WHO methods and data sources for mean haemoglobin and anaemia estimates in women of reproductive age and pre-school age children 2000-2019

Department of Nutrition and Food Safety

World Health Organization

Geneva, Switzerland

March 2021

CONTENTS

Acknowledgements

1. Introduction

Awareness about anaemia and its consequences for women's and children's health and development has increased in recent decades. In 2012, the 65th World Health Assembly approved an action plan and global targets for maternal, infant, and child nutrition, with commitment to halve anaemia prevalence in women of reproductive age by 2025 (1). This goal aligns with the Sustainable Development Goals (SDG) Goal 2 to end hunger, achieve food security and improved nutrition and promote sustainable agriculture, and anaemia in women of reproductive age was added as an indicator under Goal 2 in 2020 (2). In order to monitor progress toward the SDG Goal 2 and the global nutrition goals, the World Health Organization has updated its estimates of anaemia prevalence in children aged 6-59 months and women of reproductive age (15-49 years).

These estimates were prepared following the statistical approach used for the estimates published in The global prevalence of anaemia in 2011 and in academic papers (3–5), with enhancements in terms of covariates used, based on data that became available and new evidence. The estimates represent the best estimates of WHO, based on the evidence available up to February of 2021. They may differ from official estimates of Member States. These new estimates were published in March 2021, through SDG mechanisms, as well as the WHO Global Health Observatory.

We estimated trends between 2000 and 2019 in the distributions of blood haemoglobin for children aged 6-59 months and for women of reproductive age (15-49 years), separately by pregnancy status, in 197 countries and territories. Our analysis included three steps:

1) Identifying data sources on haemoglobin and anaemia; accessing and extracting data; and systematically assessing population representativeness of data;

2) Adjusting haemoglobin for altitude; and

3) Applying a statistical model to estimate trends in blood haemoglobin distributions and their uncertainties for children and for women of reproductive age by pregnancy status.

The distributions estimated in step 3 allowed coherent and consistent estimation of mean haemoglobin and of the prevalences of total and severe anaemia. We defined total anaemia based on the World Health Organization (WHO) thresholds of haemoglobin < 110 g/L for children under 5 years of age and pregnant women, and < 120 g/L for non-pregnant women. Severe anaemia was defined as blood haemoglobin < 70 g/L for children under 5 years and pregnant women, and < 80 g/L for non-pregnant women.

These estimates have been documented following the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) (6). The locations where GATHER reporting items are reported are given in the Appendix.

2. Data identification, access and inclusion

Our data search and access strategy were designed to obtain as many sources as possible while ensuring that the sources were representative of the population at the national level or at least covered three regions within the country. The distribution of blood haemoglobin concentration in a population is commonly summarized as a percentage below a threshold, or a prevalence of anaemia. Mean haemoglobin and its standard deviation may also be reported. Anaemia thresholds typically vary by age, sex, and pregnancy status. Studies may also use different haemoglobin thresholds to define anaemia, and may report multiple anaemia severities, such as mild, moderate and severe anaemia. We accessed data in two forms: 1) anonymised individual-level haemoglobin data when available, and 2) summary statistics, including mean haemoglobin and anaemia prevalences below specific thresholds. We used anaemia prevalences with any threshold in our statistical model described in Section 4, which accounts for the specific thresholds used to define anaemia when using the data.

We included data sources if:

- blood haemoglobin was measured;
- the study reported anaemia or mean haemoglobin for preschool-age children (6-59 months of age) or women of reproductive age (15 to 49 years of age);
- a probabilistic sampling method with a defined sampling frame was used and data were representative of at least three areas within a country;
- data were collected in or after 1995;
- data were from 194 Member States or 3 territories: Puerto Rico; Taiwan, China; West Bank and Gaza Strip; and
- standard, validated data collection techniques and laboratory methodologies were used.

Measurement of haemoglobin for children younger than 6 months of age is not common in surveys because it requires a puncture to obtain blood drops. For this reason, estimates were made for children 6-59 months of age. Some data sources did not report anaemia prevalence for the target age ranges. We included data on preschool-age children up to 71 months of age as long as children 6-59 months were included in the study/survey sample. For women, we included data sources reporting on any age range 10 years and over

5

provided that women of reproductive age (15 to 49 years of age) were included in the sample. Data sources that did not cover the exact age groups of interest were given smaller weights, as described in Section 4.

We performed an additional screening if a facility-based sampling scheme was used in order to exclude data where these would not be representative of the general population. The general threshold for inclusion was 80% affiliation of the target population with the facility. For studies of children sampled from primary care physician rosters or well-child visits, we included the data if national coverage of the third dose of DTP vaccine exceeded 80% (7). For women sampled from obstetric care providers, data were included if the coverage of at least one ANC visit was greater than 80% (8). For school-based sampling of adolescent girls, the completion rate of lower secondary school for girls was required to be greater than 80% (9).

In four countries where migrants comprised more than 40% of the population in the country (Kuwait, Qatar, United Arab Emirates, and Singapore), we excluded any data that covered only national citizens (10). We also excluded subnational data sources if the subnational area was selected on a variable causally related to anaemia prevalence, e.g., malaria endemicity.

A country consultation was carried out in December, 2020, with national focal points nominated by the country's SDG focal point. Focal points were consulted on draft estimates of mean haemoglobin, prevalence of anaemia and prevalence of severe anaemia in children 6-59 months of age and women 15 to 49 years of age (estimates by pregnancy status and combined estimates), as well as on the input data (i.e. primary data such as household surveys) that were used to compute the anaemia estimates for each demographic group in each country. Countries were asked to share data sources not included in the draft input data series (e.g. published or unpublished nationally representative surveys or peer-reviewed publications with national level data).We included all data provided by national governments in the VMNIS and used them to compute final estimates provided that they fulfilled inclusion criteria. We manually identified and removed duplicated data accessed from more than one source. Our dataset closed in February, 2021.

2.1 Individual-level data

We obtained anonymised individual-level data from health-examination surveys and household surveys with haemoglobin measurements. Most of these sources were implemented under the umbrella of multi-country survey programs including the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), Reproductive Health Surveys (RHS) and the Malaria Indicator Surveys (MIS). We also used national health examination survey data if the data were available to us and it was not possible to obtain summary statistics for the source. From each source, we extracted the following variables (if available): age, sex,

haemoglobin concentration, pregnancy status, altitude, survey sample weight, stratum, and primary sampling unit.

We only used data for children aged 6 to 59 months and women aged 15 to 49 years. Haemoglobin concentrations recorded in survey datasets that were considered biologically implausible were excluded; i.e. haemoglobin measurements that were less than 25 g/L or greater than 200 g/L. Finally, we ensured that haemoglobin data were appropriately adjusted for altitude, as described in Section 3.

We excluded all individual-record observations for women of reproductive age who reported that they did not know their pregnancy status. Five surveys with individual-level data did not record pregnancy status as a part of their design. In these cases, we did not use individual-level data; rather, we calculated summary statistics for the whole sample and used the source in the same manner as data accessed directly as summary statistics, as described below.

2.2 Data accessed as summary statistics

WHO maintains a Micronutrients database, part of the Vitamin and Mineral Nutrition Information System (VMNIS), which contains haemoglobin concentrations and anaemia prevalence (11). Data are identified via periodic MEDLINE searches and an international network of collaborators, who uncover data sources not reported in bibliographic databases. In order to get up-to-date anaemia data, a search on bibliographic databases[1] was performed. The search was limited to humans, and the following search terms were used:

(((national) AND (survey)) OR ((population) AND (prevalence))) AND ((iron status) OR (iron deficiency) OR (anaemia) OR (anaemia) OR OR (haemoglobin) OR (haemoglobin) OR (low iron level) OR (transferrin receptor) OR (ferritin) OR (insufficient iron))

Studies are included in the WHO Micronutrients Database if there is a defined population-based sampling frame; a probabilistic sampling procedure is used and standard, validated data collection techniques and laboratory methodologies were used. For inclusion of these summary data in the analysis, we screened these summary data using our exclusion criteria. Consistent with our inclusion and exclusion criteria, we excluded summarized data sources if:

---

[1] Bibliographic databases used to identify anaemia data sources: MEDLINE, EMBASE, Web of Science, CINAHL, AGRICOLA, IBECS, SCIELO, LILACS, AIM (AFRO), IMEMR (EMRO), PAHO, WHOLIS, WPRIM (WPRO), IMSEAR, Native Health Research Database

- data were collected prior to 1995;

- we had access to the same data as individual-level records;

- non-random sampling methods were used, or sampling methods were not adequately described;

- a facility-based surveillance method was used in a country where facility affiliation was lower than 80% (as described above);

- they were representative of only one or two first administrative level(s);

- we were unable to determine whether the data were adjusted for altitude and the data were collected in a high-altitude country;

- only aggregated summary statistics that included subjects beyond an acceptable age interval were reported (i.e., data on preschool children included children over 6 years of age or data on women of reproductive age included girls under 10 years of age); or

- the study did not have data on haemoglobin concentration or anaemia prevalence in children age 6-59 months or women aged 15-49 years.

In some cases, the sample size was not reported in the WHO Micronutrients Database. In that case, we conservatively assumed a sample size of 100. For four data sources, we had some information that sample sizes were substantially larger than 100, such as number of households sampled or all-age-sex sample size, which we used to make a conservative estimate of sample size (India 2019-2020 National Family Health Survey – Phase 1, Indonesia RISKESDAS 2013, Pakistan National Nutrition Survey 2018, and China Chronic Diseases and Nutritional Health Surveillance 2015).

We excluded the following data sources which reported anaemia prevalence, mean haemoglobin or haemoglobin distribution which we considered implausible: Enquète nationale sur les indicateurs du paludisme au Tchad 2010 (ENIPT-2010), and Enquète nationale sur les indicateurs du paludisme au Tchad de 2017 (ENIPT-2017), 2005 Rwanda DHS, and 1998-1999 Bahrain National Nutrition Survey. We also excluded pre-school children's mean haemoglobin from the Iraq national micronutrient deficiencies: assessment and response 2011-2012 and prevalence of anaemia in pregnant women reported by the Afghanistan National Nutrition Survey 2004 because they we considered the values to be implausible. Finally, we excluded any population standard deviation below 5 g/L, as these are likely mislabelled standard errors.

2.3 Accounting for complex survey design

As described in Section 4, the statistical model used individual-level data when available, and summary statistics when not, to estimate the full distributions of blood haemoglobin concentration by country and year.

All the individual-level data in the analysis came from surveys that used complex survey designs. Specifically, in designing a representative survey, the target populations were usually divided into strata based on geographical regions within the country, whether place of residence was rural or urban, and/or the socio-economic characteristics of the place of residence; within each stratum, a number of clusters were randomly selected. Clusters may be villages, administrative units, or census units. Households or participants were then randomly sampled within each cluster. Depending on survey design, individuals or households in some units may have a higher probability of being selected than those in other units. To account for the differences in probability of being sampled, each observation is assigned a sample weight. These weights are calculated to make the survey data representative of the total population.

An implication of the sampling method is that the so-called effective sample size of the survey (ESS) is different from its actual sample size. This occurs primarily because the sampled individuals are from clusters that are representative but are not randomly sampled from the entire country, and hence contain less information than they would, had they been a true simple random sample of the population.

To reflect the true availability of information in each survey and in the individual level data that it provided to the statistical model, we estimated ESS based on the "*estat effects*" command of the Stata version 16.1 svy suite of commands (StataCorp, 2019). In particular, this command generates the design effect (DEFF), which is the ratio between the (usually smaller) ESS and the real sample size, e.g. a survey with 1000 subjects with a DEFF of 2.0 has an ESS of 500. The DEFF may differ by summary statistic metrics (mean vs. prevalence below 100 g/L vs. prevalence below 120 g/L) depending on how these indicators are distributed across the strata and clusters. For each survey, we calculated the DEFF as the median of those from a range of metrics, specifically, mean haemoglobin concentration and prevalence below 90, 100, 110, 120, 130 g/L. ESS was then calculated as sample size divided by DEFF.

In our statistical model, we accounted for the difference between the real and effective sample sizes and for the difference in weights for each observation by scaling the weights across all observations in a study to sum to the ESS. These scaled weights were then used to weight the likelihood contributions from each individual. In addition, surveys may over- or under-sample pregnant women relative to their fraction of the population. To account for this imbalance in the statistical model for women, we scaled the weights for each individual such that the sum of the weights for pregnant women was equal to the total ESS for the study multiplied by the proportion of pregnant women in the study; we did the same for and non-pregnant women. This ensured that the sum of weights across all women was equal to the ESS for the study and that

the relative weighting of pregnant and non-pregnant women reflected the number of women in each category in the study.

Data sources providing only summary statistics were also predominantly from surveys that used complex survey designs, but sample sizes recorded for these data sources are actual sample sizes and not the effective sample sizes. To ensure that the sample sizes used for these sources in the statistical modelling also reflect the complex survey design, we estimated ESS for each study as the actual sample size multiplied by an estimate of the DEFF. Calculating the DEFF requires individual-level data, which by definition is not available for these data sources. We used the median DEFF from all surveys with individual-level data. We then used the estimated ESS for each study in deriving the joint normal likelihood for the summary statistics from each study.

2.4 Prevalence of pregnancy

As described below, the statistical model used separate data for pregnant and non-pregnant women when data were available by pregnancy status, and accounted for the proportion of the sample that was pregnant in sources that reported combined summary statistics. We then made estimates for women of reproductive age by pregnancy status, and combined these to estimate total prevalence of anaemia in women of reproductive age. Both of these steps required national estimates of the percentage of women of reproductive age who are pregnant.

In our prior modelling exercise, we found that women are not very likely to report early pregnancies, with reporting of pregnancies becoming consistent after the 10th week of pregnancy. This early pregnancy period is also the period during which the decline in haemoglobin concentrations is steepest (4). To be consistent with the reporting behaviours and to restrict pregnancy to periods when haemoglobin concentrations are consistently low, our operational definition of pregnancy was restricted to after 8 weeks of gestation. We used an average pregnancy duration of 32 weeks (gestational weeks 8-40) to calculate the proportion of women in each country and year who were pregnant at any time. To calculate this quantity, we used country- and age-specific data on live births from the UN Population Division's 2019 Revision of the World Population Prospects and an estimate of the number of stillbirths, together with estimates of the total number of women of reproductive age (12,13).

3. Methods for adjusting haemoglobin for altitude

Haemoglobin needs are greater for those living at high altitudes due to the lower concentration of oxygen in the atmosphere (14). When altitude measurements corresponding to individual-level observations were

available, we adjusted haemoglobin concentrations using a formula developed by the US Centers for Disease Control and Prevention and commonly used in studies worldwide (14,15):

$$Hb_{adjusted} = Hb_{unadjusted} + 0.32 *(altitude * 0.0033) - 0.22 *(altitude *. 0.0033)^2 \qquad (1)$$

where haemoglobin is measured in g/L and altitude is measured in meters above sea level (m.a.s.l.). The adjustment is only applied to individuals living at altitudes over 1,000 m.a.s.l. We were unable to obtain altitude information for individual subjects for some surveys with individual record data. When the proportion of population living at altitudes above 1500 m.a.s.l. (an altitude at which there is 3 g/L effect on haemoglobin concentration) was less than 5% of total population (hereafter termed low-altitude countries), we included the source. Data from individual-record data sources without individual-level altitude in a country where more than 5% of total population lived above 1500 m.a.s.l were converted to summary statistics and adjusted as described below.

For data available as only summary statistics, we determined whether the data were from a low-altitude country. If data were from a low-altitude country, we used the data regardless of adjustment for altitude. We developed regression equations to correct unadjusted summary statistics from high-altitude countries. We pooled data from 98 DHS surveys in countries with some population living over 1000 m.a.s.l. We extracted both adjusted and unadjusted haemoglobin data for all survey participants in the target populations (women of reproductive age and children 6-59 months). We calculated mean haemoglobin and prevalence of anaemia (using each of the cut-offs 70, 100, 110, 120, and 130 g/L) for women and children using both adjusted and unadjusted data from each of the surveys. We then related altitude-adjusted mean haemoglobin concentration and prevalence of anaemia to the corresponding unadjusted values using separate regressions for each of the six metrics and for each population group, for a total of 12 regressions. For each metric we regressed the adjusted values against the unadjusted values, including as additional covariates the percent of population living over 1,000 m.a.s.l., the percent of population living over 2,000 m.a.s.l., and an interaction between the two. This regression specification was intended to mimic the quadratic relationship in the CDC adjustment (Eq. 1). In the regression for mean haemoglobin, we fixed the coefficient of unadjusted haemoglobin at one to reflect the relationship in (Eq. 1).

We used these regression relationships to predict adjusted mean haemoglobin concentration and adjusted prevalence of anaemia based on unadjusted summary statistics from high-altitude countries. We accounted for uncertainty of this step by calculating the standard regression prediction variance, which reflects both uncertainty in estimating the regression relationship and variability of individual values around the

11

regression line. This variability from the effect of predicting adjusted country-level metrics was then included in the statistical modelling as an added variance in the likelihood for each summary statistic from these sources.

4. Bayesian hierarchical mixture model

Our aim was to estimate the complete distributions of blood haemoglobin for every country and year, which would then allow calculating any relevant summary statistic. This approach allows making coherent inference on mean haemoglobin and on the prevalence of anaemia at all levels of severity. All analyses were done separately for children and women of reproductive age.

The statistical methods are described in detail in a previous publication (5). In brief, we used a Bayesian hierarchical mixture model, which uses all available data to make estimates for each country-year. In the hierarchical model, estimates for each country-year were informed by data from that country-year itself, if available, and by data from other years in the same country and in other countries, especially those in the same region with data in similar time periods. The hierarchical model shares information to a greater degree where data are non-existent or weakly informative (i.e., have large uncertainty), and a lesser degree in data-rich countries and regions. We modelled trends over time as a linear trend plus a smooth nonlinear trend, at the country, regional, and global levels. The estimates are also informed by covariates that help predict haemoglobin levels. The model included a variance term that accounted for unobserved design factors (sample design, season, haemoglobin measurement method, etc.) that lead to additional variability in the data beyond that expected due to sample size. Finally, the model accounted for the fact that subnational data and data that do not exactly cover the age ranges of interest may have larger variation than national data and data that exactly cover the age ranges of interest, respectively. We fitted the model to data from 1995 to 2020 to limit boundary effects but report results between 2000 and 2019, because there were fewer sources between 1995 and 1999 and in 2020.

Importantly, within the framework of the model described above, the set of time-varying covariates that were used to inform the model were reviewed and revised. Previous editions of WHO haemoglobin estimates used the following time-varying covariates to inform estimates: mean years of maternal education, proportion of population in urban areas, mean latitude, prevalence of sickle-cell disorders and thalassemias, mean body-mass index for women, and mean weight-for-age Z score for children. Some of these datasets are no longer updated, including prevalence of sickle-cell disorders and thalassemias and mean weight-for-age Z score in children, while new covariate datasets have become available. We tested three covariate sets. Set 0 excluded time-varying covariates. Set 1 approximated covariates used in prior estimation rounds while

accounting for current data availability, excluding prevalence of haemoglobinopathies and replacing mean weight-for-age Z score with log of shock-free child mortality (16) for children since the former two datasets are no longer updated. Set 2 included socio-demographic index (17), meat supply (kcal/capita) (18,19), mean BMI (for women) (20), and log of shock-free under-five mortality (for children). We compared Watanabe-Akaike information criterion (WAIC) – a measure of model fit – across the three covariate sets (21). How to implement WAIC in this complicated setting with temporal and spatial correlation is not well-defined, and there is some uncertainty in estimating WAIC, so we consider the WAIC results to be one piece of evidence and not conclusive. We also compared the fitted estimates using each of the three covariate sets, and compared the estimates to primary data where the estimates differed. For the women's model, WAIC was inconclusive; for the children's model, WAIC indicated that including covariates was preferred to a model with no time-varying covariates. Comparing fitted estimates using each of the three covariate sets, estimated prevalence of anaemia was similar at the global and regional levels and in countries with primary data available. Therefore, we chose the covariate set based on our assessment of plausibility of estimates in countries with no, sparse or inconsistent data. We considered the estimates generated using covariate set 2 to be most plausible.

There is increasing concern that factors associated with haemoglobin measurement can have a substantial effect on the estimated prevalence of anaemia (22–24). Factors that may influence haemoglobin measurements include different type of blood sample (e.g. venous or capillary blood) and analytical methods for measuring blood haemoglobin in household surveys. We tested an offset for blood collection by capillary puncture in the global estimation model. This approach carries some risk. Capillary puncture is more likely to be used in some settings (less wealthy countries) and time periods (more recent years). The fitted coefficient for the offset may reflect residual regional or time differences, rather than any systematic bias of capillary blood collection. We ran both the model of haemoglobin distribution in preschool-age children and the model of haemoglobin distribution in women of reproductive age with an offset for capillary blood collection. The capillary offset was estimated to be 1 g/L for pre-school children (i.e., the model estimated that capillary blood collection underestimates haemoglobin by 1 g/L), and -1.5 g/L for women (capillary blood collection overestimates haemoglobin by 1.5 g/L). WAIC indicates that including the capillary offset makes the fits worse for children, but improves model fit for women. The offset estimated in the model was smaller than expected for children (expected value in the range of 5-10 g/L) based on comparison of near-in-time survey pairs with different blood collection types. For women of reproductive age, the offset was estimated to be small, as expected. We suspect that the offset coefficient in the children's model may reflect residual differences in region/time period that were not explained by the other terms in the regression model. Inclusion of the offset also required exclusion of data sources that did not report type of blood collection.

13

Therefore, we chose not to include the offset in the final global estimation model for children. In the interest of fitting a consistent and parsimonious model for women, we also excluded the offset from the women's model.

The mixture model uses a mixture (a weighted-average) of multiple normal ("bell-shaped") densities to estimate the full haemoglobin distributions, which may themselves be skewed. We used a mixture of five normal distributions for children. For adult women, we used two five-component mixtures, one for pregnant women and another for non-pregnant women. This approach uses all data sources – those that separate pregnant and non-pregnant women, those in which pregnant and non-pregnant women are reported together, and those in which only one group was measured – to make separate estimates by pregnancy status. The differences in haemoglobin distributions between pregnant and non-pregnant women were allowed to vary by country and year. In years and countries where separate data by pregnancy status were lacking, the difference was informed based on other sources, especially those in the same region with data in similar time periods.

The model is specified as follows, with g an indicator differentiating pregnant and non-pregnant strata within a study:

$$f_{gi}(z) = \sum_{m=1}^{M+1} w_{mgi} \, \mathcal{N}\left(z \,|\, \theta_m, \sigma_m^2\right) \tag{2}$$

$$w_{mgi} = \begin{cases} \Phi(\alpha_{mgi}) \prod_{u=1}^{m-1}(1 - \Phi(\alpha_{ugi})) & \text{if } m \leq M \\ \prod_{u=1}^{M}(1 - \Phi(\alpha_{mgi})) & \text{if } m = M+1 \end{cases} \tag{3}$$

$$\alpha_{mgi} = \delta_{mj[i]}^c + (\varphi\delta^c)_{mj[i]} t_i + u_{mj[i]t_i} + \beta_m x_i + a_{mi} + b_{mi} + I_{gi}(\gamma_{mj[i]}^c + (\varphi\gamma^c)_{mj[i]} t_i + c_{mi}) \tag{4}$$

Details on the model specifications and features are provided elsewhere (5). Briefly, equation 2 describes a finite mixture of $M + 1$ normal (N) distributions (or mixture components), where the weights (ω) on the constituent normal distributions vary across studies. We specified a probit stick-breaking model for the ω's in equation 3. This transformation uses the standard normal cumulative distribution function (Φ) to transform α's that range between - ∞ and ∞ to ω's that range between 0 and 1. Specifically, the α's determine the relative weights assigned to each cluster in the following manner: starting with a 'stick' of length one, $\Phi(\alpha 1_{gi})$ is the proportion of the stick that we break off and assign to ω_{1gi}; $\Phi(\alpha_{2gi})$ is the proportion of the remaining stick of length $(1 - \omega_{1gi})$ that we break off and allocate to ω_{2i}; and so on. Larger values of α_{mgi} thus correspond to higher weights on the mth mixture component for stratum g in study i. The probit stick-breaking transformation therefore allows placing a flexible model on the α's, while ensuring that

the ω's still add to one, in such a way that large mass in one part of the haemoglobin distribution is balanced by smaller mass in others parts, and vice versa, through exchanges among the constituent mixture components.

In equation 4, α_{mgi} is defined to leverage all available information in making estimates for each country-year-stratum. $\delta^c_{mj[i]}$ is a country-by-component interaction term, determining the baseline weight placed on each of the $M + 1$ normal distributions in country j. $(\varphi\delta^c)_{mj[i]}$ is a country- and component-specific linear time effect, determining the linear parts of country j's time trend. Letting $T = 26$ be the total number of analysis years (1995, 1996,… , 2020), the T-vector $u_{mj[i]}$ captures smooth nonlinear change over time in country j and mixture component m. b_m is the effect of time-varying country-level covariates x (described above) in mixture component m. The a's are study-specific random effects, and the b's capture the extra variance of studies that included women under age 15 or over age 50 (or those that did not cover exactly 6-59 months of age in the model for children). The difference between the models for women and children is that for the former, the model includes the additional terms that are multiplied by I_{gi}, which is an indicator variable that takes the value one when stratum g contains pregnant women and -1 when stratum g contains non-pregnant women.

This indicator multiplies a country- and component-specific term, $\gamma^c_{mj[i]}$, that quantifies the overall difference between pregnant and non-pregnant women, a linear time effect for the pregnant/non-pregnant difference, $(\varphi\gamma^c)_{mj[i]}$, and study-specific errors, c_{mi}, in the difference. The difference in haemoglobin between pregnant and non-pregnant women was modelled as linear for simplicity and because there are insufficient data to reliably estimate more complex trends in difference.

The hierarchical prior distributions for the country-specific terms and specifications of the study-specific error terms are described in detail elsewhere (5), with the additional terms introduced here, $\gamma^c_{mj[i]}$, $(\varphi\gamma^c)_{mj[i]}$ and $c_{mi,}$ treated analogously to δ^c_{mj}, $(\varphi\delta^c)_{mj}$ and a_{mi}, respectively.

For data accessed as summary statistics for which pregnant and non-pregnant women are not distinguished, we took the mixture densities for pregnant and for non-pregnant women and combined them into a $(2M + 2)$-component mixture, weighting by the proportion of pregnant women estimated for that country-year, as described earlier.

The uncertainties of our estimates incorporated sampling error in each data source; non-sampling error of national data, e.g. because of issues in sample design and measurement; additional error associated with

subnational data; uncertainty due to altitude adjustments; and uncertainty due to making estimates by country and year when data were missing altogether, when only summary statistics (vs. individual-level data) were available, or when data were not available separately by pregnancy status.

We fitted the Bayesian model using the Markov chain Monte Carlo (MCMC) algorithm and obtained 3,000 samples from the parameters' posterior, in turn used to obtain 3,000 posterior samples of the population haemoglobin distributions for each country-year. With each of the 3,000 sampled distributions we calculated the population haemoglobin mean and total and severe anaemia prevalences for each country-year. Numbers of persons affected by severe and total anaemia were computed for each sample by multiplying county-year-age-sex-group population totals from the UN Population Division's 2019 Revision of the World Population Prospects by the prevalence of severe or total anaemia (13). All reported uncertainty intervals represent the 2.5th-97.5th percentiles of these 3,000 draws.

5. Comparison with other studies

The results presented here represent an update of previous WHO estimates for 1995-2011 (3,4) and 1995-2017 (25). At the global level, the estimates presented here are similar to previous WHO estimates, with new estimates for 2000 and 2016 for all population groups falling within the uncertainty interval of the previous estimate set. Differences in estimates at the country or region level can be attributed to inclusion of recently published haemoglobin data. The current estimates included 489 data sources spanning 1995-2020, while the estimates for 1995-2016 analysed 357 data sources spanning 1990-2016. The covariates used in the global model were also updated, however, sensitivity analyses demonstrated that this change had a limited effect on estimates of haemoglobin distribution globally and in countries with primary data (Section 4). The Global Burden of Disease 2019 study has also computed global, regional, and national estimates of anaemia prevalence, using different inclusion/exclusion criteria, statistical models, and time-varying covariates (26). Nevertheless, they estimated a similar prevalence of anaemia in women of reproductive age in 2019 (28.6%, 95% uncertainty interval 28.2%-28.9% *vs*. 29.9%, 27.0%-32.8%) (27). Global Burden of Disease estimates for children under the age of five years were slightly higher than ours (45.8%, 44.3%-47.8% *vs*. 39.8%, 36.0%-43.8%). This discrepancy can partially, but not entirely be explained by our exclusion of children 0-5 months of age, who have higher anaemia prevalence than older children.

6. Strengths and limitations

The strengths of this study include our extensive data search and rigorous criteria for inclusion of sources; consistent analysis by for children and women, and by pregnancy status; estimating trends by country and region; estimating the full population distributions of haemoglobin, which is consistent with epidemiological

evidence on the harms of low blood haemoglobin; and systematic estimation and reporting of uncertainty. The main limitation of our analysis is that despite the extensive data search and access, there were considerable gaps in data availability. As a result, the estimates may not capture the full variation across countries and regions, tending to "shrink" towards global means when data are sparse. This may have especially affected the estimates in high-income and upper-middle-income countries, where anaemia prevalence is low and typically addressed in a clinical setting. Nevertheless, this study benefited from an active consultation with WHO member states, which resulted in identification and inclusion of additional data sources. The consultation also raised awareness about best practices for monitoring anaemia in countries. Finally, there is increasing concern that factors associated with the measurement of blood haemoglobin in household survey can have a substantial effect on the estimated prevalence of anaemia (22–24). Influential factors may include different type of blood sample (e.g. venous or capillary blood), analytical methods for measuring blood haemoglobin in household surveys, or other factors associated with survey design and implementation. These factors have not been taken into account in this study or in previous global, regional and country estimates of anaemia prevalence (4,26). Further research is needed to determine which factors associated with survey design, implementation or haemoglobin measurement are responsible for any systematic differences among data sources, so that global models can account for the relevant factors.

References

1. World Health Organization. Comprehensive implementation plan on maternal, infant and young child nutrition [Internet]. Geneva: World Health Organization; 2014 [cited 2021 Mar 20]. Available from: http://www.who.int/nutrition/publications/CIP_document/en/

2. United Nations General Assembly. Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development Goals and targets (E/CN.3/2020/2) [Internet]. New York City,; 2020 [cited 2021 Feb 24]. Available from: https://unstats.un.org/sdgs/indicators/Global Indicator Framework after 2020 review_Eng.pdf

3. World Health Organization. The global prevalence of anaemia in 2011 [Internet]. WHO. Geneva: World Health Organization; 2015 [cited 2021 Feb 24]. Available from: https://www.who.int/nutrition/publications/micronutrients/global_prevalence_anaemia_2011/en/

4. Stevens GA, Finucane MM, De-Regil LM, Paciorek CJ, Flaxman SR, Branca F, et al. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: A systematic analysis of population-representative data. Lancet Glob Heal [Internet]. 2013 Sep [cited 2021 Feb 22];1(1). Available from: https://pubmed.ncbi.nlm.nih.gov/25103581/

5. Finucane M, Paciorek C, Stevens G, Ezzati M. Semiparametric Bayesian density estimation with disparate data sources: a meta-analysis of global childhood undernutrition. J Am Stat Assoc. 2015;110(511):889–901.

6. Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, Ezzati M, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. PLOS Med [Internet]. 2016 Jun 28 [cited 2021 Mar 24];13(6):e1002056. Available from: https://dx.plos.org/10.1371/journal.pmed.1002056

7. World Health Organization. Diptheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%) [Internet]. Geneva: World Health Organization; 2020. Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/diphtheria-tetanus-toxoid-and-pertussis-(dtp3)-immunization-coverage-among-1-year-olds-(-)

8. United Nations Children's Emergency Fund. UNICEF maternal and newborn health coverage database [Internet]. New York: UNICEF; 2020. Available from: https://data.unicef.org/topic/maternal-health/antenatal-care/

9. UNESCO Institute for Statistics. Lower secondary completion rate, female (% of relevant age group) [Internet]. Paris, France: World Bank; 2020. Available from: https://data.worldbank.org/indicator/SE.SEC.CMPT.LO.FE.ZS

10. United Nations Department of Economic and Social Affairs. International Migrant Stock 2019 (United

Nations database, POP/DB/MIG/Stock/Rev.2019). New York City: United Nations; 2019.

11. World Health Organization. Micronutrients database [Internet]. Geneva; 2021 [cited 2021 Mar 17]. Available from: https://www.who.int/vmnis/database/en/

12. Blencowe H, Cousens S, Jassir FB, Say L, Chou D, Mathers C, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: A systematic analysis. Lancet Glob Heal [Internet]. 2016 Feb 1 [cited 2021 Mar 17];4(2):e98–108. Available from: https://www.

13. United Nations Department of Economic and Social Affairs. World Population Prospects 2019 [Internet]. New York: United Nations; 2019 [cited 2021 Mar 17]. Available from: https://population.un.org/wpp/

14. Dowdle W. CDC Criteria for anemia in children and childbearing-aged women. Morb Mortal Wkly Rep. 1989;38(22):400–4.

15. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Geneva: World Health Organization; 2011.

16. United Nations Inter-Agency Group for Child Mortality Estimation (UN IGME). Levels & Trends in Child Mortality: Report 2020, Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation [Internet]. New York; 2020. Available from: childmortality.org

17. Global Burden of Disease Collaborative Network. Global Burden of Disease 2019 (GBD 2019) Covariates 1980-2019. Seattle: Institute for Health Metrics and Evaluation; 2020.

18. FAOSTAT. Food Balance Sheets [Internet]. Rome: FAOSTAT; 2019 [cited 2020 Sep 3]. Available from: http://www.fao.org/faostat/en/#data/FBS

19. Bentham J, Singh GM, Danaei G, Green R, Lin JK, Stevens GA, et al. Multidimensional characterization of global food supply from 1961 to 2013. Nat Food [Internet]. 2020 Jan 13 [cited 2021 Mar 17];1(1):70–5. Available from: https://www.nature.com/articles/s43016-019-0012-2

20. Bixby H, Bentham J, Zhou B, Di Cesare M, Paciorek CJ, Bennett JE, et al. Rising rural body-mass index is the main driver of the global obesity epidemic in adults. Nature [Internet]. 2019 May 9 [cited 2021 Mar 24];569(7755):260–4. Available from: https://doi.org/10.1038/s41586-019-1171-x

21. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput [Internet]. 2017 [cited 2021 Mar 31];27(5):1413–32. Available from: https://github.com/stan-dev/loo.

22. Rappaport AI, Karakochuk CD, Hess SY, Whitehead, RD, Namaste SML, Dary O, et al. Variability in haemoglobin concentration by measurement tool and blood source: An analysis from seven countries. J Clin Pathol [Internet]. 2020 Oct 6 [cited 2021 Feb 20]; Available from: https://jcp.bmj.com/content/early/2020/10/06/jclinpath-2020-206717

23. Hruschka DJ, Williams AM, Mei Z, Leidman E, Suchdev PS, Young MF, et al. Comparing hemoglobin distributions between population-based surveys matched by country and time. BMC Public Health. 2020;20(1):1–10.

24. SPRING. Anemia Assessment in Micronutrient and Demographic and Health Surveys: Comparisons in Malawi and Guatemala. Arlington, VA; 2018.

25. World Health Organization. Prevalence of anaemia [Internet]. Geneva, Switzerland: World Health Organization; 2017. Available from: https://www.who.int/data/gho

26. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020 Oct 17;396(10258):1204–22.

27. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. [Internet]. Seattle, United States: Institute for Health Metrics and Evaluation; 2020. Available from: http://ghdx.healthdata.org/gbd-results-tool

Appendix. GATHER checklist

Item #	Checklist item	Location reported
Objectives and funding		
1	Define the indicator(s), populations (including age, sex, and geographic entities), and time period(s) for which estimates were made.	Section 1
2	List the funding sources for the work.	Acknowledgments
Data Inputs		
For all data inputs from multiple sources that are synthesized as part of the study:		
3	Describe how the data were identified and how the data were accessed.	Section 2
4	Specify the inclusion and exclusion criteria. Identify all ad-hoc exclusions.	Section 2
5	Provide information on all included data sources and their main characteristics. For each data source used, report reference information or contact name/institution, population represented, data collection method, year(s) of data collection, sex and age range, diagnostic criteria or measurement method, and sample size, as relevant.	WHO Global Health Observatory
6	Identify and describe any categories of input data that have potentially important biases (e.g., based on characteristics listed in item 5).	Section 3
For data inputs that contribute to the analysis but were not synthesized as part of the study:		
7	Describe and give sources for any other data inputs.	Sections 2-4
For all data inputs:		
8	Provide all data inputs in a file format from which data can be efficiently extracted (e.g., a spreadsheet rather than a PDF), including all relevant meta-data listed in item 5. For any data inputs that cannot be shared because of ethical or legal reasons, such as third-party ownership, provide a contact name or the name of the institution that retains the right to the data.	WHO Global Health Observatory
Data analysis		
9	Provide a conceptual overview of the data analysis method. A diagram may be helpful.	Section 1
10	Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).	Sections 2-4
11	Describe how candidate models were evaluated and how the final model(s) were selected.	Section 4
12	Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.	Section 4
13	Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.	Section 4
14	State how analytic or statistical source code used to generate estimates can be accessed.	Acknowledgments
Results and Discussion		
15	Provide published estimates in a file format from which data can be efficiently extracted.	WHO Global Health Observatory
16	Report a quantitative measure of the uncertainty of the estimates (e.g.	WHO Global Health

	uncertainty intervals).	Observatory
17	Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.	Section 5
18	Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.	Section 6