categorization systems should come out during the investigation of optimal ranges by race within a certain BMI subgroup. If there is effect modification or different optimal ranges within normal weight women in Asia compared to other world regions, depending on the magnitude of the difference, the best solution to account for any difference will be explored.
- There were ideas proposed for statistical approaches to account for differences in BMI categorization, including the use of a mixed effect model and investigating the amount of variance due to risk,

[37] Society for Maternal-Fetal Medicine (SMFM). Electronic address: pubs@smfm.org, Martins, J. G., Biggio, J. R., & Abuhamad, A. (2020). Society for Maternal-Fetal Medicine Consult Series #52: Diagnosis and management of fetal growth restriction: (Replaces Clinical Guideline Number 3, April 2012). *American journal of obstetrics and gynecology*, *223*(4), B2–B17. https://doi.org/10.1016/j.ajog.2020.05.010

[38] Grantz, K. L., Grewal, J., Kim, S., Grobman, W. A., Newman, R. B., Owen, J., Sciscione, A., Skupski, D., Chien, E. K., Wing, D. A., Wapner, R. J., Ranzini, A. C., Nageotte, M. P., Craigo, S., Hinkle, S. N., D'Alton, M. E., He, D., Tekola-Ayele, F., Hediger, M. L., Buck Louis, G. M., … Albert, P. S. (2022). Unified standard for fetal growth: the Eunice Kennedy Shriver National Institute of Child Health and Human Development Fetal Growth Studies. *American journal of obstetrics and gynecology*, *226*(4), 576–587.e2. https://doi.org/10.1016/j.ajog.2021.12.006

[39] Kiserud, T., Benachi, A., Hecher, K., Perez, R. G., Carvalho, J., Piaggio, G., & Platt, L. D. (2018). The World Health Organization fetal growth charts: concept, findings, interpretation, and application. *American journal of obstetrics and gynecology*, *218*(2S), S619–S629. https://doi.org/10.1016/j.ajog.2017.12.010

[40] Visser, G. H. A., Nicholson, W. K., Barnea, E. R., Ramasauskaite, D., Nassar, A. H., & FIGO Safe Motherhood, Newborn Health Committee (2021). FIGO position paper on reference charts for fetal growth and size at birth: Which one to use?. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 152(2), 148–151. https://doi.org/10.1002/ijgo.13500

stratified by BMI classification. The analysis could investigate heterogeneity between the different BMI classification.

A question was raised regarding the public health implications of a global GWG standard and the risk of stigma following categorization as either 'healthy' or 'unhealthy' GWG between countries/contexts.

- The GWG Steering Committee noted that while the concern is understood, it is important to have comparability across countries to trigger action where GWG is poor. However, careful consideration is needed to determine what indicator, based on the global GWG standard, could be used to hold countries accountable for action towards healthy women and pregnancy outcomes. An appropriate indicator will be discussed and defined at a later time.
- The GWG Steering Committee noted that there will be no declarations or statements made regarding the public health use of the GWG standard.

There were requests for clarification on (1) the difference between one global chart with different cut-offs and several different reference charts and (2) who would be responsible for the development of any context-specific thresholds.

- A global chart describes how, on average, women in different parts of the world gain weight during pregnancy and is critical for global monitoring (i.e., to facilitate international comparisons) and research purposes.
- One adviser noted that the simplicity of implementation with a single global standard should not come at the cost of harm (i.e., stigma).
- WHO does not require countries to use the global standards for national implementation. During the development of national plans or interventions to promote appropriate GWG, a country may choose to use context-specific rather than global thresholds.
- Regarding the development of context-specific cut-offs, experience was shared from the implementation of the fetal growth charts. To determine whether or not they would apply the global chart to their context, many countries validated the global chart using their own data to see how well it performed against their own population.
- The GWG Steering Committee stressed that any decision to generate context-specific cut-offs or different charts will be based on the evidence available. If there is sufficient evidence for certain contexts, it may be possible for the TAG-GWG to generate subgroup-specific charts or cut-offs. Wherever this is not possible, limitations will be documented and explained.
- The GWG Steering Committee noted that there is no plan currently for the operationalization of developing context-specific cut-offs. While the GWG Standards Project may focus on investigating certain subgroups, certain countries or contexts may take on this task as needed.

An adviser noted that there may be criticism of the use of differences in adverse outcomes instead of the distribution of weight gain when defining cut-offs in the GWG standard, in part due to the limited availability of data for all adverse outcomes of inadequate GWG.

- There was agreement among advisers that the creation of any charts by subgroup should be dependent on risk of adverse outcomes rather than the distribution of weight gain during pregnancy.
- Despite the potential issues with outcome data availability, the project strives to develop a tool that, to the largest extent possible, addresses issues found across countries worldwide. All global standards and guidelines have limitations, and there will be transparency regarding any limitations in the development of a global GWG standard.

There was a concern raised about not being able to deliver a global standard relying solely on secondary data

and if the quality of the data, once available, is deemed suboptimal.

- It is not required that data from all countries be available to generate a global standard. Rather, there must be sufficient data from contexts that are relevant to the objective.
- While it is certain there will be challenges and limitations in the generation of a global standard, the project has the potential to create tools that are better than what is currently available.
- The GWG Steering Committee and advisers stressed that any and all outputs will reflect only what the evidence shows.
- While the current aim of the project is a global standard, if there is insufficient data available to generate a universal standard there are still different possible outputs that will contribute to improving the standard of care for women (e.g., references, etc.).

Session 3: Working group 1 presentation

Suzanne Phelan presented a summary of the working group 1 discussion (see **Annex III** for a list of discussion questions).

Regarding the identification of exclusionary criteria, the working group discussed identifying a list of mandatory variables inclusive of important adverse determinants or consequences of GWG (e.g., smoking, preterm birth, etc.). The mandatory variable list presented earlier by working group 1 (see Day 1 Session 3) will be revisited and potentially expanded.

Regarding the identification of non-mandatory (i.e., conditional) variables as exclusion criteria, the working group agreed that a data-driven approach will be employed after the data has been cleaned and harmonized. The process will take place as follows: 1) check the data availability, ensuring that the variable of interest is available in either 10 or 20% of low-and-middle income countries data to ensure there is no bias in the exclusion criteria that may disadvantage low-and-middle income countries; 2) if data availability is acceptable, the group will assess data quality of the measured variable; and 3) on a case-by-case basis, each variable will be investigated to determine whether using or not as an exclusion criteria shifts the percentiles in a meaningful way at a given gestational age, noting that what is defined as a meaningful difference may differ by variable.

The definition of 'meaningful' difference was discussed, but there was no clear agreement. Potential 'meaningful' definitions discussed included a 1 kg shift in BMI or weight trajectory, shifts in percentiles at 50/85/97% or if percentiles were shifted by 5 or 10%. There was no consensus within the group on whether it was necessary to have different definitions of 'meaningful' for different variables or just one metric for all variables.

Regarding how weight gain will be considered in these analyses, the working group discussed using all available data to look at weight gain trajectory within preconception BMI category and changes in percentiles over time (not tied to trimester).

The group did not discuss all question prompts (e.g., sensitivity analysis to be performed, etc.) as it is challenging to make certain decisions without first understanding what data is available.

Discussions and questions:

There was concern raised regarding not defining 'meaningful difference' in advance of any data review or analysis.

- After discussion, it was understood that the working group discussions were based on a misunderstanding of previous relevant research efforts to define meaningful difference in GWG. The previous research did not use a data-driven approach to define 'meaningful difference'. Instead, they assumed meaningful would be a 1 kg change based on clinical significance (i.e., it is expected that weight gain may shift normally within the range of 1 kg per day). The working group had framed their discussion around clinically significant differences in relation to the outcome, which is a different approach.
- A point of clarification was made that meaningful difference should not be based on the clinical significance in relation to the outcome, but rather the point at which we declare that a difference in weight gain matters from a clinical perspective (i.e., what amount of weight gain is different enough that it matters from a health perspective?).
- One TAG-GWG member commented on the potential use of *content knowledge* on the clinical importance of the difference in weight to decide what a meaningful difference is and what is important to women (e.g., asking women how much weight change is meaningful to them, or how much weight gain or loss results in a change of clothing size?).
- There was a concern raised about not relating clinically significant change to outcomes as its relation may change by outcome (i.e., a 1 kg gain might increase risk of GDM, but not other outcomes). However, an adviser noted that this concern is related to an ideal of weight examination before a diagnosis of an adverse outcome and the timing of diagnosis of adverse outcomes will not exist in the data.
- There was consensus that the definition of meaningful difference should be done a priori, and if there are any changes required once the data is available then the rationale for these changes will be well documented.

After there was clearer understanding of how to define meaningful difference, there was discussion on use of absolute or relative values.

- There were objections raised about the use of an absolute value (e.g., 1 kg) as its clinical significance may vary between trimesters. In addition, a 1 kg change may have different significance at different percentiles. Unless it is targeted to a specific end of the distribution, there may be issues with its application.
- An adviser noted that the meaningful difference should consider pre-pregnancy BMI because the difference will depend on BMI category.
- There was question regarding whether the ± 2 kg figure used by working group 2 to assess agreement should also be considered for a definition of meaningful difference, but it was noted that these two analyses have different objectives and there does not need to be consistency between them.
- There was a general consensus that relative change (e.g., 5 or 10%) should be considered for the definition of meaningful difference, and that this relative change should be anchored in either percentiles or z-scores. There is no agreement yet, however, on what this relative change threshold will be.
- An adviser noted that there must be consistency in the definition of meaningful difference across all variables.
- A request was made for more discussion on this topic, in a structured way, before a decision is made on a definition of meaningful difference.

There was a concern raised about consistency of vocabulary used across working groups and workstreams**.**

- Regarding mandatory variables, there was confusion in the use of the word mandatory to describe 'core' variables. Working group 1 has grouped mandatory, conditional, etc. variables into 'core' variables groups (groups 1, 2 and 3). The list called 'mandatory' in this discussion are really the 'core' variables.
- It was suggested that vocabulary be clarified across each step in the processes for each working group to avoid confusion.

A comment was made regarding the identification of core variables.

- It was noted that the first step must be to look at what data is available, and where. If there is a vote to select core variables before this, selection bias may be introduced. After data is available, the selection of core variables can take place and the working group can look at the distribution to see what does or does not change.

Session 4: Working group 3 presentation

Tafsir Hasan presented a summary of the working group 3 discussion (see **Annex III** for a list of discussion questions).

Regarding outliers' assessment, the first issue discussed was whether the group considers it necessary to re-assess outliers in the pooled dataset, after the heterogeneity assessment. The consensus from the group discussion was that no re-assessment/detection of outliers should take place once the heterogeneity assessment is complete. The outliers' assessment process will take place as follows:

- Step 0: Application of study-level eligibility (inclusion/exclusion) criteria
- Step 1: Exclusion of gross data entry errors
- Step 2: Outliers' detection based on the methods defined by the working group 3 for cross-sectional and longitudinal variables
- Step 3: Heterogeneity assessment.

Once the pooling is complete, if there are significant differences among the key variables found, this is an issue of heterogeneity and not outliers.

The group also discussed whether the minimum study sample eligibility criteria (n=200) would be applied after flagged outliers are removed (i.e., after outliers are removed, the study sample is now <200). The consensus of the group was that datasets would not be excluded after outlier removal unless 30%[41] or more observations are flagged as outliers.

Regarding how the non-linearity of gestational age will be incorporated into the models of outliers (in both cross-sectional and longitudinal models), the group consensus was to use fractional polynomials to avoid subjectivity[42]. The group also agreed to try to account for skewness in these models.

Regarding modelling of GWG trajectory and which distribution will be assumed, there was consensus that the group will not assume normal distribution (assume skewness) and the modelled trajectory should not follow the extreme values of distribution of the data. Instead the trajectory should follow predicted values restricted to within ± 5 SD of residuals. This will apply both to cross-sectional and longitudinal outliers.

[41] The 30% figure is based on guidance for when one might consider conducting multiple imputations.
[42] When using cubic splines you need to define the number of knots, which can be subjective.

The group did not have enough time to discuss all the heterogeneity assessment questions. However, there was some discussion on the use of SSD and the variance component and the need for sensitivity analysis to look at the trajectories of weight gain and how the removal of particular studies that are considered heterogeneous may change those trajectories. Working group 3 will discuss the remaining questions in a future working group meeting.

Discussions and questions:

There was a request for clarification on how gestational age is non-linear.

- It was noted that it refers to the variability of weight as a function of gestational age.

There was a question about how identified heterogeneity will be addressed.

- This was a question in the list of discussion prompts but working group 3 has not yet had time to discuss what will be done in case of heterogeneity (i.e., if these will be excluded, if a sensitivity analysis will be conducted, etc.). This will be discussed at future working group 3 meetings.

There was a request for clarification on when and for what the 30% threshold proposed by the working group is used.

- One of the study-level eligibility criteria is a minimum sample size of 200. Outlier detection will be conducted for all eligible studies and any identified outliers will be excluded. The 30% is in reference to 30% of the measurements of a dataset being excluded after identification of outliers. Such a high percentage of measurements flagged as outliers would indicate a data quality problem for that particular dataset.

There was a question about whether multiple imputation would be conducted once there is a pooled dataset to account for missing data.

- The working group responded that no multiple imputation will be conducted. The creation of the standard only considers gestational age and weight. There will be no missing weight data within the pooled datasets. While the outcome data may be missing, it is not possible to impute outcomes.

There was a question about the definition of biological implausibility and when this plausibility will be assessed.

- The working group responded that their discussion was focused on outliers' assessment. In step 2 of the outliers' assessment, gross data entry errors (e.g., values of 0 or 999 or a height of 4 metres) will be excluded. The group recommended to use of biologically implausible values for gestational age (e.g., 320 days is not a plausible gestational age).

It was requested that a flow chart be created to illustrate the outliers' assessment process.

- It was agreed that a flow chart or depiction of the process can be developed.

Session 5: *Modus operandi* of working groups – way forward

Giovanna Gatica-Domínguez, WHO/NFS, presented an update on planned future meetings for each working group.

As of the third TAG-GWG meeting, no meetings for working groups 1 and 2 were scheduled. However, based on discussions during Day 1 and 2, at least one meeting will be scheduled for working group 1 to continue

discussions.

For meeting timing, working group 1 agreed that any future meeting would take place on the first Thursday of a given month and working group 2 agreed that any future meetings would occur on the second Thursday of a given month. For any future meetings, working groups will be contacted and notified in advance of required meetings.

Working group 3 has two meetings scheduled across March and April and will continue meeting monthly as needed. If additional meetings are required, the working group will be notified in advance.

Discussions and questions:

It was requested that a meeting be scheduled for working group 2 to continue discussions initiated during the third TAG-GWG meeting.

- The GWG Steering Committee will share a meeting invite for working group 2.

There was a request that a third meeting be scheduled for working group 3 to continue discussions on heterogeneity.

- The GWG Steering Committee will share an invite for a third meeting in May.

There was a comment that any TAG-GWG member can join working groups 2 and 3.

- The GWG Steering Committee agreed that members can join additional working groups. Members should notify of GWG Steering Committee of a request for additional working group membership.
- It was requested that a TAG-GWG member from working group 2 with a background in statistical methods consider joining working group 3. Discussions on working group membership will continue after the third TAG-GWG meeting.

Day 2 closing remarks

Elaine Borghi, WHO/NFS, introduced Mariana Arruda and Victor Keller, both members of the GWG Steering Committee, that had connected virtually. Both members have been involved in the review of studies to assess their eligibility and will contribute to the data harmonization.

4. SUMMARY OF DAY 3 PRESENTATIONS AND DISCUSSIONS

Day 3 sessions were chaired by Suzanne Phelan. A summary of presentations and key discussion points from Day 2 was provided.

Session 1: How to call the GWG centile curves, which will be developed based on “semi-prescriptive” sample?

Elaine Borghi, WHO/NFS, presented considerations for the label/title and goals of the final project product.

As discussed in earlier meeting sessions, issues with data availability may prevent the development of a completely prescriptive sample (i.e., a sample that covers all possible outcomes for maternal and child health). However, a variety of rationale were presented for why the final project product may still be referred

to as a standard: 1) the underlying sample will be semi-prescriptive, with significant effort undertaken to define comprehensive eligibility criteria; 2) the sample will be based on multi-country, cross-contextual data (however, it is not yet clear if global coverage is possible); 3) it will be recommended by WHO, which classifies a product with centiles, cut-offs for risk and recommendations as a standard; and 4) it will describe weight gain patterns based on a sample of women who did not have adverse outcomes (not all adverse outcomes can be accounted for, but those that can be will be declared). Rationale for why the product may not be a labelled as a 'standard' were also presented: 1) data may not be available to apply all the eligibility criteria needed to derive a completely prescriptive sample; and 2) all women composing the sample may not be classified as healthy (e.g., women who are too thin or obese may be included in the sample).

The two goals of the products from this project were presented:

1. To provide tools to guide the assessment of weight gain patterns during pregnancy
2. To provide tools to identify pregnant women who are at risk of adverse pregnancy outcomes due to suboptimal weight.

A list of potential labels/titles for the product were presented for consideration: 1) WHO Gestational Weight Gain Standards; 2) WHO Reference Charts for Gestational Weight Gain; 3) WHO Reference Charts for Pregnancy Weight Gain; 4) WHO Pregnancy Weight Gain Charts; and 5) WHO Gestational Weight Gain Charts.

Careful consideration is needed in the choice of 'standard', 'reference' and 'chart' in the name of the final product as these may have different meanings for different product types and groups. The use of 'pregnancy' rather than gestational was also included as a consideration as pregnancy is less confusing for non-clinicians (e.g., women using the tool themselves)

Discussions and questions:

There were several comments on the wording of the product goals. There was a comment regarding the use of "due to" in goal 2. The phrase "due to" implies a causal linkage between adverse outcomes and 'too little' or 'too much' weight gain.

- There were suggestions to change 'due to' to either 'related to' or 'associated with'.

There was a concern regarding use of the phrase 'suboptimal weight' in goal 2. It was suggested that 'suboptimal' implies 'too little' weight gain, when the product is focused both on too little and too much weight gain. In addition, WHO is moving away from using the words 'optimum' or 'optimal'.

- Alternative wording was suggested by several advisers:
    - 'Variations in weight gain'
    - 'Weight gain outside the healthy range'
    - 'Inadequate/adequate weight gain'
    - 'Insufficient weight gain'
    - 'Inadequate or excessive weight gain'
    - 'Healthy or unhealthy weight gain'
    - 'Insufficient or excessive weight gain'

There was concern regarding the use of "assessment of weight gain" in goal 1.

- There was a suggestion to replace "assessment" with "monitoring" as it better reflects the intended use of the product (to monitor weigh gain through pregnancy during ANC visits). Another suggestion was to omit the words "the assessment of".

There was a comment on the repetition of "to provide tools to" in both goals. It was suggested to use "to provide tools to" only once as an introduction to the two specific goals.

There was a request to user test whatever the final wording of product goals and product label is to ensure it is perceived as intended by the actual end-users and beneficiaries.

There was a comment on whether the goal phrasing should also include reference to 'recommending weight gain'. The tool is both for screening and to prompt intervention.

There was a comment that it might be premature to discuss the labelling of final product (e.g., standard, reference chart, charts, etc.) as it is not yet clear what the final data will be / what data will be available.

There was comment that the term 'standard' has different implications across different country contexts.

Session 2: Use of GWG standards in WHO recommendations

Özge Tuncalp, SRH/MPH, presented the types of norm and standard documents developed and published by WHO.

'Norms and standards' is the term used to reference the different types of documents/products produced by WHO. A 'quality norms and standards' (QNS) department was created as part of the new Science Division at WHO.
During the development of this department, a mapping exercise was conducted to understand the types of products that WHO produces. From this mapping exercise, eight categories of WHO norms and standards were identified:[43]

1. *Knowledge gaps*: What are the knowledge gaps or areas of uncertainty?
2. *Classification and nomenclature*: How should interventions, conditions or diseases be named, defined or categorized?
3. *Standards:* What are the standards, cut-offs, targets or thresholds for medicines, health products, exposures or interventions?
4. *Strategy and policy:* What are the evidence-informed options and strategies that the end user can use to approach or analyse complex challenges in public health?
5. *Guideline:* What should the end user do with respect to various options for action?
6. *Implementation guidance*: How should the end user implement a policy, approach, intervention or programme?
7. *Learning materials:* Which learning materials effectively support the acquisition of knowledge, skills and competencies of the health workforce?
8. *Evaluation methodology:* What is the optimal approach to measure, assess or evaluate a risk factor, disease or condition, or the effect of an intervention, or conduct research?

Currently, using WHO norms and standards categories, the GWG project would be considered a 'standard'. However, additional product types may also be warranted to compliment the GWG standard, such as implementation guidance.

It was noted that anything produced by WHO is an 'offer' to countries for use. WHO cannot enforce countries to adopt any WHO norms or standards.

[43] "WHO handbook for the development of normative products", currently undergoing final editing and layout design.

The eight norm and standard categories were also presented on the evidence to impact continuum to help clarify what types of content is relative to each category (see **Figures 3 and 4**).

Figure 3: WHO normative guidance types in the evidence to impact continuum

Figure 4: WHO normative guidance examples per type

An overview of standards versus guidelines was presented:

- Standards: What are the standards, cut-offs, targets or thresholds for medicines, health products, exposures or interventions?
    - These products define the *standards, cut-offs, targets, or thresholds* for medicines, health products, exposures, or interventions.
    - They provide *benchmarks for evaluating* the safety, quality, efficacy, or effectiveness of these entities, and help ensure that they meet specific criteria or requirements.
- Guidelines: What should the end user do with respect to various options for action?
    - A guideline is a *systematically developed set of statements* that answer a question of evidential uncertainty on effectiveness and safety to *propose a choice* among interventions or

implementation options, with one or more evidence syntheses contributing to the assessment of *effectiveness, acceptability, feasibility, and other factors*.

Each WHO norm and standard type has a specific review process to clear before publication. The WHO process to develop guidelines (e.g., WHO recommendations on antenatal care for a positive pregnancy experience) was reviewed. WHO guidelines are considered 'living' and can be updated as new evidence becomes available.[44] Since 2016, four recommendations within the WHO ANC guideline have been updated to reflect new evidence. The publication of the global GWG curves could also inform updates of relevant recommendations.

The global GWG curves are also important for guidelines and recommendations related to non-communicable disease during pregnancy.

Discussions and questions:

There was a comment regarding the importance of keeping the intended end-users of the final GWG standard in mind (e.g., clinicians, community health workers, pregnant women, etc.).

- There was a comment to move away from what we think the end-user will want and relate to and conduct actual end-user testing.
- It was confirmed that there will be end-user testing as part of the review process within WHO.

There was a comment that many health professionals (e.g., nutritionists) use curves/charts not only to monitor/screen but also to recommend weight gain or specific interventions. It was suggested the draft project goals do not currently capture that the standard also provides optimal ranges for recommended weight gain.

- The final product will include a shaded area around the curve to illustrate an optimal GWG gain.
- An adviser stressed that there are different uses for these charts (e.g., to screen and to recommend weight gain) and this should be considered in the development of goals and product labels.

There was a request for clarification on whether 'norm' and 'standard' are interchangeable.

- The term 'norms and standards' is an umbrella term to describe a variety of product types.

There was a request for clarification on the difference between guidance and guidelines.

- WHO clarified that guidelines and guidance are quite different. Guidelines follow a specific process, go through a guideline review committee and are called either a WHO recommendation or WHO guideline. Any of the other seven categories of 'norms and standards' can be considered WHO guidance on a specific topic/issue.
- WHO confirmed that the GWG standards project is not producing guidelines.

There was a comment that the term 'standard' may differ in the field/at the country-level.

- WHO noted that even if the GWG curve would be considered a 'standard' using WHO terminology, it is not required that it is labelled as a 'standard'. There is flexibility in product labelling, and consideration of end-user perception is possible.

There was a comment that a significant amount of time and resources are required to complete the entire

[44] Vogel, J. P., Dowswell, T., Lewin, S., Bonet, M., Hampson, L., Kellie, F., Portela, A., Bucagu, M., Norris, S. L., Neilson, J., Gülmezoglu, A. M., & 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. https://doi.org/10.1136/bmjgh-2019-001683

WHO review process (inclusive of user testing, etc.).

- WHO will discuss internally how to proceed with the review process, ensuring that all requirements of the approval process are met.
- WHO suggested for discussion the inclusion of representatives from professional organizations/associations and relevant women's groups/community organizations in the review process.

Next steps

Giovanna Gatica-Domínguez, WHO/NFS, presented a table of TAG-GWG Phase 1 deliverables for the 2024-2025 period (see **Annex IV**). The deadline for the construction of the GWG standard is September 2025.

Several interim deadlines were discussed:

- Individual-level eligibility criteria was originally scheduled to be completed by March 2024. A written document has already been developed and will be updated based on the discussions from the meeting and will be circulated. It is estimated that the individual-level eligibility criteria will be finalized within one to two months of the meeting.
- The outputs for the development of methodology for construction of global GWG standards were reviewed. All outputs are scheduled to be finalized by June 2025, however, discussions with the working group will continue to see if this timeline is feasible.
- The protocol for the construction of the GWG standards is dependent on finalization of the study-level and individual-level eligibility criteria. Based on the discussions on working group 2 and 3, the protocol is under development so it may be ready for review by June 2024.
- Regarding the development and closure of the GWG database, the data identification was scheduled to be finalized in March 2024, however, it will continue as more data are submitted. Data acquisition will be finalized in September 2024 and data harmonization will begin in April 2024. A technical report will be published describing the final underlying sample for the GWG standards in February 2025.
- The GWG standards will be constructed between January and June 2025, with a report published on the final GWG standard in September 2025, including methods and products.

Discussions and questions

There was a question about plans for publication of peer reviewed articles or protocols and what journals/platforms would be considered. It was suggested to plan for any such publications in advance.

- The GWG Steering Committee responded that there should be a series of peer reviewed articles with different objectives (e.g., to summarize protocols, to summarize the use/implementation of the tool, etc.). However, no list for publications/articles exists yet. It was agreed a list should start to be developed by WHO Secretariat as a next step from this meeting.
- An adviser suggested developing a formal publication plan that outlines the anticipated outputs and the authorship guidelines.

- A GWG Steering Committee member suggested a supplement for the envisioned articles. It was requested that any ideas of potential journals be shared with WHO, noting that WHO only can publish in Plan S journals (i.e., journals that are open access).[45]

There was a question regarding the timing of the next meeting.

- The GWG Steering Committed reported that there is no fixed date yet for the fourth TAG-GWG meeting, however it will likely not place until November at the soonest (tentatively the second half of November).
- The next meeting will be virtual.
- Planning will also need to begin soon for the next in person meeting, tentatively in the first quarter of 2025.

Closing remarks

On behalf of WHO, Elaine Borghi closed the meeting by conveying message of appreciation from Franceso Branca and thanking all participants for their contributions.

[45] WHO is a member of cOAlition S and its open-access policy is in line with the principles of Plan S. For more information on the WHO policy on open access, please see: https://www.who.int/about/policies/publishing/open-access

ANNEX I. Meeting participants [46]

TECHNICAL ADVISORY GROUP ON GESTATIONAL WEIGHT GAIN (TAG-GWG)

- Professor Annick Bogaerts *(Belgium)*
  Midwife, perinatal epidemiology, lifestyle interventions (RCT), mental health
- Professor Amel Fayed*§ *(Egypt)*
  Public health and biostatistics
- Dr S. M. Tafsir Hasan, MD MSc (*Bangladesh*)
  Maternal Nutrition, public health
- Dr Kari Johansson *(Sweden)*
  Perinatal epidemiology
- Professor Lisa M. Bodnar *(United States)*
  Epidemiology, maternal nutrition, adverse pregnancy outcomes ***– not available to participate***
- Dr Cinthya G. Muñoz-Manrique *(Mexico)*
  Maternal-neonatal mortality and morbidity in high-risk women
- Professor Eric Ohuma *(Kenya)*
  Medical statistician on maternal, newborn and child health
- Dr Jodie Dodd *(Australia)*
  Obstetrician and maternal fetal medicine specialist research ***– not available to participate***
- Dr Hayfaa Wahabi* *(Sudan)*
  Obstetrics and Gynaecology and Maternal Epidemiology
- Professor Aris Papageorghiou*§ *(Germany)*
  Maternal and perinatal health, maternal disease in pregnancy, fetal diagnosis and therapy and ultrasound
- Dr Nandita Perumal *(India)*
  Perinatal epidemiology and global maternal and child health
- Professor Suzanne Phelan *(United States)*
  Kinesiology, public health, maternal and child nutritional assessment
- Dr Dayana Rodrigues Farias *(Brazil)*
  Nutritional epidemiology
- Professor Harshpal Singh Sachdev* *(India)*
  Paediatrician, paediatrics and clinical epidemiology, maternal and child nutrition
- Professor Helena Teede *(Australia)*
  Public health, epidemiology, healthy gestational weight gain and health in women of reproductive age
- Dr Molin Wang *(China)*
  Epidemiology, biostatistics, gestational weight gain assessment

GWG STEERING COMMITTEE

From the WHO Secretariat:

- Francesco Branca (NFS Director) ***– not available to participate***
- Elaine Borghi (MNF/NFS)
- Giovanna Gatica-Domínguez (MNF/NFS)
- Monica Flores-Urrutia (MNF/NFS)
- Richard Kumapley (MNF/NFS)
- Olufemi Taiwo Oladapo (SRH/MPH)
- Özge Tuncalp (SRH/MPH)
- Rimu Byadya (AHS/NFS)
- Allisyn Carol Moran (MCA/MAH) ***– not available to participate***
- Maurice Bucagu (MCA/MPH***) – not available to participate***

From external institutions:

- Gilberto Kac (Universidade Federal do Rio de Janeiro)
- Jennifer Hutcheon (University of British Columbia)
- Kathleen M. Rasmussen (Cornell University**)**
- Thais Rangel Bousquet Carrilho (University of British Columbia)
- Mariana Arruda Silva*§ (Universidade Federal do Rio de Janeiro)
- Victor Keller*§ (Universidade Federal do Rio de Janeiro)

[46] **Attended virtually; §Did not attend all three days*

EXTERNAL PARTICIPANTS

(Not members of the TAG-GWG or the GWG Steering Committee)

- Rebecca Goldstein (Monash University)*§
- Aya Mousa (Monash University)*§

ANNEX II. Meeting agenda

Chair: Helena Teede

Day 1: Tuesday, 12 March		
9:00 – 9:05 (5 min)	Welcoming	Elaine Borghi
9:05 – 9:15 (10 min)	Objectives and agenda of the meeting	Chairs: Helena Teede/Suzanne Phelan
9:15 – 9:35 (20 min)	*Session 1:* Overview of the GWG project	Gilberto Kac
9:35 – 9:45 (10 min)	Questions & Answers	Gilberto Kac/Elaine Borghi
9:45 – 10:15 (30 min)	*Session 2*: Systematic literature review on association of GWG with maternal and infant	Dr Mousa and Dr Goldstein from Monash University
10:15 – 10:35 (20 min)	Feedback from TAG-GWG	Moderator: Monica Flores-Urrutia
10:35 – 11:00 (25 min)	Break	
11:00 – 11:40 (40 min)	*Session 3:* Working Group 1 - Individual-level eligibility criteria (updates)	Suzanne Phelan
11:40 – 12:00 (20 min)	Feedback from TAG-GWG	Moderator: Giovanna Gatica-Domínguez
12:00 -13:00 (60 min)	Lunch	
13:00 – 13:30 (30 min)	*Session 4:* Working Group 2 - Identification and harmonization of databases (updates)	Nandita Perumal
13:30 – 14:00 (30 min)	Feedback from TAG-GWG	Moderator: Richard Kumapley
14:00 – 14:30 (30 min)	*Session 5:* Data acquisition (updates)	Giovanna Gatica-Domínguez
14:30 – 15:00 (30 min)	Feedback from TAG-GWG	Moderator: Thais Carrilho
15:00 – 15:20 (20 min)	Break	
15:20 – 15:50 (30 min)	*Session 6:* Working Group 3 - Methods for the development of GWG Standards (updates)	Eric Ohuma
15:50 – 16:20 (30 min)	Feedback from TAG-GWG	Moderator: Elaine Borghi
16:20 – 16:40 (20 min)	*Session 7:* WHO Guidelines for Declaration of Interests	Monica Flores-Urrutia/Elaine Borghi
16:40 – 17:00 (20 min)	Questions & answers	Moderator: Richard Kumapley

Chair: Suzanne Phelan

Day 2: Wednesday, 13 March		
9:00 – 9:15 (15 min)	Summary of Day 1 discussions	Chairs: Helena Teede/Suzanne Phelan
9:15 – 9:30 (15 min)	*Session 1:* Guide for the Working Group discussions	Gilberto Kac
9:30 – 10:30 (60min)	Working Group's discussions	Working Groups #1 and #3
10:30 – 10:50 (20 min)	Break	
10:50 – 12:00 (70 min)	Working Group discussions (cont.)	Working Groups #1 and #3
12:00 – 13:00 (60 min)	Lunch	
13:00 – 13:30 (30 min)	*Session 2:* "Considerations around separate charts for subgroups with different weight gain patterns"	Jennifer Hutcheon
13:30 – 14:30 (60 min)	Discussion	Moderator: Olufemi Oladapo
14:30 – 14:50 (20 min)	*Session 3:* Working Group #1 presentation	Working Group #1 representative
14:50 – 15:10 (20 min)	Discussion	Moderator: Monica Flores-Urrutia
15:10 – 15:20 (10 min)	Summary of the TAG-GWG recommendations	Kathleen Rasmussen
13:00 – 13:30 (30 min)	*Session 2:* "Considerations around separate charts for subgroups with different weight gain patterns"	Jennifer Hutcheon
15:20 – 15:40 (20 min)	Break	
15:40 – 16:00 (20 min)	*Session 4:* Working Group #3 presentation	Working Group #3 representative
16:00 – 16:20 (20 min)	Discussion	Moderator: Thais Carrilho
16:20 – 16:30 (10 min)	Summary of the TAG-GWG recommendations	Richard Kumapley
16:30 – 17:00 (30 min)	*Session 5: Modus operandi* of working groups – way forward	Moderator: Giovanna Gatica-Domínguez

Chair: Suzanne Phelan

Day 3: Thursday, 14 March (half-day)		
9:00 – 9:15 (15 min)	Summary of Day 2 discussions	Chairs: Helena Teede/Suzanne Phelan
9:15 – 9:30 (15 min)	*Session1:* How to call the GWG centile curves, which will be developed based on a "semi-	Elaine Borghi
9:30 – 10:30 (60 min)	Brainstorming	Moderator: Eric Ohuma
10:30 – 10:50 (20 min)	Break	
10:50 – 11:10 (20 min)	*Session 2: Uses of GWG standards in WHO* recommendations	Özge Tuncalp
11:10 – 11:30 (20 min)	Questions & Answers	Moderator: Monica Flores-Urrutia
11:30 – 11:45 (15 min)	Next steps	Giovanna Gatica-Domínguez
11:45 – 12:00 (15 min)	Closing remarks	Francesco Branca/Elaine Borghi

ANNEX III. Working groups guide, points for discussion (Day 2)

Working group #1

1. Plan of analysis to define the inclusion/exclusion criteria

Note: This analysis will be performed in the pooled cleaned dataset, after identification and removal of outliers and heterogeneity assessment (according to working group #3 discussions).

a. How are we going to consider weight gain in these analyses: total weight gain, weight gain in each trimester, weight gain trajectories?
    - In case of weight gain in each trimester, what to do when the individual has more than one measurement in the period? Consider only one? Which one?

b. Which statistical measurements will be used in these analyses (e.g., means, medians, other percentiles)?

c. What will be considered as a 'meaningful difference' when the criterion is excluded (e.g., 1 kg change, 10% change) for each BMI category?

d. Is it reasonable to use a complete-case approach in these analyses, i.e., only studies and individuals without missing variables in the criterion being tested will be considered? Alternatively, the whole dataset could be included in the analysis, but in the presence of missing data, it will not be possible to know whether the change in weight gain distribution is related to the variable itself or to differences in the characteristics of individuals with versus without missing data.

e. Are there any other (sensitivity) analysis to be performed? (E.g.: as suggested by the working group, combine pre-pregnancy diabetes + gestational diabetes)

Working group #3

1. Outliers' assessment
    a. To discuss whether the group considers it necessary to re-assess outliers in the pooled dataset after the heterogeneity assessment.
    b. To revisit the minimum sample size and number of measurements per individual after data cleaning and removal of outliers. Currently, the minimum sample for a study to be eligible for the pool is 200 and with at least two weight measurements per individual besides the initial weight. If measurements flagged as outliers are removed, will we keep these criteria for a dataset to be part of the pool after that?
    c. To discuss how the non-linearity of gestational age will be incorporated into the models of outliers (both cross-sectional and longitudinal). Will we test/use splines, restricted cubic splines, fractional polynomials, or transformations of the variable?
        - For the selected approach in c, how will the optimal fit be decided, e.g., for splines, how will we decide the number of knots to be used and their location?
2. Heterogeneity assessment
    a. To briefly discuss the possibility of using any other method to assess heterogeneity of GWG trajectories beyond those decided during Feb 19 meeting (SSD and variance component)
    b. To define the gestational age intervals to be used for comparisons across studies (e.g., 14-18, 19-24, etc.) and decide the need for adjustments (e.g.: using predicted values for means and SDs as performed in IG-21st).

c. To discuss the parameters/characteristics (e.g., means, medians, etc.) to be used for heterogeneity assessment, including their thresholds.
d. To discuss how small sample size and other quality issues will be addressed in this step. *Regarding data quality:* The GWG Steering Committee is preparing a checklist for an initial quality check after data harmonization. The checklist for each study can be revisited after heterogeneity assessment to see if the studies with a 'lower' quality are the heterogeneous ones.
e. To discuss how identified heterogeneity across studies/countries/regions will be addressed.

ANNEX IV. 2024–2025 work plan (Phase 1)

Final outputs/intermediate outputs	2024												2025								
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Individual-level eligibility criteria	X	X	X																		
Methodology for the construction of the global GWG standards, including the entire data processes and analyses						X															
Method selection to flag outliers, considering cross-sectional and/or longitudinal approaches	X	X																			
Method selection to assess the heterogeneity after pooling data from multiple sources, including the levels to be considered in this assessment (e.g., study, country, region)		X	X																		
Method selection to calculate the minimum sample size to model the GWG centiles				X																	
Selection of the statistical approach, including diagnostic and validation procedures to be adopted for each model			X	X	X																
Identify methodology to treat data imbalances, particularly by geographic distribution and other factors flagged as relevant					X																
Protocol for the construction of the GWG standards	X	X	X	X	X	X															
Development and closure of the GWG database						X															
Data identification	X	X	X																		
Data acquisition	X	X	X	X	X	X	X	X	X												
Data harmonization				X	X	X	X	X	X	X	X	X									
Technical report on the description of the final underlying sample for the standards													X	X							

Final outputs/intermediate outputs	2024												2025								
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Construction of the GWG standards (report)													X	X	X	X	X	X			
Technical report on the Global GWG standards: methods and outcomes																			X	X	X