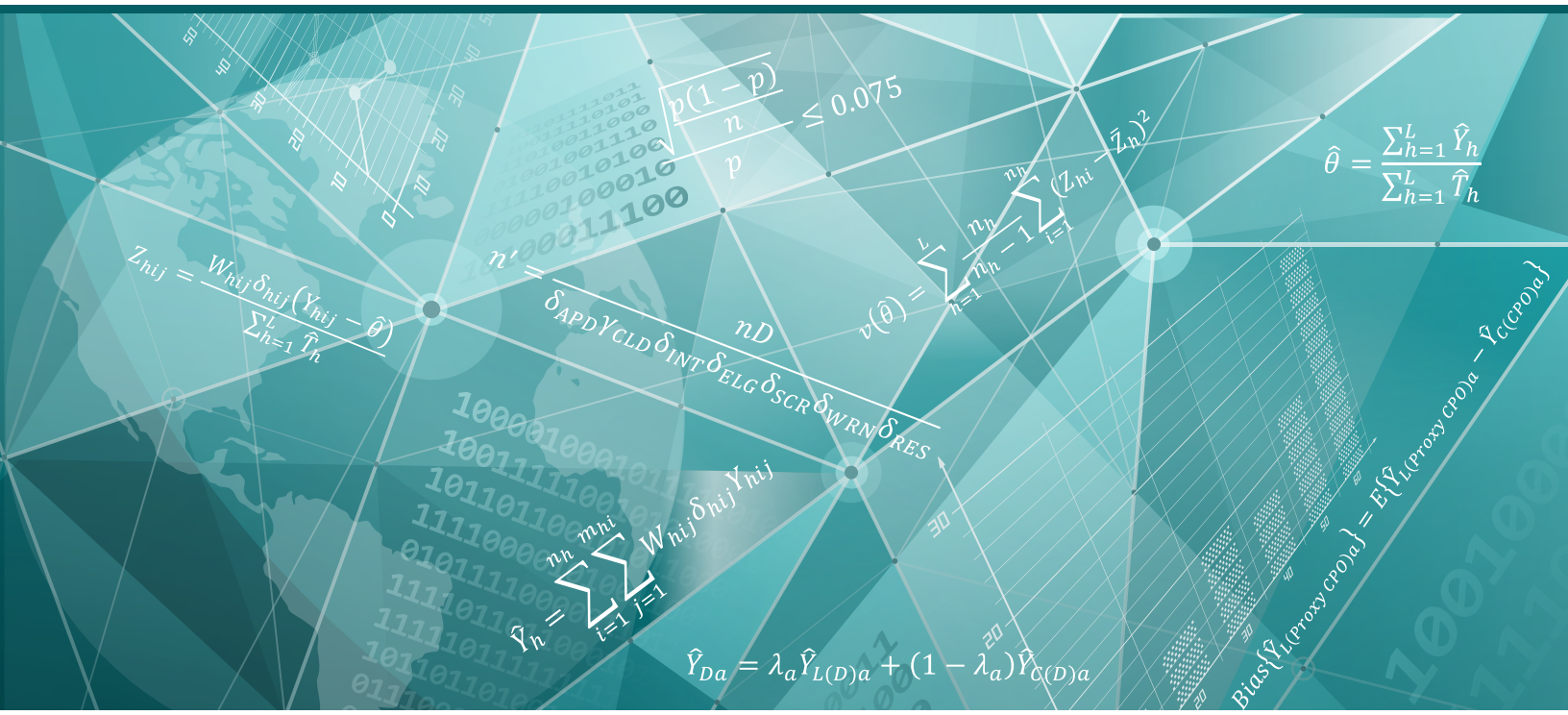
NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 206

April 2024



Assessing Laboratory Method Validations for Informing Inference Across Survey Cycles in the National Health and Nutrition Examination Survey

Data Evaluation and Methods Research

Centers for Disease Control and Prevention
National Center for Health Statistics

NCHS reports can be downloaded from: https://www.cdc.gov/nchs/products/index.htm.

Copyright information

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Suggested citation

Chuang K, Rammon J, Shin HC, Chen TC. Assessing laboratory method validations for informing inference across survey cycles in the National Health and Nutrition Examination Survey. National Center for Health Statistics. Vital Health Stat 2(206). 2024. DOI: https://dx.doi.org/10.15620/cdc:145591.

For sale by the U.S. Government Publishing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20401–0001
Printed on acid-free paper.

NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 206 April 2024

Assessing Laboratory Method Validations for Informing Inference Across Survey Cycles in the National Health and Nutrition Examination Survey

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
April 2024

National Center for Health Statistics

Brian C. Moyer, Ph.D., *Director*

Amy M. Branum, Ph.D., *Associate Director for Science*

Division of Research and Methodology

Jennifer D. Parker, Ph.D., *Director*

John R. Pleis, Ph.D., *Associate Director for Science*

Division of Health and Nutrition Examination Surveys

Alan E. Simon, M.D., *Director*

Lara J. Akinbami, M.D., *Associate Director for Science*

Contents

Acknowledgment v

Abstract 1

Introduction 1

Methods 2
- NHANES 2
- Method Validation Studies 2
- DHANES Method Validation Studies 3
- Data Simulation: Pseudo-crossover Studies 4
- Data Analysis 7

Results 10
- Main Analysis: Data Summary and Adjustment Concordance 10
- Joint Analysis: Three-by-three Concordance Measures 10
- Joint Analysis: Dichotomized Concordance Measures 12
- Joint Analysis: Logistic Regression 12
- Joint Analysis: Stratified Adjustment Concordance 12

Discussion 12

References 14

Appendix I. Method Comparison R Program 25

Appendix II. Joint Analysis: Three-by-three Concordance Measures After Omitting the Data Set Adjusted With Ordinary Least Squares Regression 31

Text Figures

1. Parameters for sample selection of 120 pseudo-crossover studies from the full 2017–2018 National Health and Nutrition Examination Survey laboratory data and the simulation of new measurement values 5
2. Sample sizes and sampling process for 120 pseudo-crossover data sets 6
3. Simulating data for 120 pseudo-crossover data sets using 2017–2018 National Health and Nutrition Examination Survey laboratory data 7
4. Analysis process for individual pseudo-crossover studies: Determining final adjustment recommendations and classifying concordance category 8
5. Distribution of unweighted percentiles for old measurement values, by simulated pseudo-crossover study and analyte: 2017–2018 National Health and Nutrition Examination Survey 11

Detailed Tables

1. Unweighted percentile summaries of selected analytes: 2017–2018 National Health and Nutrition Examination Survey and aggregated pseudo-crossover study data 16
2. Unweighted frequencies and percentages of sampling method and difference type, by analyte: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 17
3. Unweighted frequencies and percentages, by concordance category: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 18
4. Unweighted frequencies and percentages of recommended adjustments, by difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 19

Contents—Con.

5. Unweighted concordance measures of recommended adjustments, by difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 20

6. Unweighted frequencies and percentages of adjustment recommendation, by dichotomized difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 20

7. Unweighted percentages for selected concordance measures for the dichotomized difference type where nonrandom differences are the positive result: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 21

8. Unweighted coefficient estimates, odds ratios, and *p* values from reduced and full logistic regressions modeling the probability that the final adjustment recommendation matched the difference type, by selected characteristics: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 22

9. Unweighted frequencies and percentages of concordance categories, by selected analytes: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 23

10. Unweighted frequencies and percentages of concordance categories, by difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 24

Appendix Tables

I. Unweighted frequencies and percentages of difference type, by recommended adjustment, omitting data set adjusted by ordinary least squares regression: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 31

II. Unweighted percentages for selected concordance measures, by difference type, omitting data set adjusted by ordinary least squares regression: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 32

III. Unweighted frequencies and percentages of selected concordance categories, by sampling method: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies 32

Acknowledgment

The authors wish to acknowledge Renee Storandt, leading clinical laboratory scientist in the Division of Health and Nutrition Examination Surveys, for providing subject matter expertise on the laboratory components of this research and for reviewing the final manuscript.

Assessing Laboratory Method Validations for Informing Inference Across Survey Cycles in the National Health and Nutrition Examination Survey

by Kevin Chuang, M.P.H., Jennifer Rammon, M.S., Hee-Choon Shin, Ph.D., and Te-Ching Chen, Ph.D.

Abstract

Background and objectives

Laboratory tests conducted on survey respondents' biological specimens are a major component of the National Health and Nutrition Examination Survey. The National Center for Health Statistics' Division of Health and Nutrition Examination Surveys performs internal analytic method validation studies whenever laboratories undergo instrumental or methodological changes, or when contract laboratories change. These studies assess agreement between methods to evaluate how methodological changes could affect data inference or compromise consistency of measurements across survey cycles. When systematic differences between methods are observed, adjustment equations are released with the data documentation for analysts planning to combine survey cycles or conduct a trend analysis. Adjustment equations help ensure that observed differences from methodological changes are not misinterpreted as population changes. This report assesses the reliability of statistical methods used by the Division of Health and Nutrition Examination Surveys when conducting method validation studies to address concerns that adjustment equations are being overproduced (recommended too frequently).

Methods

Public-use 2017–2018 National Health and Nutrition Examination Survey laboratory data were used to simulate "new" measurements for 120 analytic method validation studies. Blinded studies were analyzed to determine the final adjustment recommendation for each study using difference plots, descriptive statistics, *t*-tests, and Deming regressions. Final recommendations were compared with simulated difference types to assess how often spurious results were observed. Concordance estimates (concordance, misclassification, sensitivity, specificity, and positive and negative predictive values) informed assessments.

Results

Adjustment equations were appropriately recommended for 75.0% of the studies, over-recommended for 5.8%, under-recommended for 15.8%, and recommended with an inappropriate technique for 3.3%. Across simulated difference types, sensitivity ranged from 65.9% to 84.4% and specificity from 74.7% to 97.5%.

Conclusions

Findings from this report suggest that the current methodology used by the Division of Health and Nutrition Examination Surveys performs moderately well. Based on these data and analyses, underadjustment was more prevalent than overadjustment, suggesting that the current methodology is conservative.

Keywords: crossover • bridging • calibration • trends • biochemical assays • National Health and Nutrition Examination Survey (NHANES)

Introduction

Clinical laboratories conduct tests on biospecimens to obtain information about the health of a subject. This often includes measuring the amount of substance in a sample, such as the concentration of glucose in plasma or folate in serum. In each case, the laboratory uses assays (analytical tests) that have been uniquely developed for that substance. Inevitably, as measurement methods improve (such as manufacturer reformulation of chemical reagents used in assays or revised detection parameters) or measurement instruments are discontinued, new methods are introduced and comparisons must be made to assess agreement between the two methods and to evaluate the potential impact of an assay change. The process of comparing measurement assays has been written about extensively (1,2). In most cases, comparison studies are meant to demonstrate that an analytic procedure is suitable for its intended purpose,

or that analytic measurement methods are comparable across study sites. They are not meant to indicate that two measurement methods are identical.

The National Health and Nutrition Examination Survey (NHANES) is a nationally representative survey conducted by the National Center for Health Statistics and features a complex, multistage design that combines interviews and health examinations. As part of its laboratory component, blood, urine, and other biological and environmental specimens are collected, processed, stored, and shipped (3). Collectively, these specimens provide data about the health of the U.S. civilian noninstitutional population (4). Specifically, NHANES data is used to determine the prevalence and risk factors of major diseases; inform development of health policies, programs, and services that affect the nation; and provide national standards. For example, blood lead data were instrumental in developing policy to eliminate lead from gasoline; survey data combined with laboratory glucose data continue to indicate that undiagnosed diabetes is a significant problem in the United States; and national programs to reduce cholesterol levels depend on NHANES laboratory data to identify at-risk populations and measure success in curtailing risk factors associated with heart disease, the nation's number one cause of death (5–7).

Whenever the Centers for Disease Control and Prevention and contracted laboratories undergo instrumental or methodological measurement changes or when contracted laboratories change, the Division of Health and Nutrition Examination Surveys (DHANES) laboratory methods workgroup (LMW) performs analytic method validation studies to evaluate how changes in methodology may influence data inference. LMW consists of clinical laboratory scientists and statisticians. Unlike traditional method validation studies, in which clinical laboratories focus on the accuracy or reliability of a measurement method, LMW evaluates the consistency of measurements from one survey cycle (that is, the 2-year data release cycle) to the next as well as consistency within a survey cycle. Specifically, LMW is concerned about analyses that combine survey cycles or trend analyses conducted with data that include cycles where measurement methods changed (such as a trend analysis across five survey cycles in which two cycles were measured using one assay and three cycles were measured using a different assay). When systematic differences are observed between measurement methods, DHANES releases adjustment equations (that is, linear prediction equations for the new measurement based on the old measurement, such as $NEW\ MEASUREMENT = INTERCEPT + SLOPE \bullet OLD\ MEASUREMENT$) for use by data analysts along with the data documentation. The goal of adjustment equations is to ensure that differences from instrumental or measurement-related changes are not falsely interpreted by analysts as changes in the U.S. population over time.

This report aims to assess the reliability of the statistical methodology used by LMW to address LMW members' concerns that adjustment equations may be overproduced (recommended too frequently). This was done by using publicly available NHANES data to simulate blinded pseudo-analytic method validation studies (pseudo-crossover studies), performing analyses on the pseudo-crossover studies using the same methodology that LMW uses, and comparing results with the "truth" from the simulations to identify if adjustment equations are being overproduced, underproduced (not recommended often enough), or appropriately produced (recommended when needed and not otherwise). Lastly, this report seeks to identify potential associations between sample characteristics (such as sample size, sampling method, and simulated difference type) and the likelihood of an appropriate (concordant) adjustment.

Methods

NHANES

NHANES is a complex, multistage probability sample of the U.S. civilian noninstitutionalized population conducted by the National Center for Health Statistics. Household interviews, standardized health examinations conducted in mobile examination centers, and laboratory tests are used to collect data. The mobile examination centers are staffed by full-time personnel, including certified phlebotomists who perform venipuncture using a standardized protocol (8). Laboratory testing of NHANES specimens is conducted by the Centers for Disease Control and Prevention in Atlanta and contracted laboratories across the country. These laboratories use state-of-the-art measurement methods based on their expertise about which bioassay is most appropriate for each analyte (that is, the substance whose chemical components are being identified and measured) in collaboration with the DHANES laboratory project officer. For health examinations, participants age 18 and older provide written consent; documented assent is obtained from participants ages 7–17 as well as written permission from a parent or other legal guardian; and for participants younger than age 7, written permission is obtained from a parent or other legal guardian.

NHANES protocols are approved by the National Center for Health Statistics Ethics Review Board. Beginning in 1999, NHANES became a continuous, cross-sectional survey based on 4-year sample designs, where nationally representative data from the continuous NHANES were collected and released in 2-year cycles between 1999 and 2018. Overall examination response rates for children and adults ranged from 76% during 1999–2000 to 49% during 2017–2018 (9).

Method Validation Studies

Analytic method validation studies (that is, method comparison, crossover, bridging, or calibration studies) compare two assays that measure the concentration of a particular analyte to assess the degree of agreement

between the two measurement methods and to identify any systematic differences. These types of studies are performed by both manufacturers and clinical laboratory professionals whenever procedural changes, such as advances in laboratory methods or discontinuation of instruments, occur. Standard methods for conducting analytic method validation studies have been published by the Clinical and Laboratory Standards Institute, Bland and Altman, Westgard QC, the Food and Drug Administration, and others (1,2,10). Ideally, a candidate assay is compared with a generally accepted standard or reference assay. However, for many comparisons, neither measurement method is the "gold standard," or produces the true measurement value without error; in this case, a candidate assay is compared with the best assay currently available. When no standard assay is available there is no "true" value to compare with, and therefore estimated differences do not necessarily indicate true bias.

DHANES Method Validation Studies

LMW performs internal analytic method validation studies whenever the Centers for Disease Control and Prevention or contracted laboratories undergo instrumental or methodological measurement changes or when contracted laboratories change. These changes can occur between or within survey cycles. One example of such a change is when manufacturers adjust the formulation of a reagent used in an assay. Other examples include changes in how analytes are detected or how machines are calibrated. As previously mentioned, LMW focuses on the consistency of measurements from one survey cycle to the next. When systematic differences are observed between measurement methods, DHANES releases adjustment equations with the data documentation for use by analysts planning to combine survey cycles or conduct a trend analysis. Use of the adjustment equations ensures that any identified differences between survey cycles are reflective of changes in the U.S. population rather than methodological measurement changes.

Before the formation of LMW in 2018, these comparisons were performed by a clinical pathologist. The current workgroup uses this same approach, which is also outlined in guidelines published by the Clinical and Laboratory Standards Institute in "EPO9c: Measurement Procedure Comparison and Bias Estimation Using Patient Samples" (2). These guidelines provide recommendations for quantifying systematic differences across measurement procedures and promote proper and effective data analysis. In brief, this approach involves:

1. Reviewing descriptive estimates (that is, means and percentiles) of the measurements provided by the old and new measurement methods,
2. Reviewing visual displays (such as scatterplots and difference plots) to compare old and new measurement methods,
3. Evaluating difference plots for patterns to determine if observed differences between the two measurement procedures are constant or proportional to the concentration on the horizontal axis (old measurement value),
4. Statistical testing (that is, *t*-tests), and
5. Regression analyses (such as ordinary least squares, constant Deming, and weighted Deming).

For method comparison studies, Deming regressions are preferred over ordinary least squares (OLS) regressions. Unlike OLS regressions, which assume no measurement error on the *x*-axis (old measurement value), Deming regressions assume measurement errors on both the *x*- and *y*-axes (old and new measurement values, respectively). This is appropriate because there is known imprecision in both measurement procedures. Deming regressions account for the measurement errors by minimizing the sum of the distances between the measured values and the regression line at an angle specified by the variance ratio (the ratio of the errors associated with each measurement). Additionally, Deming regression acknowledges the independence of errors for both measurement methods. In making these assumptions, Deming regression can avoid biased estimates of the slope. A variation of the constant Deming regression is the weighted Deming regression, which gives each point a weight inversely proportional to the square of the measurement concentration on the *x*-axis (in this case, the old measurement value) (11,12).

When the observed difference between the old measurement method and the new measurement method is constant across measurement values, emphasis is placed on the paired *t*-test results (step 4, statistical testing) and constant Deming regression results (step 5, regression analyses). When the observed difference is proportional across measurements, emphasis is placed on the independent *t*-test results, testing the relative difference (step 4, statistical testing) and weighted Deming regression results (step 5, regression analyses). In the most classic example, an observed constant difference leads to a statistically significant paired *t*-test and a statistically significant intercept in the constant Deming regression analysis. Observed proportional differences typically lead to a statistically significant independent *t*-test and a statistically significant slope term in the weighted Deming regression analysis (2).

Usually, if a result is statistically significant, an adjustment equation is recommended by DHANES and released with the data documentation. When the observed difference is constant, the recommended adjustment equation most often corresponds to the estimated constant Deming regression line. If the observed difference is proportional to the old measurement concentration, the recommended adjustment equation often corresponds to the estimated weighted Deming regression line. It is worth emphasizing that released data are not adjusted by DHANES. Instead, adjustment

equations are provided in the data documentation for analysts to use when preparing their data.

Data Simulation: Pseudo-crossover Studies

For this project, 120 data sets (pseudo-crossover studies) were created using 2017–2018 NHANES public-use laboratory data (13). The results presented in this report are based on the simulated pseudo-data sets described and are not taken directly from any previously conducted NHANES analytic method validation studies or from publicly released adjustment equations. This report does not critique specific analytic method validation studies performed by DHANES. Instead, it is intended as an illustration of how the current methodology for producing publicly released adjustment equations performs.

Analytes were chosen from a master list of previously conducted DHANES method validation studies. All analytes that were chosen had been evaluated by LMW at least once since 1999–2000 and special consideration was given to analytes that had been evaluated more than once. Additional efforts were made to include a variety of analytes with different characteristics, such as small and large numeric measurement ranges, small and large variations in the distribution of measurements, and differing clinical functions. The final chosen analytes were creatinine, ferritin, folate, high-density lipoprotein (HDL) cholesterol, insulin, and vitamin C.

For each of the six analytes, 20 simulation data sets were created as outlined in Figures 1 and 2. For each of the 120 simulation data sets, two sampling parameters were selected randomly: sampling method and sample size (Figure 1). Sampling method was chosen from a Bernoulli (2/3) distribution, with "success" corresponding to a systematic random sample and "failure" corresponding to a simple random sample. For systematic random sampling, survey participants were ordered from least to greatest by measurement value, a starting point was selected at random, and every k^{th} member was selected to be in the sample, where

$$k_i = \frac{N_A}{n_i}$$

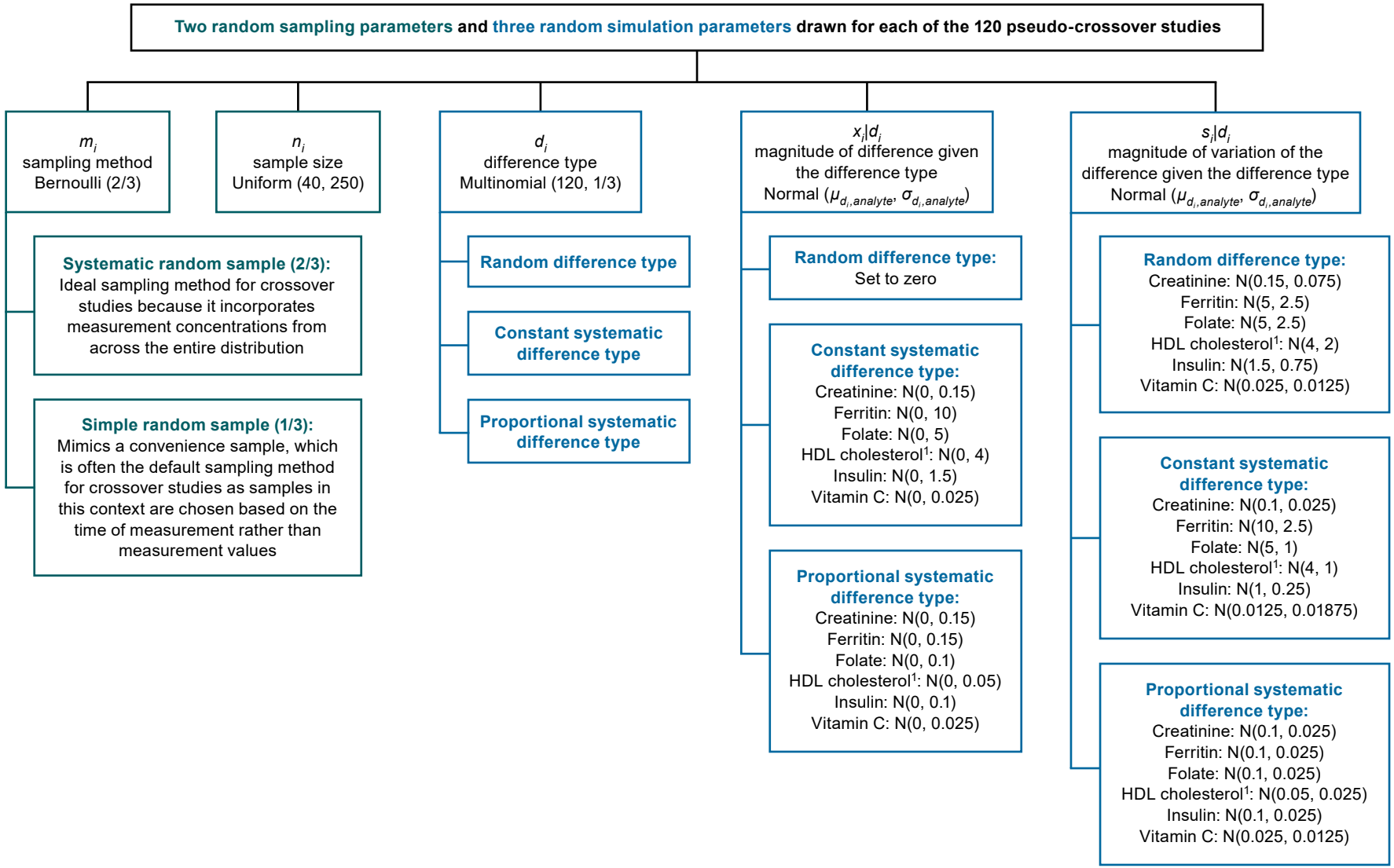
N_A = *total sample size available for analyte A*, and n_i = *selected sample size for simulated data set i* (Figure 2). Systematic random sampling is ideal for method validation studies because it incorporates measurement concentrations from across the entire distribution. However, in many cases (NHANES included), convenience samples are used instead and, in this context, where convenience samples are chosen based on the time of measurement and are not dependent on measurement values, simple random samples mimic convenience samples well. In this context, neither systematic random sampling nor convenience sampling produces nationally representative samples and so nationally representative estimates are not made. Sample size was chosen from a Uniform (40,250) distribution and was always rounded down to the nearest whole number. The parameters for sample size correspond to the guidelines provided by the Clinical and Laboratory Standards Institute report (2) for the design of measurement procedure comparison experiments and to the sample sizes generally seen in DHANES method validation studies. Samples were drawn from the 2017–2018 NHANES public-use data using PROC SURVEYSELECT in SAS 9.4 (Figure 2) (14).

Once simulation samples were chosen, three simulation parameters were randomly selected for each data set: 1) difference type (random, constant, or proportional), 2) the magnitude of the difference given the difference type, and 3) the magnitude of the variation of the difference given the difference type. Figure 1 outlines the statistical distributions used to select the simulation parameters. Difference type was selected from a multinomial distribution with an equal probability for each group to ensure that each difference type had the same probability of being selected. Parameters 2 and 3 varied by difference type and analyte. The statistical distributions for 2 and 3 were based on the difference distributions of previously conducted DHANES method validation studies, as well as the basic characteristics of normal distributions (for example, about 99% of values in the distribution are within three standard deviations of the mean).

Figure 3 indicates how the simulation parameters were used to create variables within the simulation data sets (pseudo-crossover studies). For each data set, original measurement values represented "old measurements." "New measurement" values were created by combining the old measurement value with a randomly chosen error term ($e_{i,j}$) based on the simulation parameters. For all j individuals in a data set, the term $e_{i,j}$ was chosen from a normal distribution, Normal ($x_i|d_i$, abs($s_i|d_i$)), where $x_i|d_i$ represents the selected magnitude of the difference given the difference type (randomly drawn from a separate normal distribution as indicated in Figure 1) and $s_i|d_i$ represents the selected magnitude of the variation of the difference given the difference type for data set i (randomly drawn from a separate normal distribution as indicated in Figure 1). For data sets with random or constant difference types, the new measurement value for each individual was set equal to the old measurement value plus $e_{i,j}$ ($new_{ij} = old_{ij} + e_{ij}$). For data sets with a proportional difference type, the new measurement value for each individual was set equal to the old measurement value plus the product of the old measurement value and $e_{i,j}$ ($new_{ij} = old_{ij} + (old_{ij} \bullet e_{ij})$).

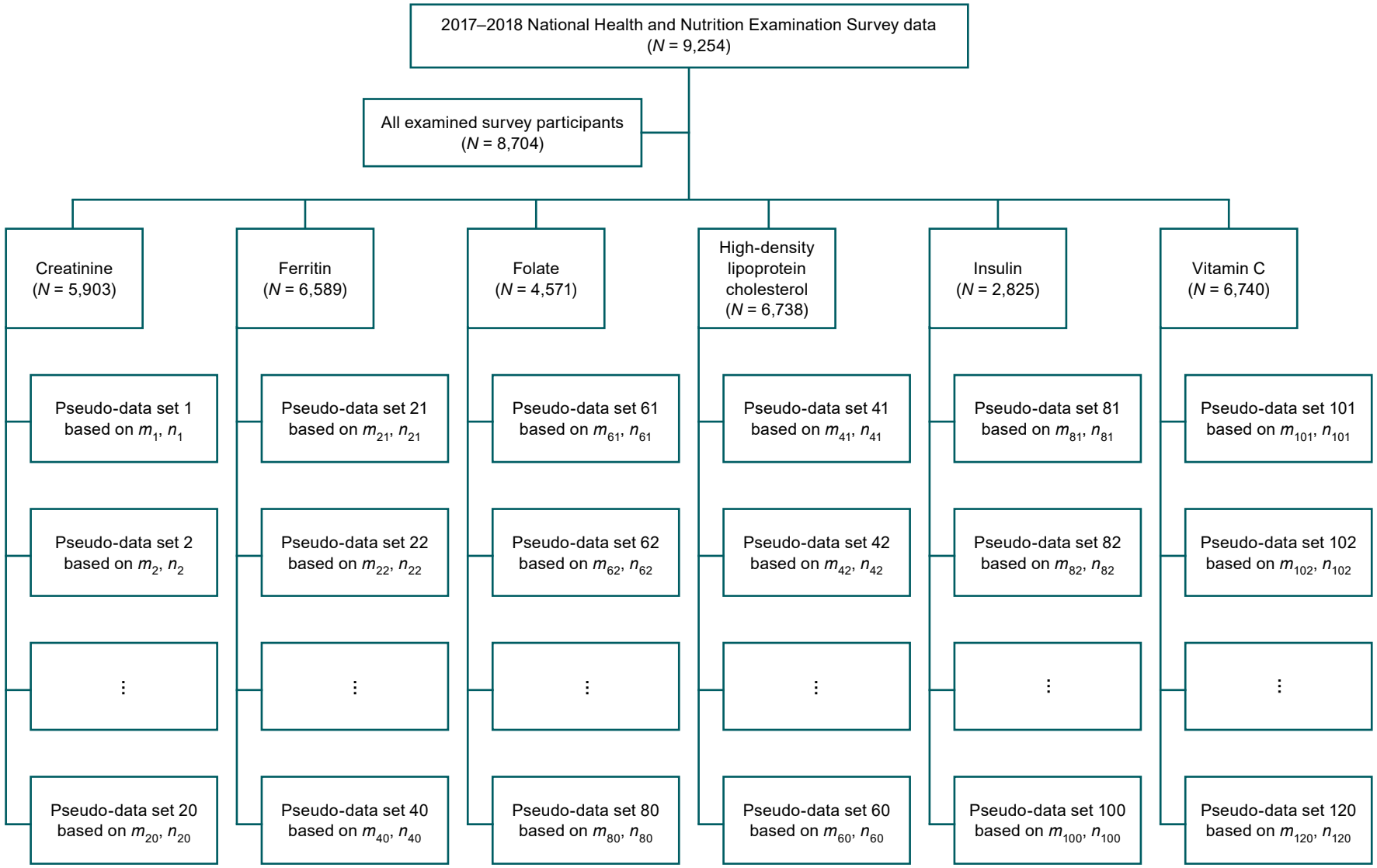
For new measurement values below the lower limit of detection, an imputed value was used. This value was the analyte-specific lower limit of detection (LLOD) divided by the square root of two and is the minimum value released by

Figure 1. Parameters for sample selection of 120 pseudo-crossover studies from the full 2017–2018 National Health and Nutrition Examination Survey laboratory data and the simulation of new measurement values



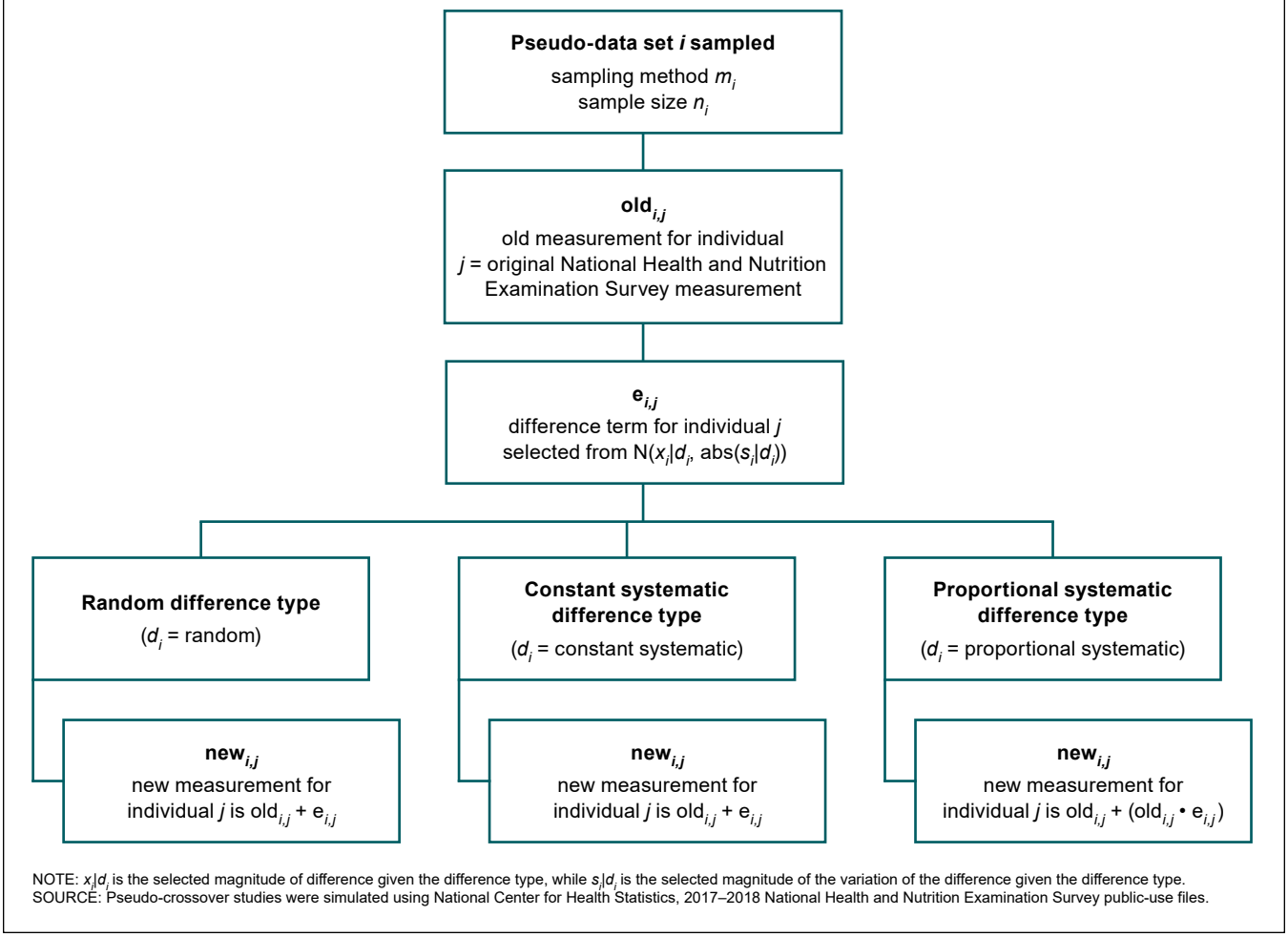
[1]HDL is high-density lipoprotein.
NOTES: Statistical distributions for $x_i|d_i$ and $s_i|d_i$ were based on the difference distributions of previously conducted Division of Health and Nutrition Examination Surveys method validation studies, as well as the basic characteristics of normal distributions (for example, approximately 99% of values in the distribution are within three standard deviations of the mean). For all j individuals in a data set i, an error term ($e_{i,j}$) was chosen from a normal distribution, $N(x_i|d_i, abs(s_i|d_i))$. Figure 3 in this report indicates how this error term was used to derive new measurement values.
SOURCE: National Center for Health Statistics.

Figure 2. Sample sizes and sampling process for 120 pseudo-crossover data sets using data from the full 2017–2018 National Health and Nutrition Examination Survey



NOTES: m_i is sampling method and n_i is sample size of pseudo-data set i. The total sample size of each analyte reflects the eligible sample as defined by the documentation (available from: https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory&CycleBeginYear=2017): creatinine (LBXSCR in Standard Biochemistry Panel), ferritin (LBXFER in Ferritin), folate (LBDFOT in Folate Forms—Total & Individual Serum), high-density lipoprotein cholesterol (LBDHDD in Cholesterol-High-Density Lipoprotein), insulin (LBXIN in Insulin), and vitamin C (LBXVIC in Vitamin C).
SOURCE: All samples are drawn from the National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Figure 3. Simulating data for 120 pseudo-crossover data sets using 2017–2018 National Health and Nutrition Examination Survey laboratory data



NOTE: $x_i|d_i$ is the selected magnitude of difference given the difference type, while $s_i|d_i$ is the selected magnitude of the variation of the difference given the difference type.
SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

DHANES in the public-use data files (13):

$$\frac{LLOD_{analyte}}{\sqrt{2}}$$

(that is, 0.070 for creatinine, 0.400 for ferritin, 0.720 for folate, 2.000 for HDL cholesterol, 0.710 for insulin, and 0.021 for vitamin C).

Data Analysis

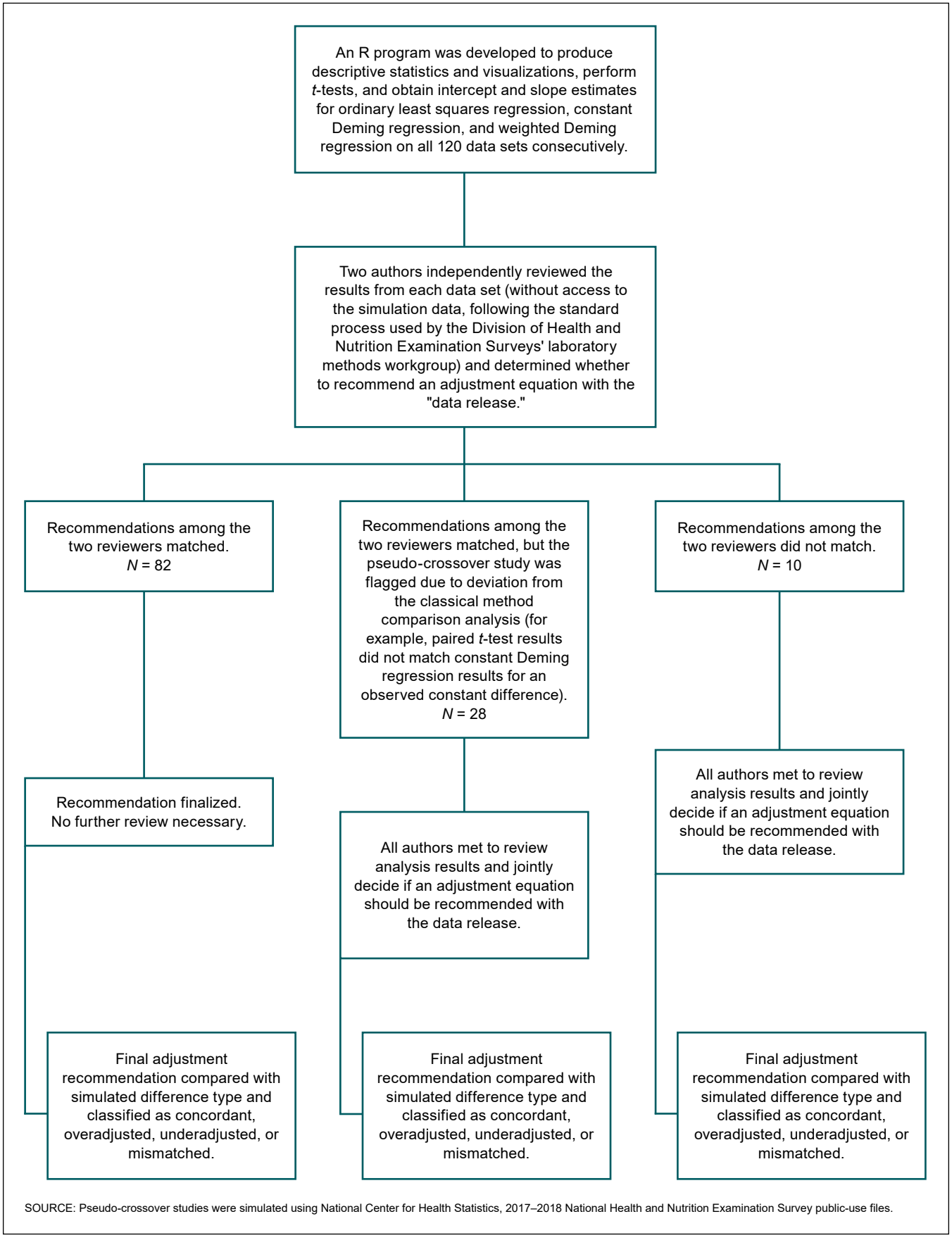
Analysis of individual pseudo-crossover studies

Figure 4 outlines the analysis process. First, a program (see Appendix I) was developed in R version 4.0.3 using the mcr (Method Comparison Regression), tidyverse, and pastecs packages that produced descriptive statistics and visualizations; performed paired *t*-tests of the constant difference and independent *t*-tests of the relative difference; and provided intercept and slope estimates for the OLS regression line, the constant Deming regression line, and the weighted Deming regression line (15–18). The program iterated through all 120 data sets consecutively. All analyses were unweighted because subsamples and estimates were not nationally representative.

Once results were obtained, two authors used the same procedure as LMW to review the results from each data set to determine if differences were random, constant, or proportional to the old measurement value and to assess if the publicly released data should be released with an adjustment equation. The two authors worked independently, using only the output, and were blinded from the simulation data. An alternative approach may have been to use statistical testing to assess releasing an adjustment equation. Statistical testing would avoid introducing researcher bias. However, the decision to manually review all results adheres to the process currently used by LMW, which intentionally relies on subject matter expertise to help inform decision making because clinical chemistry data are complex.

In cases where the authors made different recommendations or observed any sort of deviation from the classic method comparison analysis (n = 38), data sets were flagged for additional review. In some cases, data sets were flagged for

Figure 4. Analysis process for individual pseudo-crossover studies: Determining final adjustment recommendations and classifying concordance category



SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

further review due to methodological interests not directly related to this project, such as differences between OLS regression results and Deming regression results. During the further review process, all authors reviewed the data sets together to determine, by unanimous consent, whether adjustment equations should be recommended. This meeting mimics how LMW makes decisions.

After determining the final adjustment recommendations for all 120 pseudo-crossover studies, the authors compared the results with the simulated difference type to identify each data set as concordant, overadjusted (adjusted unnecessarily), underadjusted (a needed adjustment was not recommended), or mismatched (the wrong type of adjustment was recommended). Results were considered concordant if:

1. No adjustment was recommended for a simulated random difference,
2. An adjustment using the constant Deming regression equation was recommended for a simulated constant difference, or
3. An adjustment using the weighted Deming regression equation was recommended for a simulated proportional difference.

Results were considered an overadjustment if:

1. A constant Deming regression was recommended for a simulated random difference, or
2. A weighted Deming regression was recommended for a simulated random difference.

Results were considered an underadjustment if:

1. No adjustment equation was recommended for a simulated constant difference, or
2. No adjustment equation was recommended for a simulated proportional difference.

Results were considered mismatched if:

1. A constant Deming regression equation was recommended for a simulated proportional difference,
2. A weighted Deming regression equation was recommended for a simulated constant difference, or
3. An OLS regression equation was recommended for a simulated constant or proportional difference.

Joint analysis to determine concordance of final adjustments

After determining the final adjustment recommendations for all 120 pseudo-crossover studies, a joint analysis was conducted to identify adjustment patterns across the data sets. All joint analyses were conducted using SAS 9.4 (14).

First, a three-by-three concordance table was created to compare final adjustment recommendations with the simulated difference types. Standard concordance measures, including sensitivity, specificity, positive predictive value (PPV) (that is, precision), and negative predictive value (NPV), were calculated for each difference type (19). Considering each of the three difference types (Figure 1) as a class, shown as D, standard concordance measures in this context are defined as:

Sensitivity$_D$—How well does the implementation of this methodology recognize samples that belong to class D?

Specificity$_D$—How well does the implementation of this methodology recognize that a sample does not belong to class D?

PPV or precision$_D$—Given that the recommended adjustment equation corresponds to class D, what is the probability that the sample truly belongs to class D?

NPV$_D$—Given that the recommended adjustment equation does not correspond to class D, what is the probability that the sample truly does not belong to class D?

Second, to simplify the interpretation of the results, a two-by-two concordance table was created by collapsing constant and proportional difference types into one category (nonrandom difference type) and dichotomizing adjustment recommendations (adjust or not adjust). Concordance, misclassification rate, sensitivity, specificity, precision, and NPV were calculated for this table by treating the nonrandom difference type as the positive result and the random difference type as the negative result. In this context, concordance measures are defined as:

Sensitivity—How well does the implementation of this methodology correctly classify that data with a systematic (nonrandom) difference type should be released with a recommended adjustment equation?

Specificity—How well does the implementation of this methodology correctly classify that data with random differences should not be released with a recommended adjustment equation?

PPV or precision—Given that an adjustment equation is recommended, what is the probability that differences are truly systematic (nonrandom)?

NPV—Given that an adjustment equation is not recommended, what is the probability that differences are truly random?

Third, using logistic regression, potential associations were explored between sample characteristics and the likelihood of the final adjustment matching the simulated difference type. This exploratory analysis was used to identify which variables to stratify by in subsequent analyses. The results were not used for any other purpose. The outcome variable was defined as a binary indicator: Either the final adjustment recommendation matched the simulated difference type, or the final adjustment recommendation did not match the simulated difference type. Independent variables included sample size (continuous), sampling method (simple random

sample or systematic random sample), analyte (creatinine, ferritin, folate, HDL cholesterol, insulin, or vitamin C), difference type (random, constant, or proportional), and four binary distribution variables (normal or skewed) based on visual assessment and corresponding to the distribution of old measurement values, distribution of new measurement values, constant difference distribution, and proportional difference distribution. Two models were formed. The base model included sample size, sampling method, difference type, and analyte, while the full model included all independent variables.

Fourth, stratification variables were identified using results from the logistic regression model, and concordance category tabulations (that is, concordant, overadjusted, underadjusted, and mismatched) were stratified.

Results

Main Analysis: Data Summary and Adjustment Concordance

Unweighted percentile summaries and means for the full 2017–2018 NHANES data and the aggregated pseudo-crossover data are presented in Table 1 by analyte. This demonstrates that for each analyte, the distribution of sampled old measurement values in the simulation data sets aligns well with the distribution of measurement values in the full 2017–2018 NHANES data. Figure 5 presents the distribution of unweighted percentiles across individual pseudo-crossover studies by analyte. This demonstrates the similarity that the 20 simulated data sets for each analyte have with one another across the distribution of sampled old measurement values. As expected, percentile distributions across the pseudo-crossover data by analyte were similar to one another (Figure 5) and similar to the full 2017–2018 NHANES data (Table 1) with expected variation in the extremes (the minimum and maximum values).

Table 2 shows the distributions of sampling method and simulated difference type across data sets by analyte. By design, more data sets were selected using systematic random sampling than simple random sampling because systematic random sampling is the ideal method for designing crossover studies. Vitamin C happened to have more data sets selected using simple random sampling. Distribution of difference type was relatively even across the data sets by analyte, with a slightly higher percentage of HDL cholesterol and insulin data sets having a random difference type.

Following the individual analyses of the pseudo-crossover studies, 38 (31.7%) data sets were flagged for additional review as described in Figure 4. In those,

- 10 were flagged because of differing recommendations between the two independent reviewers;
- 20 were flagged because of departures from standard expectations, such as:
  - Paired *t*-test results did not match constant Deming regression results for an observed constant difference,
  - Results from an independent *t*-test of the relative difference did not match weighted Deming regression results for an observed proportional difference, or
  - Statistically significant regression results were noted for an observed random difference; and
- 8 were flagged because of unexplainable differences or miscellaneous reasons, including:
  - Unexplainable differences between OLS and constant Deming regression estimates for intercept or slope,
  - Unexplainable differences between the weighted Deming and constant Deming regression estimates for intercept or slope in cases of observed constant differences, and
  - Unusual difference patterns.

In the latter case (unexplainable or miscellaneous), data sets were flagged primarily because of research questions that are beyond the scope of this report (such as comparisons between OLS and Deming regression analyses in the context of method comparison studies), but are still of interest to the authors and LMW. For our purposes, additional review was expected to minimize bias and, therefore, was treated as the default if any concern about a particular data set remained.

Table 3 summarizes the concordance categories of the final adjustment recommendations when compared with simulated difference types. Specifically, 90 (75.0%) pseudo-crossover studies had a concordant adjustment, 19 (15.8%) data sets were underadjusted, 7 (5.8%) were overadjusted, and 4 (3.3%) had a mismatched adjustment.

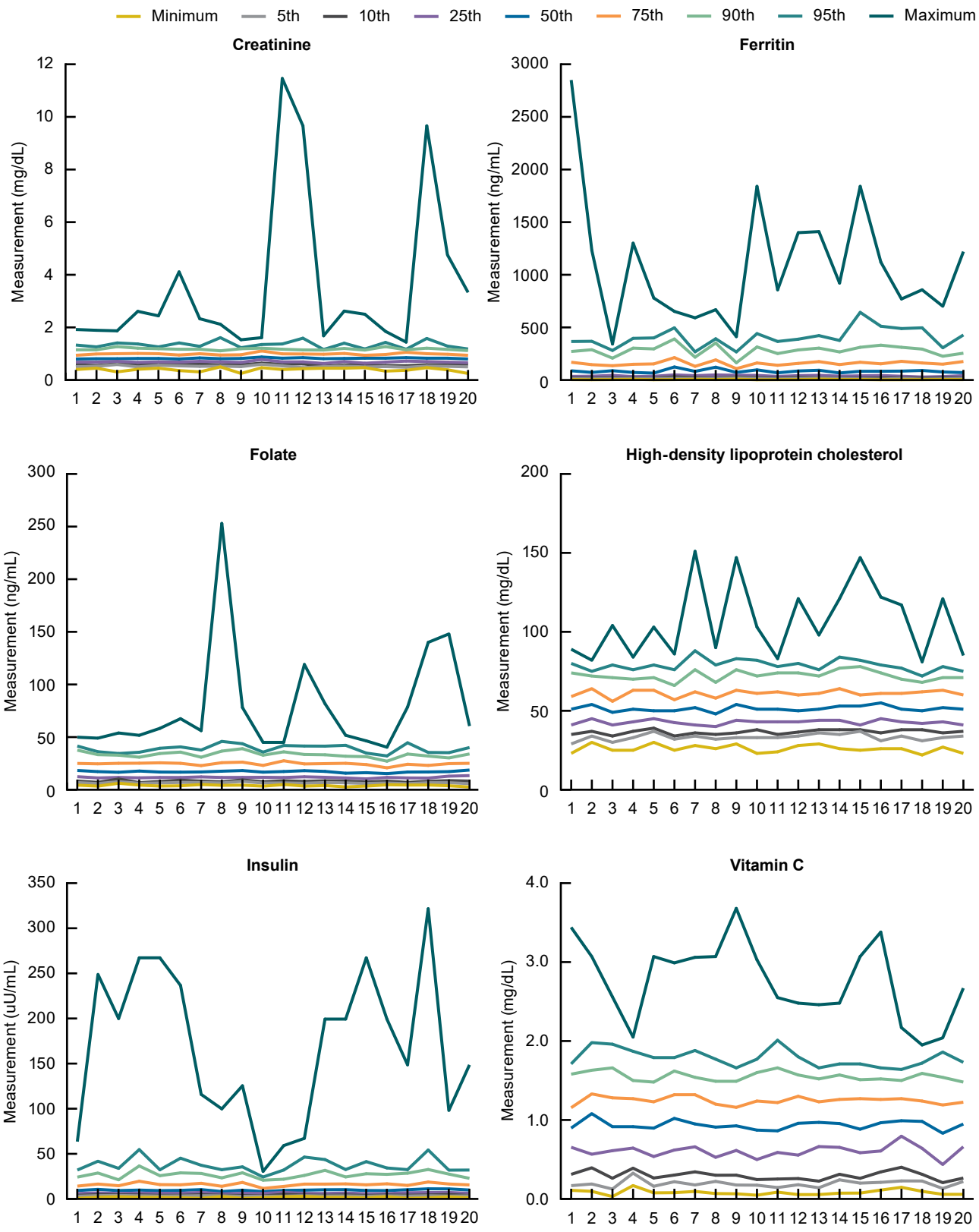
Joint Analysis: Three-by-three Concordance Measures

Table 4 presents the three-by-three concordances comparing the final adjustment recommendations with the simulated difference types. Estimates of sensitivity, specificity, PPV or precision, and NPV by difference type (Table 5) are outlined below.

1. For the random difference type (treating a random difference as a positive result), sensitivity was 84.4%, specificity was 74.7%, PPV or precision was 66.7%, and NPV was 88.9%.
2. For the constant difference type (treating a constant difference as a positive result), sensitivity and PPV or precision were 76.5%, while specificity and NPV were 90.7%.
3. For the proportional difference type (treating a proportional difference as a positive result), sensitivity was 65.9%, specificity was 97.5%, PPV or precision was 93.1%, and NPV was 84.6%.

One data set had a large standard error associated with the intercept estimate from the constant Deming regression,

Figure 5. Distribution of unweighted percentiles for old measurement values, by simulated pseudo-crossover study and analyte: 2017–2018 National Health and Nutrition Examination Survey



NOTE: Measurements presented reflect the respective units of measure of each analyte.
SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

and the recommended adjustment equation was an OLS regression. It is unusual for an OLS regression adjustment to be chosen over a constant Deming or weighted Deming regression adjustment. Therefore, for the concordance tables this data set was treated as a data set with a constant difference. In Appendix II, Tables I and II, a duplicate analysis was performed after omitting the data set adjusted with OLS. When omitting the OLS data set, concordance for the constant difference type reduces to 75.8%.

Joint Analysis: Dichotomized Concordance Measures

To potentially simplify interpretations, the dichotomized concordance table and concordance measures are presented as Table 6 and Table 7, respectively. When collapsing rows and columns, overall concordance increased to 78.3%, while overall misclassification decreased to 21.7% (Table 7). Treating a nonrandom difference as a positive result, sensitivity was 74.7%, specificity was 84.4%, PPV or precision was 88.9%, and NPV was 66.7%. These results are identical to those shown above for Table 5 for random difference type except that the dichotomized analysis identifies a nonrandom difference as a positive result (sensitivity describes the ability to classify a nonrandom difference as needing an adjustment), while Table 5 for random difference type identifies a random difference as a positive result (sensitivity describes the ability to classify a random difference as not needing an adjustment). These results provide different perspectives to the same findings.

Joint Analysis: Logistic Regression

The regression coefficients from the two logistic regression models are presented in Table 8 (with standard errors, odds ratios [OR] with 95% confidence intervals, and *p* values). Recall that in both models the outcome variable was defined as a binary indicator: 1) the final adjustment recommendation matched the simulated difference type, and 2) the final adjustment recommendation did not match the simulated difference type. In both models, a statistically significant association between analyte and the outcome variable was noted. For example, ferritin (OR_{Base} = 9.51, p_{Base} = 0.01; OR_{Full} = 8.26, p_{Full} = 0.03) was more positively associated with the probability of a concordant adjustment as compared with creatinine. An association between difference type and the outcome variable was also seen. In the full model, a proportional difference type was negatively associated with the probability of a concordant adjustment as compared with a random difference type (*OR* = 0.27, *p* = 0.03), and in the base model a similar relationship was observed (*OR* = 0.33, *p* = 0.05). Some nonsignificant results could be due to the relatively large size of the standard errors (in relation to the coefficient estimates), suggesting that the study may be underpowered for the logistic regression analysis. However, because the study was only being used as an exploratory analysis, the findings provided motivation to stratify by analyte and difference type.

Joint Analysis: Stratified Adjustment Concordance

Concordance groups by analyte and difference type are shown in Table 9 and Table 10, respectively. Concordance groups by sampling method are shown in Appendix II, Table III. Stratification by sampling method was motivated by an *a priori* interest to identify recommendations for acquiring the samples used by LMW to evaluate methodological measurement changes rather than by the logistic regression results.

Table 9 shows that creatinine and HDL cholesterol had the most overadjustments (three each), compared with one for insulin and none for the other analytes. Creatinine and vitamin C had the most underadjustments (six and five, respectively), while the other analytes each had two. These differences across analytes may not be generalizable. A larger number of pseudo-crossover studies per analyte is needed to draw statistical comparisons. Table 10 reiterates that random difference types had the highest overall concordance rate (84.4%), followed by constant difference types (73.5%), and then proportional difference types (65.9%). It also displays some of the intrinsic limitations of each difference type, namely that random difference types may not be underadjusted or mismatched, while constant and proportional difference types may not be overadjusted. Appendix II, Table III indicates that when stratifying the concordance groups by sampling method, 10.6% of the data sets acquired through simple random sampling were underadjusted compared with 19.2% of the data sets acquired through systematic random sampling.

Discussion

This report presents findings obtained from an analysis aimed at evaluating the reliability of the statistical methodology used by the DHANES LMW for method validation studies. These studies provide an analytical framework for comparing laboratory procedures and help ensure appropriate data interpretation. LMW is most concerned with providing information to data users who plan to combine survey cycles or conduct a trend analysis using data that includes two different measurement procedures or laboratories. Using the 2017–2018 NHANES public-use files to simulate 120 pseudo-crossover studies, this report shows analysis of each study individually and then compares adjustment recommendations across the studies to assess how often adjustment equations were appropriately recommended, overproduced, underproduced, or mismatched based on the simulated difference type.

This analysis suggests that the current methodology used by LMW to assess and produce adjustment equations

performs moderately well, with room for improvement. Specifically, adjustment equations were recommended when necessary 75.0% of the time, overadjusted 5.8% of the time, underadjusted 15.8% of the time, and mismatched 3.3% of the time (including the OLS-adjusted data set).

Regarding the initial concern that adjustment equations are recommended too frequently, this analysis displayed the opposite: Underadjustment was more prevalent than overadjustment. Just like Type I errors (false positive: reject null hypothesis even when it is true) are often preferred over Type II errors (false negative: fail to reject null hypothesis even when it is false) in traditional statistical analyses, and false positives are preferred over false negatives when screening for disease, in this context, a tendency to underadjust is preferred over a tendency to overadjust.

The higher prevalence of underadjustment compared with overadjustment suggests that the current methodology is conservative (that is, in instances where the decision to adjust or not adjust is unclear, the analysts conducting this study have a higher tendency to proceed without recommending an adjustment equation than to recommend an adjustment equation).

Standard concordance measures further confirm that, based on this analysis, the current methodology used by LMW to assess and produce adjustment equations is conservative. Based on the observed sensitivity values for this analysis, the methodology has an ability to correctly classify a random difference 84.4% of the time, correctly classify a constant difference 76.5% of the time, and correctly classify a proportional difference 65.9% of the time (Table 5). In other words, the likelihood of falsely identifying a nonrandom difference as random (25.3%) (1 - *specificity*) is much higher than the likelihood of falsely identifying a random or proportional difference as a constant difference (9.3%) or a random or constant difference as a proportional difference (2.5%).

Considering all results collectively, such that the null hypothesis corresponds to no statistically significant results or no adjustment and the alternative hypothesis corresponds to at least one statistically significant result or recommended adjustment, then a tendency to underadjust is like failing to reject the null hypothesis when the alternative hypothesis is correct and corresponds to a Type II error. When preferring Type II errors over Type I errors, high specificity and high PPV are desired.

In the dichotomized analysis, specificity indicates that the percentage of unbiased data sets (random difference type) with no adjustment recommendation (the negative result) was 84.4%, while sensitivity indicates the percentage of biased data sets (constant or proportional difference type) with an adjustment recommendation (the positive result) was 74.7%. Similarly, PPV or precision indicates that the proportion of adjusted data sets with a nonrandom difference type was 88.9%, and NPV indicates that the proportion of unadjusted data sets with a random difference type was 66.7%. These high specificity and PPV or precision measures are assuring.

The stratified analyses show that certain analytes are more prone to being overadjusted, while other analytes are more likely to be underadjusted. These tendencies are not fully understood and require further evaluation of the underlying characteristics for each analyte to understand how they might affect the analytes examined in future method validation studies. It could be that these observations are manifestations of the way the data were simulated. For example, observed differences across analytes may be attributed to the underlying distributions used to sample the magnitude of the difference given the difference type and the magnitude of the variation of the difference given the difference type. Moreover, two of the three analytes with data sets that resulted in an overadjustment (HDL cholesterol and insulin) also had a larger proportion (50%) of pseudo-crossover studies simulated with a random difference type (Table 2), which by design may only be adjusted appropriately or overadjusted (not underadjusted) (Table 10). As a result, observed differences across analyte may be attributed to differences in the distribution of difference type across data sets by analyte. However, definitive conclusions cannot be made because only 20 pseudo-crossover studies per analyte were simulated.

Finally, based on the analysis in Appendix II, Table III, 10.6% of the data sets acquired through simple random sampling were underadjusted compared with 19.2% of the data sets acquired through systematic random sampling. Correspondingly, 8.5% of the data sets acquired through simple random sampling were overadjusted compared with 4.1% of the data sets acquired through systematic random sampling. In practice, convenience samples are often used for crossover studies and simple random samples are more like convenience samples than systematic random samples. Therefore, if a larger proportion of crossover studies being conducted by LMW are based on convenience samples rather than systematic random samples, this finding could potentially explain the concern that data sets are being overadjusted.

A major limitation of this research is that potential reviewer bias may have affected the prevalence of underadjustments observed in this study due to reviewers' prior knowledge of the study's aims. However, as mentioned previously, LMW relies on manual review of statistical results, as opposed to mechanical decisions based strictly on statistical significance. As shown by the 38 data sets that were flagged for additional review, few data sets rigidly align with the classic example of a constant or proportional difference distribution, and human review is still essential for providing a full understanding of each method validation study and subsequent adjustment decisions. Clinical chemistry data are complex, and subject matter expertise regarding the way the data were measured or will be used in future analyses is important.

Another limitation could be the simulation model. As mentioned previously, it is unclear if some of the associations identified between sample characteristics and concordance could be attributed to the sampling distributions used to simulate the model. An alternative approach would be to choose a different sampling distribution for the error terms, such as a half-normal distribution. Using half-normal distributions instead of normal distributions would likely produce a larger proportion of error terms close to zero and a smaller proportion of error terms in the tails of the distribution. Another approach would be to consider multiple sampling distributions for each analyte. A third approach would be to simulate both the old and the new measurement values, as opposed to using observed NHANES data to obtain the old measurement values. This would provide an opportunity to be more systematic in choosing the magnitudes of difference across each analyte and would potentially allow for more uniqueness across data sets.

Finally, reviewers must consider the implications of using conservative methodology to assess and produce adjustment equations. When necessary, adjustment equations help ensure that differences from instrumental or measurement-related changes are not falsely interpreted by analysts as changes in the U.S. population over time. Adjustment equations are preferable over adjusted data because not all analysts will combine data from before and after a methodological change. However, some laboratory statisticians criticize the use of adjustment equations because each old measurement value is replaced by only one new measurement value, so adjustment equations may fail to account for important sources of variability, leading to underestimated standard errors, confidence intervals that are too narrow, and potentially incorrect *p* values (20). Just as failing to adjust a potentially influential difference between two measurement procedures could lead to false inference, overuse of adjustment equations also presents risk. Therefore, to consider the implications of conservative methodology on both past and future analyses of the NHANES laboratory data, reviewers must recognize that a higher tendency to adjust would not necessarily lead to more accurate analysis results.

Findings from this report could be used to inform the downstream effects of making inappropriate adjustment decisions. Specifically, this data could be used to assess the impact of making an overadjustment, an underadjustment, or a mismatched adjustment on analyses of the full 2017–2018 NHANES data set combined with other survey cycles (such as for trend analysis). This could help confirm if a conservative approach is truly preferable and help identify specific implications for failing to recommend a necessary adjustment equation or equivocally recommending an unnecessary adjustment equation.

References

1. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 8(2):135–60. 1999.
2. Clinical and Laboratory Standards Institute. EP09c: Measurement procedure comparison and bias estimation using patient samples. 3rd ed. 2018.
3. National Center for Health Statistics. National Health and Nutrition Examination Survey: Laboratory procedures manual. 2011.
4. National Center for Health Statistics. National Health and Nutrition Examination Survey: About NHANES. 2023. Available from: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm.
5. Centers for Disease Control and Prevention. Update: Blood lead levels—United States, 1991–1994. MMWR Morb Mortal Wkly Rep 46(7):141–6. 1997. Available from: https://stacks.cdc.gov/view/cdc/27040. Erratum: 46(26):607. 1997. Available from: https://stacks.cdc.gov/view/cdc/27059.
6. Mendola ND, Chen T-C, Gu Q, Eberhardt MS, Saydah S. Prevalence of total, diagnosed, and undiagnosed diabetes among adults: United States, 2013–2016. NCHS Data Brief, no 319. Hyattsville, MD: National Center for Health Statistics. 2018.
7. Carroll MD, Fryar CD. Total and high-density lipoprotein cholesterol in adults: United States, 2015–2018. NCHS Data Brief, no 363. Hyattsville, MD: National Center for Health Statistics. 2020.
8. National Center for Health Statistics. National Health and Nutrition Examination Survey: MEC laboratory procedures manual. 2017.
9. National Center for Health Statistics. National Health and Nutrition Examination Survey: Response rates. Available from: https://wwwn.cdc.gov/nchs/nhanes/ResponseRates.aspx.
10. Food and Drug Administration. Analytical procedures and methods validation for drugs and biologics: Guidance for industry. 2015.
11. Linnet K. Evaluation of regression procedures for methods comparison studies. Clin Chem 39(3):424–32. 1993.
12. Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. Clin Chem 44(5):1024–31. 1998.
13. National Center for Health Statistics. National Health and Nutrition Examination Survey: 2017–2018 laboratory data—continuous NHANES. Available from: https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory&Cycle=2017-2018.

14. SAS Institute, Inc. SAS/STAT 15.1 user's guide. Cary, NC: SAS Institute, Inc. 2018.

15. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2020. Available from: https://www.R-project.org/.

16. Potapov S, Model F, Schuetzenmeister A, Manuilova E, Dufey F, Raymaekers J. mcr: Method comparison regression. R package version 1.3.1 [computer software]. 2022. Available from: https://CRAN.R-project.org/package=mcr.

17. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. J Open Source Softw 4(43):1686. 2019.

18. Grosjean P, Ibanez F, Etienne M. pastecs: Package for analysis of space-time ecological series. R package version 1.3.21 [computer software]. 2018. Available from: https://CRAN.R-project.org/package=pastecs.

19. Beleites C, Salzer R, Sergo V. Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to the grading of astrocytoma tissues. Chemometr Intell Lab Syst 122:12–22. 2013.

20. Sternberg M. Multiple imputation to evaluate the impact of an assay change in national surveys. Stat Med 36(17):2697–719. 2017.

Table 1. Unweighted percentile summaries of selected analytes: 2017–2018 National Health and Nutrition Examination Survey and aggregated pseudo-crossover study data

Unweighted percentile and analyte	Sample size	Mean	Minimum value	5th	10th	25th	Median	75th	90th	95th	Maximum value
Full 2017–2018 National Health and Nutrition Examination Survey, observed											
Creatinine	5,903	0.88	0.25	0.53	0.58	0.68	0.82	0.98	1.16	1.31	12.74
Ferritin	6,589	133.39	1.04	12.80	19.10	36.60	80.70	165.00	295.00	414.00	5190.00
Folate	4,571	19.22	1.44	6.49	8.12	11.40	16.80	24.30	32.70	39.00	253.00
High-density lipoprotein cholesterol	6,738	53.39	10.00	34.00	37.00	43.00	51.00	61.00	72.00	80.00	189.00
Insulin	2,825	14.67	0.71	3.32	4.25	6.38	10.04	16.47	27.37	37.30	485.10
Vitamin C	6,740	0.95	0.02	0.18	0.29	0.60	0.95	1.25	1.55	1.76	14.60
Aggregated pseudo-crossover studies, simulated											
Creatinine	2,496	0.88	0.40	0.53	0.59	0.69	0.82	0.99	1.17	1.32	3.57
Ferritin	2,855	129.40	4.37	14.22	20.49	39.65	84.30	157.56	278.49	393.49	1088.00
Folate	2,662	19.38	4.05	6.69	8.20	11.78	17.12	24.53	33.15	38.66	78.69
High-density lipoprotein cholesterol	2,731	53.04	26.00	33.55	36.66	42.78	51.35	60.91	72.04	78.46	106.75
Insulin	2,691	14.74	1.65	3.38	4.24	6.42	9.72	15.85	26.61	36.47	168.15
Vitamin C	2,987	0.94	0.08	0.20	0.30	0.61	0.94	1.25	1.54	1.76	2.76

NOTES: All pseudo-data sets were selected independently; therefore, some individuals may be sampled in multiple pseudo-crossover studies. Sample sizes of aggregated pseudo-crossover studies indicate the number of total measurements represented across the 20 data sets for each analyte.

SOURCES: National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files. Pseudo-crossover studies were simulated using 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 2. Unweighted frequencies and percentages of sampling method and difference type, by analyte: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Analyte	Simulated sampling method		Simulated difference type		
	Simple random sample	Systematic random sample	Random	Constant	Proportional
	Frequency (percent)				
Creatinine	9 (45.0)	11 (55.0)	7 (35.0)	4 (20.0)	9 (45.0)
Ferritin	3 (15.0)	17 (85.0)	7 (35.0)	5 (25.0)	8 (40.0)
Folate	7 (35.0)	13 (65.0)	7 (35.0)	5 (25.0)	8 (40.0)
High-density lipoprotein cholesterol	8 (40.0)	12 (60.0)	10 (50.0)	6 (30.0)	4 (20.0)
Insulin	8 (40.0)	12 (60.0)	10 (50.0)	4 (20.0)	6 (30.0)
Vitamin C	12 (60.0)	8 (40.0)	4 (20.0)	10 (50.0)	6 (30.0)
Total	47 (39.2)	73 (60.8)	45 (37.5)	34 (28.3)	41 (34.2)

NOTES: Sampling method for each data set was chosen randomly from a Bernoulli distribution (2/3) with "success" corresponding to a systematic random sample. The simulated difference type for each data set was randomly chosen from a multinomial distribution (120, 1/3). Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 3. Unweighted frequencies and percentages, by concordance category: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance category	Frequency (percent)
Concordant adjustment[1]	90 (75.0)
Overadjustment[2]	7 (5.8)
Underadjustment[3]	19 (15.8)
Mismatched adjustment[4]	4 (3.3)
Total	120 (99.9)

[1]Final adjustment recommendation matched the simulated difference type.
[2]Adjustment was recommended but not needed.
[3]No adjustment was recommended when one was needed.
[4]Adjustment was recommended when one was needed, but adjustment type did not match the simulated difference type (including one data set adjusted by ordinary least squares regression).

NOTE: Total percentage does not sum to 100 because of rounding.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 4. Unweighted frequencies and percentages of recommended adjustments, by difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Simulated difference type	Recommended adjustment			Total
	No adjustment	Constant Deming	Weighted Deming	
	Frequency (percent)			
Random	38 (84.4)	6 (13.3)	1 (2.2)	45 (99.9)
Constant	7 (20.6)	26 (76.5)	1 (2.9)	34 (100.0)
Proportional	12 (29.3)	2 (4.9)	27 (65.9)	41 (100.1)

NOTES: Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration. Adjustments using the constant Deming and weighted Deming regressions incorporate measurement error for both the new and old measurement methods. One pseudo-data set was adjusted using ordinary least squares regression and was combined with the data sets that were adjusted using constant Deming regression. Row percentages are presented. Total row percentages for random and proportional difference types do not sum to 100 because of rounding.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 5. Unweighted concordance measures of recommended adjustments, by difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance measures	Simulated difference type		
	Random	Constant	Proportional
	Percent		
Sensitivity[1]	84.4	76.5	65.9
Specificity[2]	74.7	90.7	97.5
Positive predictive value or precision[3]	66.7	76.5	93.1
Negative predictive value[4]	88.9	90.7	84.6

[1]True positives divided by (true positives plus false negatives).
[2]True negatives divided by (false positives plus true negatives).
[3]True positives divided by (true positives plus false positives).
[4]True negatives divided by (false negatives plus true negatives).

NOTE: Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 6. Unweighted frequencies and percentages of adjustment recommendation, by dichotomized difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Simulated difference type	Recommended adjustment	
	Adjustment	No adjustment
	Frequency (percent)	
Nonrandom	56 (46.7)	19 (15.8)
Random	7 (5.8)	38 (31.7)

NOTES: Percentages derived using table total as denominator. Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration. Nonrandom simulated difference type combines constant and proportional simulated difference types. The adjustment category combines constant Deming regression adjustment with the recommended weighted Deming regression adjustment. Adjustments using the constant Deming and weighted Deming regressions incorporate measurement error for both new and old measurement methods.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 7. Unweighted percentages for selected concordance measures for the dichotomized difference type where nonrandom differences are the positive result: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance measure	Percent
Concordance	78.3
Misclassification	21.7
Sensitivity[1]	74.7
Specificity[2]	84.4
Positive predictive value (PPV) or precision[3]	88.9
Negative predictive value (NPV)[4]	66.7

[1]True positives divided by (true positives plus false negatives).
[2]True negatives divided by (false positives plus true negatives).
[3]True positives divided by (true positives plus false positives).
[4]True negatives divided by (false negatives plus true negatives).

NOTES: Nonrandom difference type comprises both the constant and proportional difference types. The calculated concordance and misclassification rate do not match presented rates (see Table 6 in this report) because of rounding.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 8. Unweighted coefficient estimates, odds ratios, and *p* values from reduced and full logistic regressions modeling the probability that the final adjustment recommendation matched the difference type, by selected characteristics: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Characteristic	Frequency	Base model			Full model		
		β (standard error)	Odds ratio (95% confidence interval)	*p* value	β (standard error)	Odds ratio (95% confidence interval)	*p* value
Intercept	120	1.42 (0.75)	4.14 (0.95, 17.99)	0.06	1.10 (0.93)	3.00 (0.49, 18.59)	0.24
Sample size	120	-0.003 (0.004)	0.99 (0.99, 1.01)	0.50	-0.003 (0.004)	1.00 (0.99, 1.00)	0.35
Analyte							
Creatinine	20	Reference	Reference	Reference	Reference	Reference	Reference
Ferritin	20	2.25 (0.90)	9.51 (1.61, 56.06)	*0.01	2.11 (0.94)	8.26 (1.30, 52.40)	*0.03
Folate	20	1.30 (0.74)	3.66 (0.86, 15.66)	0.08	1.11 (0.80)	3.02 (0.63, 14.45)	0.17
High-density lipoprotein cholesterol	20	0.31 (0.69)	1.37 (0.35, 5.29)	0.65	0.32 (0.71)	1.38 (0.34, 5.57)	0.65
Insulin	20	1.55 (0.80)	4.71 (0.98, 22.64)	0.05	1.40 (0.84)	4.04 (0.78, 20.86)	0.10
Vitamin C	20	0.99 (0.73)	2.69 (0.64, 11.23)	0.18	0.93 (0.80)	2.52 (0.53, 12.01)	0.25
Sampling method							
Simple random sample	47	Reference	Reference	Reference	Reference	Reference	Reference
Systematic random sample	73	-0.43 (0.50)	0.65 (0.25, 1.73)	0.39	-0.51 (0.51)	0.60 (0.22, 1.65)	0.32
Simulated difference type							
Random	45	Reference	Reference	Reference	Reference	Reference	Reference
Constant	34	-0.72 (0.60)	0.49 (0.15, 1.59)	0.23	-0.72 (0.62)	0.48 (0.15, 1.62)	0.24
Proportional	41	-1.11 (0.57)	0.33 (0.11, 1.01)	0.05	-1.32 (0.62)	0.27 (0.08, 0.90)	*0.03
Distribution of old measurement							
Normal	11	…	…	…	Reference	Reference	Reference
Skewed	109	…	…	…	-0.69 (1.50)	0.50 (0.03, 9.42)	0.64
Distribution of new measurement							
Normal	14	…	…	…	Reference	Reference	Reference
Skewed	106	…	…	…	1.35 (1.45)	3.85 (0.23, 65.45)	0.35
Distribution of constant difference							
Normal	50	…	…	…	Reference	Reference	Reference
Skewed	70	…	…	…	0.31 (0.54)	1.36 (0.48, 3.89)	0.56
Distribution of proportional difference							
Normal	33	…	…	…	Reference	Reference	Reference
Skewed	87	…	…	…	-0.001 (0.53)	1.00 (0.35, 2.84)	0.99

… Category not applicable.
* Statistically significant at $\alpha = 0.05$ when compared with reference category.

NOTES: Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration. Intercept represents the log-odds that the final adjustment matches the simulated difference type, versus not matching the simulated difference type when all predictors are equal to the reference (that is, creatinine, simple random sample, random difference type, etc.).

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 9. Unweighted frequencies and percentages of concordance categories, by selected analytes: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance category	Creatinine	Ferritin	Folate	High-density lipoprotein cholesterol	Insulin	Vitamin C
			Frequency (percent)			
Concordant adjustment[1]	11 (55.0)	18 (90.0)	16 (80.0)	13 (65.0)	17 (85.0)	15 (75.0)
Overadjustment[2]	3 (15.0)	0 (0.0)	0 (0.0)	3 (15.0)	1 (5.0)	0 (0.0)
Underadjustment[3]	6 (30.0)	2 (10.0)	2 (10.0)	2 (10.0)	2 (10.0)	5 (25.0)
Mismatched adjustment[4]	0 (0.0)	0 (0.0)	2 (10.0)	2 (10.0)	0 (0.0)	0 (0.0)
Total	20 (100.0)	20 (100.0)	20 (100.0)	20 (100.0)	20 (100.0)	20 (100.0)

[1]Final adjustment recommendation matched the simulated difference type.
[2]Adjustment was recommended but not needed.
[3]No adjustment was recommended when one was needed.
[4]Adjustment was recommended when one was needed, but the adjustment type did not match the simulated difference type (including one data set adjusted by ordinary least squares regression).

NOTE: Percentages were derived using column totals as denominators.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table 10. Unweighted frequencies and percentages of concordance categories, by difference type: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance category	Random	Constant	Proportional
		Frequency (percent)	
Concordant adjustment[1]	38 (84.4)	25 (73.5)	27 (65.9)
Overadjustment[2]	7 (15.6)	…	…
Underadjustment[3]	…	7 (20.6)	12 (29.3)
Mismatched adjustment[4]	…	2 (5.9)	2 (4.9)
Total	45 (100.0)	34 (100.0)	41 (100.1)

… Category not applicable.
[1]Final adjustment recommendation matched the simulated difference type.
[2]Adjustment was recommended but not needed.
[3]No adjustment was recommended when one was needed.
[4]Adjustment was recommended when one was needed, but the adjustment type did not match the simulated difference type (including one data set adjusted by ordinary least squares regression).

NOTES: Percentages derived using column totals as denominators. Differences between new and old measurement values were simulated to be randomly distributed, constant, or proportional to the concentration.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Appendix I. Method Comparison R Program

```
### Packages ----
library(tidyverse)
library(mcr)
library(pastecs)

### Reading Files ----

##Set working directory to the specified location of data files. This location should not contain any additional files.
setwd("<FILE DIRECTORY>")

##Read in data files (CSV format)
my_data <- list.files(pattern="*.csv")

pseudo <- lapply(my_data, function(i){
  x <- read_csv(i, col_types=cols()) #Reads data files
  x = x[, c(<c1,c2>)] #C1 and C2 denote column numbers for old and new measurements
  x$file = i #Creates an indicator for the original file name
  x #Returns the data
})

##Name data frames within the list using their file name
names(pseudo) <- tools::file_path_sans_ext(basename(my_data))

### Obtaining Descriptive Statistics ----
### NOTE: Outputs as CSV files to the working directory

##Old Measurement
  #Mean, SD, SE, Var, 95% CI
    old_desc <- map(pseudo, ~stat.desc(.$old, basic=F), options(scipen=100, digits=3))
  #Output Mean, SD, SE, Var, 95% CI
    write.csv(old_desc, "desc_old.csv")
```

```
  #Geometric Mean
    old_geo <- map(pseudo, ~exp(mean(log(.$old))))
  #Output Geometric Mean
    write.csv(old_geo, "geomean_old.csv")
  #Min and Max
    old_min <- map(pseudo, ~min(.$old)) #Min
    old_max <- map(pseudo, ~max(.$old)) #Max
    old_minmax <- rbind(old_min, old_max) #Combines data frames
  #Output Min and Max
    write.csv(old_minmax, "minmax_old.csv")
  #Percentiles
    old_pct <- map(pseudo, ~quantile(.$old, c(.05,.1,.25,.5,.75,.9,.95)))
  #Output Percentiles
    write.csv(old_pct, "pct_old.csv")

##New Measurement
  #Mean, SD, SE, Var, 95% CI
    new_desc <- map(pseudo, ~stat.desc(.$new, basic=F), options(scipen=100, digits=3))
  #Output Mean, SD, SE, Var, 95% CI
    write.csv(new_desc, "desc_new.csv")
  #Geometric Mean
    new_geo <- map(pseudo, ~exp(mean(log(.$new))))
  #Output Geometric Mean
    write.csv(new_geo, "geomean_new.csv")
  #Min and Max
    new_min <- map(pseudo, ~min(.$new)) #Min
    new_max <- map(pseudo, ~max(.$new)) #Max
    new_minmax <- rbind(new_min, new_max) #Combine data frames
  #Output Min and Max
    write.csv(new_minmax, "minmax_new.csv")
  #Percentiles
    new_pct <- map(pseudo, ~quantile(.$new, c(.05,.1,.25,.5,.75,.9,.95)))
  #Output Percentiles
    write.csv(new_pct, "pct_new.csv")

### Calculating Constant and Proportional Differences ----
### NOTE: Outputs as CSV files to the working directory
```

```
##Constant Difference
  #Add column for constant difference
    pseudo <- map(pseudo, ~mutate(.,cdiff=new-old))
  #Mean, SD, SE, Var, and 95% CI  for constant difference
    des_cdiff <- map(pseudo, ~stat.desc(.$cdiff, basic=F))
  #Output Mean, SD, SE, Var, and 95% CI for proportional difference
    write.csv(des_cdiff, "des_cdiff.csv")

##Proportional Difference
  #Add column for proportional difference
    pseudo <- map(pseudo, ~mutate(.,pdiff=cdiff/old))
  #Mean, SD, SE, Var, and 95% CI  for proportional difference
    des_pdiff <- map(pseudo, ~stat.desc(.$pdiff, basic=F))
  #Output Mean, SD, SE, Var, and 95% CI for proportional difference
    write.csv(des_pdiff, "des_pdiff.csv")

### Correlation between old and new ----

  map(pseudo, ~cor.test(.$new, .$old))

### Descriptive Plots/Visualizations ----
### NOTE: Plots output as PDF files to the working directory

##Histograms (New Measurement)
pdf("histograms_new.pdf")
  map(pseudo,~ggplot(.,aes(x=new))+
  geom_histogram(bins=20, color="black", fill="red")+
  labs(title="Distribution of New Measurement", subtitle=.$file, x="CONCENTRATION", y="FREQUENCY")+
  theme(plot.title=element_text(hjust=0.5), plot.subtitle=element_text(hjust=1)))
dev.off()

##Histograms (Old Measurement)
pdf("histograms_old.pdf")
  map(pseudo,~ggplot(.,aes(x=old))+
  geom_histogram(bins=20, color="black", fill="red")+
  labs(title="Distribution of Old Measurement", subtitle=.$file, x="CONCENTRATION", y="FREQUENCY")+
  theme(plot.title=element_text(hjust=0.5), plot.subtitle=element_text(hjust=1)))
dev.off()
```

```
##Histograms (Constant Difference)
pdf("histograms_cdiff.pdf")
  map(pseudo,~ggplot(.,aes(x=cdiff))+
  geom_histogram(bins=20, color="black", fill="red")+
  labs(title="Distribution of Constant Difference", subtitle=.$file, x="CONCENTRATION", y="FREQUENCY")+
  theme(plot.title=element_text(hjust=0.5), plot.subtitle=element_text(hjust=1)))
dev.off()

##Histograms (Proportional Difference)
pdf("histograms_pdiff.pdf")
  map(pseudo,~ggplot(.,aes(x=pdiff))+
  geom_histogram(bins=20, color="black", fill="red")+
  labs(title="Distribution of Proportional Difference", subtitle=.$file, x="CONCENTRATION", y="FREQUENCY")+
  theme(plot.title=element_text(hjust=0.5), plot.subtitle=element_text(hjust=1)))
dev.off()

##Scatterplots (Old Measurement vs New Measurement)
pdf("scatterplots.pdf")
  map(pseudo, ~ ggplot(.) + geom_point(aes(x=old, y=new))+
  labs(title="OLD versus NEW Measurement", subtitle=.$file, x="OLD", y="NEW")+
  theme(plot.title=element_text(hjust=0.5), plot.subtitle=element_text(hjust=1), aspect.ratio=1)+
  expand_limits(x=0, y=0) + scale_x_continuous(expand=c(0,0), limits=c(0,NA)) +
scale_y_continuous(expand=c(0,0), limits=c(0,NA)) +
  geom_abline(intercept=0, slope=1, color="red"))
dev.off()

### Regression Models and Difference Plots----

##OLS
model_OLS <- map(pseudo, ~mcreg(.$old, .$new, error.ratio=1, method.reg="LinReg",
       method.ci="analytical", mref.name="OLD", mtest.name="NEW", na.rm=TRUE))

##Deming
model_DM <- map(pseudo, ~mcreg(.$old, .$new, error.ratio=1, method.reg="Deming",
       method.ci="jackknife", mref.name="OLD", mtest.name="NEW", na.rm=TRUE))

##Weighted Deming
model_WD <- map(pseudo, ~mcreg(.$old, .$new, error.ratio=1, method.reg="WDeming",
```

```
    method.ci="jackknife", mref.name="OLD", mtest.name="NEW", na.rm=TRUE))

### Outputting Regression Parameters ----
### NOTE: Outputs intercept and slope estimates, standard errors, and confidence intervals as CSV files to the working directory

##OLS
 OLS_est <- map(model_OLS, ~print(.@para))
 write.csv(OLS_est, "OLS_est.csv")

##Deming
 DM_est <- map(model_DM, ~print(.@para))
 write.csv(DM_est, "DM_est.csv")

##Weighted Deming
 WD_est <- map(model_WD, ~print(.@para))
 write.csv(WD_est, "WD_est.csv")

### Outputting Difference Plots ----
### NOTE: Plots output as PDF files to the working directory

#Constant Difference Plots
 pdf("cdiff_plots.pdf")
  map(model_DM, ~plotDifference(., plot.type=1, xlab="Concentration", ylab="Constant Difference"))
 dev.off()

#Proportional Difference Plots
 pdf("pdiff_plots.pdf")
  map(model_WD, ~plotDifference(., plot.type=2, xlab="Concentration", ylab="Proportional Difference"))
 dev.off()

#Ranked Constant Difference Plots
 pdf("ranked_cdiff_plots.pdf")
  map(model_DM, ~plotDifference(., plot.type=5, xlab="Rank", ylab="Constant Difference"))
 dev.off()

#Ranked Proportional Difference Plots
 pdf("ranked_pdiff_plots.pdf")
```

```
  map(model_DM, ~plotDifference(., plot.type=6, xlab="Rank", ylab="Proportional Difference"))
 dev.off()

###Statistical Tests & Regression Plots ----
###NOTE: Plots output as PDF files to the working directory

#Output OLS Regression Plots
 pdf("reg_plots_OLS.pdf")
  map(model_OLS, ~plot(.,))
 dev.off()

#Paired T-test (Constant Difference Observed)
 map(pseudo, ~t.test(.$new, .$old, paired=TRUE))

#Output Constant Deming Regression Plots
 pdf("reg_plots_DM.pdf")
  map(model_DM, ~plot(.,))
 dev.off()

#Independent T-test (Proportional Difference Observed)
 map(pseudo, ~t.test(.$pdiff))

#Output Weighted Deming Regression Plots
 pdf("reg_plots_WD.pdf")
  map(model_WD, ~plot(.,))
 dev.off()
```

Appendix II. Joint Analysis: Three-by-three Concordance Measures After Omitting the Data Set Adjusted With Ordinary Least Squares Regression

Table I. Unweighted frequencies and percentages of difference type, by recommended adjustment, omitting data set adjusted by ordinary least squares regression: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Simulated difference type	Recommended adjustment			Total
	No adjustment	Constant Deming	Weighted Deming	
	Frequency (percent)			
Random	38 (84.4)	6 (13.3)	1 (2.2)	45 (99.9)
Constant	7 (21.2)	25 (75.8)	1 (3.0)	33 (100.0)
Proportional	12 (29.3)	2 (4.9)	27 (65.9)	41 (100.1)

NOTES: Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration. Adjustments using the constant Deming and weighted Deming regressions incorporate measurement error for both the new and old measurement methods. One pseudo-data set was adjusted using ordinary least squares regression and was omitted from this table. Row percentages are presented. Total percentages for random and proportional difference types do not sum to 100 because of rounding.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table II. Unweighted percentages for selected concordance measures, by difference type, omitting data set adjusted by ordinary least squares regression: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance measures	Random	Constant	Proportional
Sensitivity[1]	84.4	75.8	65.9
Specificity[2]	74.3	90.7	97.4
Positive predictive value or precision[3]	66.7	75.8	93.1
Negative predictive value[4]	88.7	90.7	84.4

[1]True positives divided by (true positives plus false negatives).
[2]True negatives divided by (false positives plus true negatives).
[3]True positives divided by (true positives plus false positives).
[4]True negatives divided by (false negatives plus true negatives).

NOTE: Differences between new and old measurement values were simulated to be randomly distributed, constant to the concentration, or proportional to the concentration.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Table III. Unweighted frequencies and percentages of selected concordance categories, by sampling method: 2017–2018 National Health and Nutrition Examination Survey simulated pseudo-crossover studies

Concordance category	Simple random sampling	Systematic random sampling
	Frequency (percent)	
Concordant adjustment[1]	37 (78.7)	53 (72.6)
Overadjustment[2]	4 (8.5)	3 (4.1)
Underadjustment[3]	5 (10.6)	14 (19.2)
Mismatched adjustment[4]	1 (2.1)	3 (4.1)
Total	47 (99.9)	73 (100.0)

[1]Final adjustment recommendation matched the simulated difference type.
[2]Adjustment was recommended but not needed.
[3]No adjustment was recommended when one was needed.
[4]Adjustment was recommended when one was needed, but the adjustment type did not match the simulated difference type (including one data set adjusted by ordinary least squares regression).

NOTES: Simple random sampling mimics a convenience sample, which is often the default sampling method for crossover studies as samples in this context are chosen based on the time of measurement rather than measurement values. Systematic random sampling is the ideal sampling method for crossover studies because it incorporates measurement concentrations from across the entire distribution. Percentages are derived using column totals as denominators. For simple random sampling, percentages do not sum to 100% because of rounding.

SOURCE: Pseudo-crossover studies were simulated using National Center for Health Statistics, 2017–2018 National Health and Nutrition Examination Survey public-use files.

Vital and Health Statistics Series Descriptions

Active Series

Series 1. **Programs and Collection Procedures**
Reports describe the programs and data systems of the National Center for Health Statistics, and the data collection and survey methods used. Series 1 reports also include definitions, survey design, estimation, and other material necessary for understanding and analyzing the data.

Series 2. **Data Evaluation and Methods Research**
Reports present new statistical methodology including experimental tests of new survey methods, studies of vital and health statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Reports also include comparison of U.S. methodology with those of other countries.

Series 3. **Analytical and Epidemiological Studies**
Reports present data analyses, epidemiological studies, and descriptive statistics based on national surveys and data systems. As of 2015, Series 3 includes reports that would have previously been published in Series 5, 10–15, and 20–23.

Discontinued Series

Series 4. **Documents and Committee Reports**
Reports contain findings of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002; these are now included in Series 2 or another appropriate series.

Series 5. **International Vital and Health Statistics Reports**
Reports present analytical and descriptive comparisons of U.S. vital and health statistics with those of other countries. The last Series 5 report was published in 2003; these are now included in Series 3 or another appropriate series.

Series 6. **Cognition and Survey Measurement**
Reports use methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999; these are now included in Series 2.

Series 10. **Data From the National Health Interview Survey**
Reports present statistics on illness; accidental injuries; disability; use of hospital, medical, dental, and other services; and other health-related topics. As of 2015, these are included in Series 3.

Series 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**
Reports present 1) estimates of the medically defined prevalence of specific diseases in the United States and the distribution of the population with respect to physical, physiological, and psychological characteristics and 2) analysis of relationships among the various measurements. As of 2015, these are included in Series 3.

Series 12. **Data From the Institutionalized Population Surveys**
The last Series 12 report was published in 1974; these reports were included in Series 13, and as of 2015 are in Series 3.

Series 13. **Data From the National Health Care Survey**
Reports present statistics on health resources and use of health care resources based on data collected from health care providers and provider records. As of 2015, these reports are included in Series 3.

Series 14. **Data on Health Resources: Manpower and Facilities**
The last Series 14 report was published in 1989; these reports were included in Series 13, and are now included in Series 3.

Series 15. **Data From Special Surveys**
Reports contain statistics on health and health-related topics from surveys that are not a part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002; these reports are now included in Series 3.

Series 16. **Compilations of Advance Data From Vital and Health Statistics**
The last Series 16 report was published in 1996. All reports are available online; compilations are no longer needed.

Series 20. **Data on Mortality**
Reports include analyses by cause of death and demographic variables, and geographic and trend analyses. The last Series 20 report was published in 2007; these reports are now included in Series 3.

Series 21. **Data on Natality, Marriage, and Divorce**
Reports include analyses by health and demographic variables, and geographic and trend analyses. The last Series 21 report was published in 2006; these reports are now included in Series 3.

Series 22. **Data From the National Mortality and Natality Surveys**
The last Series 22 report was published in 1973. Reports from sample surveys of vital records were included in Series 20 or 21, and are now included in Series 3.

Series 23. **Data From the National Survey of Family Growth**
Reports contain statistics on factors that affect birth rates, factors affecting the formation and dissolution of families, and behavior related to the risk of HIV and other sexually transmitted diseases. The last Series 23 report was published in 2011; these reports are now included in Series 3.

Series 24. **Compilations of Data on Natality, Mortality, Marriage, and Divorce**
The last Series 24 report was published in 1996. All reports are available online; compilations are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782

Tel: 1–800–CDC–INFO (1–800–232–4636)
TTY: 1–888–232–6348
Internet: https://www.cdc.gov/nchs
Online request form: https://www.cdc.gov/info
For e-mail updates on NCHS publication releases, subscribe online at: https://www.cdc.gov/nchs/email-updates.htm.

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782–2064

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

FIRST CLASS MAIL
POSTAGE & FEES PAID
CDC/NCHS
PERMIT NO. G-284

For more NCHS Series Reports, visit:
https://www.cdc.gov/nchs/products/series.htm

Series 2, No. 206
CS346754