Establishing Pediatric and Adult RBC Reference Intervals With NHANES Data Using Piecewise Regression

Victor L. Fulgoni, III, PhD,[1] Sanjiv Agarwal, PhD,[2] Mark D. Kellogg, PhD,[3] and Harris R. Lieberman, PhD[4]

From the [1]Henry M. Jackson Foundation, Bethesda, MD; [2]Oak Ridge Institute for Science and Education, Belcamp, MD; [3]Department of Laboratory Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA; and [4]Military Nutrition Division, US Army Research Institute of Environmental Medicine, Natick, MA.

Key Words: Hemoglobin; Hematocrit; MCHC; Mean cell volume; RDW

Am J Clin Pathol February 2019;151:128-142

DOI: 10.1093/AJCP/AQY116

ABSTRACT

Objectives: *To develop age- and sex-specific RBC reference intervals using the National Health and Nutrition Examination Survey (NHANES) 1999 to 2012, a large nationally representative, population-based, cross-sectional database (n = 44,328).*

Methods: *Comprehensive medical data were used to define a "healthy" population. Reference intervals for RBC count, hemoglobin, hematocrit, mean cell hemoglobin, mean cell hemoglobin concentration, mean cell volume, and red cell distribution width were computed using piecewise regression, an evidence-based statistical procedure that identifies breakpoints.*

Results: *The derived reference intervals were sex specific, unlike many current standards, and more precise for individuals of different ages, especially for children, adolescents, and elderly individuals, as additional breakpoints were detected for these groups. Suggested reference values for hematocrit and hemoglobin of older adult males were substantially lower than current values.*

Conclusions: *The reference intervals provided here, based on a large, nationally representative healthy population, contribute to the ongoing transition to precision medicine.*

RBC laboratory values are among the most important clinical laboratory tests used by health care providers for diagnosing diseases, monitoring various medical conditions, and assessing nutritional status.[1] Anemia, which is diagnosed with data from the CBC, is a global health problem affecting about 25% of the population worldwide, with highest prevalence among preschool-aged children and pregnant women.[2] In a survey of female US Army personnel, the prevalence of iron deficiency anemia immediately following basic combat training has been reported to be about 21%, with higher prevalence among Hispanic and African-American military personnel than among white military personnel.[3] Anemia is typically identified by the presence of low hemoglobin levels; however, other CBC measures, in particular RBC parameters, are surrogate measures of iron availability and are essential for optimal classification, monitoring, and management of anemia.[4,5] Measures collected as part of the CBC parameters are also used as a screening tool for a wide range of conditions and diseases, including infection, inflammation, neoplasia, and systemic disease and are also markers for specific types of leukemia.[6-8]

RBC reference intervals determine the upper and lower limits for normal test results and are typically defined as 95% (±2.5% upper and lower limits) of the measured values from a healthy population. They provide baseline values that are used to interpret clinical laboratory test results and for diagnosis and disease management.[7,9] RBC parameters vary with demographic factors such as age, sex, race/ethnicity, and lifestyle factors such as smoking.[10,11] Existing reference values are derived from a broad range of sources, including published research,

individual laboratory studies, manufacturers' studies, and convenience samples of various populations.[12] According to the Clinical Laboratory Standards Institute (CLSI) guidelines, reference interval values can be developed based on as few as 120 healthy reference individuals, a minimal requirement to achieve statistical significance.[13] A critical issue for establishment of valid reference values is collection of data from a well-defined "healthy" population. However, there is no universal definition of "healthy,"[10] and comprehensive, systematic data defining an individual's health status may not be available in clinical databases. Furthermore, because demographic and lifestyle factors like age, sex, race/ethnicity, and smoking habits affect RBC values, larger sample sizes and stratification of the population studied may be required.

The National Health and Nutrition Examination Survey (NHANES) is a continuous, very large, nationally representative, population-based, cross-sectional survey conducted by the Centers for Disease Control and Prevention (CDC) to monitor health and dietary status of Americans.[14] Because NHANES surveys are conducted using the same state-of-the-art techniques and standardized procedures, extremely large data sets (>60,000) can be obtained by combining multiple years of data. To date, only a limited number of studies have used NHANES data to derive reference intervals. Data from NHANES III (1988-1994) were used to derive CBC reference intervals by age, sex, and race/ethnicity.[11] However, these are based on data collected over 25 years ago and should be updated given changes in the US population, including its race/ethnic composition and average body mass index (BMI) as well as changes in laboratory methods.

Various approaches have been used to determine reference intervals. Piecewise regression, also called segmented or spline regression, is a form of regression analysis that fits multiple linear models across the full range of a variable. It is an objective procedure to determine inflection points in a variable of interest and is used to find breakpoints that are values above or below which statistical differences are present. The technique of piecewise regression was explicitly designed to detect discontinuities in data. Piecewise regression formally tests whether there is a nonlinearity in continuous data and whether an abrupt transition explains the nonlinearity.[15] This method is appropriate for development of reference values because it is an objective statistical procedure frequently used to identify breakpoints in variables of interest. The piecewise technique has been used in nutrition research to demonstrate nonlinear relationships between BMI and mortality and also between calcium intake and bone mineral density using NHANES III data.[16,17] This technique is also used by the CDC to adjust certain NHANES outcome variables (specifically laboratory measures) when assessment methodologies have changed or drifted over time.[18]

We hypothesized piecewise regression could be used to analyze pooled NHANES survey data to develop age-appropriate RBC reference intervals. The very large sample size (n = >60,000) in NHANES from years 1999 to 2012 allowed us to fit models by year of age within and across both sexes. We concomitantly used the comprehensive health data in NHANES to ensure we were studying a "healthy" population.

Materials and Methods

Study Population

Data from seven NHANES cycles (1999-2000, 2001-2002, 2003-2004, 2005-2006, 2007-2008, 2009-2010, and 2011-2012) were extracted for the current analysis. The study participants were free-living children (age 1-19 years, n = 30,545) and adults (age 20-79 years, n = 34,922) who participated in the seven NHANES cycles and had complete RBC results. Volunteers were excluded from analysis if they reported the following medical conditions: anemia, coronary heart failure, coronary heart disease, angina, acute myocardial infarction, cancer, gout, emphysema, hepatic disease, celiac disease, stroke, thyroid condition, hepatitis C or HIV, or those taking daily aspirin (data for aspirin use were available in NHANES for 1999-2002 and 2011-2012 therefore subjects who used aspirin could only be excluded from these three cycles) or hematologic prescription drugs (ie, erythropoiesis-stimulating agents [erythropoietin, epoetin alfa, darbepoetin alfa] and colony-stimulating agents [filgrastim]).[14] Pregnant or lactating females and volunteers who had donated blood in the last 3 months were also excluded.[14] **Table 1** provides the number of volunteers eliminated for each exclusion condition. A total of 6,714 children aged 1 to 19 years and 14,425 adults aged 20 to 79 years were excluded from the analysis for reported health conditions or use of drugs that could alter RBC levels. The final total number of "healthy" participants included in the analysis was 23,831 children aged 1 to 19 years and 20,497 adults aged 20 to 79 years. All participants or proxies provided written informed consent, and the Research Ethics Review Board at the National Center for Health Statistics approved the survey protocol.

Hematologic Measurements

Blood samples were collected in anticoagulant-treated vacuum tubes and analyzed in duplicate

DOI: 10.1093/ajcp/aqy116

Table 1
Exclusion Criteria and the Number of Subjects Excluded[a]

	Children Aged 1-19 Years		Adults Aged 20-79 Years	
Exclusion Criteria	**Number Excluded**	**Total Excluded**	**Number Excluded**	**Total Excluded**
None or incomplete CBC	5,981	5,981	3,223	3,223
Pregnancy	181	6,147	1,351	4,422
Lactation	35	6,180	252	4,645
Anemia	433	6,526	1,305	5,656
Coronary heart failure	0	6,526	947	6,383
Coronary heart disease	0	6,526	1,198	7,097
Angina	0	6,526	902	7,437
Acute myocardial infarction	0	6,526	1,272	7,795
Cancer	0	6,526	2,637	9,641
Gout	0	6,526	669	10,005
Emphysema	0	6,526	651	10,269
Liver condition	0	6,526	1,200	11,034
Celiac disease	7	6,531	27	11,052
Stroke	0	6,531	1,098	11,460
Thyroid condition	0	6,531	2,561	12,897
Hepatitis C	8	6,539	641	13,198
HIV	3	6,542	116	13,269
Daily aspirin use	0	6,542	1,754	13,987
Hematologic prescription drug use[b]	0	6,542	21	13,989
Donated blood	180	6,714	573	14,425
Total before exclusions		30,545		34,922
Total excluded		6,714		14,425
Total after exclusions		23,831		20,497

[a]Data from National Health and Nutrition Examination Survey 1999 to 2012.
[b]Including erythropoiesis stimulating agents (erythropoietin, epoetin alfa, darbepoetin alfa) and colony stimulating agents (filgrastim).

on a Beckman Coulter (Brea, CA) MAXM instrument. Standard Beckman Coulter procedures for calibrating, counting, and sizing, including automatic diluting and mixing for sample processing, and a single beam photometer for hemoglobinometry, were used. RBC count and hemoglobin were measured directly; hematocrit, mean cell hemoglobin (MCH), and mean cell hemoglobin concentration (MCHC) were computed; and mean cell volume (MCV) and red cell distribution width (RDW) were derived from their histograms. Detailed specimen collection and processing instructions, as well as quality control measures used, are provided in the NHANES Laboratory/Medical Technologists Procedures Manual.[19]

Statistical Measures

Analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC) and SUDAAN release 11.0 (Research Triangle Institute, Research Triangle Park, NC) using primary sampling units and strata to account for the complex sampling plan of NHANES. Appropriate sample weights were used to adjust for oversampling of selected groups and survey nonresponse as recommended in NHANES analytical guidelines.[20] Initially, means and percentiles of hematologic variables were created using SAS PROC SURVEYMEANS with default options; this procedure uses a nonparametric estimation approach. Piecewise regression was then used to determine breakpoints across ages and define population subgroups.[15] Significant breakpoints were identified when the stepwise regression analysis detected a discontinuity at the $P \leq .01$ level. A brief description of the piecewise regression procedure follows. In adults suppose f(x) is a function on [20,79] (the adult age ranges), which is initially assumed constant across the 5-year intervals [20,25], [26,30], [31,35], [36,40], …, [71,75], [76,79].

The following independent variables for a regression were created:

$$X1 = 1 \text{for } x \geq 20$$
$$X2 = 1 \text{for } x \geq 26 \text{and } 0 \text{ otherwise}$$
$$X3 = 1 \text{for } x \geq 31 \text{and } 0 \text{ otherwise}$$
$$X4 = 1 \text{for } x \geq 36 \text{ and } 0 \text{ otherwise}$$
$$\ldots$$
$$\ldots$$
$$X12 = 1 \text{for } x \geq 76 \text{ and } 0 \text{ otherwise.}$$

Any step function on [20,79] that is constant on the subintervals [20,25], [26,30], …, [76,79] may be defined as a linear combination of the functions X1, X2, …, X12 defined above. Then to fit a step function constant on the intervals [20,25], [26,30], …, [76,79] to a dependent variable (2.5th and 97.5th percentiles of hematologic parameters)

DOI: 10.1093/ajcp/aqy116

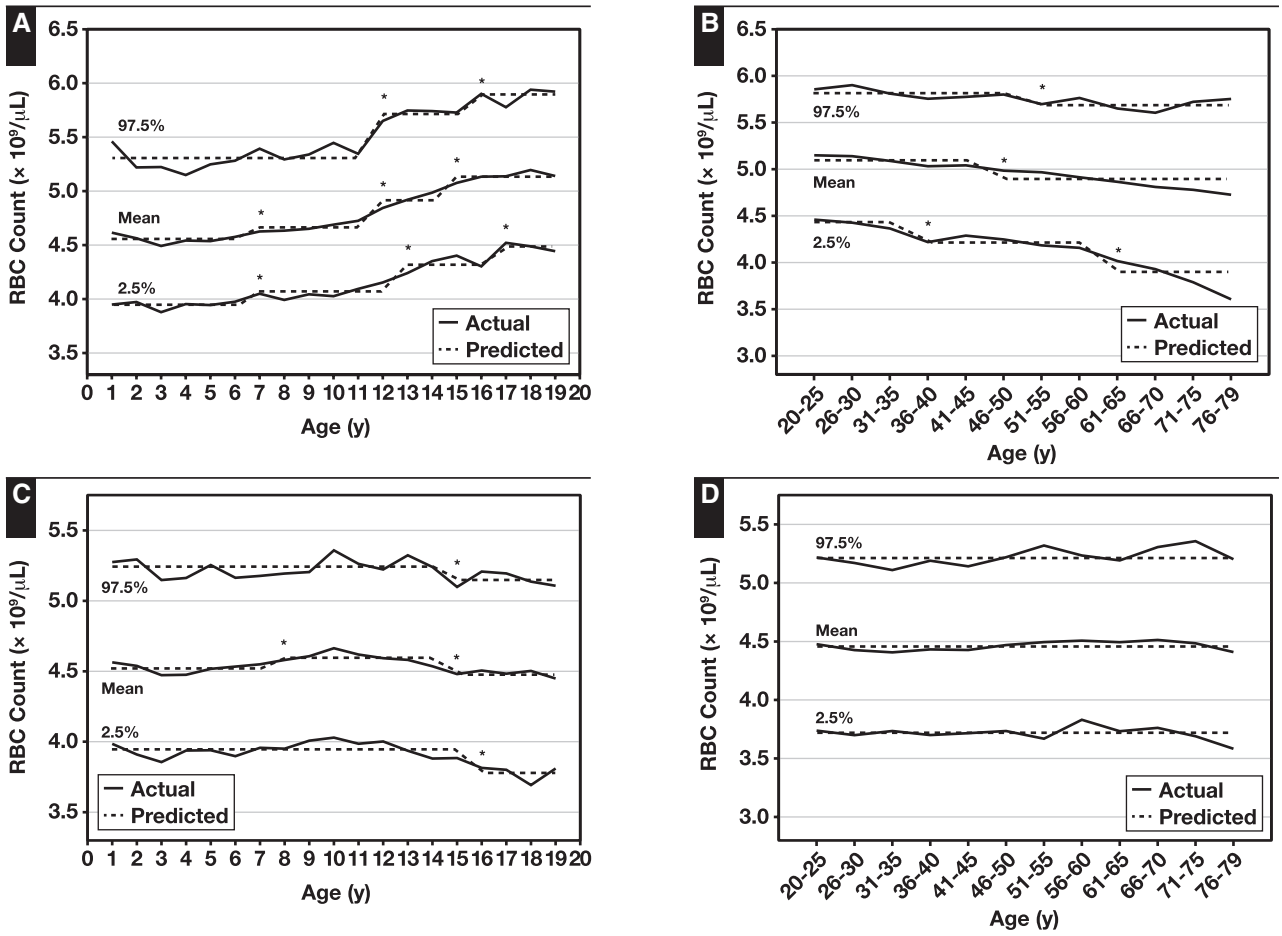

Figure 1 Actual and predicted (computed via piecewise regression) mean, lower reference limit (2.5th percentile values) and upper reference limit (97.5th percentile values) for RBC counts in male (**A**) and female (**C**) children and male (**B**) and female (**D**) adults.

at integer ages 20 to 79, one can regress the percentiles of hematologic parameters on the independent variables (age groupings) X1, X2, …, X12. This regression was performed with the 12 x-variables X1-X12 using SAS PROC REG with "selection = stepwise," "slentry = .01," and "slstay = .01" ($P \leq .01$) and weights equal to the number of subjects used to estimate the cutoff for each age. This weight approximates the relative values of the inverse of the variance for each of the estimated cutoffs. As an example,

$$\begin{aligned} &\text{if } f(x) = 2.5 \text{ for } 20 \leq x \leq 35 \\ &\quad \text{and} = 6.0 \text{ for } 36 \leq x \leq 75 \\ &\quad \text{and} = 8.0 \text{ for } 76 \leq x \leq 79 \\ &\text{then } f(x) = 2.5 * X1 + 3.5 * X4 + 2.0 * X12. \end{aligned}$$

In the same manner, any function constant on the above intervals can be written as:

$$b1 * X1 + b2 * X2 + \ldots + b12 * X12$$

where b1-b12 are regression coefficients. These b1-b12 coefficients are estimated in the regression eliminating those that are not statistically significant and the significant coefficients identify the breakpoints. In children, age groups were defined as each year of age.

The particular regression fit chosen is the best least squares fit for the dependent values using stepwise regression and leaving out those x variables that are not statistically significant. The weights are used in the regression to take into account that the estimates of the dependent variable cutoffs have different standard errors at the different ages and assume the relative variances are proportional to the inverse sample sizes for the estimates at each age.

The percentages of the population below and above existing (current) and the piecewise regression-derived reference intervals were estimated using age- and sex-specific cutoff values. The current reference intervals (cutoff points) for RBC count, hemoglobin, hematocrit,

DOI: 10.1093/ajcp/aqy116

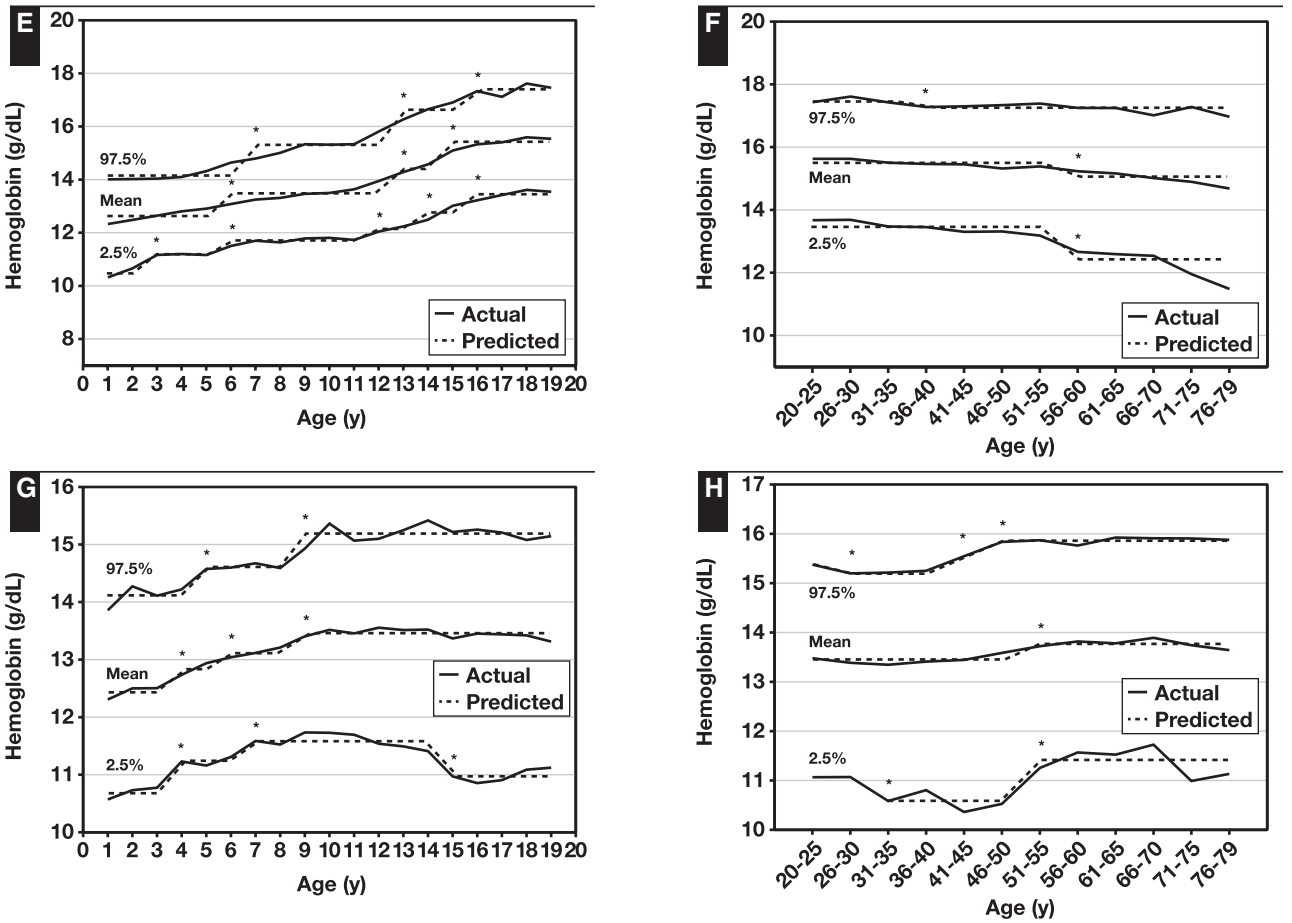

Figure 1 (cont) Also for hemoglobin levels in male (**E**) and female (**G**) children and male (**F**) and female (**H**) adults. Data from National Health and Nutrition Examination Survey 1999 to 2012 (supplemental data). *Significant breakpoint detected using piecewise regression, $P \leq .01$.

MCV, and RDW were obtained from the Mayo Clinic website in November 2016.[21]

Results

Figure 1 displays the mean, lower reference limit (LRL, 2.5th percentile), and upper reference limit (URL, 97.5th percentile) of RBC count for children and adults and the results of the piecewise regression analyses (predicted parameters). The piecewise regression analysis demonstrated an age-related increase in both LRL and URL values for male children with six separate reference intervals (ie, 1-6, 7-11, 12-12, 13-15, 16-16, 17-19 years of age). However, for female children, both LRL and URL values decreased with age and yielded only three reference intervals (ie, 1-14, 15-15, 16-19 years of age) **Table 2** and supplemental data (all supplemental materials can be found at *American Journal of Clinical Pathology* online in downloadable Excel spreadsheets). Among male adults, both LRL and URL significantly decreased with age with four separate reference intervals detected (ie, 20-35, 36-50, 51-60, 61-79 years of age). For adult females, the reference interval did not change with age **Table 3** and supplemental data.

Figure 1 also displays sex-specific pediatric and adult reference intervals for hemoglobin. In the piecewise regression analysis, the reference interval values increased with age for male and female children (for female children the LRL increased for those up to age 14 years and decreased for individuals 15-19 years) and decreased with age among male adults. For female adults, the LRL decreased for 31 to 50 years of age and then increased for 51 to 79 years of age but the URL increased slightly. However, there were fewer breakpoints for male adults than for female adults and children (Table 2 and Table 3 and supplemental data). The LRL of hemoglobin for older males (aged 56-79) was

DOI: 10.1093/ajcp/aqy116

Table 2
Current and Piecewise Regression-Derived Reference Intervals (Lower Reference Limit 2.5th Percentile Value and Upper Reference Limit 97.5th Percentile Value) for Children Aged 1 to 19 Years[a]

	Age, y	Current Reference Intervals[b]	Age, y	Piecewise Regression-Derived Reference Intervals
RBC count, million cells/μL				
Male	0.5-1	3.70-6.00	1-6	3.95-5.31
	2	4.10-5.10	7-11	4.06-5.31
	3-5	4.10-5.30	12-12	4.06-5.72
	6-11	4.20-5.10	13-15	4.32-5.72
	12-15	4.40-5.50	16-16	4.32-5.88
			17-19	4.48-5.88
Female	0.5-1	3.70-6.00	1-14	3.94-5.24
	2	4.10-5.10	15-15	3.94-5.15
	3-5	4.10-5.20	16-19	3.78-5.15
	6-11	4.10-5.30		
	12-15	4.10-5.20		
Hemoglobin, g/dL				
Male	0.5-1	10.5-13.5	1-2	10.5-14.2
	2	11.0-14.0	3-5	11.2-14.2
	3-5	11.0-14.5	6-6	11.7-14.2
	6-11	12.0-14.0	7-11	11.7-15.3
	12-15	12.8-16.0	12-12	12.1-15.3
			13-13	12.1-16.6
			14-15	12.8-16.6
			16-19	13.4-17.4
Female	0.5-1	10.5-13.5	1-3	10.7-14.1
	2	11.0-14.0	4-4	11.2-14.1
	3-5	11.8-14.7	5-6	11.2-14.6
	6-11	12.0-14.5	7-8	11.6-14.6
	12-15	12.2-14.8	9-14	11.6-15.2
			15-19	11.0-15.2
Hematocrit, %				
Male	0.5-1	33.0-40.0	1-4	31.8-41.3
	2	33.0-42.0	5-5	34.3-41.3
	3-5	33.0-43.0	6-12	34.3-44.8
	6-11	35.8-42.4	13-14	36.5-48.6
	12-15	37.3-47.3	15-15	39.2-48.6
			16-19	39.2-50.8
Female	0.5-1	33.0-40.0	1-3	31.6-41.6
	2	33.0-42.0	4-4	32.7-41.6
	3-5	35.0-44.0	5-5	32.7-42.9
	6-11	35.7-43.0	6-8	34.3-42.9
	12-15	36.3-43.4	9-14	34.3-44.7
			15-19	32.9-44.7
Mean cell volume, fL				
Male	0.5-5	74.0-89.0	1-2	68.0-86.9
	6-11	76.5-90.6	3-8	74.7-89.7
	12-15	81.4-91.9	9-12	74.7-90.9
			13-13	74.7-93.8
			14-15	77.9-93.8
			16-19	77.9-95.7
Female	0.5-5	74.0-89.0	1-2	68.6-87.0
	6-11	78.5-90.4	3-4	75.3-89.2
	12-15	79.9-92.3	5 - 11	75.3-91.2
			12-16	75.3-94.8
			17-19	75.3-96.6
Mean cell hemoglobin, pg[c]				
Male			1-2	22.5-29.8
			3-6	25.0-31.2
			7-13	25.0-31.6
			14-14	25.0-32.2
			15-15	26.1-32.2
			16-19	26.1-32.9

(continued)

DOI: 10.1093/ajcp/aqy116

Table 2 (cont)

	Age, y	Current Reference Intervals[b]	Age, y	Piecewise Regression-Derived Reference Intervals
Female			1-2	22.3-30.3
			3-3	25.0-30.3
			4-11	25.0-31.6
			12-19	25.0-32.9
Mean cell hemoglobin concentration, g/dL[c]				
Male			1-12	32.5-36.0
			13-19	32.5-35.7
Female			1-19	32.4-35.9
Red cell distribution width, %				
Male	<2	Not established	1-2	11.4-16.3
	2	12.0-14.5	3-19	11.4-14.0
	3-5	12.0-14.0		
	6-11	12.0-14.0		
	12-15	11.6-13.8		
Female	<2	Not established	1-2	11.2-16.3
	2	12.0-14.5	3-12	11.2-13.8
	3-5	12.0-14.0	13-19	11.2-15.3
	6-11	11.6-13.4		
	12-15	11.2-13.5		

[a]Data from National Health and Nutrition Examination Survey 1999 to 2012 (supplemental data).
[b]Values from Mayo Clinic.[21]
[c]Current reference intervals are not provided by Mayo Clinic.[21]

12.4 g/dL, lower than the current reference value, which is 13.5 g/dL.[21]

The sex-specific pediatric and adult reference intervals for hematocrit obtained via piecewise regression analysis are provided in **Figure 2**, Table 2 and Table 3, and supplemental data. The reference interval values generally increased with age for both male and female children except for female children aged 15 to 19 years where the LRL decreased. The reference values for adult males and females had three and two breakpoints, respectively. The male URL did not change with age but the LRL decreased with age and was substantially lower for males aged 71 to 79. The female LRL and URL increased with age.

The age- and sex-specific reference intervals for MCV and MCH increased with age for both children and adults except for the LRL for MCH, which decreased in male adults at age 76 to 79 years (Figure 2 and **Figure 3**, Table 2 and Table 3, and supplemental data). For MCHC, the reference interval decreased slightly at age 13 years for male children but did not change with age for female children, and in male and female adults changes did not occur until ages 56 and 61 years, respectively (Figure 3, Table 2 and Table 3, and supplemental data). The sex-specific reference intervals for RDW decreased with age for both male and female children (but increased at age 13 years) but increased with age for both male and female adults (**Figure 4**, Table 2 and Table 3, and supplemental data).

Table 4 provides estimates of the NHANES 1999 to 2012 population below and above the reference intervals (eg, a higher percentage indicates more of the population falls outside reference values). When comparing the piecewise regression-derived and current reference intervals, there was generally less than a 5% difference in the population above or below the reference intervals. However, there was a greater than 5% difference in the population below the piecewise regression-derived and current LRL for RDW (for both sexes combined, male and female children, and adults), for hematocrit (for both sexes combined and female children), for MCV (for male children), and for hemoglobin (female adults), and above URL for hemoglobin (for both sexes combined and male children) and for MCV (for male adults). For many of these variables, more than 10% of the NHANES population was outside the current reference intervals (Table 4).

Discussion

Piecewise regression analyses conducted using the NHANES 1999 to 2012 data yielded two or more age-dependent pediatric and adult reference intervals for multiple RBC parameters. The reference intervals derived were sex specific and more precise for individuals of different ages than current pediatric and adult reference

DOI: 10.1093/ajcp/aqy116

Table 3
Current and Piecewise Regression-Derived Reference Intervals (Lower Reference Limit 2.5th Percentile Value and Upper Reference Limit 97.5th Percentile Value) for Adults Aged 20-79 Years[a]

Variable	Age, y	Current Reference Intervals[b]	Age, y	Piecewise Regression-Derived Reference Intervals
RBC count, million cells/μL				
Male	Adults	4.32-5.72	20-35	4.42-5.82
			36-50	4.23-5.82
			51-60	4.23-5.69
			61-79	3.90-5.69
Female	Adults	3.90-5.03	20-79	3.72-5.20
Hemoglobin, g/dL				
Male	Adults	13.5-17.5	20-35	13.5-17.5
			36-55	13.5-17.3
			56-79	12.4-17.3
Female	Adults	12.0-15.5	20-25	11.1-15.4
			26-30	11.1-15.2
			31-40	10.6-15.2
			41-45	10.6-15.5
			46-50	10.6-15.9
			51-79	11.4-15.9
Hematocrit, %				
Male	Adults	38.8-50.0	20-35	40.2-51.0
			36-60	38.8-51.0
			61-70	37.0-51.0
			71-79	35.0-51.0
Female	Adults	34.9-44.5	20-45	32.5-45.2
			46-50	32.5-46.8
			51-79	34.2-46.8
Mean cell volume, fL				
Male	Adults	81.2-95.1	20-50	81.2-97.3
			51-79	81.2-100.0
Female	Adults	81.6-98.3	20-40	75.9-97.4
			41-55	71.6-99.2
			56-79	79.7-99.2
Mean cell hemoglobin, pg[c]				
Male			20-35	27.1-33.4
			36-50	27.1-34.0
			51-75	27.1-34.5
			76-79	24.8-34.5
Female			20-40	24.7-33.7
			41-79	24.7-34.3
Mean cell hemoglobin concentration, g/dL[c]				
Male			20-55	32.4-36.0
			56-79	32.4-35.7
Female			20-60	32.1-35.8
			61-79	32.1-35.6
Red cell distribution width, %				
Male	Adults	11.8-15.6	20-45	11.3-13.8
			46-50	11.5-13.8
			51-70	11.5-14.8
			71-79	11.7-14.8
Female	Adults	11.9-15.5	20-45	11.3-15.6
			46-79	11.4-15.6

[a]Data from National Health and Nutrition Examination Survey 1999 to 2012 (supplemental data).
[b]Values from Mayo Clinic.[21]
[c]Current reference intervals are not provided by Mayo Clinic.[21]

intervals. For most variables, similar proportions of the population were outside the current and piecewise regression-derived reference intervals. Piecewise regression-derived pediatric reference intervals are age dependent and yield different sets of values for younger and older children. As expected from a biological growth and development perspective, there were definite breakpoints around onset of puberty in reference intervals for all the RBC parameters, yielding different reference interval values for children and adolescents, unlike

DOI: 10.1093/ajcp/aqy116

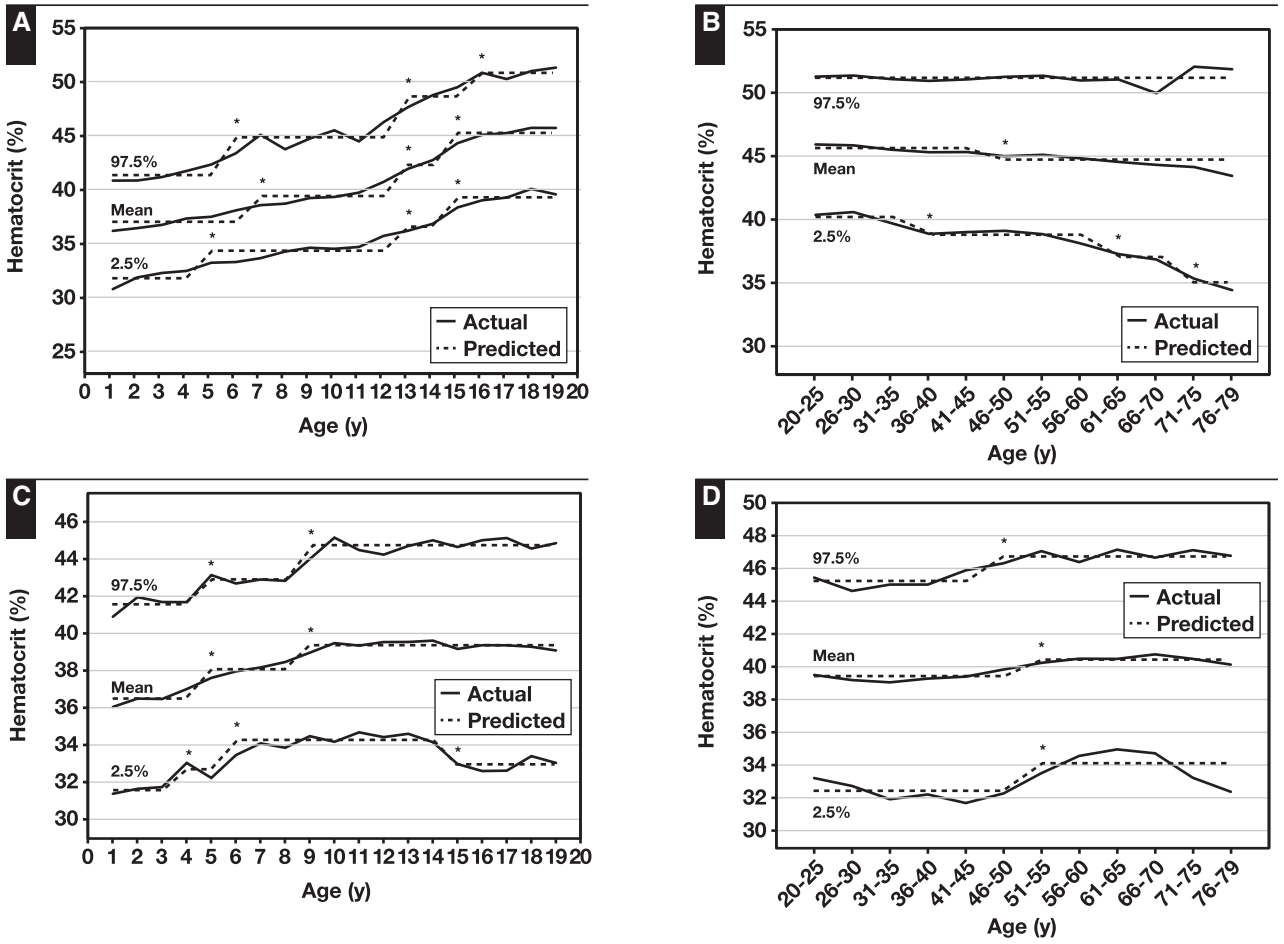

Figure 2 Actual and predicted (computed via piecewise regression) mean, lower reference limit (2.5th percentile values) and upper reference limit (97.5th percentile values) for hematocrit in male (**A**) and female (**C**) children and male (**B**) and female (**D**) adults.

current reference values. The piecewise regression analysis of the NHANES dataset also yielded separate RBC count, hematocrit, hemoglobin, MCV, MCH, MCHC, and RDW pediatric reference intervals for male and female children, unlike current standards; however, statistical breakpoints identified by piecewise regression are not necessarily biologically relevant.

Typically, current adult reference intervals are a single set of values for all adults, while the piecewise regression derived values generally provided separate reference interval values for young and older adults. Although not every statistical breakpoint identified by piecewise regression may be biologically relevant, in some cases the differences are substantial and may have clinical implications, which should be determined by additional research. For example, for a male aged 56 to 79 years, the piecewise regression derived LRL for hemoglobin was 12.4 g/dL, but the current reference values are not age based and the LRL for all adult males is 13.5 g/dL. For another key hematologic parameter, hematocrit, there was also a difference in older males compared to their younger peers. In males aged 71 to 79 years, the LRL derived by piecewise regression was 35.0% while the current LRL is not age based and is 38.8%. LRLs are generally more critical and clinically relevant, but in some cases exceeding URLs can suggest the presence of disease states. Age-related differences in several hematologic parameters were also reported in large studies conducted with Canadian and Italian populations, and both indicate there are age-related declines in hematologic status.[22,23] Although lower LRLs for hemoglobin and hematocrit in older males are suggested by piecewise regression, before adopting these reference values, research should be conducted to ensure these new reference intervals represent values expected for healthy individuals of this age and do not result in missed diagnoses of various diseases or compromised function. Because NHANES includes

DOI: 10.1093/ajcp/aqy116

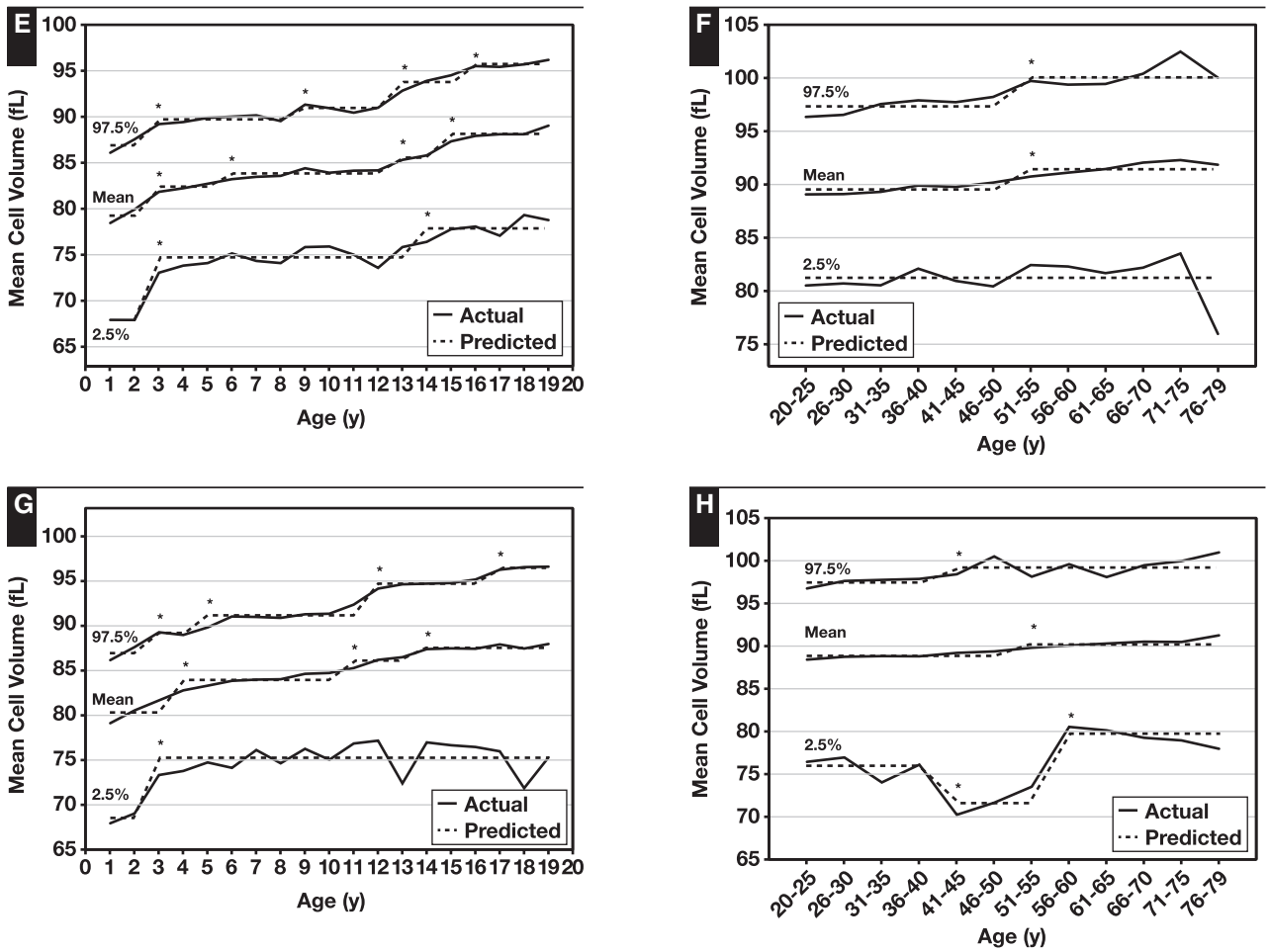

Figure 2 (cont) Also for mean cell volume levels in male (**E**) and female (**G**) children and male (**F**) and female (**H**) adults. Data from National Health and Nutrition Examination Survey 1999 to 2012 (supplemental data). *Significant breakpoint detected using piecewise regression, $P \leq .01$.

comprehensive data on the physical and mental health of participants, it could be used for such studies.

By definition, reference intervals should include 95% of the population. However, more than 10% of the pediatric and adult NHANES population were outside the respective current reference intervals for RDW. Although RDW values vary based on instrumentation and method of calculation,[24] our findings are consistent with others that report age and sex differences in RDW.[25,26]

Current reference intervals are derived from a wide variety of sources including published studies, patient data, individual laboratory studies, manufacturer's recommendations, or medical staff recommendations and are based on small and not necessarily nationally representative population samples.[12] Reference intervals based on published data or manufacturer information sheets can be outdated and/or nonrepresentative of a broader population. Patient data, while readily obtained, may not be representative of healthy individuals and, as a consequence, may be inaccurate. Additionally, there is the possibility of variation in hematology methods, especially with regard to diagnostic sensitivity and specificity across studies. Adult reference intervals are validated using internal laboratory data in most cases, but pediatric reference intervals are usually adopted from external sources without internal validation.[12]

In the United States, efforts to harmonize reference intervals are far behind programs in the United Kingdom, the Nordic region, and Australasia.[27] As such, local laboratories are left on their own to establish appropriate reference intervals for their institution. Given the difficulty determining reference intervals, most local laboratories primarily use reference intervals taken from textbooks or other literature and ignore potential differences related to instrumentation bias.[28,29] While these textbook intervals are a start, they often lack data on the population

DOI: 10.1093/ajcp/aqy116

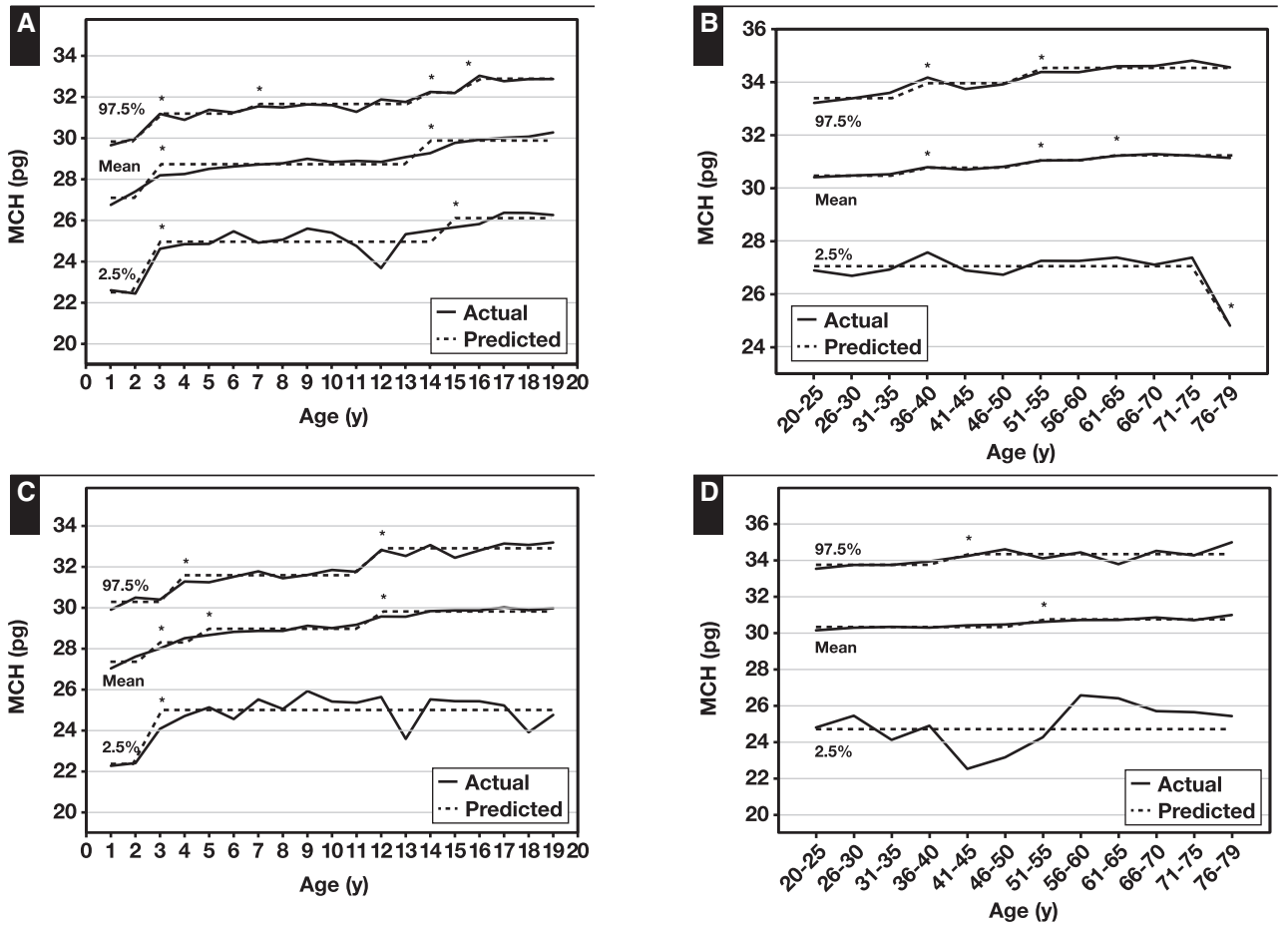

Figure 3 Actual and predicted (computed via piecewise regression) mean, lower reference limit (2.5th percentile values) and upper reference limit (97.5th percentile values) for mean cell hemoglobin (MCH) in male (**A**) and female (**C**) children and male (**B**) and female (**D**) adults.

from which they were derived and usually are based on a single instrument/methodology. The data presented here are based on a very large nationally representative population and therefore provide very accurate estimates of reference intervals. However, given the use of a single instrument platform (Beckman Coulter MAXM), local laboratories still need to verify the intervals published in this manuscript are transferrable to their platform and population. In addition to following guidelines provided by the CLSI[13] to verify a published interval, in combination with mining local datasets, a local laboratory can also effectively use the intervals suggested here.[30]

It is recommended that reference intervals be developed based on a minimum of 120 healthy reference subjects.[13] However, "healthy" is difficult to define, and there is considerable variation between laboratories in criteria used to define "healthy." The NHANES surveys are designed to assess the health and nutritional status of the US population using validated, standardized, state-of-the-art sample collection and analysis techniques. We used precise, detailed criteria to exclude NHANES participants with medical conditions and select "healthy" participants for analysis (Table 1). As we excluded over 21,000 subjects (nearly 22% of the NHANES pediatric population and 41% of the adult population) for self-reported health issues it is unlikely any unhealthy individuals remaining in our sample substantially affected our findings. The large sample size of NHANES provided a high degree of statistical power in determining age- and sex-specific partitions in reference intervals. A recent study using adult NHANES data from 2011 to 2012 (n = 3,077) compared reference intervals across ethnic subgroups but did not evaluate differences between different age groups.[31]

Use of appropriate statistical techniques is critical for developing reference intervals. Simple nonparametric as well as parametric methods with suitable transformation (eg, logarithmic, power, or some other function) were recommended by CLSI.[13] Hoffman developed a statistical technique to indirectly derive reference intervals from database values based on specified

DOI: 10.1093/ajcp/aqy116

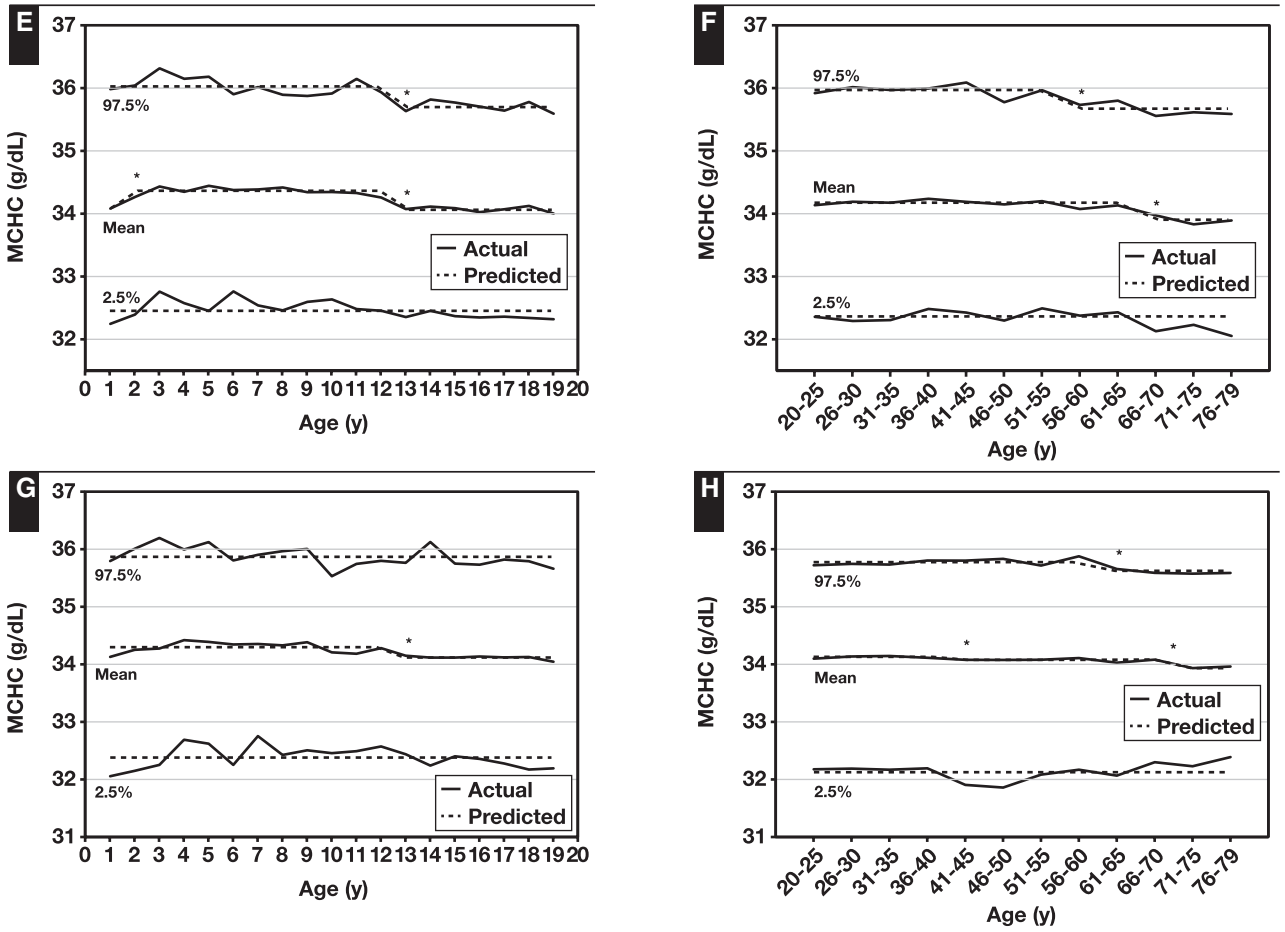

Figure 3 (cont) Also for mean cell hemoglobin concentration (MCHC) in male (**E**) and female (**G**) children and male (**F**) and female (**H**) adults. Data from National Health and Nutrition Examination Survey 1999 to 2012 (supplemental data). *Significant breakpoint detected using piecewise regression, $P \leq .01$.

mathematical assumptions.[32] It was suggested Hoffmann's indirect method, which uses Chauvenet criteria to identify and exclude outliers, is a robust statistical approach.[28] We used piecewise regression to objectively identify statistically significant breakpoints in RBC parameters across age and sex groups. This approach, which is a formal statistical procedure for standardized identification of breakpoints for different age/sex groups, may be preferable to less quantitative procedures, especially when applied to well-defined subpopulations. This technique has been used to objectively demonstrate the nonlinear relationship between calcium intake and bone mineral density,[17] between BMI and mortality,[16] and more recently between sedentary behavior and mortality.[33] While other procedures for curve fitting may provide a better overall fit of the raw data, piecewise regression was chosen for this study because it objectively identifies breakpoints and, as a result, provides greater specificity.

Limitations of our study include possible bias from use of self-reported health information to define a "healthy" population. There is also the possibility of analytical drift over the 14 years of NHANES data used, but standardized laboratory procedures are consistently used in NHANES and if necessary values are adjusted for any change over time. It should also be noted that the new reference values presented may only be relevant when the same analytical procedures used by NHANES are employed. We compared our reference ranges to that used by the Mayo Clinic in late 2016, and recently the Mayo Clinic updated their reference values to include ranges for more age/sex groups, consistent with our findings. In addition, the large sample size provided by NHANES could yield small but statistically significant differences that may not be of biological significance. It should also be noted that data presented for closely related parameters, like hematocrit and hemoglobin, sometimes have different breakpoints for the same population. An explanation for such differences is not readily apparent but may be due to a variety of factors, such as differences in

DOI: 10.1093/ajcp/aqy116

Table 4
Percent of Population Below and Above the Age- and Sex-Specific Reference Values for Children and Adults[a]

		Estimated % Population Below the Lower Value of Reference Interval			Estimated % Population Above the Upper Value of Reference Interval		
Variables	Sex	Current	Piecewise Regression-Derived	Difference	Current	Piecewise Regression-Derived	Difference
Children, age 1-19 y							
RBC count	Combined	6.85	2.42	–4.43	4.44	2.58	–1.86
	Male	6.82	2.50	–4.32	6.46	2.55	–3.91
	Female	6.89	2.33	–4.55	2.25	2.62	0.36
Hemoglobin	Combined	5.48	2.21	–3.27	8.77	3.11	–5.65
	Male	4.16	2.29	–1.87	11.46	3.22	–8.24
	Female	6.91	2.13	–4.79	5.86	2.99	–2.87
Hematocrit	Combined	8.87	2.51	–6.35	5.44	2.72	–2.72
	Male	6.60	2.43	–4.17	6.00	2.95	–3.06
	Female	11.32	2.59	–8.72	4.82	2.48	–2.35
Mean cell volume	Combined	6.60	2.43	–4.17	4.43	2.73	–1.71
	Male	7.50	2.45	–5.05	3.46	2.69	–0.77
	Female	5.63	2.42	–3.22	5.49	2.77	–2.72
Mean cell hemoglobin	Combined		2.47			2.95	
	Male		2.45			2.90	
	Female		2.49			3.01	
Mean cell hemoglobin concentration	Combined		2.22			2.78	
	Male		2.09			2.82	
	Female		2.36			2.73	
Red cell distribution width	Combined	15.08	2.02	–13.07	4.29	2.59	–1.70
	Male	15.42	1.53	–13.88	3.59	2.57	–1.02
	Female	14.73	2.54	–12.19	5.04	2.61	–2.43
Adults, age 20-79 y							
RBC count	Combined	4.93	2.54	–2.39	5.03	2.70	–2.32
	Male	3.26	2.57	–0.69	3.89	2.73	–1.16
	Female	6.65	2.51	–4.14	6.20	2.68	–3.52
Hemoglobin	Combined	5.15	2.26	–2.89	2.41	3.08	0.67
	Male	2.90	2.16	–0.75	1.86	3.37	1.50
	Female	7.46	2.37	–5.10	2.97	2.78	–0.19
Hematocrit	Combined	4.31	2.46	–1.86	5.76	2.73	–3.04
	Male	2.19	2.35	0.17	5.35	2.81	–2.54
	Female	6.50	2.57	–3.94	6.19	2.64	–3.55
Mean cell volume	Combined	4.80	2.54	–2.26	6.51	2.63	–3.88
	Male	2.45	2.52	0.07	10.39	2.71	–7.68
	Female	7.22	2.56	–4.66	2.52	2.55	–0.03
Mean cell hemoglobin	Combined		2.61			2.51	
	Male		2.63			2.54	
	Female		2.59			2.48	
Mean cell hemoglobin concentration	Combined		2.35			2.97	
	Male		2.25			2.88	
	Female		2.46			3.07	
Red cell distribution width	Combined	10.95	2.05	–8.90	1.59	2.63	1.04
	Male	10.07	2.07	–7.99	0.52	2.57	2.05
	Female	11.86	2.03	–9.83	2.69	2.69	0.00

[a]Data from National Health and Nutrition Examination Survey 1999 to 2012.

variability and precision of the assays for the parameters in question. A major strength of our study is the use of a large nationally representative population-based sample, use of numerous variables to help define a "healthy" population, and the use of piecewise regression to develop objectively/statistically defined breakpoints. Additionally, to ensure transparency and expedite their use, the data presented here are provided in the supplemental data.

Future research should evaluate whether separate reference ranges are necessary for other well-defined groups (ie, by race/ethnicity), and for those living in specific geographic regions of the US, such as high altitude environments. In the future, when the computational

DOI: 10.1093/ajcp/aqy116

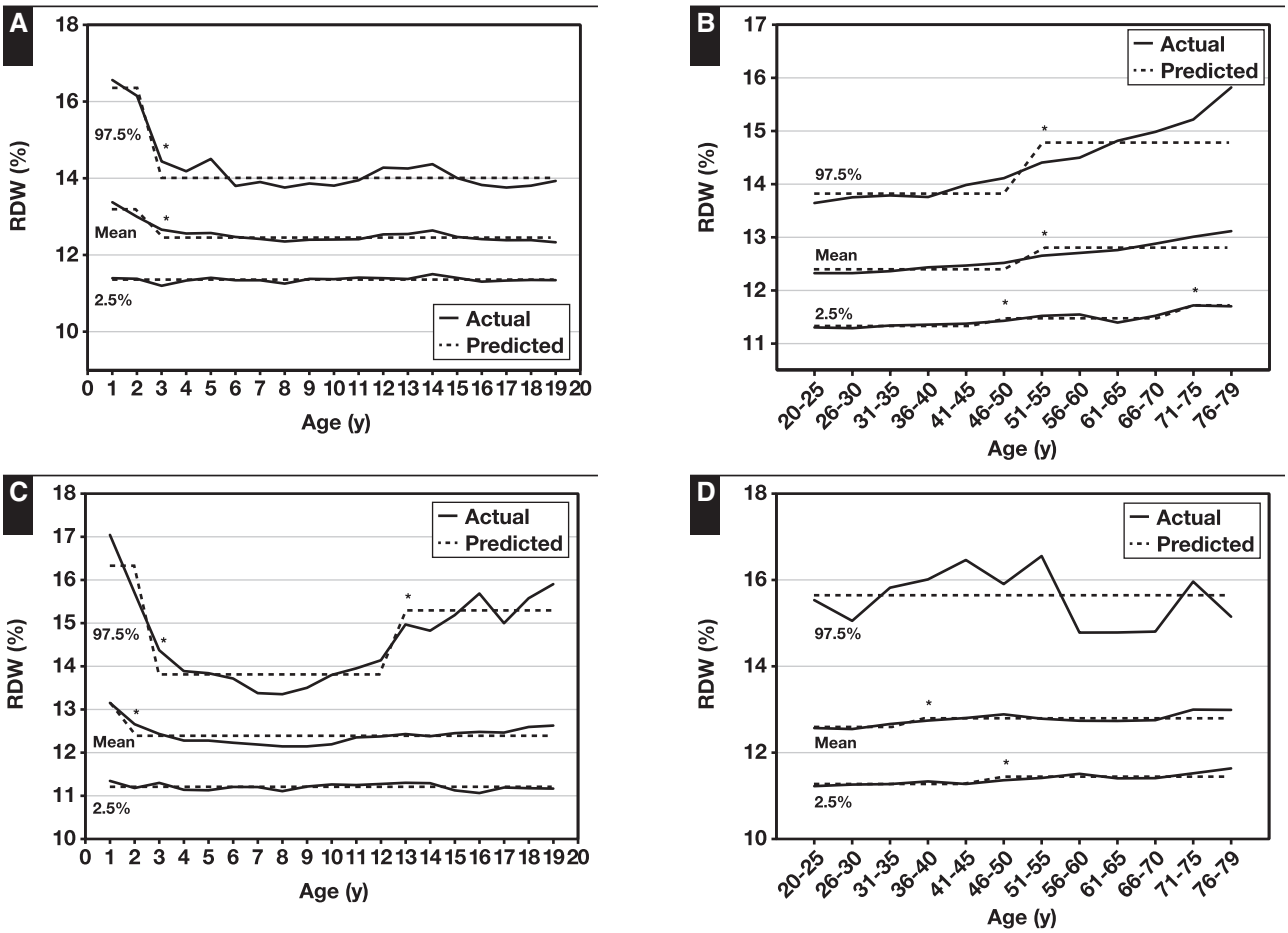

Figure 4 Actual and predicted (computed via piecewise regression) mean, lower reference limit (2.5th percentile values) and upper reference limit (97.5th percentile values) for red cell distribution width (RDW) in male (**A**) and female (**C**) children and male (**B**) and female (**D**) adults. Data from National Health and Nutrition Examination Survey 1999 to 2012 (supplemental data). *Significant breakpoint detected using piecewise regression, $P \leq .01$.

infrastructure for personalized medicine exists, customized reference ranges for very specific subsets of individuals (eg, age, sex, race/ethnicity, and/or other defining specific characteristic), could be provided.

Corresponding author: Harris R. Lieberman, PhD; harris.r.lieberman.civ@mail.mil.

Acknowledgments: We thank Lauren Thompson and Marques A. Wilson for their help with figures and tables.

Disclaimer: The views, opinions, and findings in this report are those of the authors and should not be construed as an official Department of Defense or Army position, policy, or decision, unless so designated by other official documentation. Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations. The investigators have adhered to the policies for protection of human subjects as prescribed in DOD Instruction 3216.02 and the research was conducted in adherence with the provisions of 32 CFR Part 219.

Funding: This work was supported by US Army Medical Research and Materiel Command (USAMRMC) and Department of Defense Center Alliance for Nutrition and Dietary Supplement Research; and by an appointment to the Research Participation Program at the US Army Medical Research Institute of Environmental Medicine, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and USAMRMC.

Conflict of interest statement: Victor L. Fulgoni, III, and Sanjiv Agarwal are nutrition consultants and as such provide services to industry.

References

1. Horvath AR. From evidence to best practice in laboratory medicine. *Clin Biochem Rev*. 2013;34:47-60.
2. de Benoist B, McLean E, Egli I, et al., eds. *Worldwide Prevalence of Anaemia 1993-2005, WHO Global Database on Anaemia*. Geneva, Switzerland: World Health Organization; 2008.

DOI: 10.1093/ajcp/aqy116

3. McClung JP, Marchitelli LJ, Friedl KE, et al. Prevalence of iron deficiency and iron deficiency anemia among three populations of female military personnel in the US army. *J Am Coll Nutr*. 2006;25:64-69.
4. Brugnara C, Mohandas N. Red cell indices in classification and treatment of anemias: from M.M. Wintrobes's original 1934 classification to the third millennium. *Curr Opin Hematol*. 2013;20:222-230.
5. Buttarello M. Laboratory diagnosis of anemia: are the old and new red cell parameters useful in classification and treatment, how? *Int J Lab Hematol*. 2016;38(suppl 1):123-132.
6. Cleveland Clinic. Diagnostic and testing, complete blood count, 2014. http://my.clevelandclinic.org/health/diagnostics/hic_complete_blood_count. Accessed September 12, 2016.
7. Leach M. Interpretation of the full blood count in systemic disease—a guide for the physician. *J R Coll Physicians Edinb*. 2014;44:36-41.
8. Spell DW, Jones DV Jr, Harper WF, et al. The value of a complete blood count in predicting cancer of the colon. *Cancer Detect Prev*. 2004;28:37-42.
9. Dixon LR. The complete blood count: physiologic basis and clinical usage. *J Perinat Neonatal Nurs*. 1997;11:1-18.
10. Ceriotti F, Hinzmann R, Panteghini M. Reference intervals: the way forward. *Ann Clin Biochem*. 2009;46:8-17.
11. Cheng CK, Chan J, Cembrowski GS, et al. Complete blood count reference interval diagrams derived from NHANES III: stratification by age, sex, and race. *Lab Hematol*. 2004;10:42-53.
12. Friedberg RC, Souers R, Wagar EA, et al; College of American Pathologists. The origin of reference intervals. *Arch Pathol Lab Med*. 2007;131:348-357.
13. Clinical and Laboratory Standards Institute. *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory, Approved Guideline, CLSI Document C28-A3*, 3rd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
14. Centers for Disease Control and Prevention, National Center for Health Statistics. National health and nutrition examination survey. http://www.cdc.gov/nchs/nhanes/about_nhanes.htm. Accessed January 9, 2016.
15. Ficetola GF, Denoel M. Ecological thresholds: an assessment of methods to identify abrupt changes in species-habitat relationships. *Ecography*. 2009;32:1075-1084.
16. Keith SW, Allison DB. A free-knot spline modeling framework for piecewise linear logistic regression in complex samples with body mass index and mortality as an example. *Front Nutr*. 2014;2014:00016.
17. Breitling LP. Calcium intake and bone mineral density as an example of non-linearity and threshold analysis. *Osteoporos Int*. 2015;26:1271-1281.
18 .National Center for Health Statistics. Analytical note for 25-hydroxyvitamin D data analysis using NHANES III (1988-1994), NHANES 2001-2006, and NHANES 2007-2010. http://wwwn.cdc.gov/nchs/nhanes/VitaminD/AnalyticalNote.aspx; 2015. Accessed September 14, 2016.
19. Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination Survey. Laboratory procedures manual, 2011. https://www.cdc.gov/nchs/data/nhanes/nhanes_11_12/2011-12_laboratory_procedures_manual.pdf. Accessed January 15, 2017.
20. Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination Survey: analytic guidelines, 2011-2012. https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/analytic_guidelines_11_12.pdf. Accessed January 9, 2016.
21. Mayo Clinic. CBC with differential, blood. http://www.mayomedicallaboratories.com/test-catalog/Clinical+and+Interpretive/9109; 2016. Accessed November 3, 2016.
22. Adeli K, Raizman JE, Chen Y, et al. Complex biological profile of hematologic markers across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian Health Measures Survey. *Clin Chem*. 2015;61:1075-1086.
23. Tettamanti M, Lucca U, Gandini F, et al. Prevalence, incidence and types of mild anemia in the elderly: the "health and anemia" population-based study. *Haematologica*. 2010;95:1849-1856.
24. Lippi G, Pavesi F, Bardi M, et al. Lack of harmonization of red blood cell distribution width (RDW): evaluation of four hematological analyzers. *Clin Biochem*. 2014;47:1100-1103.
25. Hoffmann JJ, Nabbe KC, van den Broek NM. Effect of age and gender on reference intervals of red blood cell distribution width (RDW) and mean red cell volume (MCV). *Clin Chem Lab Med*. 2015;53:2015-2019.
26. Alis R, Fuster O, Rivera L, et al. Influence of age and gender on red blood cell distribution width. *Clin Chem Lab Med*. 2015;53:e25-e28.
27. Tate JR, Koerbin G, Adeli K. Opinion paper: deriving harmonised reference intervals - global activities. *EJIFCC*. 2016;27:48-65.
28. Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? *Am J Clin Pathol*. 2010;133:180-186.
29. McCafferty R, McHugh J, Regan I, et al. National Clinical Program for Pathology. Reference Interval Harmonization Project Group. 2nd draft survey report on reference intervals for the full blood count in the Republic of Ireland. https://rcpi-live-cdn.s3.amazonaws.com/wp-content/uploads/2017/09/Reference-Intervals-for-the-FBC-in-ROI-for-NCPP-2017.pdf. Accessed July 3, 2018.
30. Jones GR. Validating common reference intervals in routine laboratories. *Clin Chim Acta*. 2014;432:119-121.
31. Lim E, Miyamura J, Chen JJ. Racial/ethnic-specific reference intervals for common laboratory tests: a comparison among Asians, Blacks, Hispanics, and White. *Hawaii J Med Public Health*. 2015;74:302-310.
32. Hoffmann RG. Statistics in the practice of medicine. *JAMA*. 1963;185:864-873.
33. Lee PH. Examining non-linear associations between accelerometer-measured physical activity, sedentary behavior, and all-cause mortality using segmented Cox regression. *Front Physiol*. 2016;7:272.

DOI: 10.1093/ajcp/aqy116