

A.1. Confidence Interval Estimates

This section describes hypotheses used to test the primary objectives and the sample sizes needed to have adequate precision for the 95% confidence intervals used to assess these hypotheses.

Hypotheses associated with the first endpoint discussed in Section 8.3.1:

Null hypothesis 1A: The true difference d between Mercy TAPE predictions and the true weight satisfies: $|d| \geq d_{0A}$.

Alternative hypothesis 1A: $|d| < d_{0A}$.

Null hypothesis 1B: The true difference d between the Mercy TAPE predictions and the Mercy Method predictions satisfies: $|d| \geq d_{0B}$.

Alternative hypothesis 1B: $|d| < d_{0B}$.

A choice of $d_{0B} = 1.25$ gives an error around 3% for a participant whose weight is near a possible overall average of 40-50 kg, and an error of around 10% for participants over 10 kg. Also, the relationship between TAPE and Method should be no worse than the relationship between TAPE and the true weight seen in Table 1, and a confidence interval for the Method ME in Table 1 will be within the equivalence region of $[-1.25, 1.25]$ when the sample size is adequate. Similarly, based on the observed ME values in Table 1, an equivalence region of $[-1.75, 1.75]$ was chosen for the relationship between TAPE and the true weight. This region gives an error below 5% for participants around the average weight and would prevent equivalence from being determined for the methods with high (below -1 or above 1) MEs in Table 1.

For the calculation of confidence interval half-width, the error caused by the deviation of the predictions from the true weight is assumed to be additive and normally distributed for computational simplicity, although these assumptions will be tested when analyzing the data.

Based on the RMSE in Table 1, the sample standard deviation of the differences is assumed to be near 3.65. Values greater than 3.65 should be assumed for testing the first primary hypothesis, while values less than 3.65 may be possible when comparing Mercy Method and either Mercy TAPE.

The half widths of 95% confidence intervals calculated using the t-distribution are shown in Table A.1. The half-width gives the length between the parameter estimate and the end of the confidence interval. It can be seen that a sample size of 150 participants will be adequate to show

equivalence when the observed ME between TAPE and the true weight is between -1 and 1 even when the standard deviation is relatively high.

Table A.1. Half-widths for 95% confidence intervals for the pairwise difference between two methods of estimating weight at different values of the sample size (N) and standard deviation (Std. Dev.) of the paired differences.

N	Half-Widths			
	Std. Dev. 2.75	Std. Dev. 3.65	Std. Dev. 4.5	Std. Dev. 5.5
25	1.14	1.51	1.86	2.27
50	0.78	1.04	1.28	1.56
75	0.63	0.84	1.04	1.27
100	0.55	0.72	0.89	1.09
125	0.49	0.65	0.80	0.97
150	0.44	0.59	0.73	0.89
200	0.38	0.51	0.63	0.77
250	0.34	0.46	0.56	0.69
300	0.31	0.42	0.51	0.63
350	0.29	0.38	0.47	0.58
400	0.27	0.36	0.44	0.54
450	0.26	0.34	0.42	0.51
500	0.24	0.32	0.40	0.48

Hypotheses associated with the second endpoint discussed in Section 8.3.1:

Null hypothesis 2A: The concordance correlation coefficient (CCC) for the relationship between the Mercy Tape prediction and the true weight satisfies: $CCC \leq \rho_A$.

Alternative hypothesis 2A: $CCC > \rho_A$.

Null hypothesis 2B: The concordance correlation coefficient (CCC) for the relationship between the Mercy Tape prediction and the Mercy Method prediction satisfies: $CCC \leq \rho_B$.

Alternative hypothesis 2A: $CCC > \rho_B$.

The equivalence bound defined by $\rho_B = 0.9$ was chosen to require the CCC for Mercy TAPE to be close to the straight line for the relationship between Mercy TAPE and Method, while the relationship between Mercy TAPE and the true weight will be comparable, if not superior, to other methods shown in Table 1 if the CCC is above $\rho_A = 0.85$.

Table A.2 shows assumptions about the distribution of the distribution of the Mercy TAPE, Mercy Method, and true weights used to calculate the confidence intervals in Table A.3. These assumed standard deviations consider the possibility that the TAPE predictions have less variability than the true weight; this would be true if, for example, the TAPE predictions underestimate high weights or overestimate low weights. Note that assumption A or B is more likely than assumption C when considering the relationship between TAPE and the true weight. Also, Figure 1 results suggest that the difference in standard deviations between the true weight and Mercy TAPE should be small, which suggests that these are conservative assumptions. Calculations are performed using the asymptotic normality of the Z-transformation [20].

Table A.2. Assumptions used to calculate confidence intervals in Table A.3.

	Assumption			
	A	B	C	D
Pearson Correlation	0.99	0.975	0.95	0.9
Mean Error	0.5	1	2	2
Std Dev: True or Mercy Method Weight	25	25	25	25
Std Dev: TAPE Prediction	22.5	20	20	20
CCC	0.984	0.950	0.923	0.875

Table A.3. Lower bounds of 95% confidence intervals at different sample sizes (N) using assumptions from Table A.2.

N	Assumption			
	A	B	C	D
25	0.970	0.911	0.850	0.750
50	0.975	0.925	0.878	0.797
75	0.977	0.931	0.888	0.814
100	0.978	0.934	0.894	0.824
125	0.979	0.936	0.897	0.830
150	0.980	0.937	0.900	0.835
200	0.980	0.939	0.903	0.841
250	0.981	0.941	0.906	0.845
300	0.981	0.941	0.907	0.847
350	0.981	0.942	0.909	0.850
400	0.982	0.943	0.910	0.851
450	0.982	0.943	0.910	0.853
500	0.982	0.944	0.911	0.854

Hypotheses associated with the third endpoint discussed in Section 8.3.1:

Null hypothesis 3A: The proportion p of participants with Mercy TAPE predictions within 10% of the true weight satisfies: $p \leq p_{0A}$.

Alternative hypothesis 3A: $p > p_{0A}$.

Null hypothesis 3B: The proportion p of participants with Mercy TAPE predictions within 10% of the Mercy Method satisfies: $p \leq p_{0B}$.

Alternative hypothesis 3B: $p > p_{0B}$.

The value of p_0 for both hypotheses should be at least 0.5; the equivalence region for the relationship between Mercy TAPE and the true weight is based on what is acceptable based on the performance of other methods in Table 1.

Table A.4. Lower bounds of exact 95% confidence intervals for the proportion of subjects with predicted weights within 10% at different sample sizes (N) and observed proportions.

N	<u>Lower Bounds</u>			
	Proportion 0.6	Proportion 0.65	Proportion 0.7	Proportion 0.75
25	0.39	0.44	0.49	0.54
50	0.45	0.50	0.55	0.61
75	0.48	0.53	0.58	0.64
100	0.50	0.55	0.60	0.65
125	0.51	0.56	0.61	0.67
150	0.52	0.57	0.62	0.67
200	0.53	0.58	0.63	0.68
250	0.54	0.59	0.64	0.69
300	0.54	0.59	0.65	0.70
350	0.55	0.60	0.65	0.70
400	0.55	0.60	0.65	0.71
450	0.55	0.60	0.66	0.71
500	0.56	0.61	0.66	0.71

A.2. Power Calculations

In order to evaluate the power of the study, true and predicted weights were simulated for children from each of the 17 age groups. Genders were sampled uniformly. The mean and standard deviation of the true weights for each age group and gender were based on historical NHANES data. Several scenarios were used to model the relationship between the true weights and the predicted weights. The parameter values used for each scenario are shown in Table A.

Scenarios 1-4: True weights were simulated from a normal distribution with age- and gender-dependent parameters. Conditional upon the true weight X_i , the predicted weight Y_i was simulated from the linear regression model

$$\text{Model 1: } Y_i = b + b_1 X_i + \text{error}_i.$$

Here, the error term is normally distributed with mean 0. In scenarios 1-2, the slope $b_1=1$, but the error is assumed to vary across all age and gender strata so that the correlation is close to the assumed value for each stratum. In scenarios 3-4, the error is assumed to have constant variance. The variance was selected to ensure that the correlation between the true and predicted weights is close to the assumed value for 9 year old males, but this implies that the correlation will decrease as the age group decreases. As shown in Table A.5, the overall correlation is higher. These correlation values are sufficiently high to ensure that the CCC is within the equivalence region.

Scenarios 5-6: True and predicted weights were simulated from a bivariate normal distribution with means μ_{True} and $\mu_{\text{True}} + b$, where b is the bias in the predicted weights compared to the true weights. Standard deviations are σ_{True} and σ_{Pred} with $\sigma_{\text{True}} > \sigma_{\text{Pred}}$, and the correlation is ρ .

Using this sampling model, the predicted weight for subject i , Y_i , can again be written as an additive function of the true weight X_i :

$$\text{Model 2: } Y_i = X_i + b + \text{error}_i.$$

As in scenarios 1-4, the error term is normally distributed, but in scenarios 7-8, it is negatively correlated with the true weight. This error structure may be true when the predicted weight tends to under predict the weight of a child with relatively high weight and over predict the weight of a child with relatively low weight.

Table A.5. Parameters used for each simulation scenario. Only standard deviation ratios that are not determined by other parameters are shown. Values of population correlations (ρ) and proportions estimated using Monte Carlo methods.

Scenario: Model	Within-Stratum ρ	Population ρ	Population Proportion	Bias (b)	Slope (b_1)	$\sigma_{\text{Pred}} / \sigma_{\text{True}}$
1: Model 1	0.92	0.989	0.82	-0.5	1	
2: Model 1	0.89	0.986	0.76	-0.8	1	
3: Model 1	0.92 for 9 Y.O. Males	0.993	0.73	0.4	0.95	
4: Model 1	0.89 for 9 Y.O. Males	0.990	0.66	0.9	0.875	
5: Model 2	0.92	0.990	0.86	-0.5		0.9
6: Model 2	0.89	0.986	0.76	-0.8		0.825

The following plots show percentiles of bounds for the 95% confidence intervals calculated from 1000 simulations at each sample size from 4-35 participants per age group. Plots for the proportion and CCC show the 10th and 20th percentiles of the lower bounds of the simulated 95% confidence intervals. If the 10th percentile is within the equivalence region, then the estimated probability is at least 0.9 that an observed confidence interval will be within the equivalence region. Plots of the mean error show the 80th and 90th percentiles for the absolute value of the most extreme (greatest in absolute value) confidence interval bound. Here, if the 90th percentile is within the equivalence region, then the estimated probability is also at least 0.9 that an observed confidence interval will be within the equivalence region.

Figures A.1 and A.2 show expected changes in the confidence interval for the mean error as N increases. The scenario in Figure A.1 reflects the hypothesized relationship between Mercy TAPE and the true weight, and Figure A.2 shows the hypothesized relationship between the Mercy TAPE and Mercy Method predicted weights. It can be seen that 90% of the simulated confidence intervals for the mean error is within the equivalence region when the sample size is at least 10 per age group (N=170 total), and the power is estimated to be above 0.9 at this sample size.

Figure A.1. Percentiles of the simulated confidence interval bounds for the mean error. Dashed line shows the boundary of equivalence region for the relationship between Mercy TAPE and the true weights.

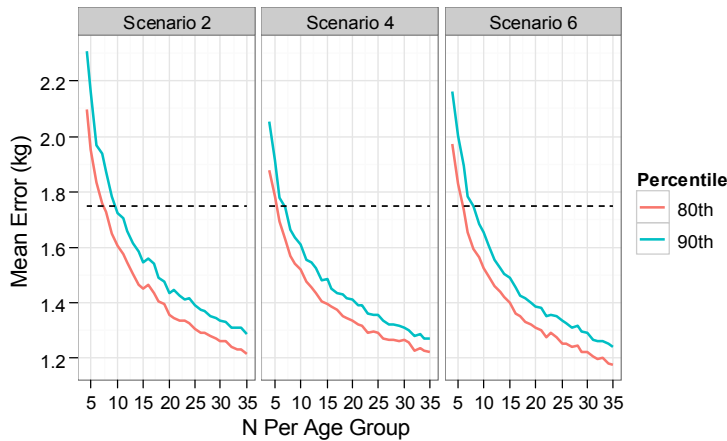
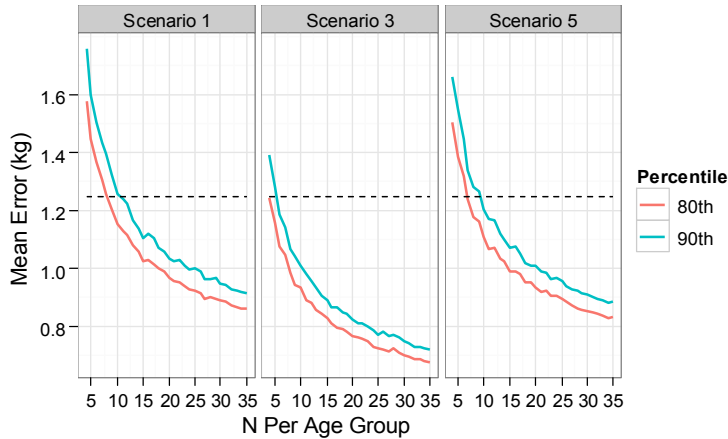


Figure A.2. Percentiles of the simulated confidence interval bounds for the mean error. Dashed line shows the boundary of equivalence region for the relationship between Mercy TAPE and Mercy Method.



Figures A.3 and A.4 show expected changes in the confidence interval for the proportion of predicted weights within 10% of the true weight as N increases. The scenario in Figure A.3 reflects the hypothesized relationship between Mercy TAPE and the true weight, and Figure A.4 shows the hypothesized relationship between the Mercy TAPE and Mercy Method predicted weights. It can be seen that over 90% of the simulated confidence intervals for the proportion are within the equivalence region when the sample size is at least 10 per age group.

Figure A.3. Percentiles of the simulated confidence interval bounds for the proportion of predicted weights within 10% of the true weight. Dashed line shows the boundary of equivalence region for the relationship between Mercy TAPE and the true weight.

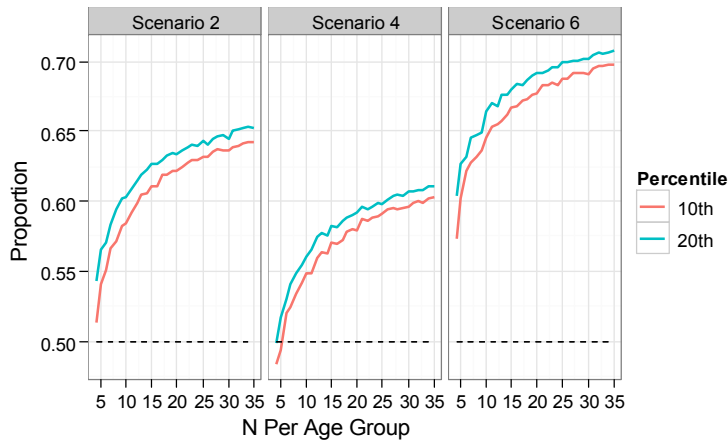
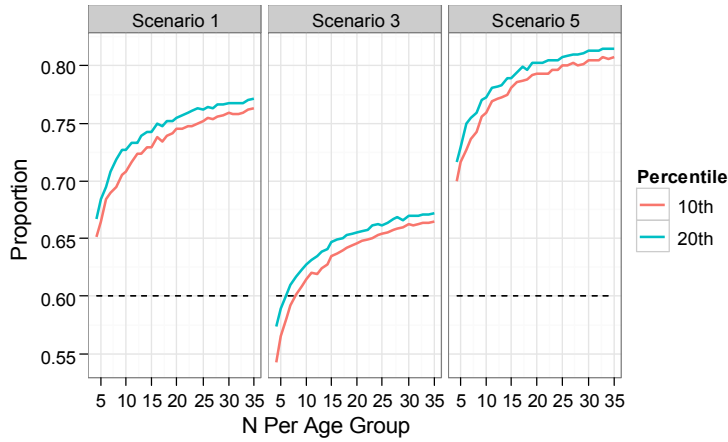


Figure A.4. Percentiles of the simulated confidence interval bounds for the proportion of predicted weights within 10% of the true weight. Dashed line shows the boundary of equivalence region for the relationship between Mercy TAPE and Mercy Method.



The following plots show expected changes in the confidence interval for CCC as N increases. The scenario in Figure A.5 reflects the hypothesized relationship between Mercy TAPE and the true weight, and Figure A.6 shows the hypothesized relationship between the Mercy TAPE and Mercy Method predicted weights. It can be seen that the confidence interval for the CCC is expected to be well within the equivalence region regardless of the sample size.

Figure A.5. Percentiles of the simulated confidence interval bounds for the CCC.

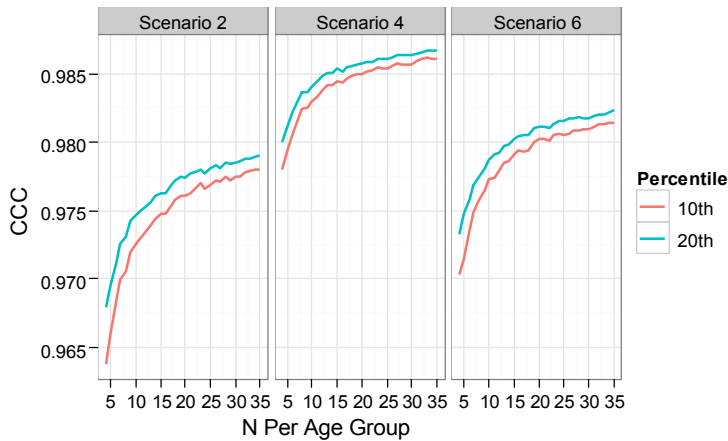


Figure A.6. Percentiles of the simulated confidence interval bounds for the CCC.

