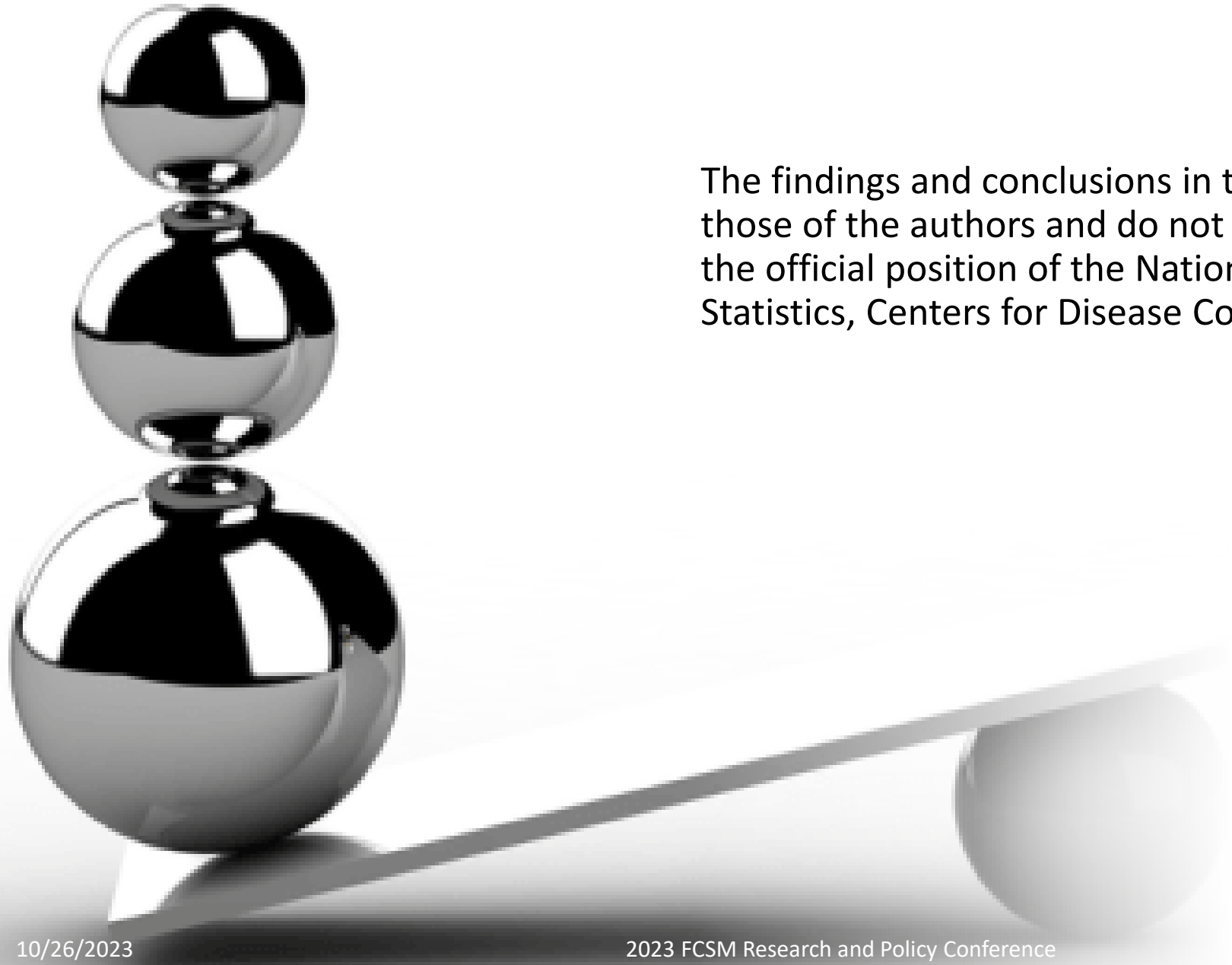
Measuring the robustness of method comparison techniques using data from the National Health and Nutrition Examination Survey

Authors: Jennifer Rammon[1], Kevin Chuang[2], Te-Ching Chen[3], Hee-Choon Shin[1]

[1] Division of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention

[2] Based on work conducted as an Oak Ridge Institute for Science and Education (ORISE) fellow with the Division of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention

[3] Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention



The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Background

Analytic method validation studies

- Clinical laboratories use chemical assays (investigative procedures) to identify and measure the amount of substance (i.e., analyte) in a sample
  - E.g., concentration of glucose in serum, folate in plasma
- Analytic method validation studies are performed when advances or changes in laboratory measurement methods or instrumentation occur *(Bland and Altman, Westgard, Clinical and Laboratory Standards Institute)*
  - Compares two assays to assess the **potential impact of assay changes**
  - In most cases, neither method is the "gold standard" or true value of the measurement
  - Instead, comparisons are meant to show that an assay is suitable for its intended purpose or that the two assays are comparable

National Health and Nutrition Examination Surveys (NHANES) laboratory component

- Biological specimens (e.g., blood, urine) are collected and processed in the field, then shipped to testing laboratories
- Testing results provide data about the health and nutritional status of the civilian noninstitutionalized U.S. population
- Analytic measurements are made by CDC laboratories in Atlanta and contracted laboratories throughout the country utilizing state of the art methods
- Based on their expertise, CDC and contracted laboratories determine which bioassay is most suitable for each analyte in collaboration with the DHANES laboratory project officer

NHANES Laboratory Methods Workgroup (LMW)

- The NHANES LMW performs internal method validation studies whenever CDC and contract laboratories undergo instrumental or methodological measurement changes or when contract laboratories change
- Specifically, analytic method validation studies are used to **evaluate how changes in methodology may influence data inference** and address concerns about the **consistency of measurements across survey cycles**
- When systematic differences are observed, DHANES **releases adjustment equations** with the public data file documentation to help analysts planning to **combine survey cycles or conduct trends analyses**
- Adjustment equations help **ensure that differences from instrumental or measurement-related changes are not falsely interpreted as changes in the U.S population** over time

Objective

- This project seeks to **assess the reliability of the statistical methodology used by the NHANES LMW to address current concerns that adjustment equations are being overproduced or recommended too frequently**

- Methodological approach:
  - Simulated pseudo analytic method validation studies using publicly available 2017-2018 NHANES data
  - Pseudo-studies were analyzed by a team of statisticians
  - Results were compared to simulation “truth”

Methods

NHANES

https://www.cdc.gov/nchs/nhanes/index.htm

- **Complex, multistage probability sample** of the civilian noninstitutionalized U.S. population
- Conducted by the **National Center for Health Statistics** (NCHS)
- Includes household questionnaires, standardized health examinations conducted in Mobile Examination Centers (MEC), and two 24-hour dietary recalls
- The MECs are staffed by full-time personnel, including **phlebotomists who perform venipuncture using standardized protocols**
- A series of cross-sectional surveys were conducted periodically from 1960 through 1994
- Beginning in 1999, NHANES became a **continuous survey**: a series of cross-sectional nationally representative surveys and exams that were collected and released in 2-year cycles between 1999 and 2018
  - 2019-2020 data were released in a combined 2017-March 2020 Pre-Pandemic dataset

Conducting analytic method validation studies

1. Review **descriptive estimates** of measurements provided by old and new methods (separately)
   - Means, medians, percentiles
2. Compare old and new measurements through **visual displays**
   - Scatterplots, difference plots
3. Evaluate difference type using **difference plots**
   - Differences are **constant** to old measurement
   - Differences are **proportional** to old measurement

Data format:

ID	Old	New
1	57	47
2	31	25
3	40	34
4	85	78
5	57	52
…	…	…
50	60	49
51	34	29
52	56	47

Conducting analytic method validation studies (continued)

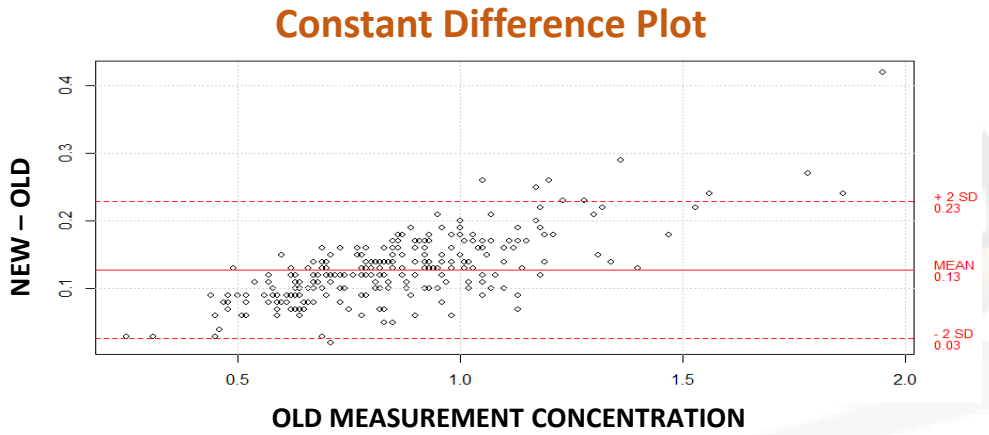
4. Conduct **statistical testing**
   - paired t-test
   - independent t-test of the relative difference
5. Perform **regression analysis**
   - Deming regression preferred over ordinary least squares (OLS) regression as it incorporates imprecision from both measurement procedures by accounting for observation errors on both the x- and y- axis
   - Weighted Deming regressions give each point a weight inversely proportional to the square of the concentration on the x-axis

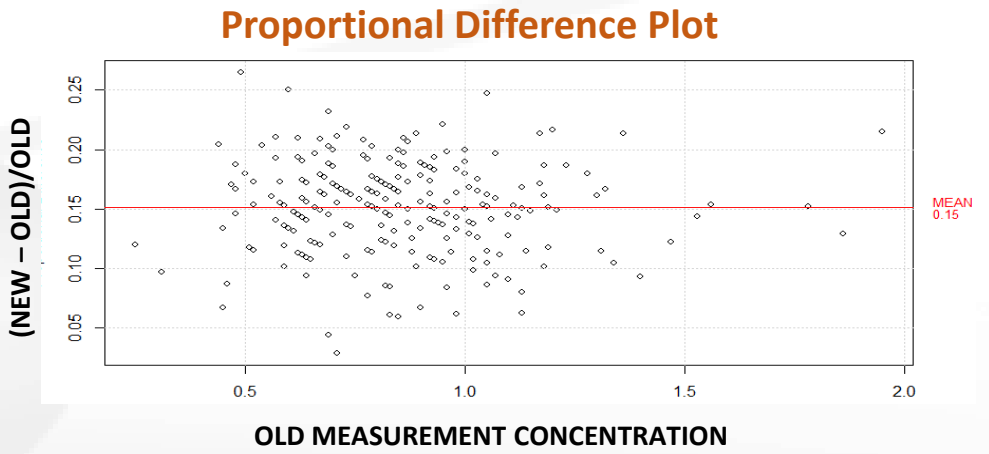
Data format:

ID	Old	New
1	57	47
2	31	25
3	40	34
4	85	78
5	57	52
…	…	…
50	60	49
51	34	29
52	56	47

Conducting analytic method validation studies (continued)

- Classic **observed constant difference**
  - Paired t-test and constant Deming intercept are statistically significant
  - Constant Deming slope is statistically insignificant
- Classic **observed proportional difference**
  - Independent t-test and weighted Deming slope are statistically significant
  - Weighted Deming intercept is statistically insignificant





Simulating the pseudo cross-over studies

- **120 datasets** were created using **2017–2018 NHANES** public-use laboratory data
- Analytes chosen from a master list of **previously conducted** DHANES **method validation studies** with special consideration given to analytes evaluated in multiple method validation studies
- **Efforts made to include a variety of analytes** with different characteristics: small and large numeric measurements, small and large variations, and differing clinical functions
- Final analytes included **creatinine, ferritin, HDL cholesterol, folate, insulin**, and **vitamin C**

Simulating the pseudo cross-over studies (continued)

- **Random** error
  - *NEW = OLD + ERROR*
- **Constant** systematic difference
  - *NEW = OLD + DIFFERENCE*
  - Error incorporated into the difference term
- **Proportional** systematic difference
  - *NEW = OLD + (OLD x DIFFERENCE)*
  - Error incorporated into the difference term
- Minimum imposed on all new measurements: $\frac{analyte\ limit\ of\ detection}{\sqrt{2}}$
  - Standard imputed value
  - Matches minimum released by NHANES for all 'old' measurements

ID #	Old Measurement	New Measurement
SEQN	2017–2018 NHANES	New
002	40	39
003	62	65
004	56	57
…	…	…
086	49	51
087	54	47

Data analysis

- **Two statisticians reviewed the results from each dataset** using the standard method comparison techniques followed by DHANES LMW
  - Worked independently of one another and without access to any of the simulation data
  - Classified differences as random, constant, or proportional
  - Determined whether the ‘publicly-released pseudo-data’ should be released with an adjustment equation or not
- In cases where the two recommendations differed, reviewers observed any sort of deviation from a standard method comparison analysis, or noted something of interest, datasets were flagged for further review
- Further review involved a **team of four statisticians,** who evaluated the results together to reach a final adjustment decision

Joint data analysis

- Upon determining the final adjustment recommendations for all 120 pseudo-crossover studies
  - **Recommendations were compared to the simulated difference type** to create a three-by-three concordance table
  - **Classified** all recommendations as:
    - **Concordant adjustment –** Final adjustment decision matched simulated difference type
    - **Overadjustment -** An adjustment was recommended when (based on simulated difference type) one was not needed
    - **Underadjustment -** An adjustment was not recommended when (based on simulated difference type) one was needed
    - **Mismatched adjustment -** Final adjustment decision did not match the simulated difference type
  - Logistic regression
  - **Stratified** results **by analyte, difference type,** and **sampling method**

Down-stream analyses

The effects of adjustment equations on subsequent analyses of combined NHANES datasets

- **Age-adjusted HDL cholesterol** among adults 20 years and older
- Trends from 1999–2000 through 2017–2018
- For each pseudo-crossover dataset, new measurement values were simulated for **all 2017–2018 NHANES participants** based on the drawn simulation parameters for that dataset.
- Three analyses were compared
  - **Original data:** 1999–2018, no simulations (gold standard for measures of association, i.e., regression estimates)
  - **Adjusted data:** 1999–2016 measurements adjusted to be compatible with 2017–2018 measurements
  - **Unadjusted data:** 1999–2016 measurements compared with 'new' 2017–2018 measurement values

Results

Disclaimer: All results shown here are based on the simulated pseudo-datasets described. They are not based on any publicly released NHANES adjustment equations.

Final adjustment by simulated difference type

Simulated difference type

Final adjustment recommendation	Random (%)	Constant (%)	Proportional (%)
No Adjustment	**38 (31.7)**	**7 (5.8)**	**12 (10.0)**
Deming regression	**6 (5.0)**	**25 (20.8)**	**2 (1.7)**
Weighted Deming regression	**1 (0.8)**	**1 (0.8)**	**27 (22.5)**
Ordinary least squares regression	0 (0.0)	**1 (0.8)**	0 (0.0)

NOTE: Percentages do not sum to 100% due to rounding errors

Concordance

Note: Percents don't sum to 100 due to rounding error

	Frequency (%)
Concordant Adjustment	**90 (75.0)**
Overadjustment	**7 (5.8)**
Underadjustment	**19 (15.8)**
Mismatched Adjustment	**4 (3.3)**

Concordant Adjustment An adjustment was recommended when one was needed and matched the simulated difference type

Overadjustment: An adjustment was recommended when one was not needed

Underadjustment: An adjustment was not recommended when one was needed

Mismatched Adjustment – An adjustment was recommended when one was needed, but did not match the simulated difference type

Final adjustment by simulated difference type (2x2 table)

Simulated Error Type

Final Adjustment	Constant/ Proportional (%) (+)	Random (%) (-)
Adjust (+)	**56 (46.7)**	**7 (5.8)**
Don't Adjust (-)	**19 (15.8)**	**38 (31.7)**

NOTE: Mismatched adjustment category from previous slides is included in the first cell (adjusted constant/proportional difference) for dichotomized analysis

NOTE: Percentages do not sum to 100% due to rounding errors

Concordance: 78.3%

Misclassification: 21.7%

Sensitivity: 74.7%

Specificity: 84.4%

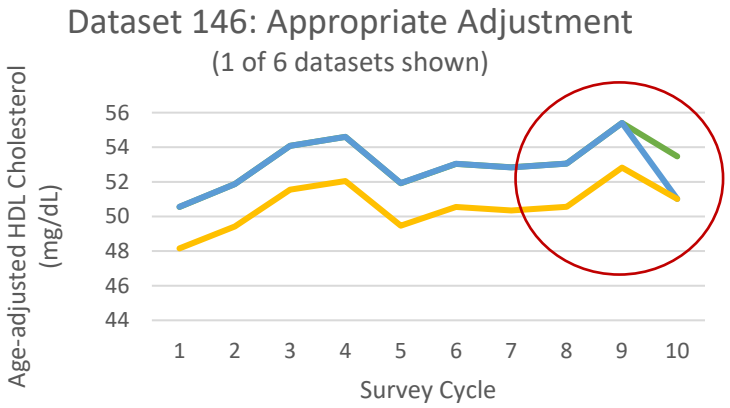
Precision/Positive Predicted Value: 88.9%

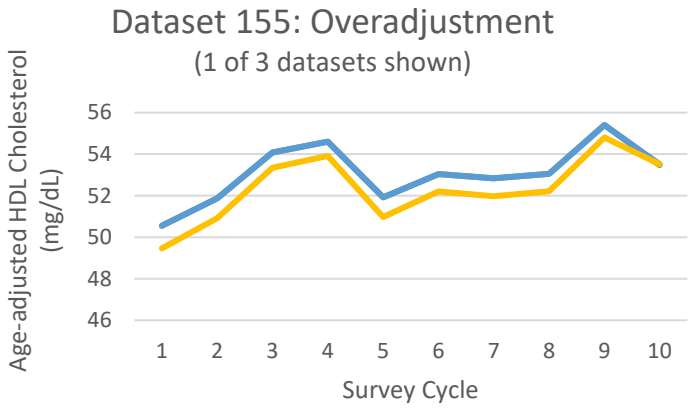
Negative Predicted Value: 66.7%

Trends in age-adjusted HDL cholesterol
based on the original 2017–2018 NHANES data

Survey Cycle	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	2015-2016	2017-2018
Sample size	4,003	4,563	4,367	4,363	5,163	5,509	4,762	5,172	4,998	4,779
Mean HDL (mg/dL)	50.55	51.87	54.08	54.6	51.92	53.04	52.84	53.05	55.4	53.47

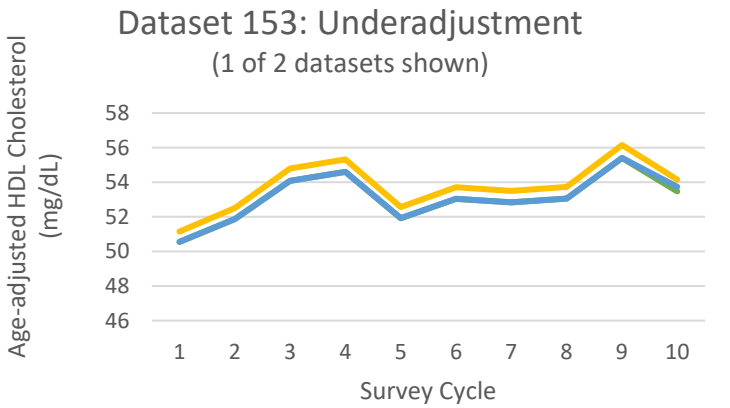
Trends in age-adjusted HDL cholesterol (continued)

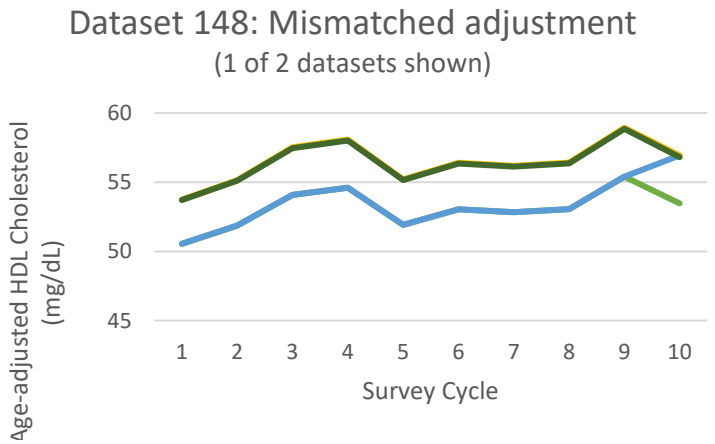




Datasets:

- Original NHANES
- Unadjusted
- Mismatched adjusted
- Ideally adjusted





Trends in age-adjusted HDL cholesterol (continued)

Dataset	Trends slope estimate based on 1999-2016 unadjusted estimates		Trends slope estimate based on 1999-2016 adjusted estimates	
	β	P-Value	β	P-Value
Original NHANES: Gold Standard			0.25	0.0001
Dataset 146: Appropriate adjustment	0.10	0.09	0.24	0.0001
Dataset 155: Overadjustment	0.25	0.0001	0.32	<0.0001
Dataset 153: Underadjustment	0.26	<0.0001	0.26	0.0001
Dataset 148: Mismatched Adjustment	0.45	<0.0001	0.26	0.0001
Dataset 148: Ideal Adjustment			0.27	<0.0001

Standard error terms for β coefficients were 0.06–0.07 throughout (adjusted and unadjusted).

Conclusions

Conclusions

- Based on these simulations, the **statistical methodology** being utilized by DHANES **performs reasonably well**
- Specifically, for these simulations, data was **adjusted appropriately 75% of the time**
- In cases where the data was not adjusted appropriately, there was a **higher tendency to underadjust than overadjust**
- For these simulations, **when data was adjusted appropriately** substantial differences were observed between analyses performed with the adjusted data versus analyses performed with the unadjusted data, **suggesting a high need to adjust**
- For these simulations**, adjustment 'errors' tended to have minimal impacts on down stream analyses**
  - For datasets that were 'overadjusted' or 'underadjusted', there were no substantial differences observed between analyses performed with the adjusted data versus analyses performed with the unadjusted data
  - For datasets with a 'mismatched' adjustment, there were no substantial differences observed between analyses performed with mismatched adjusted data versus analyses performed with ideally adjusted data.

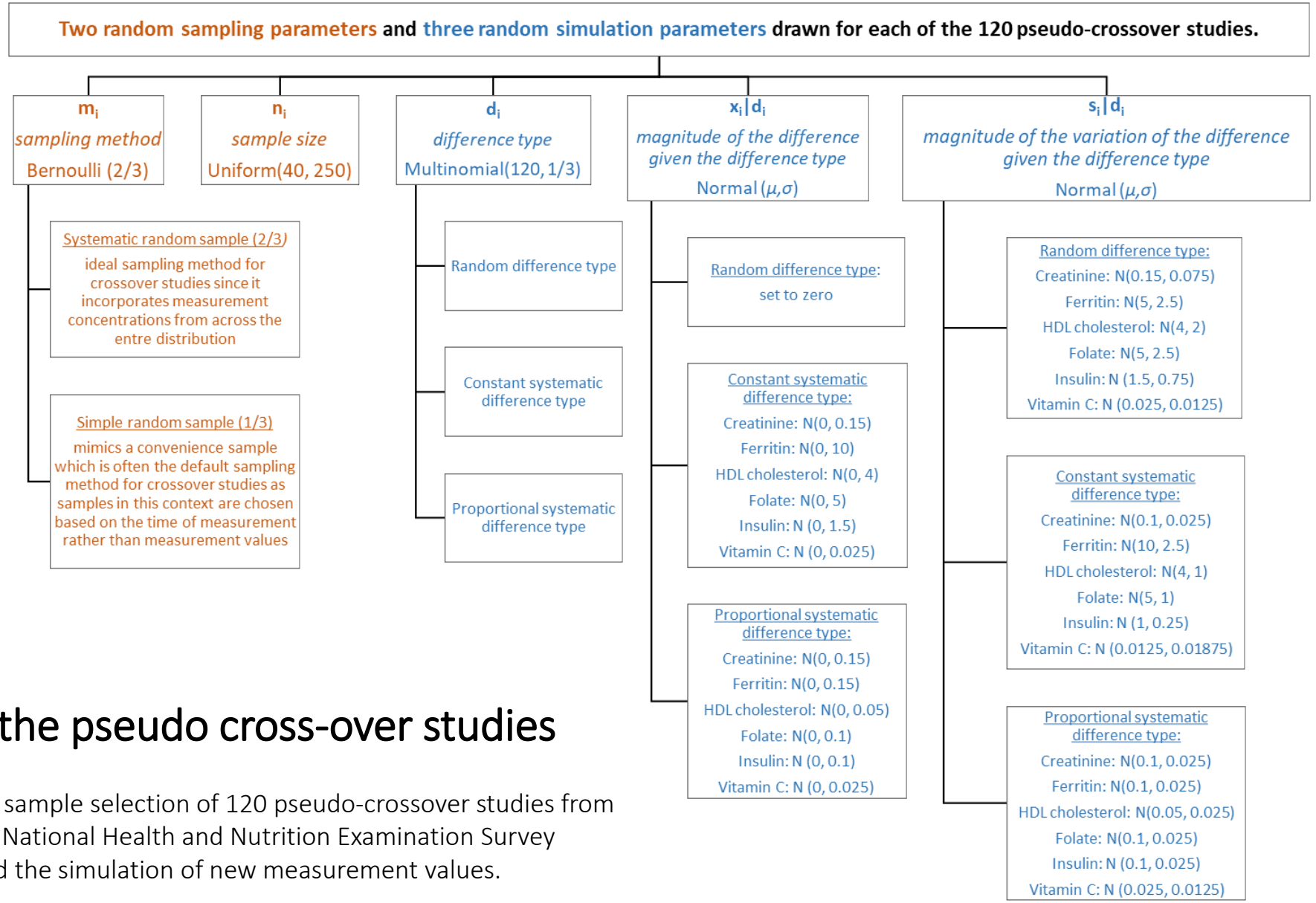
Questions?

Jennifer Rammon

LMI3@CDC.GOV

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

Forthcoming NCHS Vital Health and Statistics Series Report (2024).



Simulating the pseudo cross-over studies

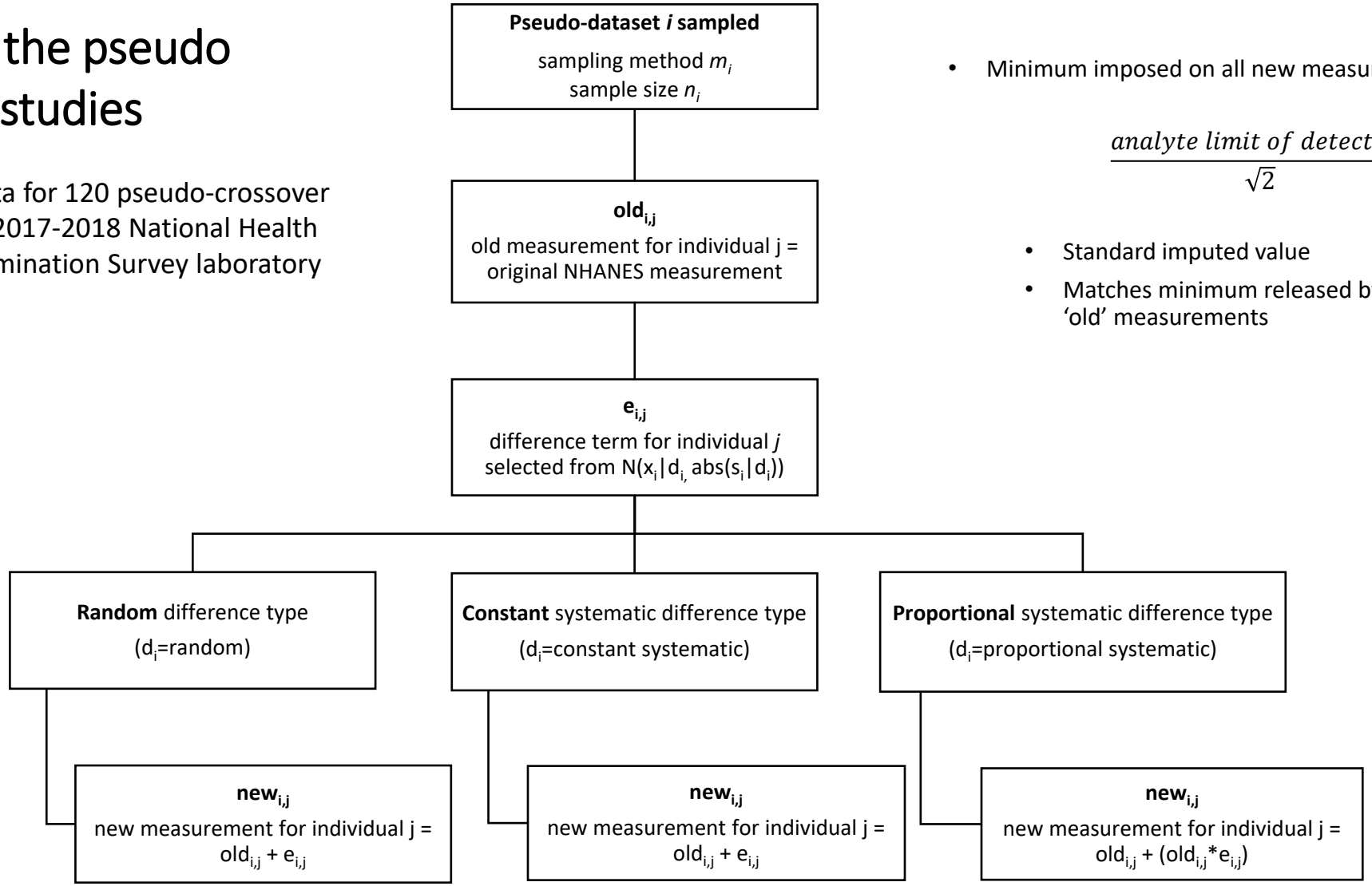
Parameters for the sample selection of 120 pseudo-crossover studies from the full 2017-2018 National Health and Nutrition Examination Survey laboratory data and the simulation of new measurement values.

Simulating the pseudo cross-over studies

Sample sizes and sampling process for the 120 pseudo-crossover dataset

Simulating the pseudo cross-over studies

Simulating the data for 120 pseudo-crossover datasets utilizing 2017-2018 National Health and Nutrition Examination Survey laboratory data

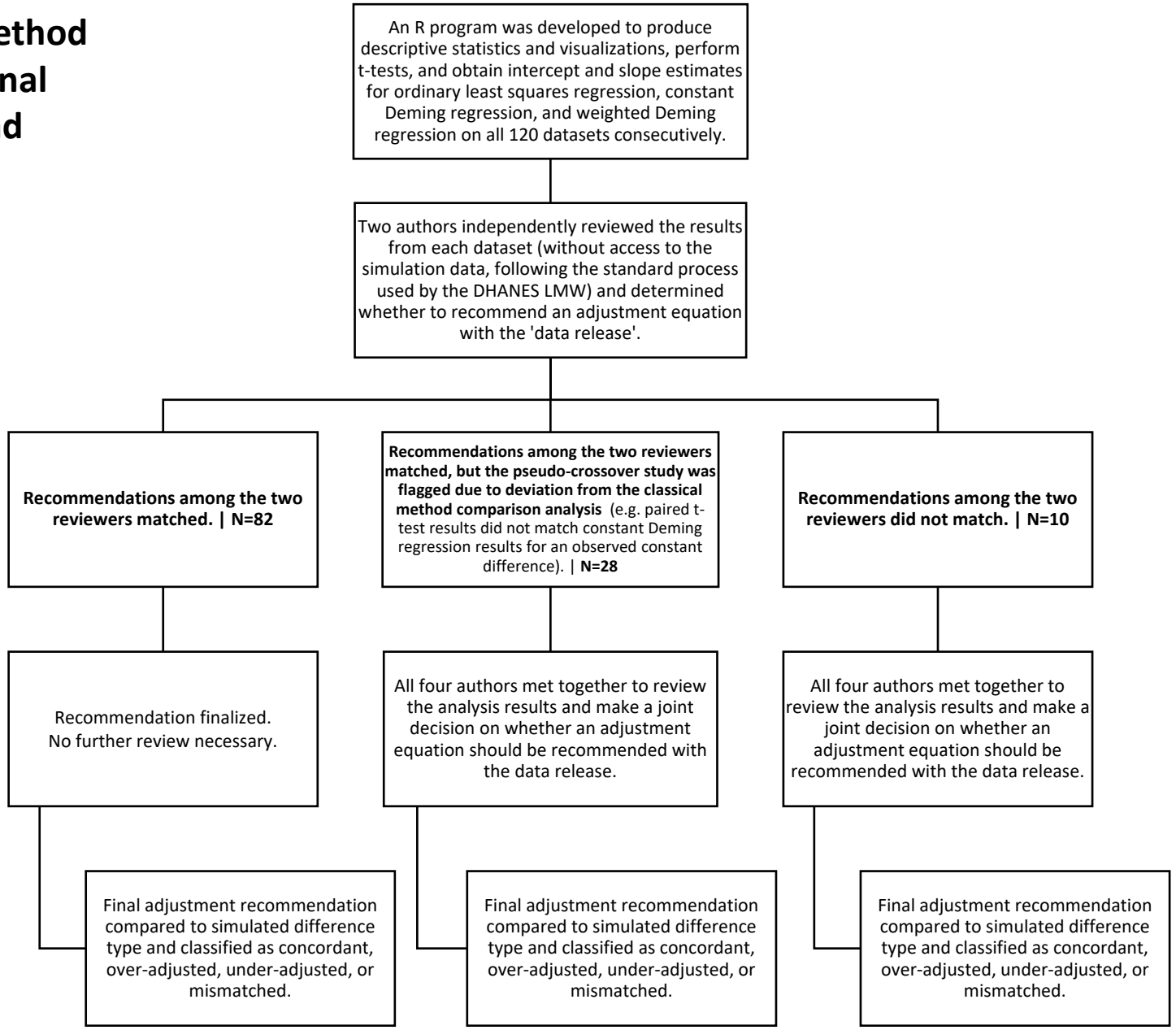


- Minimum imposed on all new measurements:

$$\frac{analyte\ limit\ of\ detection}{\sqrt{2}}$$

  - Standard imputed value
  - Matches minimum released by NHANES for all 'old' measurements

Analysis process for individual method validation studies: determining final adjustment recommendations and classifying concordance category



Concordance by sampling method

	Systematic Random Sampling (%)	Simple Random Sampling (%)
Concordant Adjustment	**37 (78.7)**	**53 (72.6)**
Overadjustment	**4 (8.5)**	**3 (4.1)**
Underadjustment	**5 (10.6)**	**14 (19.2)**
Mismatched Adjustment	**1 (2.1)**	**3 (4.1)**

NOTE: Percentages shown correspond to column totals
NOTE: Percentages do not sum to 100% due to rounding errors

Concordance by difference type

	Random (%)	Constant (%)	Proportional* (%)
Concordant Adjustment	**38 (84.4)**	**25 (73.5)**	**27 (65.9)**
Over-Adjustment	**7 (15.6)**		
Under-Adjustment		**7 (20.6)**	**12 (29.3)**
Mismatched Adjustment		**2 (5.9)**	**2 (4.9)**

NOTE: Percentages shown correspond to column totals
NOTE: Grey cells indicate adjustment type is not possible with specified error type
NOTE: Percentages do not sum to 100% due to rounding errors

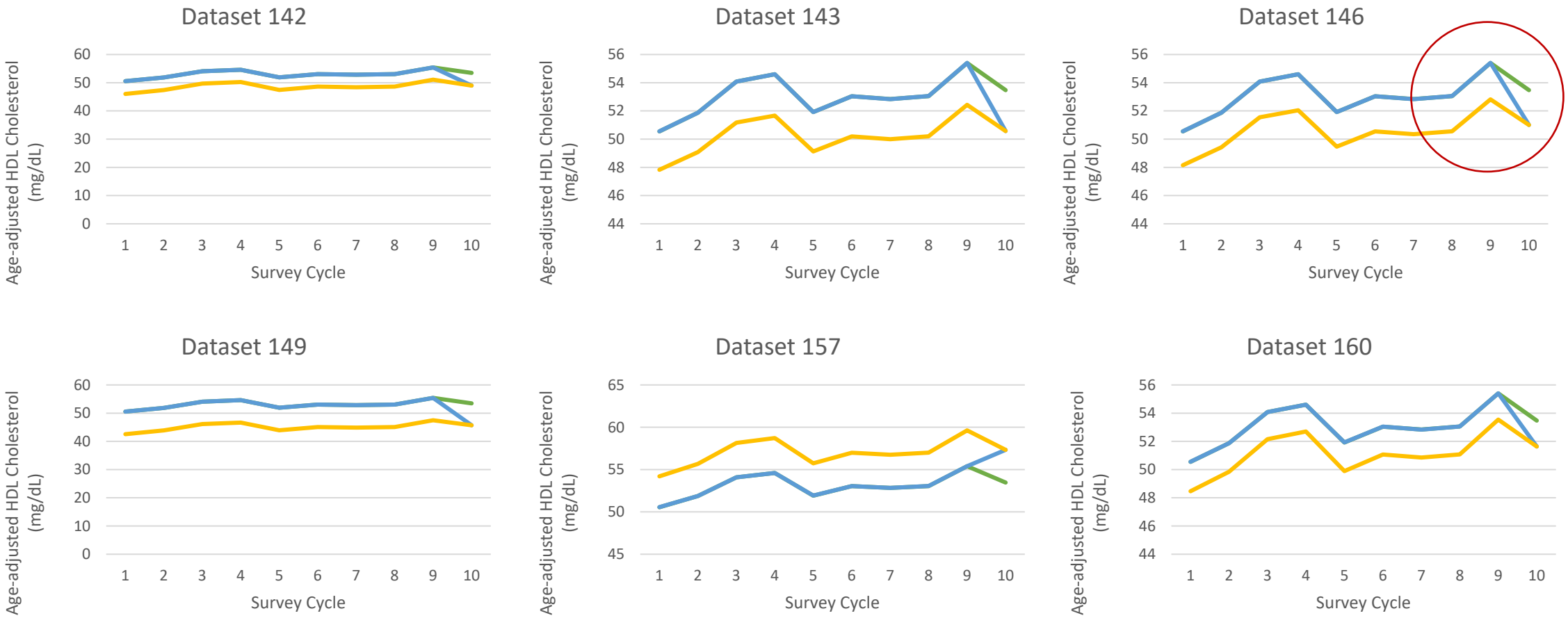
Concordance by analyte

	Creatinine (%)	Ferritin (%)	Folate (%)	HDL (%)	Insulin (%)	Vitamin C (%)
Concordant Adjustment	**11 (55)**	**18 (90)**	**16 (80)**	**13 (65)**	**17 (85)**	**15 (75)**
Over-Adjustment	**3 (15)**	0 (0.0)	0 (0.0)	**3 (15)**	**1 (5)**	0 (0.0)
Under-Adjustment	**6 (30)**	**2 (10)**	**2 (10)**	**2 (10)**	**2 (10)**	**5 (25)**
Mismatched Adjustment	0 (0.0)	0 (0.0)	**2 (10)**	**2 (10)**	0 (0.0)	0 (0.0)

NOTE: Percentages shown correspond to column totals
NOTE: Percentages do not sum to 100% due to rounding errors

Appropriate Adjustment

Includes all HDL cholesterol pseudo-crossover studies for which an adjustment was recommended and the final adjustment recommendation matched the simulated difference type.



Appropriate Adjustment

Includes all HDL cholesterol pseudo-crossover studies for which an adjustment was recommended and the final adjustment recommendation matched the simulated difference type.

Gold standard:
Slope estimate for original data: 0.25 (SE=0.06)
P-value for original data: 0.0001

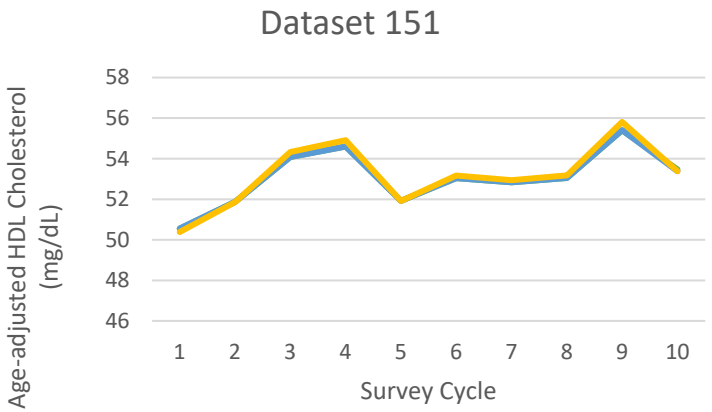
Dataset	Trends slope estimate based on 1999-2016 unadjusted estimates		Trends slope estimate based on 1999-2016 adjusted estimates	
	β	P-Value	β	P-Value
142	-0.01	0.84	0.25	0.0001
143	0.08	0.20	0.23	0.0001
146	0.10	0.09	0.24	0.0001
149	-0.20	0.002	0.26	<0.0001
157	0.47	<0.0001	0.27	0.0001
160	0.14	0.02	0.27	<0.0001

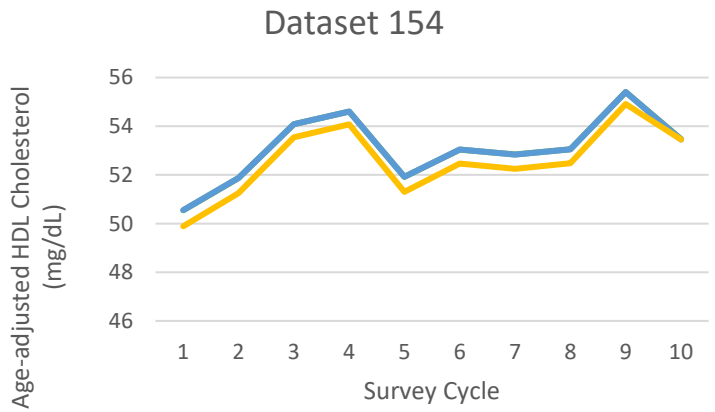
Standard error terms for β coefficients were 0.06–0.07 throughout (adjusted and unadjusted).

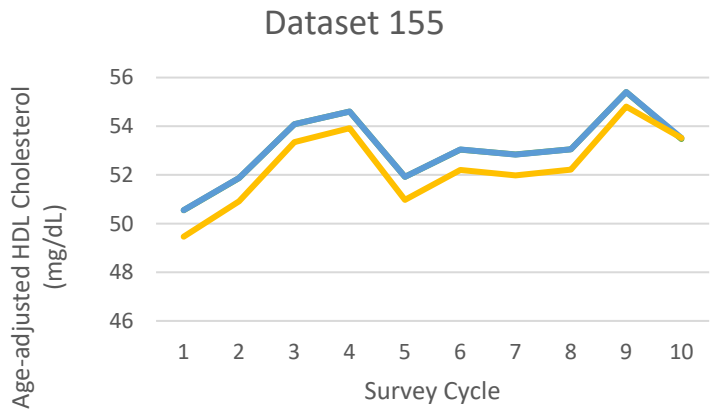
Overadjustment

Includes all HDL cholesterol pseudo-crossover studies for which the final adjustment recommendation included an adjustment, when in fact the simulated difference type was random.

Original NHANES data
Unadjusted data
Adjusted data







Overadjustment

Includes all HDL cholesterol pseudo-crossover studies for which the final adjustment recommendation included an adjustment, when in fact the simulated difference type was random.

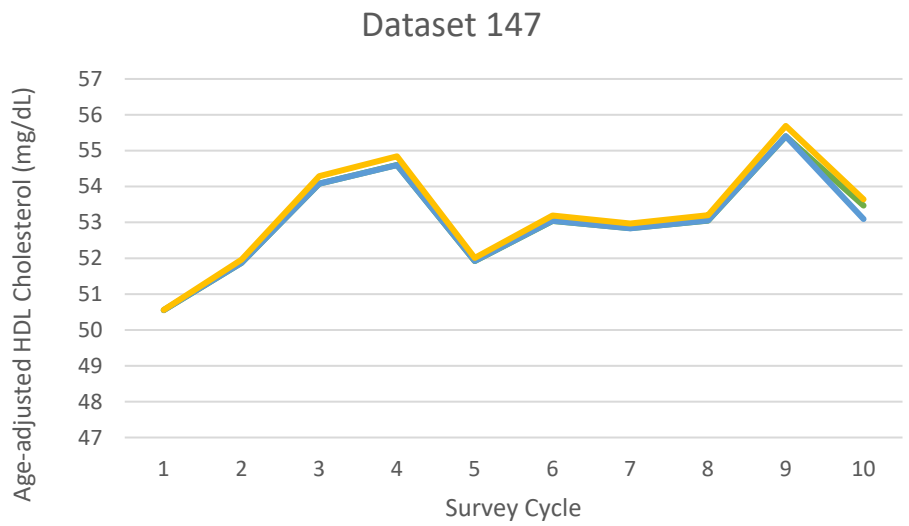
Gold standard:
Slope estimate for original data: 0.25 (SE=0.06)
P-value for original data: 0.0001

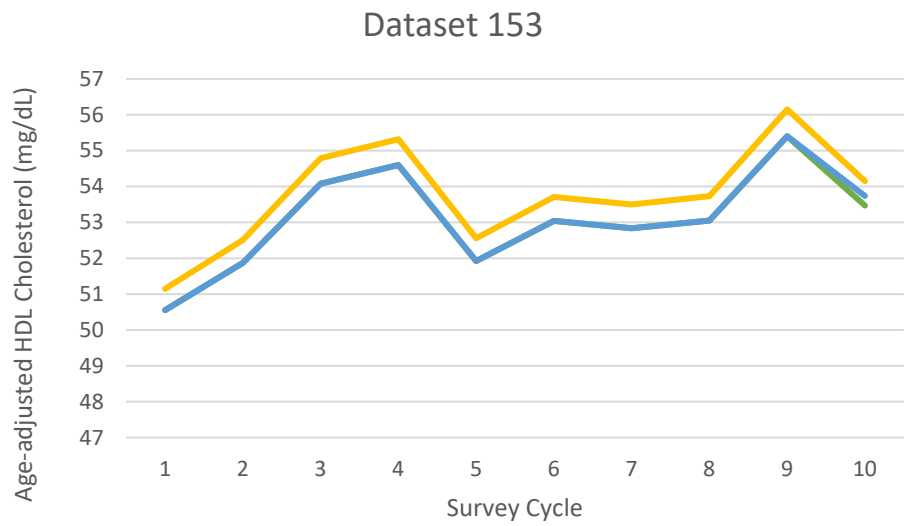
Dataset	Trends slope estimate based on 1999-2016 unadjusted estimates		Trends slope estimate based on 1999-2016 adjusted estimates	
	β	P-Value	β	P-Value
151	0.24	0.0002	0.26	0.0002
154	0.25	0.0001	0.29	<0.0001
155	0.25	0.0001	0.32	<0.0001

Standard error terms for β coefficients were 0.06-0.07 throughout (adjusted and unadjusted).

Underadjustment

Includes all HDL cholesterol pseudo-crossover studies for which the simulated difference type was systematic, but no adjustment equation was recommended.





Underadjustment

Includes all HDL cholesterol pseudo-crossover studies for which the simulated difference type was systematic, but no adjustment equation was recommended.

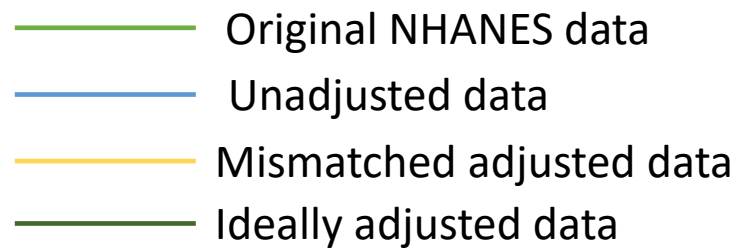
Gold Standard:
Slope estimate for original data: 0.25 (SE=0.06)
P-value for original data: 0.0001

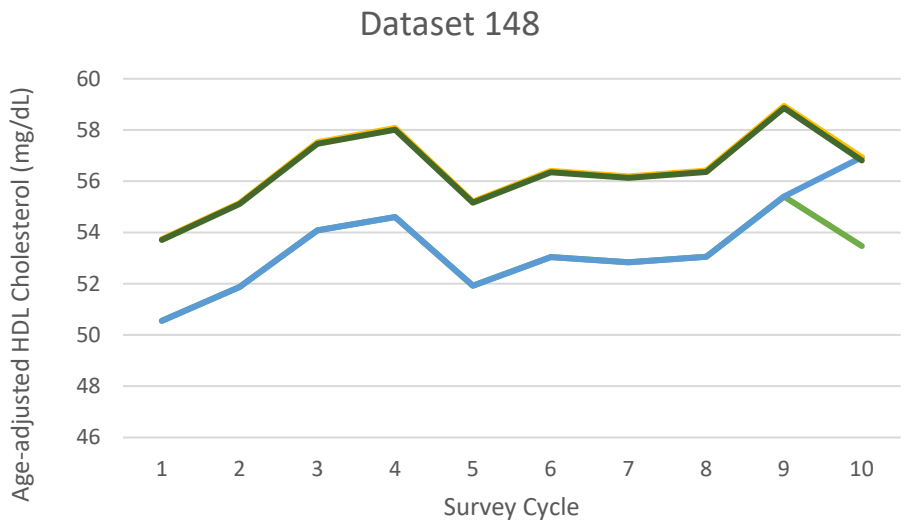
Dataset	Trends slope estimate based on 1999-2016 unadjusted estimates		Trends slope estimate based on 1999-2016 adjusted estimates	
	β	P-Value	β	P-Value
147	0.23	0.0003	0.26	0.0001
153	0.26	<0.0001	0.26	0.0001

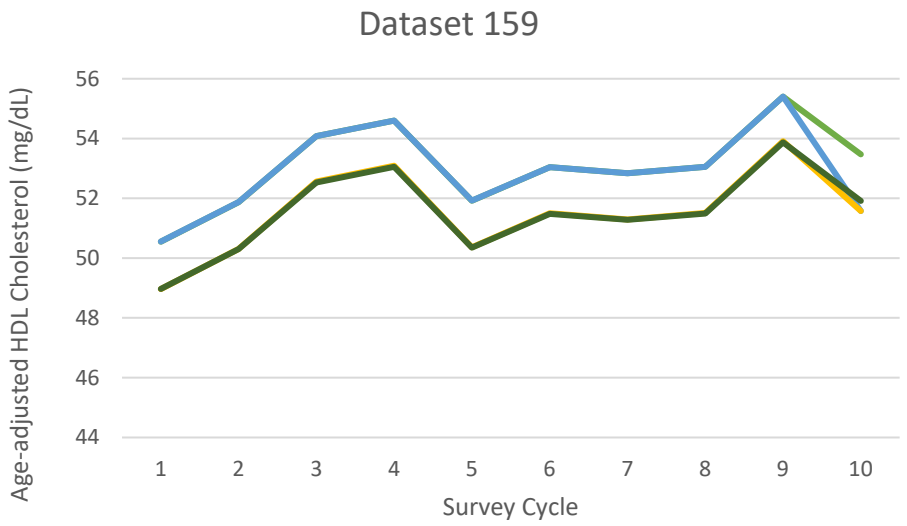
Standard error terms for β coefficients were 0.06-0.07 throughout (adjusted and unadjusted).

Mismatched adjustment

Includes all HDL cholesterol pseudo-crossover studies for which an adjustment was recommended when it should have been, but the simulated difference type mismatched the final recommendation type.







Mismatched adjustment

Includes all HDL cholesterol pseudo-crossover studies for which an adjustment was recommended when it should have been, but the simulated difference type mismatched the final recommendation type.

Gold Standard:
Slope estimate for original data: 0.25 (SE=0.06)
P-value for original data: 0.0001

Dataset	Trends slope estimate based on 1999-2016 unadjusted estimates		Trends slope estimate based on 1999-2016 mismatched adjusted estimates		Trends slope estimate based on 1999-2016 ideally adjusted estimate	
	β	P-Value	β	P-Value	β	P-Value
148	**0.45**	<0.0001	0.26	0.0001	0.27	0.0001
159	**0.14**	0.0260	0.25	0.0001	0.23	0.0002

Standard error terms for β coefficients were 0.06-0.07 throughout (adjusted and unadjusted).