

# Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005

Erin McLean<sup>1</sup>, Mary Cogswell<sup>2</sup>, Ines Egli<sup>3</sup>, Daniel Wojdyla<sup>4</sup> and Bruno de Benoist<sup>1,\*</sup>

<sup>1</sup>Department of Nutrition for Health and Development, World Health Organization, 20 Avenue Appia, CH-1211, Geneva 27, Switzerland; <sup>2</sup>Division of Nutrition and Physical Activity, Centers for Disease Control and Prevention (CDC), USA; <sup>3</sup>Institute of Food Science and Nutrition, Swiss Federal Institute of Technology, Zurich, Switzerland; <sup>4</sup>Escuela de Estadística, Universidad Nacional de Rosario, Argentina

Submitted 28 May 2007; Accepted 6 March 2008

## Abstract

**Objective:** To provide current global and regional estimates of anaemia prevalence and number of persons affected in the total population and by population subgroup.

**Setting and design:** We used anaemia prevalence data from the WHO Vitamin and Mineral Nutrition Information System for 1993–2005 to generate anaemia prevalence estimates for countries with data representative at the national level or at the first administrative level that is below the national level. For countries without eligible data, we employed regression-based estimates, which used the UN Human Development Index (HDI) and other health indicators. We combined country estimates, weighted by their population, to estimate anaemia prevalence at the global level, by UN Regions and by category of human development.

**Results:** Survey data covered 48·8% of the global population, 76·1% of preschool-aged children, 69·0% of pregnant women and 73·5% of non-pregnant women. The estimated global anaemia prevalence is 24·8% (95% CI 22·9, 26·7%), affecting 1·62 billion people (95% CI 1·50, 1·74 billion). Estimated anaemia prevalence is 47·4% (95% CI 45·7, 49·1%) in preschool-aged children, 41·8% (95% CI 39·9, 43·8%) in pregnant women and 30·2% (95% CI 28·7, 31·6%) in non-pregnant women. In numbers, 293 million (95% CI 282, 303 million) preschool-aged children, 56 million (95% CI 54, 59 million) pregnant women and 468 million (95% CI 446, 491 million) non-pregnant women are affected.

**Conclusion:** Anaemia affects one-quarter of the world's population and is concentrated in preschool-aged children and women, making it a global public health problem. Data on relative contributions of causal factors are lacking, however, which makes it difficult to effectively address the problem.

**Keywords**  
Anaemia  
Haemoglobin  
Nutritional status  
Iron deficiency

Anaemia, one of the most common and widespread disorders in the world, is a public health problem in both industrialised and non-industrialised countries. In 2002, the WHO estimated that anaemia resulting from iron deficiency was one of the ten most important factors contributing to the global burden of diseases and that it increases morbidity and mortality in preschool-aged children and pregnant women<sup>(1)</sup>. Anaemia is defined as a decrease in the concentration of circulating red blood cells or in the haemoglobin concentration and a concomitant impaired capacity to transport oxygen. It has multiple precipitating factors that can occur in isolation but more frequently co-occur<sup>(2)</sup>. These factors may be genetic, such as haemoglobinopathies; infectious, such as malaria, intestinal helminths and chronic infection; or nutritional, which includes iron deficiency as well as

deficiencies of other vitamins and minerals, such as folate, vitamins A and B<sub>12</sub>, and copper<sup>(2)</sup>.

Because iron deficiency makes a large contribution to anaemia, global efforts to reduce the anaemia burden have largely been directed towards increasing intake of iron through supplementation, food fortification and diversification of the diet. To assess the iron status of the population or the response to an intervention to prevent and control iron deficiency, haemoglobin concentration has often been used in surveys as a proxy indicator for iron status under the assumption that anaemia is always associated with iron deficiency, even if many other possible causes are present. These surveys have rarely measured iron deficiency or any of the other factors that contribute to the development of anaemia and therefore the contributing factors frequently remain unknown.

\*Corresponding author: Email debenoistb@who.int

Previous estimates of the prevalence of anaemia in the world were reported on population subgroups in 1982 and 1992<sup>(3,4)</sup> and on all population groups in 1985<sup>(5)</sup> and 2001<sup>(6)</sup>. With the exception of the most recent estimates, however, which included data up to 1995<sup>(6)</sup>, these reports did not include data collected after 1990. Moreover, the 1985 report by DeMaeyer and Tegman<sup>(5)</sup> did not include data for China, which represents 20% of the global population.

The objective of the present paper is to provide current global and regional estimates of the prevalence of anaemia and of the number of persons affected based on surveys conducted between 1993 and 2005 for the 192 Member States of the WHO. As a result of the vastly different methodologies used, these estimates are not quantitatively comparable to previous estimates.

## Methods

### Data source

We based the current estimates on data available in the WHO Global Database on Anaemia, a part of the Vitamin and Mineral Nutrition Information System (VMNIS) (<http://www.who.int/vmnis>). This database includes data on haemoglobin concentration and the prevalence of anaemia presented by country in a standardised, easily accessible format.

To establish the WHO Global Database on Anaemia, we systematically searched and collected data from the scientific literature (Medline and WHO regional databases) and through a broad network of partners, including WHO regional and country offices, UN organisations, ministries of health, research and academic institutions and non-governmental organisations. We augmented these resources by manual searching of articles published in non-indexed medical and professional journals and reports from principal investigators.

To include data in the WHO Global Database on Anaemia, we required a complete original survey report with details of the sampling method used. In a few cases, we accepted data provided in writing directly by ministries of health with detailed methodology, even without a formal published report. We included surveys representative of any administrative level and any population group in the WHO Global Database on Anaemia if they:

- were population based or facility based (for pregnant women, newborns, preschool- and school-aged children),
- were cross-sectional or baseline values from an intervention programme,
- measured haemoglobin concentration from capillary, venous or cord blood using quantitative photometric methods or automated cell counters and
- reported the prevalence of anaemia or mean haemoglobin concentrations.

We excluded surveys that measured only clinical signs of anaemia or the haematocrit and contacted study authors for clarification or additional information when necessary. The administrative level of a survey is national when the sample is nationally representative, or sub-national when the sample is representative of a given administrative level, namely, region, state (first administrative boundary), district (second administrative boundary) or local. Infrequently, surveys could be national even though some regions had to be left out for security or other reasons.

### Data selection

For this analysis, we used the following four variables in selecting data from the WHO Global Database on Anaemia on haemoglobin concentration and/or the prevalence of anaemia: the time frame of the survey, the administrative level for which the survey was representative (national or subnational), the sample size and the population groups surveyed.

The time frame for the estimates was from January 1993 to December 2005, and surveys that took place during this time period and were published by 31 December 2005 were eligible. As of that date, 696 surveys that reported on data collected between 1993 and 2005 were available. We used the publication date when the period of data collection was not specified.

We used data from the most recent national survey in preference to subnational surveys of more recent vintage. For one country, where an area had been left out of a national survey because of security concerns, available data from the missing region (weighted by the general population estimate for that area) were pooled with the national data to provide an estimate for the country. The estimate was determined by using the most recent census data from the country. The surveys were conducted within 1 year of each other and adding the missing region changed the overall estimate by only 0.1%. If two national surveys were conducted in the same year, as was the case for two population groups from one country, we pooled the survey results into a single summary measure, weighted by the sample size of the two surveys. The difference between the estimates in the two surveys was 5–15%, depending on the population group. In the absence of national data, we used surveys that were representative at the first administrative-level boundary if two or more surveys at this level were available for the population group and country concerned within the acceptable time frame. We pooled the results into a single summary measure, weighted by the total general population for that region or state, based on the most recent and available census data between 1993 and 2005, without considering the age range covered by the survey. We did not use local or district-level surveys in these estimates because they have the potential for more bias.

As a general rule, we excluded prevalence data based on a sample of fewer than 100 persons. Given a sample of 100 and a confidence level of 95%, the error around a prevalence estimate of 50% would be  $\pm 10$  percentage points; a smaller sample would have an even larger error. A few exceptions were made, however. National surveys with fewer than 100 but more than fifty participants were accepted but only where the results were being extrapolated to fewer than 500 000 people or to pregnant women.

For this analysis, we defined population subgroups as follows: preschool children below 5 years, school-aged children aged 5·00–14·99 years, pregnant women of any age, non-pregnant women aged 15·00–49·99 years, men aged 15·00–59·99 years and the elderly, aged 60+ years. Where possible, we excluded children below 0·5 years from preschool-aged children because an appropriate haemoglobin cut-off for this age group has not been determined<sup>(6)</sup>. We did not provide a separate population estimate for women aged 50·00–59·99 years, as these women are rarely surveyed. We did, however, include an estimate for women aged 50·00–59·99 years in our estimate of the global anaemia burden. The methods for accomplishing this are detailed later. Infrequently, if data were not disaggregated, we included all women in the estimate for non-pregnant women even if we did not know whether pregnant women were included. Where surveys provided data disaggregated by physiological status, lactating women and non-pregnant women were combined for the population subgroup non-pregnant women.

We used data disaggregated by the ages that were closest to the defined age ranges for the population subgroups. If the age range overlapped two population subgroups, we placed the survey with the subgroup where there was a greater overlap in age. When the age range was unavailable, we used the mean age of the sample to classify the data. If this was unavailable and the age range equally spanned two population subgroups, we used the population-specific haemoglobin concentration threshold to classify the data. If data represented less than 20% of the age range of a population group, we did not include the survey.

#### ***Prevalence of anaemia for countries with survey data***

Normal haemoglobin distributions vary with age, sex and physiological status, for example, pregnant (varies by trimester) and non-pregnant<sup>(7)</sup>. We used WHO haemoglobin thresholds to classify persons living at sea level as anaemic: children 0·50–4·99 years and pregnant women, 110 g/l; children 5·00–11·99 years, 115 g/l; children 12·00–14·99 years and non-pregnant women  $\geq 15$ ·00 years, 120 g/l; men  $\geq 15$ ·00 years, 130 g/l<sup>(6)</sup>. Statistical and physiological evidence indicate that haemoglobin distributions vary with smoking<sup>(8)</sup> and altitude<sup>(9)</sup> and, therefore, we used the prevalence of anaemia corrected for

these factors when provided by the survey. We did not accept any other corrections. For severe anaemia, we included surveys that used a haemoglobin cut-off of 70 g/l, which was used by almost all of the surveys that reported the prevalence of severe anaemia.

For surveys that classified persons by the WHO anaemia threshold, we used the reported prevalence data without any additional calculations. If prevalence was not reported, or was reported for a different threshold, we estimated the prevalence using one of the several methods, all of which assumed a normal distribution of haemoglobin concentrations. This would slightly overestimate the prevalence of anaemia in populations where it is high because population curves of haemoglobin concentrations would be skewed to the lower values. We used the following methods to estimate the prevalence of anaemia in order of preference:

1. We used the mean and SD of the haemoglobin concentration to estimate the proportion of persons falling below the appropriate haemoglobin cut-off for the population subgroup ( $n$  20). We validated this by assessing the correlation between the estimated and predicted prevalence of anaemia in surveys from the database where a mean, an SD and a prevalence for the WHO age- and sex-specific cut-off were provided. This relationship was plotted ( $n$  508), and for most surveys, the two figures were extremely close ( $r^2 = 0\cdot95$ ,  $P < 0\cdot001$ ) for all four cut-offs (haemoglobin concentration  $< 110$ , 115, 120, 130 g/l). Overall, predicted prevalence overestimated actual prevalence by 3·8 percentage points. For 6·5% of the surveys, estimated prevalence overestimated actual prevalence by 10 percentage points or more, and in these surveys overestimation averaged 16·3%.
2. When no SD was provided, but prevalence for a non-WHO cut-off and mean haemoglobin concentration were available ( $n$  3), we used these two figures to calculate the SD of the haemoglobin concentration by assuming a normal distribution within the population and deriving the  $Z$ -score for the prevalence in order to back-calculate the SD [ $SD = (\text{provided cut-off} - \text{mean haemoglobin}) / Z\text{-score for given prevalence}$ ]. Following this calculation, the mean and SD were used as above to derive the prevalence for the WHO cut-off.
3. Finally, for surveys ( $n$  23) that did not present the mean and SD or the prevalence at the recommended threshold, we estimated the prevalence of anaemia from the prevalence at an alternative threshold. We assumed that an average SD for the same population subgroup would be close to the actual SD in the survey. We calculated the mean SD of the haemoglobin concentration for each population subgroup from the surveys included in the estimates, which had data available for participants within the defined age range

of the population subgroup (mean sd: preschool-aged children: 13.79 g/l, school-aged children: 11.29 g/l, non-pregnant women: 13.65 g/l, pregnant women: 14.04 g/l, men: 14.53 g/l). We estimated the population mean haemoglobin concentration from the prevalence at the cut-off provided in the survey and the assumed sd and created a table to look up the anaemia prevalence at the recommended threshold.

Severe anaemia is at the lowest end of the haemoglobin distribution and varies much more than the prevalence of all anaemia. Thus, we could not use a normal distribution curve for haemoglobin to estimate prevalence where data were lacking, and no estimates of the prevalence of severe anaemia were generated for countries with data for other haemoglobin thresholds.

#### *Utilising aggregated and disaggregated data for country estimates*

When no disaggregated data were available and the prevalence of anaemia was reported for school-aged children using one non-WHO cut-off where two should have been used, we adjusted the prevalence for the WHO cut-off that applied to the group in the majority. Similarly, when two non-WHO cut-offs were used for one group, we adjusted the prevalence by assuming that the one that applied to the group in the majority had been used for the entire group.

In several cases, we combined data provided separately, such as data for women by physiological status or any population subgroup disaggregated by age. We combined the prevalence estimates, weighted by sample size, and if this information was missing for one of the groups, we assumed that it had the average number of participants of the other groups. If information on sample size was missing from all the data pooled, we gave them equal weight.

#### *95% Confidence intervals*

We considered each estimate of the prevalence of anaemia ( $\hat{p}$ ) as representative of the whole country, whether from national or subnational data. We constructed the 95% CI in the logit scale (logit transformation of the estimated prevalence of anaemia) and back-transformed it to the original scale to provide an interval estimation of the prevalence. We approximated the estimate of the variance in the logit scale as  $\text{var}(\hat{p}^*) \cong [n\hat{p}(1-\hat{p})]^{-1}$ , where  $\hat{p}^*$  is the logit transformation of the estimated prevalence  $\log(\hat{p}/(1-\hat{p}))$  and  $n$  is the size of the sample<sup>(10,11)</sup>. We used a design effect of 2 to compute the CI because most surveys used for the estimates employed cluster sampling but did not provide a design effect. Based on surveys that did provide their design effect, we considered that 2 was a good estimate of the typical effect in the surveys used. Finite population corrections were negligible given the small sampling fraction in all the countries. In a few country surveys

where sample size was unknown ( $n$  13), we assumed  $n$  was 100.

#### ***Estimated prevalence of anaemia for countries without national or eligible subnational data***

For countries without a national or eligible subnational estimate, we predicted the country's prevalence of anaemia from regression equations using the UN Human Development Index (HDI), which is a composite indicator of a life expectancy index, an education index and a wealth index<sup>(12)</sup>, and health indicators from the World Health Statistics database. Separate prediction equations for each population subgroup were based on countries with data on the prevalence of anaemia for that subgroup. We started with the most recent available HDI (2002) for the regression models because development and health are most often intertwined. We forced the models to include HDI and selected the model based on the adjusted  $R^2$  statistic. We solved problems with multicollinearity using the variance inflation factor and removing variables with a value  $>5$ . In all population groups, the models with covariates in addition to HDI improved the prediction of anaemia. In the elderly, the covariates added to HDI were not statistically significant, but the adjusted  $R^2$  improved by  $>40\%$  ( $P = 0.198$ ).

For the seventeen countries where the HDI was not available, we fitted a regression model using two of the same components and one proxy indicator for education (average years of schooling in adults instead of adult literacy and gross enrollment in school) to the group of countries with HDI estimates and derived an estimated HDI score<sup>(13-15)</sup>. The percentage of variation explained by the components was high (96%). Subsequently, we used HDI and estimated HDI to predict the prevalence of anaemia using a multiple regression model. Variables that we considered for inclusion were general health indicators available for almost all WHO Member States (at least 190) and are listed in Table 1<sup>(16)</sup>.

**Table 1** Potential variables for anaemia prediction equations (WHO Statistics, 2002)

Annual population growth rate (%)
Population in urban areas (%)
Immunization coverage for measles in under-1-year-olds (%)
Immunization coverage for DTP3 in under-1-year-olds (%)
Total expenditure on health (as % of GDP)
General government expenditure on health (as % of total government expenditure)
Per capita total expenditure on health (in international dollars)
Life expectancy at birth (females)
Life expectancy at birth (males)
Healthy life expectancy at birth (females)
Healthy life expectancy at birth (males)
Adult mortality rate (females)
Adult mortality rate (males)
Under-5 mortality rate
Neonatal mortality

DTP3, diphtheria, tetanus and pertussis; GDP, gross domestic product.

Worldwide anaemia prevalence

We performed a diagnostic analysis to assess outliers and identify observations with a large impact on the regression coefficients. No more than one outlier was present in each population subgroup, and the effect of removing them was negligible. The percentage of observations with a large impact on the estimated coefficients was approximately 6–11%, but it was slightly higher in the model for school-aged children.

Where only explanatory variables were known, we estimated the prevalence of anaemia by using the prediction equations (Table 2). For one country, none of the covariates were available, and thus we did not generate a country-level estimate.

For severe anaemia, we found that the prevalence was related to HDI in a curvilinear manner such that the proportion of severe anaemia was much higher in countries with low HDI scores, but we did not have sufficient data to develop a reliable prediction model of the prevalence of anaemia and so we did not make estimates of the prevalence of severe anaemia for countries without survey data.

95% Confidence intervals

We computed point estimates and 95% prediction intervals for the prevalence of anaemia by using the logit transformations in the regression models<sup>(17)</sup> and then back-transforming them to the original scale<sup>(18)</sup>.

Estimates not used in regression models

After completion of the estimates, we conducted a consultation with each of the WHO Member States to allow them to review their data before publication. During this process, it appeared that seven countries had reports that had been missed for the estimates for one or more population groups but that were published within the time frame for the estimates (before December 2005). Of these, we had used an older estimate for one country and regression-based estimates for the remaining six. We did not regenerate the regression models with the new data, but these seven estimates were replaced. In addition, we replaced estimates for three countries for one population group because errors were identified, usually a typing error in the report. The change in the estimate was 0.1% in two cases and 3.8% in the third. The impact of these changes on the global and regional estimates was negligible, 0.1% and <1.0%, respectively.

Prevalence of anaemia for all population groups

To estimate the prevalence of anaemia in the entire population of a nation or an area within a nation, we pooled the number of people affected in each population subgroup and divided by the total population to derive the prevalence. The only segment of the population missing from these estimates was women aged 50–59 years. For this segment of the population, we applied the estimate for the elderly, for several reasons. The median

Table 2 Equations used to predict the prevalence of anaemia for the specified population subgroups in countries with no eligible survey data

Population group	No. of countries	Equation to predict anaemia prevalence	R <sup>2</sup>	P value for model
Preschool-aged children*	82	$3.5979 - 4.9093 \times \text{HDI} - 0.0657 \times \text{Exp on health} - 0.0003 \times \text{Exp on health per capita} - 0.0009 \times \text{Adult female mortality}$	0.550	<0.0001
Pregnant women	60	$2.7783 - 2.8352 \times \text{HDI} - 0.0085 \times \text{Imm DTP3} - 0.0004 \times \text{Exp on health per capita} - 0.0017 \times \text{Adult Male mortality}$	0.323	<0.0001
Non-pregnant women	79	$0.9475 - 2.3447 \times \text{HDI} + 0.1643 \times \text{Population growth rate} - 0.0697 \times \text{Exp on health}$	0.453	<0.0001
School-aged children	35	$1.4248 - 2.6894 \times \text{HDI} + 0.0087 \times \text{Urban population} - 0.0129 \times \text{Imm measles} - 0.0005 \times \text{Exp on health per capita}$	0.583	<0.0001
Men	32	$0.0991 - 4.6160 \times \text{HDI} + 0.0209 \times \text{Imm DTP3} - 0.0828 \times \text{Gov exp on health}$	0.577	<0.0001
Elderly	13	$-1.6693 + 0.2872 \times \text{HDI} - 0.1359 \times \text{Exp on health} + 0.0047 \times \text{Adult male mortality}$	0.385	0.0627

HDI, United Nations Human Development Index; Exp, expenditure; Imm DTP3, immunization for diphtheria, tetanus and pertussis. \*Population subgroups: Preschool-aged children (0.00–4.99 years), Pregnant women, Non-pregnant women (15.00–49.99 years), School-aged children (5.00–14.99 years), Men (15.00–59.99 years), Elderly (≥60.00 years).

age of menopause in women is approximately 50.5 years<sup>(19)</sup>, suggesting that for the majority of women in this age group, losses of iron from menstruation have stopped. Further, the data from the National Health and Nutrition Examination Survey (NHANES) in the United States were compared by one of the authors of the paper (M.C.) between women aged 20–49 years, 50–59 years and 60+ years, and women aged 50–59 years had a haemoglobin distribution that was more similar to women aged 60+ years than to those aged 20–49 years. In addition, the distribution of C-reactive protein (CRP) was more similar between women aged 50–59 years and 60+ years. The proportion of anaemia attributable to elevated CRP in women aged 50–59 years was more similar to those aged 20–49 years.

### **Combining national estimates**

For each population subgroup, we combined prevalence for countries grouped by geographical location or level of development, based on HDI, using a weighted average of the estimates for the countries. Weights were proportional to the number of persons in the population subgroup in each country. We derived confidence limits for combined prevalence estimates for the countries by using the estimated variance of a weighted average. We calculated the number of persons suffering from anaemia in a given population subgroup by multiplying the estimated prevalence (point and confidence limits) in the subgroup by the total population in that subgroup.

For one country, the indicators used for the regression-based estimates were not available, and so we did not generate an estimate for that country. Thus, for regional, global and development group estimates, we applied the estimate for this country's UN subregion, which was well covered by survey data.

### *Anaemia by category of development*

We classified countries by category of development, using the standard UN groupings, based on the HDI: high (HDI > 0.800), medium (0.500 ≤ HDI ≤ 0.800) and low (HDI < 0.500). For the seventeen countries with no official HDI score, we used their regression-based estimate of HDI to classify them.

### **Population coverage, proportion of population and the number of persons with anaemia**

#### *Population coverage*

We produced estimates only for WHO Member States, but these countries together represent 99.8% of the global population. The population covered by survey data for a given grouping of countries (global, grouped by UN region or level of development) was calculated for each grouping as the sum of the number of persons in the population subgroup in countries with survey data divided by the total number of persons in the population subgroup in the WHO Member States of the specific

grouping. Coverage when including countries with a regression-based 'proxy' estimate is not presented because it was similar for all population subgroups and included all countries except for one (99.7–99.9% of the WHO member state population depending on the population subgroup).

### *Proportion of population and the number of people with anaemia*

We estimated the number of people with anaemia in each population subgroup for each country and each grouping of countries based on each country's proportion of the population with anaemia. We multiplied the proportion of the population subgroup with anaemia by its national population to determine the number of persons with anaemia at the country level and provided the 95% CI as a measure of uncertainty. The population figures are for the 2006 projection from the 2004 revision of the UN population estimates<sup>(20)</sup>. We derived population figures for pregnant women from the total number of births (time period 2005–10) by assuming one child per woman per year, not taking into account spontaneous and induced abortions. For fifteen countries with a small total population (they represented 0.01% of all women), birth data were not provided in tabulations of the UN population division, and here we estimated the number of pregnant women by applying a WHO regional average of births per reproductive-age woman (15.00–49.99 years) to the total number of reproductive-age women.

## **Results**

### **Coverage**

All countries, except for one, were covered by actual data or by regression-based estimates. Data from national or subnational surveys covered almost three-quarters of the global population of preschool-aged children and non-pregnant women and almost 70% of the population of pregnant women but less than half of the populations of school-aged children, men and the elderly (Table 3). Only the estimates for preschool-aged children included subnational estimates (three countries, 1.9% of the population), but national estimates still covered more than 70% of this population. Because of the low coverage for school-aged children, men and the elderly throughout the world, estimates on the prevalence of anaemia for the entire population were generated at the global level and by the level of development, but not by country. The global estimate of anaemia is based on coverage from national and eligible subnational surveys of almost half of the world's population.

### **Prevalence of anaemia**

The global figures and number of persons affected are displayed in Fig. 1 for each population subgroup.

**Table 3** Percentage of population covered and number of countries with anaemia survey data

Category	PreSAC*	NPW	PW	SAC	Men	Elderly	All
Global (192)†	76.1 (84)‡	73.5 (82)	69.0 (64)	33.0 (36)	40.2 (34)	39.1 (13)	48.8
UN regions§							
Africa (53)	76.7 (30)	63.6 (26)	65.3 (25)	18.6 (10)	32.0 (14)	1.8 (1)	40.7
Asia (47)	82.1 (30)	88.8 (34)	80.9 (21)	37.0 (11)	47.6 (13)	54.1 (7)	58.0
Europe (41)	19.2 (5)	23.9 (5)	0.9 (1)	12.9 (3)	15.9 (2)	8.7 (2)	14.9
L. America and the Caribbean (33)	70.5 (15)	37.5 (12)	38.4 (14)	28.9 (8)	0.1 (1)	0.0 (0)	22.9
N. America (2)	92.4 (1)	89.9 (1)	92.8 (1)	91.3 (1)	89.9 (1)	89.6 (1)	84.3
Oceania (16)	5.1 (3)	16.5 (4)	4.7 (2)	15.1 (3)	15.6 (3)	15.1 (2)	13.8
Level of development							
High (59)	50.5 (8)	59.3 (12)	15.1 (4)	51.3 (11)	50.0 (10)	46.2 (7)	48.8
Medium (94)	80.5 (51)	76.5 (46)	73.2 (36)	36.4 (19)	40.6 (14)	39.1 (6)	51.0
Low (39)	67.9 (23)	68.1 (21)	58.0 (20)	7.1 (5)	18.4 (8)	0.0 (0)	37.6

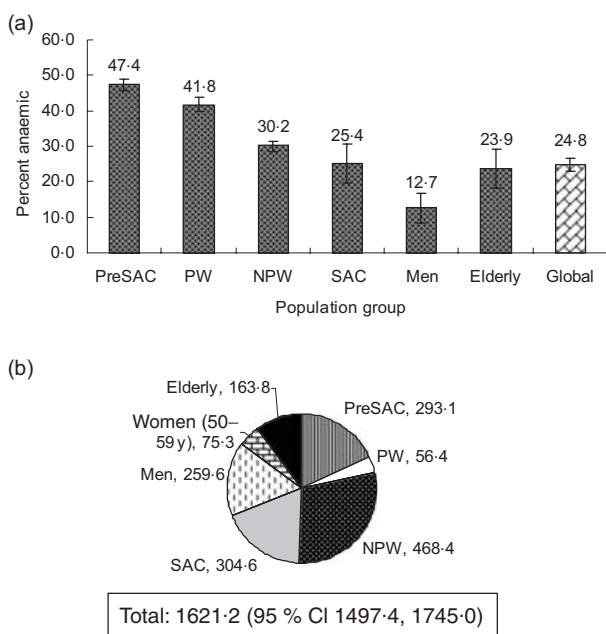
\*Population subgroups: PreSAC, preschool-aged children (0.00–4.99 years); NPW, non-pregnant women (15.00–49.99 years); PW, pregnant women; SAC, school-aged children (5.00–14.99 years); Men (15.00–59.99 years); Elderly ( $\geq 60.00$  years)

†Number of countries in each grouping.

‡Total number of countries with data, no figure is provided for All since each country may be partially covered by some population groups, but few countries have data on all six population groups and no countries have data for 50–59-year-old women.

§UN regions: Africa, Asia, Europe, Latin America and the Caribbean (LAC), Northern America (NA), Oceania.

||Level of development as classified by the UN Human Development Index (HDI): high (HDI score:  $>0.800$ ), medium (HDI score:  $0.500–0.800$ ), low (HDI score:  $<0.500$ ).



**Fig. 1** (a) Global prevalence of anaemia (%) and (b) number of individuals (millions) affected in different population groups (population subgroups: PreSAC, preschool-aged children (0.00–4.99 years); PW, pregnant women; NPW, non-pregnant women (15.00–49.99 years); SAC, school-aged children (5.00–14.99 years); Men (15.00–59.99 years); Elderly ( $\geq 60.00$  years); the number of individuals affected includes a figure for women aged 50.00–59.00 years, which is based on the estimate of anaemia prevalence in the elderly)

The global prevalence of anaemia is 24.8% (95% CI 22.9, 26.7%), and 1.62 billion people (95% CI 1.50, 1.74 billion) are affected. In numbers, non-pregnant women are the most affected (468.4 million, 95% CI 446.2, 490.6), but the highest prevalence is in preschool-aged children (47.4%, 95% CI 45.7, 49.1%).

For preschool-aged children, non-pregnant and pregnant women, estimates by UN regions are presented in Table 4. Africa has the highest prevalence of anaemia for all three population groups, but the greatest number of people affected are in Asia, where 58.0%, 56.1% and 68.0% of the anaemia burden in preschool-aged children, pregnant women and non-pregnant women, respectively, exists. The majority of these people live in south-central Asia (data not shown).

When the countries are considered by category of development, the prevalence of anaemia decreases from the low to the high category, but the majority of those affected by anaemia live in countries in the medium-development category because these countries account for 68.2% of the global population, while the countries classified in the low- and high-development categories account for only 13.1% and 18.6% of the global population, respectively. On a global scale, 9.1%, 25.7% and 42.8% of the population in countries in the high-, medium- and low-development categories are affected, respectively, resulting in 111 million (95% CI 102, 120 million), 1.1 billion (95% CI 1.0, 1.3 billion) and 367 million (95% CI 336, 398 million) people in these groups suffering from anaemia. The prevalence of anaemia and the incremental burden of anaemia by human development category are displayed for each of the population subgroups and the entire population in Fig. 2.

For preschool-aged children, data on the prevalence of severe anaemia covered 38.5% of the population and less than 50% of the population in all regions except North America, which had only one country with data, and the estimates were unreliable. In pregnant women, the data on severe anaemia are even fewer, with only 14.1% of the global population of pregnant women covered by such data. Therefore, separate estimates for severe anaemia in these two subgroups of concern could not be made. In the

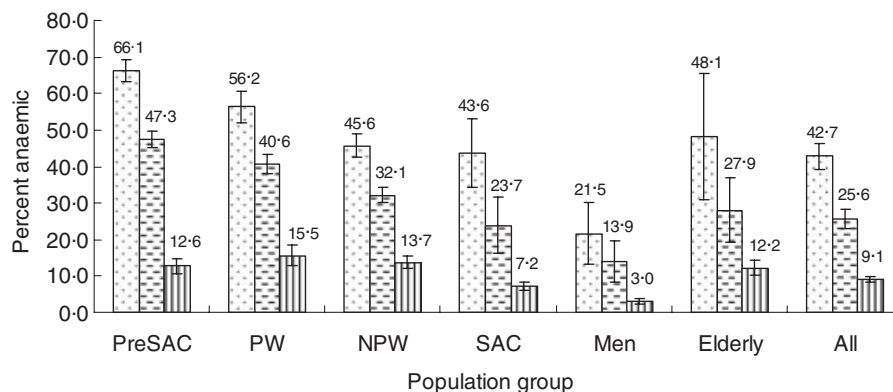
**Table 4** Anaemia in preschool-aged children, non-pregnant women and pregnant women

UN region*	Pre-SAC†		NPW		PW	
	Prevalence (%)	No. affected	Prevalence (%)	No. affected	Prevalence (%)	No. affected
Africa	64.6 (61.7, 67.5)‡	93.2 (89.1, 97.4)	44.4 (40.9, 47.8)	82.9 (76.5, 89.4)	55.8 (51.9, 59.6)	19.3 (18.0, 20.7)
Asia	47.7 (45.2, 50.3)	170.0 (161.0, 178.9)	33.0 (31.3, 34.7)	318.3 (302.0, 334.6)	41.6 (39.0, 44.2)	31.7 (29.7, 33.6)
Europe	16.7 (10.5, 23.0)	6.1 (3.8, 8.4)	15.2 (10.5, 19.9)	26.6 (18.4, 34.9)	18.7 (12.3, 25.1)	1.4 (0.9, 1.8)
LAC	39.5 (36.0, 43.0)	22.3 (20.3, 24.3)	23.5 (15.9, 31.0)	33.0 (22.4, 43.6)	31.1 (21.8, 40.4)	3.6 (2.5, 4.7)
NA	3.4 (2.0, 4.9)	0.8 (0.4, 1.1)	7.6 (5.9, 9.4)	6.0 (4.6, 7.3)	6.1 (3.4, 8.8)	0.3 (0.2, 0.4)
Oceania	28.0 (15.8, 40.2)	0.7 (0.4, 1.0)	20.2 (9.5, 30.9)	1.5 (0.7, 2.4)	30.4 (17.0, 43.9)	0.2 (0.1, 0.2)

\*UN regions: Africa, Asia, Europe, Latin America and the Caribbean (LAC), Northern America (NA) and Oceania.

†Population subgroups: Pre-SAC, preschool-aged children (0.00–4.99 years); NPW, non-pregnant women (15.00–49.99 years); PW, pregnant women.

‡95% confidence interval.



**Fig. 2** Prevalence of anaemia by categories of development (UN Human Development Index groupings of countries by category of development: ■, High (HDI > 0.800); □, Medium (0.500 ≤ HDI ≤ 0.800); ▨, Low (HDI < 0.500)) for population subgroups and all individuals (population subgroups: PreSAC, preschool-aged children (0.00–4.99 years); PW, pregnant women; NPW, non-pregnant women (15.00–49.99 years); SAC, school-aged children (5.00–14.99 years); Men (15.00–59.99 years); Elderly (≥60.00 years); All includes preceding population groups and women aged 50.00–59.99 years))

countries with data for preschool-aged children, severe anaemia averages 5.9% as a proportion of all anaemia, but this figure ranges from an average of 2.0% in the four countries with data in Latin America and the Caribbean to an average of 8.8% in the seventeen countries covered in Africa. By country, prevalence ranges from 0.4% in one country in Asia to 20.4% in a country in Africa.

## Discussion

These estimates represent the most recent and accurate data available on the worldwide prevalence of anaemia. We found that between 1993 and 2005, anaemia affected one in four persons globally: pregnant women and young children are at greatest risk. Geographically, people living in Asia and Africa are at greatest risk: almost two-thirds of preschool-aged children living in Africa are anaemic.

Assessing global progress is difficult because the methodologies used for these and previous estimates vary significantly. We believe, however, that our estimates include three major improvements. First, our global estimate includes nationally representative data for China, which accounts for 20% of the global population; the

earlier global estimate of anaemia did not include China<sup>(5)</sup>.

Second, in the past, few nationally representative surveys were available, and the estimates were based primarily on data from regional, state and local surveys. Clearly, national surveys more accurately represent the total population, especially as regional, state and local surveys may be conducted in locations with an unusually low or high prevalence of anaemia. In some instances, survey locations are chosen because of a particular concern about a health condition or economic change, and thus the findings may overestimate the prevalence of anaemia for the entire country. At other times, areas may be selected because of accessibility, and they may be better off economically than remote areas of a country. The use of national surveys should help eliminate bias in either direction. For our estimates, we used almost all nationally representative surveys, and the percentage of the population covered by these surveys remained high for preschool-aged children, pregnant women and non-pregnant women.

Finally, we used regression estimates for countries without data from eligible surveys; we found that a substantial proportion of the variation in anaemia within a

population subgroup was explained by the economic and health indicators of the countries used for the regression analysis (32–55%). Previous estimates of anaemia for countries with missing data used information from neighbouring countries or applied the anaemia estimate from all countries with data to those without data for the specified population subgroup. If we made the latter assumption, our prevalence estimates would have been slightly higher in preschool-aged children, non-pregnant and pregnant women (1.2–2.3 percentage points), and somewhat lower in school-aged children, men and the elderly (1.9–8.5 percentage points), where there are fewer data. In comparison with our estimate of 25%, DeMaeyer estimated that 30% of the world's population was anaemic around 1980 (1960–85). The methods used to derive DeMaeyer's global estimate are unclear, although it seems to be based on an extrapolation of the estimates for preschool-aged children, school-aged children, men and women. By subgroup, DeMaeyer's estimates (which excluded China) were 43% for preschool-aged children, 35% for all women and 51% in pregnant women. If we exclude China from our estimates, our corresponding estimates are 52%, 34% and 44%. As stated above, variation in methods and our larger number of nationally representative data may explain the difference in the estimates, rather than an actual change in anaemia status.

In 1992, WHO published global prevalence figures for 1988 of 37%, 51% and 35% for all women, pregnant women and non-pregnant women, respectively. These estimates include subnational data for China. Again, the current estimates of 31%, 42% and 30% may not be different when the substantial difference in methods is considered.

Our estimates are subject to limitations. First, because surveys less frequently include data on school-aged children, men and the elderly, in some regions the number of countries with data on these population groups was limited or non-existent. Therefore, we present only global estimates for these groups, but even these estimates should be interpreted with some caution.

Second, many assumptions had to be made to derive our global estimates. All surveys were treated equally, but in actuality surveys vary greatly in quality in their selection of samples, presence or absence of adjustment for smoking and altitude, and numerous other factors. Some estimates covered only a portion ( $\geq 20\%$ ) of the population subgroup, which may have resulted in an artificially high or low estimate being applied to that population subgroup, as some surveys of preschool-aged children focus on younger children, who have a higher prevalence of anaemia, while others focus on older children. For only three countries, subnational data representative of at least two first administrative-level divisions were treated similarly to national data, because data covering a substantial segment of the population for a country were preferable to proxy estimates. Even so, because the surveys are not

nationally representative, they may underestimate or overestimate the actual anaemia prevalence, as described previously.

For some countries we calculated prevalence using assumptions about the distribution of haemoglobin concentration because the prevalence of anaemia was not reported using the appropriate threshold for this concentration. Overall, this may have led to a slight overestimation of prevalence, because haemoglobin was assumed to be normally distributed for these calculations even though it is not.

A design effect of 2 was applied to all surveys because in the few surveys that provided a design effect, the average size was 1.6–1.8. However, individual surveys may have had design effects larger or smaller than this figure. This may have resulted in narrower or wider CI for the regional and global estimates.

The estimates based on grouping by HDI development are based on prevalence estimates for countries, some of which used regression equations that were based on HDI; this is a limitation of these estimates. The coverage of these groups by actual data, however, was high for preschool-aged children and non-pregnant women as well as for pregnant women in the low and medium categories of development, and they do provide useful information. For example, it is notable that there is more than a fivefold increase in anaemia prevalence in preschool-aged children from the high to low category of development, and this is based on a substantial amount of actual prevalence data. Also of interest is that, for the high category of development, prevalence in the elderly is similar to that of preschool-aged children or non-pregnant women, with comparable survey coverage. This may indicate that prevalence in the elderly for the other categories of development, where survey coverage is poor, is underestimated.

Finally, the estimates for pregnant women do not take into account the trimester at the time of assessment. Women in the first trimester may have a lower risk of anaemia than non-pregnant women because menstruation has stopped and the increases in blood cell volume and the growth of the fetus and placental tissues are minimal. In the second and third trimester, however, increases in fetal growth and expansion in red blood cell mass increase the risk of anaemia *v.* that borne by non-pregnant women. Thus, variation in the gestational age at the time of measurement may account for differences in prevalence estimates by country. For example, if a substantial proportion of women were assessed in their first trimester, the prevalence of anaemia among pregnant women may be lower than that among non-pregnant women. Another possible reason for variations in the prevalence of anaemia among pregnant women may be variation in consumption of supplements of multivitamins or minerals. In some countries, pregnant women may be more likely than non-pregnant women to consume these supplements<sup>(21)</sup>.

Anaemia results from the interaction of several causal factors that may vary from one population to another. Clearly, knowledge of the cause of anaemia is required to fully interpret data on prevalence and to design appropriate interventions to reduce it. Only a few surveys looked at causation, and the ones that did focused on iron deficiency. Even so, there are few data available on iron status, which makes it difficult to generate estimates of iron deficiency or of iron deficiency anaemia. We know, however, that anaemia may reflect iron status where iron deficiency is its main cause and therefore may be used appropriately as a proxy indicator. For instance, data from some countries, like the USA, where the prevalence of anaemia is extremely low (3.1–6.9%) suggest that it reflects the impact of the increased consumption of iron due to iron fortification of commercial foods as they contribute 20–25% to the total iron intake<sup>(22)</sup>.

In conclusion, the data available now are more plentiful and more representative than they were for any of the previous estimates, and thus we may have the most accurate picture to date of the prevalence of anaemia. Still, countries without data should be encouraged to survey their population in order to have a more accurate picture of prevalence and should also be encouraged to include assessment of helminth infection, malaria and iron status to better understand the aetiology of anaemia within their country. These estimates of prevalence are valuable because they allow the comparison of anaemia status among countries in high-risk groups and permit tracking of the progress of various countries in eliminating this scourge. They also provide useful information to assess how effective the current strategies are to control anaemia, but this information needs to be interpreted with caution. Indeed, the majority of the available surveys did not collect data on primary causes, so that their usefulness for deciding on the most effective strategies to combat anaemia is limited. Ideally, these estimates will draw the attention of the public health community to the need to assess the prevalence of factors that contribute to the development of anaemia, not only iron deficiency, but also parasitic and infectious diseases, and to determine how these causes vary by geography, level of development and other social and economic factors. This will make it easier to design more effective interventions that integrate and take into account all of these factors.

### Acknowledgements

*Sources of funding:* We would like to acknowledge CDC grant 5 U50 DP423754-04 as a source of funding for this work.

*Conflict of interest:* The authors declare that they have no conflict of interest.

*Authors' contributions:* E.M. worked on the databank that provided data for the paper, performed data analysis

and wrote the paper. M.C. provided input into the analysis of the data and the writing of the report. I.E. worked on the databank that supplied the data for the estimates, provided input for the analysis and reviewed the paper. D.W. performed statistical analyses. B.B. initiated the databank, provided input for data analysis and contributed to the writing of the paper.

*Disclaimer:* Two of the authors are staff members of the World Health Organization. They alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions, policy or views of the World Health Organization. Moreover, the findings and conclusions in this report are those of the authors and do not necessarily represent the views of the CDC.

### References

1. World Health Organization (2002) The World Health Report 2002: Reducing risks, promoting healthy life. [http://www.who.int/whr/2002/en/whr02\\_en.pdf](http://www.who.int/whr/2002/en/whr02_en.pdf) (accessed April 2007).
2. World Health Organization (2004) Focusing on anaemia: towards an integrated approach for effective anaemia control. Joint statement by the World Health Organization and the United Nations Children's Fund. <http://whqlibdoc.who.int/hq/2004/anaemiastatement.pdf> (accessed April 2007).
3. Royston E (1982) The prevalence of nutritional anaemia in women in developing countries: a critical review of available information. *World Health Stat Q* **35**, 52–91.
4. World Health Organization (1992) The prevalence of anaemia in women: a tabulation of available information. WHO/MCH/MSM/92.2. [http://whqlibdoc.who.int/hq/1992/WHO\\_MCH\\_MSM\\_92.2.pdf](http://whqlibdoc.who.int/hq/1992/WHO_MCH_MSM_92.2.pdf) (accessed April 2007).
5. DeMaeyer E & Adiels-Tegman M (1985) The prevalence of anaemia in the world. *World Health Stat Q* **38**, 302–316.
6. World Health Organization (2001) Iron deficiency anaemia: assessment, prevention, and control. A guide for programme managers. WHO/NHD/01.3. [http://www.who.int/nutrition/publications/en/ida\\_assessment\\_prevention\\_control.pdf](http://www.who.int/nutrition/publications/en/ida_assessment_prevention_control.pdf) (accessed April 2007).
7. Koller O (1982) The clinical significance of hemodilution during pregnancy. *Obstet Gynecol Surv* **37**, 649–652.
8. Nordenberg D, Yip R & Binkin NJ (1990) The effect of cigarette smoking on hemoglobin levels and anemia screening. *JAMA* **264**, 1556–1559.
9. Hurtado A, Merino C & Delgado E (1945) Influence of anoxemia on haematopoietic activities. *Arch Intern Med* **75**, 284–323.
10. Lohr SL (1998) *Sampling Design and Analysis*, 1st ed. Pacific Grove, CA: Duxbury Press.
11. Wackerly D, Mendenhall W & Scheaffer RL (2001) *Mathematical Statistics with Applications*, 6th ed. Pacific Grove, CA: Duxbury Press.
12. Human Development Indicators (2004) In *Human Development Report 2004: Cultural Liberty in Today's Diverse World*, pp. 139–250 [C Murphy, B Ross-Larson, editors]. New York: United Nations Development Programme; available at <http://www.un.org/special-rep/ohrlls/ldc/Global-Reports/HDR%202004.pdf>
13. World Health Organization (2004) The World Health Report: 2004: Changing History. [http://www.who.int/whr/2004/en/report04\\_en.pdf](http://www.who.int/whr/2004/en/report04_en.pdf) (accessed April 2007).

14. World Health Organization (2000). The World Health Report: 2000: Health systems: improving performance. [http://www.who.int/whr/2000/en/whr00\\_en.pdf](http://www.who.int/whr/2000/en/whr00_en.pdf) (accessed April 2007).
15. Mathers CD & Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* **3**, e442.
16. World Health Organization (2005) World Health Statistics 2005. <http://whqlibdoc.who.int/publications/2005/9241593261.pdf> (accessed April 2007).
17. Neter J, Kutner MH, Wasserman W & Nachtsheim CJ (1996) *Applied Linear Statistical Models*, 4th ed. New York: McGraw-Hill/Irwin.
18. Allison PD (2001) *Logistic Regression Using the SAS System*. Indianapolis, IN: WA (Wiley-SAS).
19. Whelan EA, Sandler DP, McConaughy DR & Weinberg CR (1990) Menstrual and reproductive characteristics and age at natural menopause. *Am J Epidemiol* **131**, 625–632.
20. United Nations Population Division (2005). World Population Prospects – the 2004 revision. New York.
21. Cogswell ME, Kettel-Khan L & Ramakrishnan U (2003) Iron supplement use among women in the United States: science, policy and practice. *J Nutr* **133**, 1974S–1977S.
22. Hurrell RF & Jacob S (1996) Role of the food industry in iron nutrition: Iron intake from industrial food products. In *Iron Nutrition in Health and Disease*, pp. 339–347 [L Hallberg and NG Asp, editors]. London: John Libbey & Co.