

CONTINUED NHANES USE FOR STANDARD AVERAGE BODY WEIGHT

20 December 2024

ABSTRACT

For use in aircraft weight and balance, a previously defined method to determine standard average weights for body mass of passengers and crew in a manner consistent with guidance from the United States (U.S.) Federal Aviation Administration (FAA), especially Advisory Circular (AC) 120-27 "Aircraft Weight and Balance Control," is applied to data available in 2024 from the Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (NHANES). Application of the method to data collected before and after the SARS-CoV-2/COVID-19 pandemic is discussed, and trends are illustrated. Results are applicable to Operation Specification (OpSpec) approval using survey-derived average weight values.

AUTHORS

Charles Ostick, Manager, Flight Operations Engineering, Alaska Airlines
William Yingling, Senior Aircraft Performance Engineer, AeroData

REVIEW COMMITTEE

Kristine Henning, Senior Principal Engineer, Performance Engineering, Delta Air Lines (Chair)
Ravin Agarwal, Director Performance and Operations Engineering, National Airlines
Logan Jones, PhD, Enterprise Architect, NAVBLUE, an Airbus Services Company
Richard Martin, Senior Manager Aircraft Performance Engineering, Sun Country Airlines
Chris Meterko, Lead Engineer, Flight Ops Engineering, United Airlines
Cary Robins, President, American Aeronautics
Gail Zittel, Associate Technical Fellow, Weight Engineering, Boeing Commercial Airplanes

OWNERSHIP & DISCLAIMER

This document remains the sole property of the Society of Aircraft Performance and Operations Engineers (SAPOE).

SAPOE is a member-based organization promoting the safety and efficiency of flight through knowledge of aircraft performance and weight and balance principles.

Except for brief quotations with appropriate citation, copies may be made and distributed only of the complete document, including cover page and this disclaimer.

SAPOE, the authors, and reviewers assume no liability whatsoever and make no warranty of any kind. Anyone who uses this report, or the source code and data contained in it, is solely responsible for their own operations and for any outcome of such use.

CONTENTS

PROBLEM STATEMENT	2
NHANES IN 2020 AND YEARS SINCE	2
CALCULATIONS USING SAMPLE CODE	4
ANALYSIS AND DISCUSSION	10
UPDATED SUMMARY VALUES	11
REFERENCES.....	12
APPENDIX A: CONSIDERING THE FUTURE.....	13
APPENDIX B: UPDATED SAMPLE CODE.....	17
APPENDIX C: SAPOE 2020 (Reprint).....	18

PROBLEM STATEMENT

In response to changes published by the United States (U.S.) Federal Aviation Administration (FAA) in 2019 and 2020, the Society of Aircraft Performance and Operations Engineers (SAPOE) published a paper titled “NHANES USE FOR STANDARD AVERAGE BODY WEIGHT,” dated 14 DEC 2020 (hereafter, “SAPOE 2020”), which defined a method to determine standard average weights for body mass of passengers and crew. Like FAA guidance before it, the SAPOE 2020 method relies heavily on data from the Centers for Disease Control and Prevention (CDC) obtained in the National Health and Nutrition Examination Survey (NHANES).

Since then, CDC NHANES has continued to collect and publish body-mass data, but its data collection schedule and survey design have been modified due to the SARS-CoV-2/COVID-19 pandemic (hereafter, “Covid”). Application of the SAPOE 2020 method to the new data, for use in current aviation operations, requires a review of these changes.

Familiarity with the terms and definitions in SAPOE 2020 is assumed. Formulas from SAPOE 2020 are used extensively, identified here as (2020-N) where (N) is the formula number in the original paper. This paper is referred to as “SAPOE 2024” with newly defined formulas identified using the convention (2024-X) for clarity.

NHANES IN 2020 AND YEARS SINCE

NHANES is a rigorous scientific study of health and nutrition among the U.S., non-institutionalized, civilian population whose scope includes numerous body measurements. Academically rigorous documentation, including Sample Design, Weighting Process, and Analytic Guidelines, is available in a series of papers at <https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx>.

Note: NHANES generally (not exclusively) uses “weight” to mean a numeric scaling factor, as in the common meaning of “weighted average,” rather than a measure of gravitational mass. To avoid confusion, this section carefully uses “mass” as the object of gravity and follows the NHANES convention for “weight” despite this paper’s use elsewhere of U.S. aviation vernacular, including “weight and balance” for the effect of gravity.

From 1999 until Covid, NHANES was conducted in cycles, each covering two complete calendar years. Cycles were numbered consecutively from 1. At the time of publication of SAPOE 2020, the most recently available data were from cycle 10, collected in 2017 and 2018. SAPOE 2020 followed NHANES Analytic Guidelines in combining each odd-

number cycle with the following even-number cycle to create a four-year cycle. SAPOE 2020 excluded cycles 1 and 2 from its summary because their design requires a different method of variance analysis than all later cycles.

Continuing its established pattern, NHANES started collecting data for cycle 11 at the beginning of 2019, but this effort was discontinued by CDC in March 2020 in response to Covid. Since the data collected for cycle 11 is incomplete, CDC will not publish it alone. However, this data was published in combination with data from cycle 10 as a special cycle, covering 3.2 years, which CDC numbered “66.” NHANES Analytic Guidelines indicate survey design weights for cycle 66 were recalculated to be as representative of the population as possible despite incomplete data. However, it also warns that this makes comparison of cycle 66 to other cycles unsuitable for trend analysis. (The recalculated design weights for cycle 66 are labelled WTMECPRP by CDC to distinguish them from WTMEC2YR or the combined WTMEC4YR weights.)

Continuing the now disrupted pattern, cycle 12 would have started at the beginning of 2021 and proceeded through the end of 2022. However, the start of data collection was delayed, and CDC elected to preserve the collection period of two whole years rather than the calendar-year cadence. The resulting cycle 12 covers 2.0 years beginning in August 2021 and ending in August 2023. NHANES does not recommend combining post-Covid data with pre-Covid data.

The following table summarizes these cycle differences:

Table 2024-1: NHANES Data Cycles

Cycle Number	Data Published	Body Measures File Name	Data Collection Begin	Data Collection End	Combined Cycle Design	Note
1	Jun 2002	BMX	Jan 1999	Dec 2000	Designed in advance	
2	May 2004	BMX_B	Jan 2001	Dec 2002	to combine 1 and 2	
3	Nov 2005	BMX_C	Jan 2003	Dec 2004	Designed in advance	
4	Nov 2007	BMX_D	Jan 2005	Dec 2006	to combine 3 and 4	
5	Sep 2009	BMX_E	Jan 2007	Dec 2008	Designed in advance	
6	Sep 2011	BMX_F	Jan 2009	Dec 2010	to combine 5 and 6	
7	Sep 2013	BMX_G	Jan 2011	Dec 2012	Designed in advance	
8	Oct 2015	BMX_H	Jan 2013	Dec 2014	to combine 7 and 8	
9	Sep 2017	BMX_I	Jan 2015	Dec 2016	Designed in advance	
10	Feb 2020	BMX_J	Jan 2017	Dec 2018	to combine 9 and 10	
11	Never	None	Jan 2019	Mar 2020	N/A	Data collection suspended due to Covid.
66	May 2021	P_BMX	Jan 2017	Mar 2020	Redesign combines 10 and 11	Design weights were recalculated after data collection to combine with cycle 10 despite incomplete data.
12	Sep 2024	BMX_L	Aug 2021	Aug 2023	TBA	Start of data collection delayed. Data combination method not yet announced.

NHANES Analytic Guidelines specifies that trend analysis must use a continuous time scale to account for changes in both cycle duration and the cadence of cycle start. Data may be referenced to the midpoint of the data collection period, which is given as 2017.0 for the combined 2015–2018 cycle, 2018.6 for cycle 66, and 2022.7 for cycle 12. (These are decimal fractions of a year, distinct from the “YYYY/MM” format used for whole months in date labels.)

It is important that data variables have consistent definitions when compared across survey cycles. The data variables in cycle 12 data files are consistent with those used by SAPOE 2020 except for age masking. Specifically, this means that to prevent any person surveyed from being identifiable from the data collected about them, data for anyone older than 80 is coded as if they were 80 years old, and data for anyone who was both pregnant and under the age of 20 is coded as if they were 20 years old. NHANES documentation draws attention to pregnancy and age masking because it may move data between categories used by CDC when it creates growth charts from NHANES data for use in other medical contexts. If age, gender, and crew categories defined in SAPOE 2020 are used, no data is moved from one category to another, so the age masking does not change any result. For example, data for any

pregnant teen will be in the category “Adult,” as defined by FAA, whether coded as 16 years old or as 20 years old. (There is no data for any “Child,” as defined by FAA, who is also pregnant in any NHANES data set examined.)

CALCULATIONS USING SAMPLE CODE

Calculations are demonstrated here using “R,” a data analysis language accessible to many data analysts and operations engineers, printed in a fixed-width font with executable code shown in **blue**, descriptive comments in **green**, and any output generated in **orange**. Code fragments should be executed in the order presented. The complete R script described is included in Appendix B to this paper.

Since the data variable definitions and file formats for NHANES cycle 12 are consistent with cycles 3 through 10, sample code from SAPOE 2020 will generate correct results for cycle 12 with only two minimal modifications.

The first required modification replaces the single call to `importDataCycles` to import the latest raw data files:

```
# Load the data into memory once each session
RawData <- importDataCycles(12)
```

The second required modification is one added line in function `fmtWTLB`, before `print.data.frame`, to update the years used to label the latest data:

```
cycleData <- mutate(cycleData,
  Years=ifelse(12==cycle4, '2021/08-2023/08', as.character(Years)))
```

Further reorganization of the sample code allows much improved reuse and clarity without modifying the fundamental implementation. Differences from SAPOE 2020, Appendix A, to SAPOE 2024, Appendix B, are discussed here. Unmodified output from the script in Appendix B is shown below. These data values exactly match output in the SAPOE 2020 paper, substantiating that the code reorganization does not change the values in the results.

First, since the NHANES cycles have become more complex and less predictable, all the metadata describing each cycle is grouped into a function called `nhanesCycleDefinitions`. This function must be modified to add future cycles, but it is hoped that such modifications, such as the parity for combination of cycles, can be limited to updating this one location.

For trend analysis, the metadata defined includes the **duration** of each cycle in decimal years and a continuous time scale **trendYear**. Using time relative to a contemporary date preserves all effects of cycle duration and of delayed cycle start, but it better represents available precision for calculation. (For example, the precision implied by rounding 2022.7 to tenths of a year is approximately $\pm 2.5\%$ over a two-year period, not the $\pm 0.0025\%$ implied by counting significant digits in $2.0227 \cdot 10^{+3}$ years.) The choice of the midpoint of the 2015–2018 combined cycle as the reference for relative time is arbitrary and allows convenient comparison to SAPOE 2020.

The function `importRawDataFiles` makes importing raw data files more convenient by downloading only missing cycles without having to comment out code each time the script is run. This function also encapsulates file-naming conventions that differ subtly from the cycle numbering. It would need to be modified if the website locations of the data files change, or if the conventions for variable names inside the data files change.

The function `enrichDataSet` adds unit conversion and category definitions not defined by NHANES but which are needed for this analysis, especially **faaAge**. This function would need to be modified if the conventions for variable names inside the data files change.

To emphasize when the identical method is applied to multiple population subsets, the function `sapoeNHANES` combines calculations SAPOE 2020 identified as `Step0` through `Step5`, together with `NN` for estimating population ratios using Formula (2020-7). This function should not need to change unless the method definition is changed.

The formatting function `fmtWTLB` behaves as before with accommodation for changes in function `sapoeNHANES`. The formatted output columns, defined either in SAPOE 2020 or above, are summarized in the following table.

Table 2024-2: Output Columns and Units

<i>Label</i>	<i>Meaning</i>	<i>Units</i>
<code>cycle</code>	NHANES raw or combined cycle number	none
<code>label</code>	NHANES cycle description by year and month	"YYYY/MM" string (not a number)
<code>duration</code>	NHANES raw or combined cycle duration	decimal years
<code>trendYear</code>	SAPOE 2024 relative time for trend analysis	decimal years
<code>cdcAge</code>	CDC age category for NHR	none
<code>faaAge</code>	SAPOE 2020 age category for FAA	none
<code>gender</code>	CDC defined gender, also used for FAA	none
<code>counts</code>	NHANES sample size	1
<code>WTLB</code>	body mass (unclothed, except where indicated)	pounds
<code>se</code>	standard error, using Taylor series linearization	pounds
<code>n</code>	SAPOE 2020 filtered sample size, n_f	1
<code>NP</code>	SAPOE 2020 population quantity, N	1
<code>W</code>	SAPOE 2020 average body mass, W	pounds
<code>W_*</code>	SAPOE 2020 operational body mass, $[W]$	pounds
<code>S</code>	FAA Standard Deviation, using Formula (2020-4)	pounds
<code>te%</code>	FAA Tolerable Error, using Formula (2020-5)	unitless % (not pounds)

Sample code should not be used for NHANES cycles 1 or 2 because Analytic Guidelines recommends variance analysis using a jackknife method, with specified resampling weights, for these cycles but recommends Taylor series linearization using `library(survey)` for all later cycles.

Once these functions are defined, and `setwd` is pointed to the appropriate system-specific location for data files, the following code and output demonstrate matching values to those on pages 11 and 12 of SAPOE 2020, which were compared to statistics published by CDC in National Health Statistics Reports (NHR) in 2018.

```
RawData <- importRawDataFiles(c(3:10,66,12))

Cycles2 <- nhanesCycleDefinitions(FALSE)
WTLB2YR <- enrichDataSet( RawData, Cycles2 )

# Define survey design parameters for overall dataset, per NHANES Tutorials
NHANES2<-svydesign(data=WTLB2YR, id=~SDMVPSU, strata=~SDMVSTRA, weights=~svywt, nest=TRUE)

# confirm definitions above here are correct for full NHANES by comparing to
# https://www.cdc.gov/nchs/data/nhr/nhr122-508.pdf
svyWTLB( subset( NHANES2, NotPregnant ), ~cycle+cdcAge+gender ) %>% fmtWTLB(Cycles2) %>%
  filter( cdcAge=='Adult' & cycle<10 ) # NHR Table 2 has Adults in these years
```

	<code>cycle</code>	<code>label</code>	<code>duration</code>	<code>trendYear</code>	<code>cdcAge</code>	<code>gender</code>	<code>counts</code>	<code>WTLB</code>	<code>se</code>
1	3	2003/01-2004/01	2	-13	Adult	Male	2247	193.5	1.0
2	4	2005/01-2006/01	2	-11	Adult	Male	2242	196.0	1.7
3	5	2007/01-2008/01	2	-9	Adult	Male	2755	194.7	1.4

4	6	2009/01-2010/01	2	-7	Adult	Male	2896	196.3	1.4
5	7	2011/01-2012/01	2	-5	Adult	Male	2591	194.4	1.4
6	8	2013/01-2014/01	2	-3	Adult	Male	2645	197.0	1.2
7	9	2015/01-2016/01	2	-1	Adult	Male	2584	197.8	1.9
8	3	2003/01-2004/01	2	-13	Adult	Female	2201	164.2	1.6
9	4	2005/01-2006/01	2	-11	Adult	Female	2129	165.1	1.7
10	5	2007/01-2008/01	2	-9	Adult	Female	2805	166.2	1.3
11	6	2009/01-2010/01	2	-7	Adult	Female	3039	166.3	0.9
12	7	2011/01-2012/01	2	-5	Adult	Female	2602	167.1	1.3
13	8	2013/01-2014/01	2	-3	Adult	Female	2823	169.8	1.3
14	9	2015/01-2016/01	2	-1	Adult	Female	2757	170.5	1.7

(In this and subsequent output, the first column is an automatically generated row number for reference only.)

The following code and output demonstrate matching values to those on page 14 of SAPOE 2020, which were the output from Part 1 under “Method Definition,” SAPOE 2020, using combined-cycle data.

```
Cycles4 <- nhanesCycleDefinitions(TRUE )
WTLB4YR <- enrichDataSet( RawData, Cycles4 )
```

```
Passengers <- sapoeNHANES( WTLB4YR )
fmtWTLB( Passengers, Cycles4 )
```

	cycle	label	duration	trendYear	faaAge	gender	n	NP	W	S	te%
1	4	2003/01-2006/01	4.0	-12.0	Infant	None	1494	2.9	21.8	5.1	1.2
2	6	2007/01-2010/01	4.0	-8.0	Infant	None	1382	2.9	21.1	4.9	1.2
3	8	2011/01-2014/01	4.0	-4.0	Infant	None	1167	2.4	20.9	5.2	1.4
4	10	2015/01-2018/01	4.0	0.0	Infant	None	1118	2.4	21.1	5.0	1.4
5	66	2017/01-2020/03	3.2	1.6	Infant	None	820	2.3	21.2	5.2	1.7
6	4	2003/01-2006/01	4.0	-12.0	Child	None	4087	15.0	62.0	26.4	1.3
7	6	2007/01-2010/01	4.0	-8.0	Child	None	4156	14.8	61.5	26.7	1.3
8	8	2011/01-2014/01	4.0	-4.0	Child	None	4303	14.6	62.4	26.9	1.3
9	10	2015/01-2018/01	4.0	0.0	Child	None	3692	13.7	62.9	26.3	1.3
10	66	2017/01-2020/03	3.2	1.6	Child	None	2835	13.5	63.6	27.2	1.6
11	4	2003/01-2006/01	4.0	-12.0	Adult	Male	5940	39.7	186.1	36.1	0.5
12	6	2007/01-2010/01	4.0	-8.0	Adult	Male	6291	40.1	186.5	37.3	0.5
13	8	2011/01-2014/01	4.0	-4.0	Adult	Male	5879	40.4	186.8	37.1	0.5
14	10	2015/01-2018/01	4.0	0.0	Adult	Male	5689	40.8	189.6	40.5	0.6
15	66	2017/01-2020/03	3.2	1.6	Adult	Male	4559	41.0	190.5	40.1	0.6
16	4	2003/01-2006/01	4.0	-12.0	Adult	Female	6321	42.4	155.9	35.1	0.6
17	6	2007/01-2010/01	4.0	-8.0	Adult	Female	6435	42.2	157.3	35.3	0.5
18	8	2011/01-2014/01	4.0	-4.0	Adult	Female	6125	42.6	159.3	36.4	0.6
19	10	2015/01-2018/01	4.0	0.0	Adult	Female	5937	43.1	161.6	37.2	0.6
20	66	2017/01-2020/03	3.2	1.6	Adult	Female	4680	43.2	162.7	38.4	0.7

The following code and output are the survey-derived averages for body mass calculated in accordance with Part 1 under “Method Definition,” SAPOE 2020, using two-year cycle data instead of combined-cycle data.

```
Passengers <- sapoeNHANES( WTLB2YR )
fmtWTLB( Passengers, Cycles2 )
```

	cycle	label	duration	trendYear	faaAge	gender	n	NP	W	S	te%
1	3	2003/01-2004/01	2	-13.0	Infant	None	709	2.8	22.0	5.2	1.7
2	4	2005/01-2006/01	2	-11.0	Infant	None	783	3.0	21.5	4.9	1.6
3	5	2007/01-2008/01	2	-9.0	Infant	None	700	2.9	20.8	4.8	1.7

4	6	2009/01-2010/01	2	-7.0	Infant	None	684	2.9	21.2	5.0	1.8
5	7	2011/01-2012/01	2	-5.0	Infant	None	567	2.3	20.6	5.3	2.1
6	8	2013/01-2014/01	2	-3.0	Infant	None	599	2.5	21.2	5.2	2.0
7	9	2015/01-2016/01	2	-1.0	Infant	None	606	2.5	21.1	4.8	1.8
8	10	2017/01-2018/01	2	1.0	Infant	None	513	2.4	21.1	5.4	2.2
9	12	2021/08-2023/08	2	5.7	Infant	None	230	2.3	21.3	4.5	2.7
10	3	2003/01-2004/01	2	-13.0	Child	None	1921	15.2	63.4	27.5	1.9
11	4	2005/01-2006/01	2	-11.0	Child	None	2167	14.9	60.8	25.4	1.8
12	5	2007/01-2008/01	2	-9.0	Child	None	2061	14.8	60.9	26.4	1.9
13	6	2009/01-2010/01	2	-7.0	Child	None	2095	14.8	62.1	27.0	1.9
14	7	2011/01-2012/01	2	-5.0	Child	None	2142	14.9	62.6	27.5	1.9
15	8	2013/01-2014/01	2	-3.0	Child	None	2167	14.4	62.4	26.6	1.8
16	9	2015/01-2016/01	2	-1.0	Child	None	2057	14.0	62.8	26.5	1.8
17	10	2017/01-2018/01	2	1.0	Child	None	1637	13.5	63.2	26.4	2.0
18	12	2021/08-2023/08	2	5.7	Child	None	1439	13.8	65.8	29.1	2.3
19	3	2003/01-2004/01	2	-13.0	Adult	Male	2992	39.9	185.1	35.1	0.7
20	4	2005/01-2006/01	2	-11.0	Adult	Male	2953	39.6	187.3	37.4	0.7
21	5	2007/01-2008/01	2	-9.0	Adult	Male	3043	40.0	185.2	36.3	0.7
22	6	2009/01-2010/01	2	-7.0	Adult	Male	3246	40.2	187.9	38.1	0.7
23	7	2011/01-2012/01	2	-5.0	Adult	Male	2890	40.4	186.2	36.4	0.7
24	8	2013/01-2014/01	2	-3.0	Adult	Male	2995	40.4	187.4	37.8	0.7
25	9	2015/01-2016/01	2	-1.0	Adult	Male	2905	40.6	188.0	39.6	0.8
26	10	2017/01-2018/01	2	1.0	Adult	Male	2788	41.0	191.6	41.5	0.8
27	12	2021/08-2023/08	2	5.7	Adult	Male	2931	41.2	189.1	39.6	0.8
28	3	2003/01-2004/01	2	-13.0	Adult	Female	3117	42.1	155.3	34.0	0.8
29	4	2005/01-2006/01	2	-11.0	Adult	Female	3197	42.5	156.3	36.0	0.8
30	5	2007/01-2008/01	2	-9.0	Adult	Female	3097	42.4	157.1	35.1	0.8
31	6	2009/01-2010/01	2	-7.0	Adult	Female	3345	42.1	157.7	35.7	0.8
32	7	2011/01-2012/01	2	-5.0	Adult	Female	2906	42.4	158.6	35.5	0.8
33	8	2013/01-2014/01	2	-3.0	Adult	Female	3217	42.7	160.0	37.3	0.8
34	9	2015/01-2016/01	2	-1.0	Adult	Female	3059	42.9	161.7	36.9	0.8
35	10	2017/01-2018/01	2	1.0	Adult	Female	2880	43.2	161.9	37.8	0.9
36	12	2021/08-2023/08	2	5.7	Adult	Female	3476	42.8	163.0	38.6	0.8

The following code and output are newly calculated sample values and adjustments demonstrating application of Part 2 under “Method Definition,” SAPOE 2020. Operators should use adjustments that are most appropriate to their systems.

```
Wx <- left_join(
  Passengers %>% select(cycle,faaAge,gender,W ) %>%
    pivot_wider(names_from=c(faaAge,gender),values_from=W, names_prefix='W_'),
  Passengers %>% select(cycle,faaAge,gender,NP) %>%
    pivot_wider(names_from=c(faaAge,gender),values_from=NP,names_prefix='N_'),
  by=c('cycle')
) %>% rename( N_Infant=N_Infant_None, W_Infant=W_Infant_None ) %>%
mutate(
  W_Adult_NHANES = F13( N_Adult_Male, N_Adult_Female, W_Adult_Male, W_Adult_Female ),
  W_Adult_5050 = F13( 50, 50, W_Adult_Male, W_Adult_Female ),
  dW_Infant_NHANES = N_Infant / ( N_Adult_Female + N_Adult_Male ) * W_Infant,
  'W_Adult_NHANES+Infant' = W_Adult_NHANES + dW_Infant_NHANES,
  'W_Adult_5050+Infant' = W_Adult_5050 + dW_Infant_NHANES,
)
Wx %>% select(cycle, W_Adult_NHANES,W_Adult_5050,dW_Infant_NHANES) %>% fmtWTLB(Cycles2)
Wx %>% select(cycle,'W_Adult_NHANES+Infant','W_Adult_5050+Infant') %>% fmtWTLB(Cycles2)
```


	cycle	label	duration	trendYear	W_Adult_NHANES	W_Adult_5050	dW_Infant_NHANES
1	3	2003/01-2004/01	2	-13.0	169.8	170.2	0.8
2	4	2005/01-2006/01	2	-11.0	171.3	171.8	0.8
3	5	2007/01-2008/01	2	-9.0	170.7	171.1	0.7
4	6	2009/01-2010/01	2	-7.0	172.5	172.8	0.8
5	7	2011/01-2012/01	2	-5.0	172.0	172.4	0.6
6	8	2013/01-2014/01	2	-3.0	173.3	173.7	0.6
7	9	2015/01-2016/01	2	-1.0	174.5	174.8	0.6
8	10	2017/01-2018/01	2	1.0	176.3	176.7	0.6
9	12	2021/08-2023/08	2	5.7	175.8	176.0	0.6

	cycle	label	duration	trendYear	W_Adult_NHANES+Infant	W_Adult_5050+Infant
1	3	2003/01-2004/01	2	-13.0	170.6	171.0
2	4	2005/01-2006/01	2	-11.0	172.1	172.6
3	5	2007/01-2008/01	2	-9.0	171.4	171.9
4	6	2009/01-2010/01	2	-7.0	173.2	173.6
5	7	2011/01-2012/01	2	-5.0	172.6	173.0
6	8	2013/01-2014/01	2	-3.0	174.0	174.4
7	9	2015/01-2016/01	2	-1.0	175.1	175.4
8	10	2017/01-2018/01	2	1.0	176.9	177.3
9	12	2021/08-2023/08	2	5.7	176.3	176.6

The following code and output are the survey-derived averages for body mass calculated in accordance with Part 3 under “Method Definition,” SAPOE 2020_L using two-year cycle data for Male and Female Pilots.

```
Pilots <- subset( WTLB2YR, 23<=RIDAGEYR & 65>RIDAGEYR ) %>% mutate(faaAge='Pilot')
fmtWTLB( sapoeNHANES( Pilots ), Cycles2 )
```

	cycle	label	duration	trendYear	faaAge	gender	n	NP	W	S	te%
1	3	2003/01-2004/01	2	-13.0	Pilot	Male	1321	48.6	190.6	32.1	0.9
2	4	2005/01-2006/01	2	-11.0	Pilot	Male	1444	48.8	194.2	35.5	0.9
3	5	2007/01-2008/01	2	-9.0	Pilot	Male	1776	49.3	191.2	34.4	0.8
4	6	2009/01-2010/01	2	-7.0	Pilot	Male	1873	48.8	194.8	36.5	0.8
5	7	2011/01-2012/01	2	-5.0	Pilot	Male	1714	49.0	192.5	34.9	0.9
6	8	2013/01-2014/01	2	-3.0	Pilot	Male	1795	49.3	193.7	35.4	0.8
7	9	2015/01-2016/01	2	-1.0	Pilot	Male	1705	49.1	193.6	38.0	0.9
8	10	2017/01-2018/01	2	1.0	Pilot	Male	1585	49.0	199.0	41.0	1.0
9	12	2021/08-2023/08	2	5.7	Pilot	Male	1565	49.3	195.8	38.6	1.0
10	3	2003/01-2004/01	2	-13.0	Pilot	Female	1489	51.4	161.2	35.7	1.1
11	4	2005/01-2006/01	2	-11.0	Pilot	Female	1649	51.2	163.1	37.4	1.1
12	5	2007/01-2008/01	2	-9.0	Pilot	Female	1850	50.7	162.9	35.7	1.0
13	6	2009/01-2010/01	2	-7.0	Pilot	Female	2045	51.2	162.8	36.8	1.0
14	7	2011/01-2012/01	2	-5.0	Pilot	Female	1783	51.0	164.3	35.1	1.0
15	8	2013/01-2014/01	2	-3.0	Pilot	Female	1923	50.7	167.0	39.1	1.0
16	9	2015/01-2016/01	2	-1.0	Pilot	Female	1886	50.9	167.3	38.4	1.0
17	10	2017/01-2018/01	2	1.0	Pilot	Female	1744	51.0	166.9	39.2	1.1
18	12	2021/08-2023/08	2	5.7	Pilot	Female	1979	50.7	171.0	41.0	1.1

The following code and output are newly calculated sample values for several $M_P:F_P$ ratios, demonstrating application of Part 3 under “Method Definition,” SAPOE 2020_L using two-year cycle data for Pilots.

```
fmtWTLB( sapoeNHANES( Pilots ) %>% select(cycle,gender,W) %>%
  pivot_wider(names_from=c(gender),values_from=W,names_prefix='W_P__') %>%
  mutate( W_P_50=F13( 50,50, W_P_Male,W_P_Female ),
    W_P_99=F13( 99, 1, W_P_Male,W_P_Female ) ) , Cycles2 )
```


	cycle	label	duration	trendYear	W_P_Male	W_P_Female	W_P_50	W_P_99
1	3	2003/01-2004/01	2	-13.0	190.6	161.2	175.9	190.3
2	4	2005/01-2006/01	2	-11.0	194.2	163.1	178.6	193.9
3	5	2007/01-2008/01	2	-9.0	191.2	162.9	177.0	190.9
4	6	2009/01-2010/01	2	-7.0	194.8	162.8	178.8	194.4
5	7	2011/01-2012/01	2	-5.0	192.5	164.3	178.4	192.2
6	8	2013/01-2014/01	2	-3.0	193.7	167.0	180.4	193.4
7	9	2015/01-2016/01	2	-1.0	193.6	167.3	180.4	193.3
8	10	2017/01-2018/01	2	1.0	199.0	166.9	182.9	198.6
9	12	2021/08-2023/08	2	5.7	195.8	171.0	183.4	195.6

The following code and output are the survey-derived averages for body mass calculated in accordance with Part 3 under “Method Definition,” SAPOE 2020₂ using two-year cycle data for Male and Female Flight Attendants (F/A).

```
FlightAttendants <- filter( WTLB2YR, 21<=RIDAGEYR ) %>% mutate(faaAge='F/A')
fmtWTLB( sapoeNHANES( FlightAttendants ), Cycles2 )
```

	cycle	label	duration	trendYear	faaAge	gender	n	NP	W	S	te%
1	3	2003/01-2004/01	2	-13.0	F/A	Male	2030	48.3	189.2	33.1	0.8
2	4	2005/01-2006/01	2	-11.0	F/A	Male	2048	48.0	192.2	35.3	0.8
3	5	2007/01-2008/01	2	-9.0	F/A	Male	2503	48.1	190.0	34.6	0.7
4	6	2009/01-2010/01	2	-7.0	F/A	Male	2634	48.4	192.4	36.5	0.7
5	7	2011/01-2012/01	2	-5.0	F/A	Male	2327	48.5	190.4	34.5	0.7
6	8	2013/01-2014/01	2	-3.0	F/A	Male	2404	48.3	191.7	35.5	0.7
7	9	2015/01-2016/01	2	-1.0	F/A	Male	2338	48.3	192.5	37.5	0.8
8	10	2017/01-2018/01	2	1.0	F/A	Male	2292	48.4	196.6	40.1	0.8
9	12	2021/08-2023/08	2	5.7	F/A	Male	2461	48.5	193.3	37.8	0.8
10	3	2003/01-2004/01	2	-13.0	F/A	Female	2173	51.7	158.7	33.9	0.9
11	4	2005/01-2006/01	2	-11.0	F/A	Female	2228	52.0	160.1	36.2	0.9
12	5	2007/01-2008/01	2	-9.0	F/A	Female	2602	51.9	160.3	34.6	0.8
13	6	2009/01-2010/01	2	-7.0	F/A	Female	2803	51.6	161.1	36.1	0.8
14	7	2011/01-2012/01	2	-5.0	F/A	Female	2378	51.5	161.9	34.8	0.9
15	8	2013/01-2014/01	2	-3.0	F/A	Female	2597	51.7	163.2	36.5	0.9
16	9	2015/01-2016/01	2	-1.0	F/A	Female	2524	51.7	164.2	36.4	0.9
17	10	2017/01-2018/01	2	1.0	F/A	Female	2405	51.6	165.1	37.6	0.9
18	12	2021/08-2023/08	2	5.7	F/A	Female	3010	51.5	166.4	38.3	0.8

The following code and output are newly calculated sample values for several $M_{FA}:F_{FA}$ ratios, demonstrating application of Part 3 under “Method Definition,” SAPOE 2020₂ using two-year cycle data for Flight Attendants.

```
fmtWTLB( sapoeNHANES( FlightAttendants ) %>% select(cycle,gender,W) %>%
  pivot_wider(names_from=c(gender),values_from=W,names_prefix='W_FA_') %>%
  mutate( W_FA_50=F13( 50,50, W_FA_Male,W_FA_Female ),
    W_FA_99=F13( 99, 1, W_FA_Male,W_FA_Female ) ) , Cycles2 )
```

	cycle	label	duration	trendYear	W_FA_Male	W_FA_Female	W_FA_50	W_FA_99
1	3	2003/01-2004/01	2	-13.0	189.2	158.7	174.0	188.9
2	4	2005/01-2006/01	2	-11.0	192.2	160.1	176.2	191.9
3	5	2007/01-2008/01	2	-9.0	190.0	160.3	175.2	189.7
4	6	2009/01-2010/01	2	-7.0	192.4	161.1	176.8	192.1
5	7	2011/01-2012/01	2	-5.0	190.4	161.9	176.2	190.1
6	8	2013/01-2014/01	2	-3.0	191.7	163.2	177.4	191.4
7	9	2015/01-2016/01	2	-1.0	192.5	164.2	178.4	192.3
8	10	2017/01-2018/01	2	1.0	196.6	165.1	180.9	196.3

9 12 2021/08-2023/08 2 5.7 193.3 166.4 179.8 193.0

The output above includes all NHANES data points used in the analysis that follows.

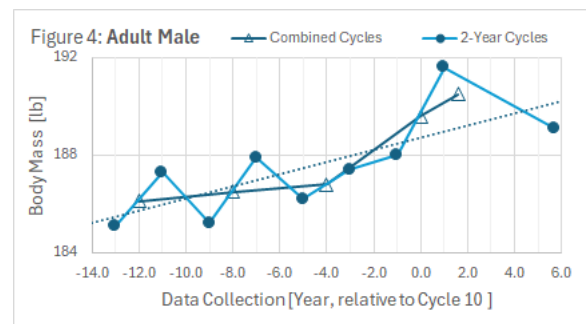
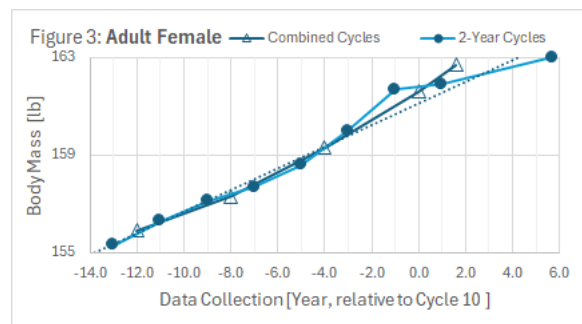
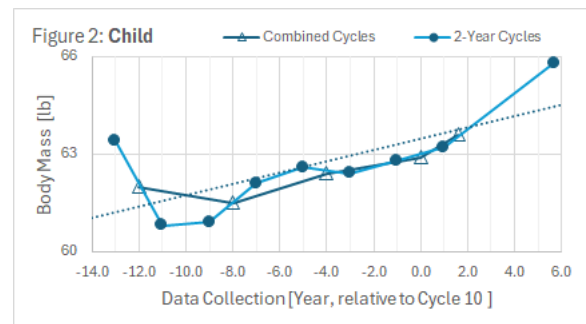
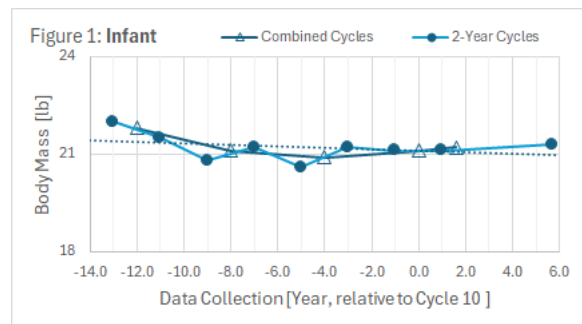
ANALYSIS AND DISCUSSION

The most significant change from SAPOE 2020 to SAPOE 2024 is the use of two-year cycle 12 after Covid while combined cycles were used prior to Covid. Each is consistent with Analytic Guidelines, but the difference merits examination.

To compare these, Figure 1 through Figure 4 below plot the average values computed using two-year cycle data and those computed using combined cycles, against the relative time scale defined above as **trendYear**. For visual reference only, the dotted line in each Figure is a least-squares trend line for the set of average values using two-year cycles, beginning with cycle 3 and including cycles 10 and 12.

There is significantly more variation between average values from the two-year cycles than from the combined cycles. In every category, the combined-cycle values are consistent with the depicted trend. Most of the values from the combined cycles are closer to the trend line calculated from the two-year cycle values than the two-year values themselves. This makes it clear that the most significant effect of using combined cycles was to reduce variation between cycles. Returning to the use of combined cycles may have similar value if Analytic Guidelines in the future defines an appropriate method of combining data post-Covid, as it does for combining pre-Covid data.

Survey-Derived Average Values for Body Mass of Passengers by Category



Finally, despite an overall increasing trend, the average value for Adult Male decreased from cycle 10 to cycle 12. Since in three of the four categories the average value has changed in the opposite direction of the overall trend multiple times in the past, this decrease is consistent with past variation and does not itself suggest an error.

UPDATED SUMMARY VALUES

The following tables summarize results found under “Method Definition,” SAPOE 2020₂ applied to NHANES combined cycles from 2003 to 2018 and the two-year cycle for August 2021–August 2023.

Table 2024-3: Survey-Derived Average Passenger Weights from NHANES Data (LB)

NHANES Cycle	Adult Male			Adult Female			Child			Infant		
	n_f	W_M	σ	n_f	W_F	σ	n_f	W_C	σ	n_f	W_I	σ
2003-2006	5940	186.1	36.1	6321	155.9	35.1	4087	62.0	26.4	1494	21.8	5.1
2007-2010	6291	186.5	37.3	6435	157.3	35.3	4156	61.5	26.7	1382	21.1	4.9
2011-2014	5879	186.8	37.1	6125	159.3	36.4	4303	62.4	29.9	1167	20.9	5.2
2015-2018	5689	189.6	40.5	5937	161.6	37.2	3692	62.9	26.3	1118	21.1	5.0
2021/08-2023/08	2931	189.1	39.6	3476	163.0	38.6	1439	65.8	29.1	230	21.3	4.5

Table 2024-4: Survey-Derived Average Crewmember Weights from NHANES Data (LB)

NHANES Cycle	Pilot, Male			Pilot, Female			Flight Attendant, Male			Flight Attendant, Female		
	n_f	$W_{P,M}$	σ	n_f	$W_{P,F}$	σ	n_f	$W_{FA,M}$	σ	n_f	$W_{FA,F}$	σ
2003-2006	2766	192.5	33.8	3136	162.1	36.5	4080	190.6	34.2	4400	159.3	35.0
2007-2010	3650	193.0	35.5	3893	162.8	36.2	5133	191.3	35.6	5403	160.6	35.2
2011-2014	3513	193.1	35.1	3709	165.9	37.4	4724	191.1	35.0	4986	162.8	35.9
2015-2018	3288	196.2	39.5	3631	167.2	38.9	4626	194.5	38.7	4928	164.7	37.1
2021/08-2023/08	1565	195.8	38.6	1979	171.0	41.0	2461	193.3	37.8	3010	166.4	38.3

Table 2024-5: Adjusted Body Weights for Sample M:F Ratios (LB)

NHANES Cycle	[W_x] Not Adjusted for Infants						[W_x] Adjusted for Infants	
	Pilots		Flight Attendants		Adults	Adults	Adults	Adults
M:F	50:50	99:1	50:50	99:1	NHANES	50:50	NHANES	50:50
2003-2006	177.3	192.2	175.0	190.3	170.5	171.0	171.3	171.7
2007-2010	177.9	192.7	175.9	190.9	171.6	171.9	172.3	172.7
2011-2014	179.5	192.8	176.9	190.8	172.7	173.0	173.3	173.6
2015-2018	181.7	195.9	179.6	194.2	175.3	175.6	175.9	176.2
2021/08-2023/08	183.4	195.6	179.8	193.0	175.8	176.0	176.3	176.6

REFERENCES

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), National Health and Nutrition Examination Survey Data, Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 1999–2023, www.cdc.gov/nchs/nhanes.

Brief Overview of Sample Design, Nonresponse Bias Assessment, and Analytic Guidelines for NHANES August 2021–August 2023, September 20, 2024, accessed September 24, 2024.

<https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/overviewbrief.aspx?Cycle=2021-2023>

T. C. Chen, J. Clark, M. K. Riddles, L. K. Mohadjer, T. H. I. Fakhouri, National Health and Nutrition Examination Survey, 2015–2018: Sample Design and Estimation Procedures, National Center for Health Statistics, Vital Health Stat 2(184), 2020.

Federal Aviation Administration, Advisory Circular 120-27 “Aircraft Weight and Balance Control”: (original) 1968; 120-27C 1995; 120-27D 2004; 120-27E 2005; 120-27F 2019.

C. D. Fryar, D. Kruszon-Moran, Q. Gu, C. L. Ogden. Mean body weight, height, waist circumference, and body mass index among adults: United States, 1999–2000 through 2015–2016, National Health Statistics Reports; no 122. Hyattsville, MD: National Center for Health Statistics. 2018. <https://www.cdc.gov/nchs/data/nhsr/nhsr122-508.pdf>

T. Lumley. Complex sampling and R, University of Washington Biostatistics and “useR” Conference, Rennes, France, 2009. <http://faculty.washington.edu/tlumley/tutorials/survey-user.pdf>

National Health and Nutrition Examination Survey (NHANES): Anthropometry Procedures Manual, January 2017.

National Health and Nutrition Examination Survey: Analytic Guidelines, 2011–2014 and 2015–2016, December 14, 2018.

B. Stierman, J. Afful, M. D. Carroll, T. C. Chen, O. Davy, S. Fink, et al., National Health and Nutrition Examination Survey 2017–March 2020. Prepandemic data files—development of files and prevalence estimates for selected health outcomes. National Health Statistics Reports; no 158, Hyattsville, MD: National Center for Health Statistics, 2021. DOI: <https://dx.doi.org/10.15620/cdc:106273>. (<https://stacks.cdc.gov/view/cdc/106273>)

A. L. Terry, M. M. Chiappa, J. McAllister, D. A. Woodwell, J. E. Graber. Plan and operations of the National Health and Nutrition Examination Survey, August 2021–August 2023, National Center for Health Statistics, Vital Health Stat 1(66), 2024. DOI: <https://dx.doi.org/10.15620/cdc:151927>. (<https://stacks.cdc.gov/view/cdc/151927>)

APPENDIX A: CONSIDERING THE FUTURE

In many operational planning and design contexts, it would be useful to forecast the survey-derived average for body mass in effect at some future date. The value in effect is typically the most recently published value at the time in question. It may not be the value that will be determined in analysis following completion of the data-collection period ongoing at the time in question, especially since that value cannot yet be known while data collection is not yet complete. (At the planned midpoint of the cycle 11 data-collection period, in January 2020, it was not yet known that the cycle would not be completed.) However, it is reasonable to expect that the survey-derived average for body mass in effect might grow at the same rate as the average for body mass calculated according to the method used above. In other words, a useful forecast might be a solution to a differential equation describing the empirical data, but from a distinct initial condition.

In Figure 1 through Figure 4 above, it appears upon inspection that a simple constant slope is a good model for growth for each passenger category over time, despite short-term variation. Logically, body mass cannot grow in perpetuity, so future analysis should look for an eventual change in growth rate. However, as noted above, in each category the change from cycle 10 to cycle 12 is consistent with past variation, so no change in derivative beyond past variation is evident yet. Using this model, all forecasts have the simple form:

$$S = \frac{dW}{dT} \quad (2024-1)$$

where S is a constant growth rate determined from past data, and thus:

$$W = W_0 + S \cdot (T - T_0). \quad (2024-2)$$

The choice of initial condition (T_0, W_0) is of interest.

It is common to simplify forecasting by using a single effective passenger weight, called $[W_X]$ in SAPOE 2020. Such a simplified model is obtained by assuming all passengers are 50:50 Adults and differentiating Formula (2020-12) to obtain:

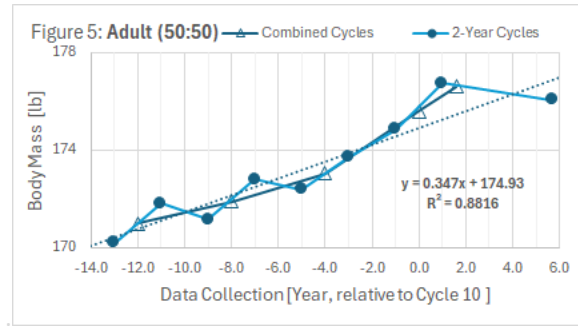
$$[W_X] = \frac{1}{2}(W_M + W_F) \quad (2020-12)$$

$$S = \frac{d}{dT} [W_X] = \frac{1}{2} \left(\frac{dW_M}{dT} + \frac{dW_F}{dT} \right). \quad (2024-3)$$

Values for dW_M/dT and dW_F/dT can be determined separately for Male and Female body mass from Figure 3 and Figure 4 above, respectively, or from $[W_X]$ directly as shown in Figure 5. As above, the dotted line in Figure 5 is a least-squares trend line for the set of average values computed using two-year cycles, beginning with cycle 3 and including cycles 10 and 12. The equation depicted in Figure 5 is not the final proposed forecast model, but does show:

$$S = 0.347 \text{ lb/year} \quad (2024-4)$$

(The precision for S is limited by body-mass precision and by rounding dates to tenths of a year.)



Alternatively, the value of S could be determined by differentiating Formula (2020-10) or Formula (2020-11), instead of Formula (2020-12), and applying operator-specific data for N_i , N_G , N_F , N_M , and N_A defined in those formulas. If separate forecast models are needed for Infant and Child passengers, a separate value for S could also be determined from Figure 2 above and used in the same way as demonstrated here. For Infants, Figure 1 above suggests assuming a derivative of zero, or a constant value of 21 lb.

The initial condition might be selected from any published survey-derived average for body mass. The most relevant cases are averages based on NHANES data, for which the authors have reconstructed the following history.

1. The first known publication of passenger weights based on NHANES is FAA AC 120-27D, dated August 11, 2004. This gave a “Summer Passenger” value of 190 lb, summer clothing weight of 5 lb, and carry-on allowance of 16 lb, implying a body mass of 169 lb for $[W_X]$. (The “Winter Passenger” was 5 lb heavier due to assumed winter clothing of 10 lb, giving the same body mass.)
2. FAA AC 120-27E was published June 11, 2005, with the same values as AC 120-27D.
3. Correspondence shared with the authors, dated July 2006, shows an analysis of the 2003–2004 cycle data and a comparison to past analysis of the 1999–2000 cycle data. Both analyses appear to be consistent with the SAPOE 2020 method, but the authors have not found any publication of these results.
4. Since AC 120-27F was published without values, the next known published values are in SAPOE 2020, which gave $[W_X]$ value for a 50:50 Adult of 175.6 lb. The late 2020 publication date rounds to 2021.0.
5. The calculations above in this paper show the most recent $[W_X]$ value for a 50:50 Adult of 176.0 lb, which is labelled “SAPOE 2024.” The late 2024 publication date rounds to 2025.0.
6. In accordance with AC 120-27F, it is anticipated that the SAPOE 2024 value will remain in effect for at least 36 months after it is published.

This history is visualized in the following table and plot. Like in Figure 1 through Figure 5, the horizontal axis time values used are continuous decimal years relative to a recent date. The reference date is chosen for convenience as the publication of SAPOE 2020, when NHANES cycle 2015–2018 was the most recently available combined cycle.

Initial Condition	T_0	Relative Year	W_0
AC 120-27D (August)	2004.6	-16.3	169.0
AC 120-27E (June)	2005.4	-15.5	169.0
SAPOE 2020 (December)	2021.0	0.0	175.6
SAPOE 2024 (December)	2025.0	4.0	176.0
SAPOE 2024 +36 months	2028.0	7.0	176.0

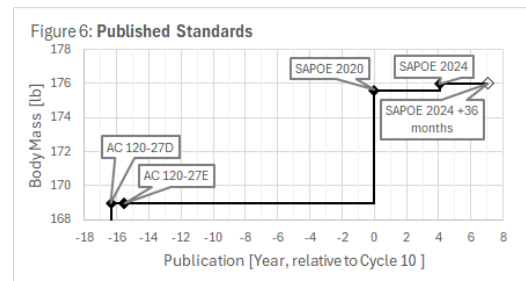
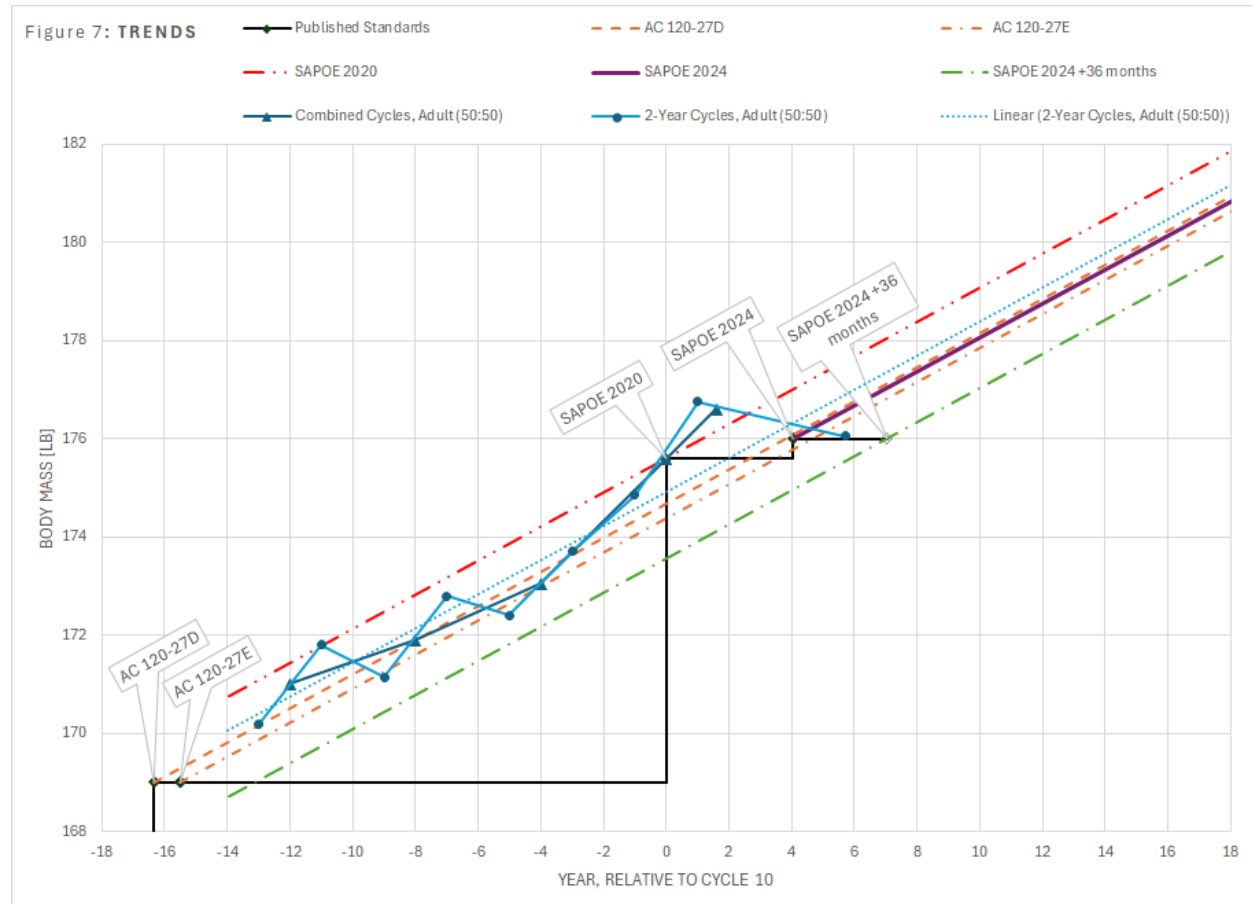


Figure 7 shows forecast models defined by Formula (2024-2) for each of the initial conditions considered. In addition, two-year cycle average values from Figure 5 are plotted on the same relative time scale, effectively comparing the most recently available data to the published values.

Comparing forecast models for the oldest and newest initial conditions highlights the consistency of these models over the entire duration of NHANES use. The model projected forward from the most recent initial condition is from SAPOE 2024 (shown in solid purple). The forward projection from the oldest initial condition, AC 120-27D (shown in orange short dashes), differs by less than 0.1 lb. The forward projection from initial condition AC 120-27E (shown in orange short-dash-dots) is just 0.2 lb lower.

The trend line from the overlay of two-year cycle data (shown in blue dots) is parallel to the model, due to the model definition, and is just 0.2 lb higher than the forecast from SAPOE 2024.

The highest forecast model (shown in red long-dash-double-dots) is from SAPOE 2020 and is 1.0 lb higher than from SAPOE 2024. The lowest forecast model (shown in green long-dash-dots) is from the anticipated effectivity of SAPOE 2024 for 36 months after publication and is 1.0 lb lower. Together these bound the variation from all but one average value from each set of data cycles, providing a convenient, if coarse, estimate of future variation.



Balancing the possibilities evaluated, the most reasonable forecast is thus from the initial condition in SAPOE 2024:

$$[W_x] = 176.0 \text{ lb} + 0.347 \frac{\text{lb}}{\text{year}} \cdot (T - 2025.0) \pm 1.0 \text{ lb} \quad (2024-5)$$

Care must be taken to evaluate the assumptions made in defining this forecast and whether these are applicable to any scenario to which the forecast is applied. Such a decision, and its outcome, is strictly the responsibility of the individual who applies it.

No attempt has been made here to determine how far into the future this forecast might be useful.

It is hoped that this forecast has value for the design of airplanes and components intended for future use. In addition, this forecast might be useful if future NHANES data collection or analysis is disrupted or delayed, or to reduce the effect of the greater variation seen when using two-year cycles compared to past use of combined cycles.

APPENDIX B: UPDATED SAMPLE CODE

Running this script requires the R language core, which is freely available for a variety of platforms from [\[www.r-project.org\]](http://www.r-project.org). Use is simplified by a graphical interface, such as RStudio Desktop (available as of this writing at no cost from [\[rstudio.com\]](http://rstudio.com)). Packages required for installation are [dplyr](#), [foreign](#), [survey](#), and [tidyr](#).

A small font and large page are used to assist copying and pasting from this document into plain text.

```
#####
# Script for "CONTINUED NHANES USE FOR STANDARD AVERAGE BODY WEIGHT"
# (C) 2024 Society of Aircraft Performance and Operations Engineers (SAPOE)
# Modified from code included in 'NHANES USE FOR STANDARD AVERAGE BODY WEIGHT'
# (C) 2020 Society of Aircraft Performance and Operations Engineers (SAPOE)
#
# SAPOE is a member-based organization promoting the safety and efficiency of
# flight through knowledge of aircraft performance and weight and balance
# principles.
#
# Except for brief quotations with appropriate citation, copies may be made
# and distributed only of the complete document including cover page and
# this disclaimer.
#
# SAPOE, the authors, and reviewers assume no liability whatsoever and make
# no warranty of any kind. Anyone who uses this report, or the source code
# and data contained in it, is solely responsible for their own operations
# and for any outcome of such use.
#
#-----
# These libraries need to be installed, along with the base R language
#
# The 'dplyr' library simplifies and improves readability, especially with
# 'bind_rows()' to append rows of data to an existing data frame,
# 'mutate()' to add calculated columns to an existing data frame,
# 'group_by()' to define categories within the data, and
# 'summarise()' (note spelling) to calculate summary statistics
library(dplyr)
#
# The 'tidyr' library simplifies and improves readability with
# pivot() and pivot_wider() for combining and summarizing data tables
library(tidyr)
#
# The 'foreign' library reads SAS Transport(XPT) files
library(foreign)
#
# The 'survey' library allows use of multi-stage survey design parameters
# to calculate weighted mean and standard error estimates recommended by
# Analytic Guidelines and NHANES tutorials
library(survey)
#
#-----
# Define a helper to keep all the meta data defining each cycle consistent
# This must be updated after cycle 12 for future NHANES conventions.
nhanesCycleDefinitions <- function( combine )
{
  # NHANES defined values for all the cycles with consistent data definitions
  # 66 is a pre-covid set combining cycle 10 and the discontinued cycle 11
  allCycles = c(3:10,66,12)
  if( !combine ) # use cycles as-is, except exclude the pre-covid combined set
  {
    cn <- allCycles[allCycles != 66 ] # exclude the 3.2-year pre-covid cycle
    cl <- rep( 2, length(cn) ) # all remaining cycles are two years
    dataMod <- function(df)
    {
      filter( df, 66 != SDO5RVYR ) %>% # remove all data from excluded cycles
      mutate(
        cycle = SDO5RVYR, # don't combine cycle numbers
        svywt = WTMEC2YR # don't combine design parameters
      )
    }
  } else # combine cycles only as recommended in Analytic Guidelines
  {
    cn <- c(4,6,8,10,66) # even pairs pre-covid (TBD post-covid)
    cl <- ifelse( 66 == cn, 3, 2, 4 ) # each pair is four years, except 66
    dataMod <- function(df)
    {
      filter( df, 12 != SDO5RVYR ) %>% # remove all data from excluded cycles
      mutate(
        cycle = ifelse( 66 == SDO5RVYR, 66, 2*ceiling(SDO5RVYR/2) ),
        svywt = ifelse( 66 == SDO5RVYR, WTMECPRP, WTMEC2YR/2 )
      )
    }
  }
  CE <- 2000+2*(cn-1) # a nominal end year, used below
  CS <- ifelse( 10 == cn, sprintf("%01-%01", CE-ceiling(cl-1),ce),
    ifelse( 66 == cn, '2017-01-2020/03',
    ifelse( 12 == cn, '2021-08-2023/08',
    'undefined' )
    )))
  # Cycle periods are no longer constant, so trend analysis vs. time must be
  # adjusted to reflect variation in both the start and duration of the data
  # collection. CDC uses the midpoint in decimal years. See References
  # '... Analytic Guidelines for NHANES August 2021-August 2023'
  # and https://www.cdc.gov/nchs/data/series/sr_02/sr_02-190.pdf
  cy <- ifelse( 10 == cn, CE - floor((cl-1)/2), # x.0 for older cycles
    ifelse( 66 == cn, 2018.6, # Mid-July 2018 for Jan 2017 to Mar 2020
    ifelse( 12 == cn, 2022.7, # August 2022 for Aug 2021 to Aug 2023
    NAN # a numeric value for 'Not Defined'
    )))
  # To compare to publications, SAPOE uses CDC recommended time, but relative to
  # the last 4-year cycle when the Dec 2020 paper was published, '2015-2018'.
  ty <- cy - cy[cn==10] # relative to cycle 10
  if( !combine ) ty <- ty+1 # the combined cycle centers 1 year earlier
  list( combined = combine, dataModifier = dataMod,
    metadata = data.frame( cycle=cn, label=CS, duration=cl, trendYear=ty )
  )
}
#
#-----
# Define a helper to keep all the source files download/import consistent
# Files naming is superficially like cycle numbering, but differs slightly
# since cycle 10 and may vary in the future independently of survey design.
importRawDataFiles <- function( cn ) # raw data cycles to import
{
  # the urls all follow a pattern, except for the combined pre-covid cycle
  y <- ifelse( 66 == cn, 2018, 2000+2*(cn-1) )
  p <- sprintf("https://www.cdc.gov/nchs/nhanes/%s-%s",y-1,y)
  D <- ifelse( 66 == cn, 'P_DEMO.XPT', sprintf("DEMO_%s.XPT",LETTERS[cn]) )
  B <- ifelse( 66 == cn, 'P_BMX.XPT', sprintf("BMX_%s.XPT",LETTERS[cn]) )
  allData <- data.frame(NULL)
  for( i in 1:length(cn) )
  {
    # download the files only if they don't already exist
    if( !file.exists(D[i]) ) download.file( paste0(p[i],D[i]), D[i], mode='wb' )
    demo <- select( read.xport(D[i]),
      # only needed columns from the DEMO file here
      any_of(c('SEQN', 'SDO5RVYR', 'RIDAGEYR', 'RIAGENDR', 'RIDEXPRG',
        'SDMVSTRA', 'SDMVPSU', 'WTMEC2YR', 'WTMECPRP'))
    )
    # add missing columns so all cycles have the same columns
    if( is.na(match('WTMEC2YR',names(demo))) ) demo <- mutate(demo,WTMEC2YR=NA)
    if( is.na(match('WTMECPRP',names(demo))) ) demo <- mutate(demo,WTMECPRP=NA)
    # exclude participants without an exam record before joining
    demo <- subset( demo, !(is.na(WTMEC2YR) & is.na(WTMECPRP)) )
    # download the files only if they don't already exist
    if( !file.exists(B[i]) ) download.file( paste0(p[i],B[i]), B[i], mode='wb' )
    bmx <- select( read.xport(B[i]),
      # only needed columns from the BMX file here
      c('SEQN','BMXWT','BMIXWT')
    )
    allData <- bind_rows(allData,left_join( demo, bmx, by = 'SEQN' ))
  }
  return(allData)
}
#
#-----
# Create a copy with calculated columns added (preserve Rawdata)
# Don't remove or delete any rows or survey design parameters are lost.
enrichDataset <- function( rawdata, cycleDefs )
{
  richdata <- mutate(rawdata,
    # 'one' is needed to count records
    one = 1,
    # make it easy to work in US units
    WTLB = BMXWT / 0.45359237, # kg to lb

```

```

    # use cut(right=FALSE) to build age intervals that include the lower
    # and exclude the upper boundary
    #
    # for FAA, Infants have not yet reached 2nd birthday.
    # Children have not yet reached 13th birthday.
    faaAge = cut(RIDAGEYR, breaks=c(-Inf, 2, 13, Inf),
      right=FALSE, labels=c( 'Infant', 'Child', 'Adult' )),
    gender=factor(ifelse(RIDAGEYR>12,RIAGENDR,0),labels=c('None','Male','Female')),
    # for comparison to CDC/NHSR, Adults are age 20 and over
    cdcAge = cut(RIDAGEYR, breaks=c(-Inf, 20, Inf),
      right=FALSE, labels=c( 'Child', 'Adult' )),
    # columns used to "filter" the data, like 'inAnalysis' in tutorials
    # NOTE: use subset(), not filter(), to "filter" or design parameters are lost
    NotMissing = ( !is.na(BMXWT) ),
    NotClothed = ( NotMissing & (is.na(BMWIT) | 3!=BMWIT) ),
    NotPregnant = ( NotMissing & (is.na(RIDEXPRG) | 1!=RIDEXPRG) )
    # for RIDEXPRG, 1=Yes, 2=No, 3=Unknown, Missing => No
  )
  cycleDefs$dataModifier( richdata ) %>% subset( NotMissing )
}
#
#-----
# Define a function to get the weighted mean and standard error estimate
# using survey design parameters and join to the unweighted count.
# Based heavily on https://www.cdc.gov/nchs/nhanes/tutorials/samplecode.aspx
# and https://www.cdc.gov/nchs/data/tutorials/08303_Pilot_R.R
# this is unmodified from 2020 and not likely to change in the future.
svyWTLB <- function(design, byFormula)
{
  # Get mean, stderr, and unweighted sample size
  c <- svyby( ~WTLB, byFormula, design, unwt=count )
  p <- svyby( ~WTLB, byFormula, design, svymean )
  select(c,-se) excludes the standard error of the count
  suppressMessages() just hides output of the implicit "by" columns
  suppressMessages( outdata <- left_join(select(c,-se), p ) )
  return(outdata)
}
#
#-----
# Define a function to get consistent output formatting
fmtWTLB <- function( results, cycleDefs )
{
  T <- right_join( cycleDefs$metadata, results, by='cycle' )
  {
    sorts <- c('cycle')
    if( !is.na(match( 'cdcAge', names(T))) ) sorts <- c('cdcAge',sorts)
    if( !is.na(match( 'faaAge', names(T))) ) sorts <- c('faaAge',sorts)
    if( !is.na(match( 'gender', names(T))) ) sorts <- c('gender',sorts)
  }
  mutate_if(T %>% arrange_at(sorts) , is.numeric,round,digits=1)
}
#
#-----
# Define a function to do exactly the METHOD defined in SAPOE's 2020 paper.
sapoeNHANES <- function( allData )
{
  # A copy of the data excluding clothed sample participants
  step0 <- allData %>% filter(NotClothed)
  # Calculate a raw weighted mean, w-bar-star (ws) using Formula (2020-1)
  step1 <- mutate(step0, ww=WTLB*svywt) %>% summarise( ws=sum(ws)/sum(svywt), .by=c(faaAge,gender,cycle) ) %>%
  # Sum weighted residuals (wr) for each row to get sigma-star (ss) using Formula (2020-2)
  step2 <- left_join(step0,step1,by=c('faaAge','gender','cycle')) %>%
  mutate( wr=svywt*(WTLB-ws)^2 ) %>% summarise( n=sum(one), ss=sqrt(sum(wr)/((n-1)/n*sum(svywt))), .by=c(faaAge,gender,cycle,ws) )
  # fmtWTLB(step2 %>% select(cycle,faaAge,gender,n,ws,ss,'te%')) # for debugging
  # Filter by excluding values more than twice ss from the raw mean
  step3 <- left_join(step0,step2,by=c('faaAge','gender','cycle')) %>%
  filter( 2 >= abs((WTLB-ws)/ss) )
  # Calculate population totals, after filter, for use in Formula (2020-7)
  NN <- step3 %>% summarise( Nt=sum(svywt), .by=cycle )
  # Calculate a weighted mean, W using Formula (3) and then Formula (2020-7)
  step4 <- mutate(step3, wr=WTLB*svywt) %>% summarise( W=sum(wr)/sum(svywt), Nc=sum(svywt), .by=c(faaAge,gender,cycle) ) %>%
  left_join(NN,cycle) %>% summarise( NP=100*Nc/Nt, .by=c(faaAge,gender,cycle,W) )
  # Sum weighted residuals (wr) for each row to get sigma (S) using Formula (2020-4)
  step5 <- left_join(step3,step4,by=c('faaAge','gender','cycle')) %>%
  mutate( wr=svywt*(WTLB-W)^2 ) %>% summarise( n=sum(one), S=sqrt(sum(wr)/((n-1)/n*sum(svywt))), .by=c(faaAge,gender,cycle,W,NP) ) %>%
  mutate( 'te%'=196*S/W/sqrt(n) )
  step5 %>% select(cycle,faaAge,gender,n,NP,W,S,'te%')
}
#
#-----
# apply sample M:F ratios using Formula (2020-13)
F13 <- function(M,F,W,M,W,F) { ( M*W_M + F*W_F )/(M+F) }
#
#-----
# set the working directory to where the raw data is or will be saved
setwd( 'C:/NHANES/' )
RawData <- importRawDataFiles(c(3:10,66,12))
Cycles2 <- nhanesCycleDefinitions(FALSE)
WTLB2YR <- enrichDataset( RawData, Cycles2 )
# Define survey design parameters for overall dataset, per NHANES Tutorials
NHANES2<-svydesign(data=WTLB2YR, id=SDMVPSU, strata=SDMVSTRA, weights=svywt, nest=TRUE)
# confirm definitions above here are correct for full NHANES by comparing to
# https://www.cdc.gov/nchs/data/nhsr/nhsr122-508.pdf
svyWTLB( subset( NHANES2, NotPregnant ), ~cycle+cdcAge+gender ) %>% fmtWTLB(Cycles2) %>%
  filter( cdcAge=='Adult' & cycle<10 ) # NHSR Table 2 has Adults in these years
#
#-----
# confirm definitions above here are correct for FAA/SAPOE definitions by calculating
# weights for past cycles and comparing output calculations shown in 2020.
Cycles4 <- nhanesCycleDefinitions(TRUE)
WTLB4YR <- enrichDataset( RawData, Cycles4 )
Passengers <- sapoeNHANES( WTLB4YR )
fmtWTLB( Passengers, Cycles4 )
#
#-----
# Then use the same process to recalculate averages from 2-year cycles
Passengers <- sapoeNHANES( WTLB2YR )
fmtWTLB( Passengers, Cycles2 )
Wx <- left_join(
  Passengers %>% select(cycle,faaAge,gender,W) %>%
  pivot_wider(names_from=c(faaAge,gender),values_from=W, names_prefix='W_'),
  Passengers %>% select(cycle,faaAge,gender,NR) %>%
  pivot_wider(names_from=c(faaAge,gender),values_from=NR,names_prefix='N_'),
  .by=c('cycle')
) %>% rename( N_Infant=N_Infant_None, W_Infant=W_Infant_None ) %>%
  mutate(
    W_Adult_NHANES = F13( N_Adult_Male, N_Adult_Female, W_Adult_Male, W_Adult_Female ),
    W_Adult_S050 = F13( W_50, W_Adult_Male, W_Adult_Female ),
    dw_Infant_NHANES = N_Infant / ( N_Adult_Female + N_Adult_Male ) * W_Infant,
    'W_Adult_NHANES+Infant' = W_Adult_NHANES + dw_Infant_NHANES,
    'W_Adult_S050+Infant' = W_Adult_S050 + dw_Infant_NHANES,
  )
Wx %>% select(cycle, W_Adult_NHANES,W_Adult_S050,dw_Infant_NHANES) %>% fmtWTLB(Cycles2)
Wx %>% select(cycle, 'W_Adult_NHANES+Infant','W_Adult_S050+Infant') %>% fmtWTLB(Cycles2)
#
#-----
# Pilots <- subset( WTLB4YR, 23<=RIDAGEYR & 65<=RIDAGEYR ) %>% mutate(faaAge='Pilot')
# fmtWTLB( sapoeNHANES( Pilots ), Cycles4 )
Pilots <- subset( WTLB2YR, 23<=RIDAGEYR & 65<=RIDAGEYR ) %>% mutate(faaAge='Pilot')
fmtWTLB( sapoeNHANES( Pilots ), Cycles2 )
fmtWTLB( sapoeNHANES( Pilots ) ) %>% select(cycle,gender,W)
  pivot_wider(names_from=c(gender),values_from=W,names_prefix='W_P_') %>%
  mutate( W_P_50=F13( W_50,W_P_Male,W_P_Female ),
    W_P_99=F13( W_99, W_P_Male,W_P_Female ) ), Cycles2 )
#
#-----
# FlightAttendants <- filter( WTLB4YR, 21<=RIDAGEYR ) %>% mutate(faaAge='F/A')
# fmtWTLB( sapoeNHANES( FlightAttendants ), Cycles4 )
FlightAttendants <- filter( WTLB2YR, 21<=RIDAGEYR ) %>% mutate(faaAge='F/A')
fmtWTLB( sapoeNHANES( FlightAttendants ), Cycles2 )
fmtWTLB( sapoeNHANES( FlightAttendants ) ) %>% select(cycle,gender,W)
  pivot_wider(names_from=c(gender),values_from=W,names_prefix='W_FA_') %>%
  mutate( W_FA_50=F13( W_50, W_FA_Male,W_FA_Female ),
    W_FA_99=F13( W_99, W_FA_Male,W_FA_Female ) ), Cycles2 )
#####
```

APPENDIX C: SAPOE 2020 (Reprint)

A reprint of the complete reference 'SAPOE 2020' follows this final page.

NHANES USE FOR STANDARD AVERAGE BODY WEIGHT

14 DEC 2020

ABSTRACT

For use in aircraft weight and balance, a method to determine standard average weights for body mass of passengers and crew is defined. The method uses data from the US Centers for Disease Control (CDC) National Health and Nutrition Examination Survey (NHANES) in a manner consistent with guidance from the Federal Aviation Administration (FAA), especially Advisory Circular (AC) 120-27 "Aircraft Weight and Balance Control". Justification for aspects of the definition is discussed. Example values using data from 2003 through 2018 are shown. Results are applicable to Operation Specifications (OpSpecs) approval using survey-derived average weight values.

AUTHORS

William Yingling, Senior Engineer Operations, JetBlue
Charles Ostick, Senior Flight Operations Engineer, Alaska Airlines

REVIEW COMMITTEE

Kristine Henning, Senior Principal Engineer, Performance Engineering, Delta Air Lines (Chair)
Mirjeta Beqiri, Ph.D., Professor of Operations Management, Gonzaga University
Michael Byham, Director Operations Engineering, American Airlines
Ryan Herzog, Ph.D., Associate Professor of Economics, Gonzaga University
Rick Martin, Acting Sr. Mgr. Aircraft Performance Engineering, Sun Country Airlines
Cary Robins, President, American Aeronautics
Donald Tiemann, Performance Engineer, Engineering, Allegiant Airlines
Gail Zittel, Senior Engineer - Weight and Balance, The Boeing Company

OWNERSHIP & DISCLAIMER

This document remains the sole property of the Society of Aircraft Performance and Operations Engineers (SAPOE).

SAPOE is a member-based organization promoting the safety and efficiency of flight through knowledge of aircraft performance and weight and balance principles.

Except for brief quotations with appropriate citation, copies may be made and distributed only of the complete document including cover page and this disclaimer.

SAPOE, the authors, and reviewers assume no liability whatsoever and make no warranty of any kind. Anyone who uses this report, or the source code and data contained in it, is solely responsible for their own operations and for any outcome of such use.

PROBLEM STATEMENT

US aviation regulations and Operations Specifications (OpSpecs A097, A098, and A099) require an approved weight and balance control system based on average, assumed, or estimated weight. For passengers and crew, establishing an average body weight by sample weighing is disruptive to operations and is viewed unfavorably as an intrusion on privacy by the US traveling public.

Since 1968, the Federal Aviation Administration (FAA) has established standard average weights for airline passengers and crew and published them in Advisory Circular (AC) 120-27 "Aircraft Weight and Balance Control". Since revision D of AC 120-27, published in 2004, these standard average weights have been based on data published by US Centers for Disease Control (CDC) National Health and Nutrition Examination Survey (NHANES).

Revision F of AC 120-27, published in May 2019, deletes the standard average weight values. FAA Policy Notice 8900.551, published June 11, 2020, makes it mandatory for US operators to establish new weight values consistent with AC 120-27F unless actual weights are used.

AC 120-27F still allows operators to use NHANES to establish new average weights for passengers and crew. However, summary publications by NHANES, including those referred to in FAA guidance, conflict with FAA definitions of terms such as "Child" and "Adult". Furthermore, CDC summary statistics exclude persons, such as pregnant women, who must be considered in average passenger weights determined using AC 120-27F. Thus, these summaries cannot be used directly. NHANES raw data, which does not make these exclusions, must be used. AC 120-27F does not define a specific acceptable method to apply the NHANES raw data.

Therefore, a method to establish average weights for passengers and crew using CDC NHANES raw data, consistent with FAA requirements and definitions, is needed.

SYSTEMS CONSIDERATIONS

The ability to apply new values in existing weight and balance control systems is also a critical requirement. A general solution for novel system designs is beyond the scope of this paper, but certain aspects of existing systems are well known and should be considered.

Age and gender categories are consistently defined in all revisions of AC 120-27, and these categories are a fundamental assumption in existing weight and balance control systems. These systems often access only the number of persons in each category and may be blocked from accessing more specific personal data even if it is available to an operator's other systems.

Weight and balance control systems differ in their ability to assign average weights to each age and gender category. For example, because past guidance did not assign a weight to infants, some systems can only account for this weight by adjusting the average value used for adults, while others can assign non-zero weight to the number of infants on board.

Changing the design of these systems can introduce significant operational and safety risk and may not be possible to accomplish in the time frame required for initial implementation of new weight values.

Therefore, the needed method must be compatible with existing weight and balance control systems and consistent with past design assumptions.

METHOD DEFINITION

A method to establish average body weights, meeting the requirements and considerations given, is defined here. No claim is made that this is the only valid method, especially if different requirements and considerations apply. Even within this method, different systems considerations may yield different resulting values.

The method is defined in three parts. The first part defines mathematical steps to determine average weights for a set of FAA-defined passenger categories. The second part defines adjustments that may be required to apply the average weights in a given operator's control systems. The third part adapts the method from Parts 1 and 2 to determine average weights for crewmembers.

How to efficiently accomplish the steps and calculations defined follows in DISCUSSION, along with justification of certain aspects of the method and comparison to other statistical techniques.

Part 1, Survey-derived Average Passenger Body Weights from NHANES Data

As allowed by AC 120-27F, values are determined from NHANES raw data. The minimum essential data is:

Table 1: Minimum Required NHANES Data Fields		
Field Name	Data File	Field Description
BMXWT	BMX	Measured body mass in kilograms
BMIWT	BMX	Body mass comment: 1 = Could not obtain; 2 = Exceeds scale capacity; 3 = Respondent Clothed; 4 = Medical appliance included
RIAGENDR	DEMO	Gender: 1 = Male; 2 = Female
RIDAGEYR	DEMO	Age in years at time of examination.
WTMEC2YR	DEMO	Full sample 2-year "weighting factor" to be applied to measured data to relate it to the larger US noninstitutionalized civilian population.
SEQN*	DEMO & BMX	Respondent sequence number: Functions as a unique record key to tie together data from DEMO and BMX files.

**SEQN is required only to join records in the BMX file to associated records in the DEMO file.*

As recommended in NHANES Analytic Guidelines and Tutorial, consecutive 2-year cycles are combined into a 4-year cycle. (The most recent 4-year cycle combines 2015-2016 and 2017-2018.) In the formulas below, WTMEC is the 4-year weighting value, which for 2003 – 2018 is half of the WTMEC2YR value in the raw data.

Consistent with AC 120-27E, Appendix 2, Paragraph 1. c., sample data without an examination record, with a missing body mass (BMXWT) value, or coded as clothed when weighed (BMIWT=3) are excluded.

From RIDAGEYR, age categories are defined to comply with FAA definitions in AC 120-27F, and earlier revisions. The "Infant" category is defined as persons who have not yet reached their 2nd birthday. The "Child" category is defined as persons with an age greater than or equal to 2 years who have not yet reached their 13th birthday. The "Adult" category is defined as persons with an age greater than or equal to 13 years.

For FAA definitions, gender is only considered for "Adults", and is given by RIAGENDR.

For each age and gender category, i.e. “Infant”, “Child”, “Adult Female” and “Adult Male”, the following steps are calculated independently. In formulas (1) and (2), n_R is the number of non-excluded records in the data set for each category.

Step 1) A raw weighted mean (\bar{w}^*) is calculated using the formula:

$$\bar{w}^* = \frac{\sum_{i=1}^{n_R} BMXWT_i \times WTMEC_i}{\sum_{i=1}^{n_R} WTMEC_i} \quad (1)$$

Step 2) Using the raw mean, a raw weighted sample standard deviation (σ^*) is calculated using the formula:

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^{n_R} WTMEC_i \times (BMXWT_i - \bar{w}^*)^2}{\frac{(n_R - 1)}{n_R} \times \sum_{i=1}^{n_R} WTMEC_i}} \quad (2)$$

Step 3) Consistent with AC 120-27E, Appendix 2, sample data with BMXWT more than twice σ^* either greater than or less than \bar{w}^* are excluded to form a filtered data set. (See DISCUSSION for elaboration.) In formulas (3) and (4), n_f is the number records in the filtered data set for each category.

Step 4) From the filtered data set, a weighted mean (W) is calculated using the formula:

$$W = \frac{\sum_{i=1}^{n_f} BMXWT_i \times WTMEC_i}{\sum_{i=1}^{n_f} WTMEC_i} \quad (3)$$

Step 5) From the filtered data set, a weighted sample standard deviation (σ) is calculated using the formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n_f} WTMEC_i \times (BMXWT_i - W)^2}{\frac{(n_f - 1)}{n_f} \times \sum_{i=1}^{n_f} WTMEC_i}} \quad (4)$$

Step 6) Due to NHANES sample design, the tolerable error percentage (e) defined in AC 120-27F should not be used to determine confidence intervals for NHANES averages. However, if an operator is required to demonstrate these values, they may be calculated using AC 120-27F, paragraph 3.3.3:

$$e = \frac{1.96 \times \sigma \times 100}{W \times \sqrt{n_f}} \quad (5)$$

The survey-derived average passenger body weights are the W values defined for Infant, Child, Adult Female, and Adult Male, labeled W_I , W_C , W_F , and W_M respectively.

Part 2, Weight Adjustments for System-specific Implementation

For application to existing systems, further adjustments may be required. In the following definitions, $[W]$ is used to represent the adjusted value used in operational systems in place of the W value calculated as defined in Part 1.

Determining these adjustments requires definition of the following quantities:

- N_I is the total number of Infants boarded
- N_C is the total number of Children boarded
- N_A is the total number of Adults boarded
- N_F is the total number of Adults boarded, known to be Female
- N_M is the total number of Adults boarded, known to be Male

N_X is the total number of Adults boarded, not known to be Female or Male

$$N_A = N_F + N_M + N_X \quad (6)$$

Values for these quantities should be determined for the operation to which the adjusted $[W]$ will be applied. For example, they might be measured by counting on representative flights for a representative time period, or estimated from existing data. Estimates for N_F , N_M , N_G , and N_I may also be obtained from the NHANES filtered data set defined in Part 1, Step 3, by dividing the total WTMEC for each category by the total WTMEC for all categories, using the formula:

$$N = \frac{\sum_{i=1}^{n_f} WTMEC_i}{\sum_{F,M,C,I} (\sum_{i=1}^{n_f} WTMEC_i)} \quad (7)$$

Additionally:

M:F is the assumed Male to Female ratio for Adults when gender is not identified; and
 $[W_X]$ is the operationally assumed weight for an Adult when gender is not identified.

Values for M:F are sometimes assigned to an operator by the FAA. For example, "50:50" is stated in the FAA template for Operation Specifications A099, but "60:40" or other ratios may be given in other contexts. (When M:F = 60:40, M = 60 and F = 40.) If M:F is not defined by the FAA, use:

$$M:F = N_M : N_F \quad (8)$$

Adjusted values $[W]$ must ensure that total weight calculated by the operational system is equal to total weight calculated assuming M:F and the W values from Part 1, which is stated in the following identity: (9)

$$N_X [W_X] + N_M [W_M] + N_F [W_F] + N_C [W_C] + N_I [W_I] = \left(N_M + \frac{M}{M+F} N_X \right) W_M + \left(N_F + \frac{F}{M+F} N_X \right) W_F + N_C W_C + N_I W_I$$

Any set of $[W]$ values that preserve the required identity (9) is consistent with this method definition.

For a common set of system constraints, where gender is always unknown ($N_M = N_F = 0$ implying $N_X = N_A$) and infants cannot be assigned a weight (implying $[W_I] = 0$), the required identity is satisfied without adjusting Child weights (implying $[W_C] = W_C$) if the Adult weight is adjusted according to:

$$[W_X] = \frac{M W_M + F W_F}{M+F} + \frac{N_I}{N_A} W_I \quad (10)$$

Furthermore, if M:F is required to be 50:50 with $[W_I] = 0$ and $[W_C] = W_C$ then:

$$[W_X] = \frac{1}{2}(W_M + W_F) + \frac{N_I}{N_A} W_I \quad (11)$$

Note that if the operational system can assign a weight to infants and M:F = 50:50, then the required identity (9) could also be satisfied with $[W_I] = W_I$ and $[W_C] = W_C$ by

$$[W_X] = \frac{1}{2}(W_M + W_F). \quad (12)$$

Part 3, Survey-derived Average Crewmember Body Weights from NHANES Data

Definitions in Part 1 and Part 2 are applied to NHANES raw data, except with different age and gender categories.

Since the FAA defined age categories of “Infant”, “Child”, and “Adult” do not apply to crewmembers, a “Pilot” category is defined as persons at least 23, and less than 65, years of age and a “Flight Attendant” category is defined as persons at least 21 years of age.

The survey-derived average crewmember body weights are the W values defined for Pilot Female, Pilot Male, Flight Attendant Female and Flight Attendant Male, labeled $W_{P,F}$, $W_{P,M}$, $W_{FA,F}$, and $W_{FA,M}$ respectively.

If the gender of crewmembers actually on each flight is not known, then $[W_P]$ for Pilots and $[W_{FA}]$ for Flight Attendants must be determined. This requires a Male to Female ratio for Pilots ($M_P:F_P$) and a Male to Female ratio for Flight Attendants ($M_{FA}:F_{FA}$). Formula (10) is restated for Pilots and Flight Attendants as

$$[W_P] = \frac{M_P W_{P,M} + F_P W_{P,F}}{M_P + F_P} \text{ and } [W_{FA}] = \frac{M_{FA} W_{FA,M} + F_{FA} W_{FA,F}}{M_{FA} + F_{FA}} \quad (13)$$

For crew members other than pilots and flight attendants, the weight for the category that more closely aligns with crew members’ duties should be used. For example, an operator might use “Pilot” weights for Flight Engineers and “Flight Attendant” weights for supernumeraries or non-certificated crewmembers.

DISCUSSION

The calculations required by METHOD DEFINITION can be accomplished in any available calculation tool, including a spreadsheet. However, more thorough analysis requires tools meant specifically for statistical work. A text-based tool allows inclusion of the analysis steps and results in this paper. “R”, an analysis tool accessible to many data analysts and operations engineers, is therefore used here.

It is hoped the style used will be readable by those accustomed to working with data, even without extensive background in the language. Full R scripts, with instructions for getting started in R, are included in an Appendix. Code fragments in this section should be executed in the order presented. For clarity, fixed-width font is used with executable code shown in **blue**, descriptive comments in **green**, and any output generated in **orange**, starting with:

```
# The 'dplyr' library simplifies and improves readability, especially with
# 'mutate()' to add calculated columns to a data table,
# 'group_by()' to define categories within the data, and
# 'summarise()' [note spelling] to calculate summary statistics
library(dplyr)
```

Note that the symbol “**%>%**” means: to the result of the commands so far on the left, next apply the action to the right.

NHANES is a rigorous scientific study of health and nutrition among US, non-institutionalized, civilian population whose scope includes numerous body size metrics. Documentation for each NHANES data release should be reviewed whenever the data is used.

- NHANES is described for the public at [www.cdc.gov/nchs/nhanes/about_nhanes.htm].
- Academically rigorous documentation, including Sample Design, Weighting Process, and Analytical Guidelines, is available in a series of papers at [www.cdc.gov/nchs/nhanes/analyticguidelines.aspx].
- A technical Tutorial for data analysis which discusses background concepts and includes practical examples is at [www.cdc.gov/nchs/nhanes/tutorials].

- The NHANES Variable Keyword Search at [wwwn.cdc.gov/nchs/nhanes/search] can be used to determine how fields of interest are identified in the NHANES data.

Note: NHANES generally (not exclusively) uses “weight” to mean a numeric scaling factor, as in the common meaning of “weighted average”, rather than a measure of gravitational mass. To avoid confusion, this section carefully uses “mass” as the object of gravity and follows the NHANES convention for “weight” despite this paper elsewhere using US aviation vernacular, including “weight and balance” for the effect of gravity.

Using documentation above, details referred to in METHOD DEFINITION and DISCUSSION were confirmed to be consistent from 2003 through 2018. Data prior to 2003 uses inconsistent definitions and this method cannot be used without changes.

NHANES uses “MEC” to refer to CDC Mobile Examination Centers in which body measurements are taken. Body mass measurements BMXWT include undergarments and an exam gown, but not other clothing, unless an exception is coded using BMIWT.

In addition to the minimal data listed in Table 1, this analysis will use values identified in Table 2.

Table 2: Additional NHANES Data Fields		
Field Name	Data File	Field Description
SDDSRVYR	DEMO	This variable represents the two-year data release cycle number.
RIDEXPRG	DEMO	Pregnancy status for females between 20 and 44 years of age at the time of MEC exam.
SDMVPSU	DEMO	Masked variance unit pseudo-primary sampling unit variable; used for variation and error analysis
SDMVSTRA	DEMO	Masked variance unit pseudo-stratum variable; used for variation and error analysis

When working with multiple cycles, SDDSRVYR identifies the cycle in each record. File names use the corresponding letter of the alphabet. For example, raw data for the 2017-2018 is identified in the data by SDDSRVYR = 10 and the letter “J”, the tenth letter of the alphabet.

The required data files, DEMO and BMX, are freely available for download from the NHANES website. For example:

DEMO: wwwn.cdc.gov/nchs/nhanes/2017-2018/DEMO_J.XPT

BMX: wwwn.cdc.gov/nchs/nhanes/2017-2018/BMX_J.XPT

The following R commands download the required data files. Once published, these files do not change, so this only needs to be done once and not every time the analysis is repeated.

```
# Define a helper to keep cycle number and year conventions consistent
cycleYears <- function( cycle2or4, cycleNumbers )
{
  yr <- 2000+2*(cycleNumbers-1)
  sprintf('%s-%s',yr-(cycle2or4-1),yr)
}

# Define a helper to download needed data files to the working directory
downloadDataCycles <- function( cycleNumbers )
{
```

```

for( cn in cycleNumbers ){
  files <- sapply(c( 'DEMO_%s.XPT', 'BMX_%s.XPT' ), sprintf, LETTERS[cn])
  for( f in files ){
    download.file(
      sprintf('https://wwwn.cdc.gov/nchs/nhanes/%s/%s', cycleYears(2,cn),f),
      f, mode='wb'
    )
  }
}
}

# Change the working directory to one the current user can save files locally
setwd( 'C:/NHANES/' )

# Download XPT files for the cycles of interest
# Do this just once (ever) and then comment out this line
downloadDataCycles(3:10)

```

The format of these files is SAS Transport [.XPT]. Many statistical software packages, including R, can read this format directly, which is preferred. Otherwise, the CDC provides a link to a free universal SAS file viewer [wwwn.cdc.gov/nchs/nhanes/sasviewer.aspx] which may be used to paste the data into more basic programs, such as a spreadsheet.

The following R commands read the required data files and load just the variables in Table 1 and Table 2 into memory. The qualification '!is.na(WTMEC2YR)' loads BMX and DEMO data for every person in the sample not missing an examination record.

```

# The 'foreign' library reads SAS Transport(XPT) files
library(foreign)

# Define a helper to load the needed data columns files in the working directory
importDataCycles <- function( cycleNumbers )
{
  allData <- data.frame(NULL)
  for( cn in cycleNumbers ){
    allData <- bind_rows(allData, left_join(
      subset(
        select(
          read.xport(sprintf('DEMO_%s.XPT', LETTERS[cn])),
          # only needed columns from the DEMO file here
          'SEQN', 'SDDSRVYR', 'RIDAGEYR', 'RIAGENDR', 'RIDEXPRG',
          'SDMVSTRA', 'SDMVPSU', 'WTMEC2YR'
        ),
        !is.na(WTMEC2YR)), # exclude participants without an exam record
      subset(
        select(
          read.xport(sprintf('BMX_%s.XPT', LETTERS[cn])),
          # only needed columns from the BMX file here
          'SEQN', 'BMXWT', 'BMIWT'
        ),
        !is.na(WTMEC2YR)
      ),
      by = 'SEQN'
    )
  }
}

```

```

    )
  }
  return(allData)
}

# Load the data into memory once each session
RawData <- importDataCycles(3:10)

```

WTMEC2YR is a weight factor determined from the inverse probability that each person in the 2-year cycle exam data sample was selected from the population. It takes into account complex survey design, including oversampling, non-response, and post-stratification adjustments and is based on population data from the US Census Bureau. This weighting factor should be applied to relate data from MEC sample records in a 2-year cycle to the study population.

NHANES Analytic Guidelines recommend merging consecutive 2-year cycles into 4-year cycles to improve accuracy. To relate data from MEC sample records in a 4-year cycle to the study population a different weight factor, called WTMEC4YR, must be calculated. NHANES samples for years 2003-2018 were designed so that the correct values of WTMEC4YR are exactly half of the published value of WTMEC2YR. This is illustrated in [<https://wwwn.cdc.gov/nchs/nhanes/tutorials/module3.aspx>]. While average passenger body mass W is defined using a 4-year cycle with $WTMEC = WTMEC4YR = WTMEC2YR/2$, for data comparison formulas (1) thru (4) and (7) are also used here for 2-year cycles with $WTMEC = WTMEC2YR$.

NHANES Analytical Guidelines also describes the sample selection method as “stratified, multi-stage probability sample design” for which exact mathematical formulas for variance estimates do not exist. Instead, the National Center for Health Statistics (NCHS) recommends “Taylor series linearization methods for variance estimation in analyses of NHANES data for 2-year cycles or combined cycles”. In R, this is provided by `library(survey)`. This method of variance estimation requires information regarding the survey strata and sampling units which indicate how the sample persons were selected from the study population. These survey design parameters are encoded in the SDMVSTRA and SDMVPSU fields. When analyzing any subset of the sample, this variance estimation requires design parameters for both the records included and those excluded. Thus, when excluding data from any analysis, new fields are added to indicate whether each record is part of the included subset instead of deleting the excluded records.

In R, the required indicator fields, along with age and gender categories as defined by FAA (and by CDC for comparison) are defined first.

```

# Create a copy with calculated columns added (preserve RawData)
# Don't remove any rows, or survey design parameters are lost.
AllData <- mutate(RawData,
  # 'one' is needed to count records
  one = 1,

  # combine 2-year cycles as recommended by Analytical Guidelines
  cycle4 = 2*ceiling(SDDSRVYR/2),
  WTMEC4YR = WTMEC2YR/2,

  # an alias to ease comparing 4- and 2-year cycles
  cycle2 = SDDSRVYR,
  # use with WTMEC2YR

```

```

# make it easy to work in US units
WTLB = BMXWT / 0.45359237, # kg to lb

# use cut(right=FALSE) to build age intervals that include the lower
# and exclude the upper boundary
#
# for FAA, Infants have not yet reached 2nd birthday.
# Children have not yet reached 13th birthday.
faaAge = cut(RIDAGEYR, breaks=c(-Inf, 2, 13, Inf),
             right=FALSE, labels=c('Infant', 'Child', 'Adult')),
gender=factor(ifelse(RIDAGEYR>12,RIAGENDR,0),labels=c('None','Male','Female')),
#
# for comparison to CDC/NHSR, Adults are age 20 and over
cdcAge = cut(RIDAGEYR, breaks=c(-Inf, 20, Inf),
             right=FALSE, labels=c('Child', 'Adult'))
)

# Add columns used to filter the data, like 'inAnalysis' in tutorials
# Don't remove any rows, or survey design parameters are lost.
AllData <- mutate(AllData,
  NotMissing = ( !is.na(BMXWT) ),
  NotClothed = ( !is.na(BMXWT) & (is.na(BMIWT) | 3!=BMIWT) ),
  NotPregnant = ( !is.na(BMXWT) & (is.na(RIDEXPRG) | 1!=RIDEXPRG) )
  # for RIDEXPRG, 1=Yes, 2=No, 3=Unknown, Missing => No
)

```

Then, the NHANES survey design parameters are defined for 2-year and 4-year cycles. [svydesign](#) and [svyby](#) are library functions which apply NCHS recommended variance estimates using the survey design parameters. User defined function [svyWTLB](#) simply saves retyping commands to use these functions, especially to combine weighted statistics with unweighted sample size. [fmtWTLB](#) provides a consistent format to simplify comparing results.

```

library(survey)

# Define survey design parameters for overall dataset
NHANES2 <- svydesign(data=AllData,
  id=~SDMVPSU, strata=~SDMVSTRA, weights=~WTMEC2YR, nest=TRUE)
NHANES4 <- svydesign(data=AllData,
  id=~SDMVPSU, strata=~SDMVSTRA, weights=~WTMEC4YR, nest=TRUE)

# Define a function to get the weighted mean and standard error estimate
# using survey design parameters and join to the unweighted count
svyWTLB <- function(design, byFormula)
{
  # Get mean, stderr, and unweighted sample size
  c <- svyby( ~WTLB, byFormula, design, unwtd.count )
  p <- svyby( ~WTLB, byFormula, design, svymean )
  # select(c,-se) excludes the standard error of the count
  # suppressMessages() just hides output of the implicit "by" columns
  suppressMessages( outData <- left_join(select(c,-se), p) )
  return(outData)
}

```

```

}

# Define a function to get consistent output formatting
fmtWTLB <- function(cycleData)
{
  c2 <- match("cycle2",names(cycleData))
  if(!is.na(c2)) cycleData <- mutate(cycleData,
                                     Years=cycleYears(2,cycle2), .after=cycle2)
  c4 <- match("cycle4",names(cycleData))
  if(!is.na(c4)) cycleData <- mutate(cycleData,
                                     Years=cycleYears(4,cycle4), .after=cycle4)
  print.data.frame( cycleData %>% mutate_if(is.numeric,round,digits=1) )
}

```

The R commands so far can be checked by comparing their results values published by the CDC in National Health Statistics Reports (NHSR). Report Number 122 gives summary data from NHANES and details how these were determined using 2-year cycles, CDC age categories, and excluding pregnant women. Table 2 is reproduced here, and the highlighted results recalculated with the following R command.

Page 6

National Health Statistics Reports

Table 2. Mean weight (pounds) among men and women aged 20 and over, by survey years, age group, and race and Hispanic origin: United States, 1999–2016

Sex, age, and race and Hispanic origin	Survey years								
	1999–2000	2001–2002	2003–2004	2005–2006	2007–2008	2009–2010	2011–2012	2013–2014	2015–2016
Men									
			Mean (standard error)						
20 and over ¹ (crude)	189.1 (1.5)	191.8 (1.1)	193.5 (1.1)	196.0 (1.7)	194.7 (1.4)	196.3 (1.4)	194.4 (1.4)	197.0 (1.2)	197.8 (1.9)
20 and over ^{1,2} (age adjusted)	189.4 (1.5)	191.2 (1.2)	193.3 (1.1)	195.7 (1.7)	194.6 (1.4)	196.2 (1.3)	194.3 (1.4)	197.0 (1.3)	197.9 (2.0)
Age group:									
20–39	185.8 (2.0)	188.1 (1.6)	190.4 (2.3)	192.1 (2.3)	189.9 (2.1)	193.1 (1.9)	190.7 (1.9)	194.4 (2.7)	196.9 (3.1)
40–59	194.3 (2.6)	197.7 (1.5)	198.4 (2.0)	203.0 (2.5)	199.9 (2.0)	202.0 (2.4)	200.4 (2.3)	200.7 (1.8)	200.9 (2.0)
60 and over	187.8 (1.7)	186.2 (1.5)	190.0 (1.7)	190.0 (1.4)	194.0 (1.8)	192.1 (1.5)	190.5 (2.5)	195.4 (2.7)	194.7 (1.9)
Race and Hispanic origin³:									
Non-Hispanic white	192.3 (1.7)	194.6 (1.3)	196.3 (1.3)	197.8 (1.6)	198.2 (1.3)	199.5 (1.4)	196.7 (1.8)	200.2 (2.0)	202.2 (2.4)
Non-Hispanic black	188.7 (1.7)	190.8 (1.9)	196.4 (3.8)	201.0 (1.9)	197.1 (2.0)	201.1 (2.4)	198.4 (2.1)	199.3 (2.6)	197.7 (2.6)
Non-Hispanic Asian	---	---	---	---	---	---	159.1 (1.6)	161.7 (1.4)	161.1 (1.3)
Hispanic ³	---	---	---	---	---	---	187.2 (1.8)	189.6 (3.1)	190.5 (1.9)
Mexican-American	177.9 (1.6)	177.5 (1.7)	180.1 (2.1)	179.3 (2.0)	184.0 (2.7)	185.6 (2.8)	189.5 (2.5)	191.9 (2.3)	190.4 (2.5)
Women									
20 and over ¹ (crude)	163.6 (1.7)	162.9 (1.2)	164.2 (1.6)	165.1 (1.7)	166.1 (1.3)	166.2 (0.9)	167.1 (1.3)	169.8 (1.3)	170.5 (1.7)
20 and over ^{1,2} (age adjusted)	163.8 (1.7)	162.8 (1.3)	164.1 (1.7)	164.8 (1.7)	166.3 (1.4)	166.1 (1.0)	167.2 (1.3)	170.1 (1.4)	170.6 (1.7)
Age group:									
20–39	161.9 (2.1)	158.9 (2.1)	160.6 (2.4)	160.5 (2.4)	166.6 (2.7)	164.3 (2.1)	165.1 (1.8)	169.4 (1.7)	167.6 (1.9)
40–59	169.4 (2.9)	168.6 (2.4)	171.0 (2.5)	172.3 (2.7)	170.2 (1.8)	167.8 (1.5)	172.3 (1.7)	175.2 (2.7)	176.4 (3.0)
60 and over	157.9 (1.3)	160.3 (1.4)	159.0 (1.3)	160.2 (1.9)	159.5 (1.1)	166.6 (1.3)	162.5 (2.3)	163.1 (1.7)	166.5 (2.6)
Race and Hispanic origin³:									
Non-Hispanic white	161.9 (2.2)	162.2 (1.3)	162.6 (2.0)	164.1 (1.9)	165.8 (2.2)	165.0 (1.2)	167.1 (1.9)	170.4 (1.8)	170.9 (2.1)
Non-Hispanic black	185.9 (2.3)	179.3 (1.9)	185.1 (2.3)	184.1 (2.3)	185.0 (3.0)	189.5 (1.9)	190.0 (2.2)	190.0 (1.9)	186.1 (2.4)
Non-Hispanic Asian	---	---	---	---	---	---	131.8 (1.7)	131.0 (1.8)	132.4 (1.1)
Hispanic ³	---	---	---	---	---	---	160.7 (1.3)	162.4 (1.3)	166.3 (2.1)
Mexican-American	157.5 (2.2)	157.5 (2.1)	162.7 (2.7)	160.8 (2.3)	160.8 (0.9)	161.6 (1.4)	165.0 (2.4)	170.0 (2.2)	171.9 (1.5)

--- Data not available.

¹Includes other races not shown separately.

²Age adjusted to the projected 2000 U.S. census population using age groups 20–39, 40–59, and 60 and over.

³Includes Mexican-American persons.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2016.

```

# confirm definitions above here are correct by comparing to
# https://www.cdc.gov/nchs/data/nhsr/nhsr122-508.pdf

fmtWTLB( svyWTLB( subset( NHANES2, NotPregnant ), ~cycle2+cdcAge+gender ) ) %>%
  filter(cdcAge=='Adult'&cycle2<10) # NHSR Table 2 has Adults only

```

	cycle2	Years	cdcAge	gender	counts	WTLB	se
1	3	2003-2004	Adult	Male	2247	193.5	1.0
2	4	2005-2006	Adult	Male	2242	196.0	1.7
3	5	2007-2008	Adult	Male	2755	194.7	1.4
4	6	2009-2010	Adult	Male	2896	196.3	1.4
5	7	2011-2012	Adult	Male	2591	194.4	1.4

6	8	2013-2014	Adult	Male	2645	197.0	1.2
7	9	2015-2016	Adult	Male	2584	197.8	1.9
8	3	2003-2004	Adult	Female	2201	164.2	1.6
9	4	2005-2006	Adult	Female	2129	165.1	1.7
10	5	2007-2008	Adult	Female	2805	166.2	1.3
11	6	2009-2010	Adult	Female	3039	166.3	0.9
12	7	2011-2012	Adult	Female	2602	167.1	1.3
13	8	2013-2014	Adult	Female	2823	169.8	1.3
14	9	2015-2016	Adult	Female	2757	170.5	1.7

(In this and subsequent output, the first column is an automatically generated row number for reference only.)

Data in the NHSR report was calculated using SUDAAN software rather than R, yet comparing the output above to highlighted values in the published table shows an almost exact match, with no difference greater than 0.1 lb. This substantiates that the raw data is loaded correctly, and calculated mean and standard error values are consistent with recommended practice by CDC/NHANES and NCHS. The same command, adjusted only to use 4-year cycles, FAA age categories, and not exclude pregnant females, produces the following values:

```
# run the same formula with faaAge, and not excluding pregnant females
#fmtWTLB( svyWTLB( subset( NHANES2, NotPregnant ), ~cycle2+cdcAge+gender ) )
  fmtWTLB( svyWTLB( subset( NHANES4, NotClothed ), ~cycle4+faaAge+gender ) )
```

	cycle4	Years	faaAge	gender	counts	WTLB	se
1	4	2003-2006	Infant	None	1578	21.5	0.2
2	6	2007-2010	Infant	None	1451	21.0	0.2
3	8	2011-2014	Infant	None	1212	20.9	0.2
4	10	2015-2018	Infant	None	1172	21.0	0.2
5	4	2003-2006	Child	None	4352	66.8	0.6
6	6	2007-2010	Child	None	4373	66.3	0.7
7	8	2011-2014	Child	None	4501	67.5	0.8
8	10	2015-2018	Child	None	3890	67.5	0.6
9	4	2003-2006	Adult	Male	6233	189.9	0.9
10	6	2007-2010	Adult	Male	6603	191.0	1.0
11	8	2011-2014	Adult	Male	6152	191.6	0.9
12	10	2015-2018	Adult	Male	5940	194.7	1.3
13	4	2003-2006	Adult	Female	6593	161.7	1.2
14	6	2007-2010	Adult	Female	6741	163.0	0.8
15	8	2011-2014	Adult	Female	6407	165.5	0.9
16	10	2015-2018	Adult	Female	6235	167.9	1.2

This method might not be obviously the same as [METHOD DEFINITION](#), but in fact the definitions are identical for mean values. (Differences in variance estimates are discussed below.) To demonstrate this, a manual implementation of Steps (1) and (2) to calculate \bar{w}^* (called ws in code) and σ^* (called ss in the code) follow.

```
# A copy of the data excluding clothed sample participants
Step0 <- AllData %>% filter(NotClothed)

# Calculate a raw weighted mean, w-bar-star (ws) using Formula (1)
Step1 <- mutate(Step0, ww=WTLB*WTMEC4YR) %>%
  group_by(faaAge,gender,cycle4) %>%
```

```

summarise( ws=sum(wv)/sum(WTMEC4YR) )

# Sum weighted residuals (wr) for each row to get sigma-star (ss) using Formula (2)
Step2 <- left_join(Step0,Step1,by=c('faaAge','gender','cycle4'))      %>%
  mutate( wr=WTMEC4YR*(WTLB-ws)^2 )                                  %>%
  group_by(faaAge,gender,cycle4,ws)                                  %>%
  summarise( n=sum(one), ss=sqrt(sum(wr)/((n-1)/n*sum(WTMEC4YR))) )  %>%
  mutate( 'te%'=196*ss/ws/sqrt(n) )

fmtWTLB(Step2 %>% select(cycle4,faaAge,gender,n,ws,ss,'te%'))

```

	cycle4	Years	faaAge	gender	n	ws	ss	te%
1	4	2003-2006	Infant	None	1578	21.5	5.6	1.3
2	6	2007-2010	Infant	None	1451	21.0	5.5	1.4
3	8	2011-2014	Infant	None	1212	20.9	5.7	1.5
4	10	2015-2018	Infant	None	1172	21.0	5.6	1.5
5	4	2003-2006	Child	None	4352	66.8	33.4	1.5
6	6	2007-2010	Child	None	4373	66.3	33.6	1.5
7	8	2011-2014	Child	None	4501	67.5	34.8	1.5
8	10	2015-2018	Child	None	3890	67.5	33.1	1.5
9	4	2003-2006	Adult	Male	6233	189.9	44.7	0.6
10	6	2007-2010	Adult	Male	6603	191.0	45.8	0.6
11	8	2011-2014	Adult	Male	6152	191.6	45.9	0.6
12	10	2015-2018	Adult	Male	5940	194.7	49.2	0.6
13	4	2003-2006	Adult	Female	6593	161.7	44.1	0.7
14	6	2007-2010	Adult	Female	6741	163.0	43.9	0.6
15	8	2011-2014	Adult	Female	6407	165.5	45.7	0.7
16	10	2015-2018	Adult	Female	6235	167.9	46.5	0.7

The calculated values for the raw weighted mean are identical to those calculated immediately above. The FAA Tolerable Error (te%) calculated from σ^* is expressed as a percentage of the mean, while the Taylor-series based standard error (se) has units of mass, so these values should not be compared directly.

Consistent with the purpose of NHANES as a health and nutrition study, this data represents individuals across the entire spectrum of health and nutrition conditions, including the most unhealthy and malnourished. Individuals at the extremes of such a health spectrum are likely to be underrepresented in, if not completely absent from, the flying public. Thus, when applied to aviation, extreme values in NHANES data should have a reduced effect. Care should be taken that both low and high outliers are excluded.

A standard technique to reduce the effect of extreme values is to exclude sample values which are more than a multiple of the sample standard deviation from the sample mean; twice the standard deviation is common practice.

In certain contexts, it is important to account for possible variation between the survey-derived average value and values for groups of actual passengers. Operational systems which account for this variation have assumed a $2\sigma^*$ filtering technique because of its publication in AC 120-27, revisions D and E. Choosing a consistent definition avoids having to redesign these systems. An example is AC 120-27F, Appendix D, which relies on the sigma value derived using AC 120-27E, Appendix 2.

The estimates used for sample standard deviation, σ^* and σ , are technically exact only for a simple, independent random sample selection. However, academic reviewers advised that such a simplification can be used for NHANES. This approximation has the advantage that it can be represented exactly by a formula and implemented using any calculation tool without requiring specific software.

A disadvantage which may prove pertinent in the future, when changes in NHANES data will have to be evaluated compared to the current values, is that formulas (2), (4), and (5) do not fully reflect the design of NHANES sample selection. They are thus unreliable for determining whether changes between cycles are due to randomness in sample selection or due to changes in the population. Such an analysis should be accomplished using recommended practices by CDC/NHANES and NCHS.

The defined method of filtering extreme values from the data is implemented in the following commands.

```
# Filter by excluding values more than twice ss from the raw mean
Step3 <- left_join(Step0,Step2,by=c('faaAge','gender','cycle4'))      %>%
  filter( 2 >= abs((WTLB-ws)/ss) )

# Calculate a weighted mean, W using Formula (3)
Step4 <- mutate(Step3, wr=WTLB*WTMEC4YR)                             %>%
  group_by(faaAge,gender,cycle4)                                       %>%
  summarise( W=sum(wr)/sum(WTMEC4YR) )

# Sum weighted residuals (wr) for each row to get sigma (S) using Formula (4)
Step5 <- left_join(Step3,Step4,by=c('faaAge','gender','cycle4'))      %>%
  mutate( wr=WTMEC4YR*(WTLB-W)^2 )                                     %>%
  group_by(faaAge,gender,cycle4,W)                                     %>%
  summarise( n=sum(one), S=sqrt(sum(wr)/((n-1)/n*sum(WTMEC4YR))) )    %>%
  mutate( 'te%'=196*S/W/sqrt(n) )

fmtWTLB(Step5 %>% select(cycle4,faaAge,gender,n,W,S,'te%'))
```

	cycle4	Years	faaAge	gender	n	W	S	te%
1	4	2003-2006	Infant	None	1494	21.8	5.1	1.2
2	6	2007-2010	Infant	None	1382	21.1	4.9	1.2
3	8	2011-2014	Infant	None	1167	20.9	5.2	1.4
4	10	2015-2018	Infant	None	1118	21.1	5.0	1.4
5	4	2003-2006	Child	None	4087	62.0	26.4	1.3
6	6	2007-2010	Child	None	4156	61.5	26.7	1.3
7	8	2011-2014	Child	None	4303	62.4	26.9	1.3
8	10	2015-2018	Child	None	3692	62.9	26.3	1.3
9	4	2003-2006	Adult	Male	5940	186.1	36.1	0.5
10	6	2007-2010	Adult	Male	6291	186.5	37.3	0.5
11	8	2011-2014	Adult	Male	5879	186.8	37.1	0.5
12	10	2015-2018	Adult	Male	5689	189.6	40.5	0.6
13	4	2003-2006	Adult	Female	6321	155.9	35.1	0.6
14	6	2007-2010	Adult	Female	6435	157.3	35.3	0.5
15	8	2011-2014	Adult	Female	6125	159.3	36.4	0.6
16	10	2015-2018	Adult	Female	5937	161.6	37.2	0.6

These are the survey-derived averages for body mass calculated in accordance with Part 1 of METHOD DEFINITION.

In Part 2 of **METHOD DEFINITION**, the operator should select the combinations of adjustments, or none, which are most appropriate to their systems. For some systems these may be calculated for each flight, season, or other period, while some systems require static adjustments updated only when OpSpecs are revised.

Estimated values of N for each FAA age/gender category can be determined from NHANES raw data.

```
# Calculate population fractions using Formula (7)
NN <- left_join(
  Step3 %>% group_by(cycle4,faaAge,gender) %>% summarise( Nc=sum(WTMEC4YR) ),
  Step3 %>% group_by(cycle4,
                    by=c('cycle4')
                    ) %>% summarise( Nt=sum(WTMEC4YR) ),
  group_by(cycle4,faaAge,gender) %>% summarise( N=100*Nc/Nt )

fmtWTLB( NN )
```

	cycle4	Years	faaAge	gender	N
1	4	2003-2006	Infant	None	2.9
2	4	2003-2006	Child	None	15.0
3	4	2003-2006	Adult	Male	39.7
4	4	2003-2006	Adult	Female	42.4
5	6	2007-2010	Infant	None	2.9
6	6	2007-2010	Child	None	14.8
7	6	2007-2010	Adult	Male	40.1
8	6	2007-2010	Adult	Female	42.2
9	8	2011-2014	Infant	None	2.4
10	8	2011-2014	Child	None	14.6
11	8	2011-2014	Adult	Male	40.4
12	8	2011-2014	Adult	Female	42.6
13	10	2015-2018	Infant	None	2.4
14	10	2015-2018	Child	None	13.7
15	10	2015-2018	Adult	Male	40.8
16	10	2015-2018	Adult	Female	43.1

From this data, calculating $[W]$ values for use in operational systems is simplest with a hand calculator. For completeness, implementation of formulas (10), (11), and (12) are also shown in R.

```
library(tidyr) # for pivot

Wx <- left_join( Step5 %>% select(cycle4,faaAge,gender,W) %>%
  pivot_wider(names_from=c(faaAge,gender),values_from=W,names_prefix='W_'),
  NN %>%
  pivot_wider(names_from=c(faaAge,gender),values_from=N,names_prefix='N_'),
  by=c('cycle4') )
rename( N_Infant=N_Infant_None, W_Infant=W_Infant_None )
mutate(
  W_Adult_NHANES = ( N_Adult_Male * W_Adult_Male +
                    N_Adult_Female * W_Adult_Female
                    ) / ( N_Adult_Male + N_Adult_Female ),
  W_Adult_5050 = 0.5*( W_Adult_Male + W_Adult_Female ), # Formula 12
```

```

    dW_Infant_NHANES = N_Infant / ( N_Adult_Female + N_Adult_Male ) * W_Infant,

    'W_Adult_NHANES+Infant' = W_Adult_NHANES + dW_Infant_NHANES, # Formula 10
    'W_Adult_5050+Infant'   = W_Adult_5050   + dW_Infant_NHANES, # Formula 11
  )

fmtWTLB( Wx %>% select(cycle4,W_Adult_NHANES,W_Adult_5050,dW_Infant_NHANES) )

  cycle4      Years W_Adult_NHANES W_Adult_5050 dW_Infant_NHANES
1         4 2003-2006         170.5         171.0           0.8
2         6 2007-2010         171.6         171.9           0.7
3         8 2011-2014         172.7         173.0           0.6
4        10 2015-2018         175.3         175.6           0.6

fmtWTLB( Wx %>% select(cycle4,'W_Adult_NHANES+Infant','W_Adult_5050+Infant') )

  cycle4      Years W_Adult_NHANES+Infant W_Adult_5050+Infant
1         4 2003-2006         171.3         171.7
2         6 2007-2010         172.3         172.7
3         8 2011-2014         173.3         173.6
4        10 2015-2018         175.9         176.2

```

In the data shown, the weight increase resulting from a “50:50” M:F ratio compared to a ratio estimated from NHANES does not exceed half a pound, which is comparable in scale to rounding to whole pounds.

Anecdotally, fewer adults travel with infants than have infant children, so using the proportion of Infants to Adults in the NHANES population almost certainly over-estimates the body mass of Infants on typical flights. Operator-specific data is expected to reduce the magnitude of adjustments for infants, which is also comparable in scale to rounding to whole pounds.

In Part 3 of METHOD DEFINITION, appropriate ages for crewmembers differ significantly from the definition of Adult in AC 120-27F. Since age requirements for crew are not universal, the definition uses typical values. For pilots the minimum comes from 14 CFR 61.153(a)(1) for an unrestricted airline transport pilot (ATP) certificate and the maximum from 15 CFR 121.383(d) for air carrier limitations on pilot age. For Flight Attendants, the minimum age to serve alcohol in any US state has been observed as a common hiring requirement and is used here.

Commands used above to implement Steps 1 through 5 are repeated here in a condensed format, using **P** for Pilots and **F** for Flight Attendants.

```

# Repeat Part 1, Steps 1 thru 5 using ages representative of Pilots
P0 <- AllData %>% filter( NotClothed & 23<=RIDAGEYR & 65>RIDAGEYR )
P1 <- mutate(P0, ww=WTLB*WTMEC4YR) %>%
  group_by(gender,cycle4) %>%
  summarise( ws=sum(ww)/sum(WTMEC4YR) )
P2 <- left_join(P0,P1,by=c('gender','cycle4')) %>%
  mutate( wr=WTMEC4YR*(WTLB-ws)^2 ) %>%
  group_by(gender,cycle4,ws) %>%
  summarise( n=sum(one), ss=sqrt(sum(wr)/((n-1)/n*sum(WTMEC4YR))) ) %>%
  mutate( 'te%'=196*ss/ws/sqrt(n) )

```

```

P3 <- left_join(P0,P2,by=c('gender','cycle4'))           %>%
  filter( 2 >= abs((WTLB-ws)/ss) )
P4 <- mutate(P3, wr=WTLB*WTMEC4YR)                       %>%
  group_by(gender,cycle4)                                %>%
  summarise( W=sum(wr)/sum(WTMEC4YR) )
P5 <- left_join(P3,P4,by=c('gender','cycle4'))           %>%
  mutate( wr=WTMEC4YR*(WTLB-W)^2 )                      %>%
  group_by(gender,cycle4,W)                              %>%
  summarise( n=sum(one), S=sqrt(sum(wr)/((n-1)/n*sum(WTMEC4YR))) ) %>%
  mutate( 'te%'=196*S/W/sqrt(n) )
fmtWTLB(P5 %>% select(cycle4,gender,n,W,S,'te%'))

```

	cycle4	Years	gender	n	W	S	te%
1	4	2003-2006	Male	2766	192.5	33.8	0.7
2	6	2007-2010	Male	3650	193.0	35.5	0.6
3	8	2011-2014	Male	3513	193.1	35.1	0.6
4	10	2015-2018	Male	3288	196.2	39.5	0.7
5	4	2003-2006	Female	3136	162.1	36.5	0.8
6	6	2007-2010	Female	3893	162.8	36.2	0.7
7	8	2011-2014	Female	3709	165.9	37.4	0.7
8	10	2015-2018	Female	3631	167.2	38.9	0.8

Repeat Part 1, Steps 1 thru 5 using ages representative of Flight Attendants

```

F0 <- AllData %>% filter( NotClothed & 21<=RIDAGEYR )
F1 <- mutate(F0, ww=WTLB*WTMEC4YR)                       %>%
  group_by(gender,cycle4)                                %>%
  summarise( ws=sum(ww)/sum(WTMEC4YR) )
F2 <- left_join(F0,F1,by=c('gender','cycle4'))           %>%
  mutate( wr=WTMEC4YR*(WTLB-ws)^2 )                      %>%
  group_by(gender,cycle4,ws)                              %>%
  summarise( n=sum(one), ss=sqrt(sum(wr)/((n-1)/n*sum(WTMEC4YR))) ) %>%
  mutate( 'te%'=196*ss/ws/sqrt(n) )
F3 <- left_join(F0,F2,by=c('gender','cycle4'))           %>%
  filter( 2 >= abs((WTLB-ws)/ss) )
F4 <- mutate(F3, wr=WTLB*WTMEC4YR)                       %>%
  group_by(gender,cycle4)                                %>%
  summarise( W=sum(wr)/sum(WTMEC4YR) )
F5 <- left_join(F3,F4,by=c('gender','cycle4'))           %>%
  mutate( wr=WTMEC4YR*(WTLB-W)^2 )                      %>%
  group_by(gender,cycle4,W)                              %>%
  summarise( n=sum(one), S=sqrt(sum(wr)/((n-1)/n*sum(WTMEC4YR))) ) %>%
  mutate( 'te%'=196*S/W/sqrt(n) )
fmtWTLB(F5 %>% select(cycle4,gender,n,W,S,'te%'))

```

	cycle4	Years	gender	n	W	S	te%
1	4	2003-2006	Male	4080	190.6	34.2	0.6
2	6	2007-2010	Male	5133	191.3	35.6	0.5
3	8	2011-2014	Male	4724	191.1	35.0	0.5
4	10	2015-2018	Male	4626	194.5	38.7	0.6
5	4	2003-2006	Female	4400	159.3	35.0	0.6

```

6      6 2007-2010 Female 5403 160.6 35.2 0.6
7      8 2011-2014 Female 4986 162.8 35.9 0.6
8     10 2015-2018 Female 4928 164.7 37.1 0.6

```

Generally, ratios for $M_P:F_P$ and $M_{FA}:F_{FA}$ may be estimated from an operator's total employed crewmembers. Sample calculations using 50:50 and 99:1 are shown for reference when implementing formula (13) with no claim that these ratios are applicable to any specific operation.

```

# for example only, apply sample M:F ratios for crewmembers using Formula (13)
F13 <- function(M,F,W_M,W_F) { ( M*W_M + F*W_F )/(M+F) }

```

```

fmtWTLB( P5 %>% select(cycle4,gender,W)                                %>%
  pivot_wider(names_from=c(gender),values_from=W,names_prefix='W_P__') %>%
  mutate( W_P_50=F13( 50,50, W_P__Male,W_P__Female ),
          W_P_99=F13( 99, 1, W_P__Male,W_P__Female )) )

```

```

cycle4    Years W_P__Male W_P__Female W_P_50 W_P_99
1         4 2003-2006    192.5     162.1  177.3  192.2
2         6 2007-2010    193.0     162.8  177.9  192.7
3         8 2011-2014    193.1     165.9  179.5  192.8
4        10 2015-2018    196.2     167.2  181.7  195.9

```

```

fmtWTLB( F5 %>% select(cycle4,gender,W)                                %>%
  pivot_wider(names_from=c(gender),values_from=W,names_prefix='W_FA_') %>%
  mutate( W_FA_50=F13( 50,50, W_FA_Male,W_FA_Female ),
          W_FA_99=F13( 99, 1, W_FA_Male,W_FA_Female )) )

```

```

cycle4    Years W_FA_Male W_FA_Female W_FA_50 W_FA_99
1         4 2003-2006    190.6     159.3  175.0  190.3
2         6 2007-2010    191.3     160.6  175.9  190.9
3         8 2011-2014    191.1     162.8  176.9  190.8
4        10 2015-2018    194.5     164.7  179.6  194.2

```

SUMMARY VALUES

The following tables show results of the METHOD DEFINITION applied to NHANES 4-year cycles from 2003 to 2018.

These values were calculated using a Microsoft Excel spreadsheet and validated using the R scripts in DISCUSSION. To the number of decimal places reported, there is no difference between calculation tools used.

Table 3: Survey-derived Average Passenger Weights from NHANES Data (LB)												
NHANES 4 Yr. Cycle	Adult Male			Adult Female			Child			Infant		
	n_f	W_M	σ	n_f	W_F	σ	n_f	W_C	σ	n_f	W_I	σ
2003-2006	5940	186.1	36.1	6321	155.9	35.1	4087	62.0	26.4	1494	21.8	5.1
2007-2010	6291	186.5	37.3	6435	157.3	35.3	4156	61.5	26.7	1382	21.1	4.9
2011-2014	5879	186.8	37.1	6125	159.3	36.4	4303	62.4	29.9	1167	20.9	5.2
2015-2018	5689	189.6	40.5	5937	161.6	37.2	3692	62.9	26.3	1118	21.1	5.0

Table 4: Survey-derived Average Crewmember Weights from NHANES Data (LB)

NHANES 4 Yr. Cycle	Pilot, Male			Pilot, Female			Flight Attendant, Male			Flight Attendant, Female		
	n_f	$W_{P,M}$	σ	n_f	$W_{P,F}$	σ	n_f	$W_{FA,M}$	σ	n_f	$W_{FA,F}$	σ
2003-2006	2766	192.5	33.8	3136	162.1	36.5	4080	190.6	34.2	4400	159.3	35.0
2007-2010	3650	193.0	35.5	3893	162.8	36.2	5133	191.3	35.6	5403	160.6	35.2
2011-2014	3513	193.1	35.1	3709	165.9	37.4	4724	191.1	35.0	4986	162.8	35.9
2015-2018	3288	196.2	39.5	3631	167.2	38.9	4626	194.5	38.7	4928	164.7	37.1

Table 5: Adjusted Body Weights for Sample M:F Ratios (LB)

NHANES 4 Yr. Cycle	[W_x] Not Adjusted for Infants					[W_x] Adjusted for Infants		
	Pilots		Flight Attendants		Adults	Adults	Adults	Adults
M:F	50:50	99:1	50:50	99:1	NHANES	50:50	NHANES	50:50
2003-2006	177.3	192.2	175.0	190.3	170.5	171.0	171.3	171.7
2007-2010	177.9	192.7	175.9	190.9	171.6	171.9	172.3	172.7
2011-2014	179.5	192.8	176.9	190.8	172.7	173.0	173.3	173.6
2015-2018	181.7	195.9	179.6	194.2	175.3	175.6	175.9	176.2

REFERENCES

Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2003-2018, www.cdc.gov/nchs/nhanes.

Chen TC, Clark J, Riddles MK, Mohadjer LK, Fakhouri THI. National Health and Nutrition Examination Survey, 2015–2018: Sample design and estimation procedures. National Center for Health Statistics. Vital Health Stat 2(184). 2020.

Federal Aviation Administration, Advisory Circular 120-27 “Aircraft Weight and Balance Control”: (original) 1968; 120-27C 1995; 120-27D 2004; 120-27E 2005; 120-27F 2019.

Fryar CD, Kruszon-Moran D, Gu Q, Ogden CL. Mean body weight, height, waist circumference, and body mass index among adults: United States, 1999–2000 through 2015–2016. National Health Statistics Reports; no 122. Hyattsville, MD: National Center for Health Statistics. 2018. <https://www.cdc.gov/nchs/data/nhsr/nhsr122-508.pdf>

Lumley, Thomas. Complex sampling and R. University of Washington Biostatistics and “user” Conference, Rennes, France. 2009. <http://faculty.washington.edu/tlumley/tutorials/survey-user.pdf>

National Health and Nutrition Examination Survey (NHANES): Anthropometry Procedures Manual, January 2017.

National Health and Nutrition Examination Survey: Analytic Guidelines, 2011-2014 and 2015-2016 December 14, 2018.

APPENDIX A

To run this script, those new to R will need to install the language <https://www.r-project.org/>, which is freely available for a variety of platforms from <https://www.r-project.org/>. Use is simplified by installing a graphical development interface, such as RStudio Desktop (available at no cost from <https://www.rstudio.com/>), and pasting the relevant code into it directly. The code may be executed line by line or by selecting “Run All”.

Packages required for installation are: `dpRr`, `forSAS`, `survey`, and `tidyr`.

Other modifications are generally not required (except for the path at `setwd()`), as the scripts are accessible to the user and are not addressed to a specific system. System specific details such as what differentiates male and female are accessible to the user and are not addressed to a specific system. A small font and large page are used to assist pasting from a document into plain text.

```
# Script for DISCUSSION section 'NAMES USE FOR STANDARD MESSAGE' (SAPOE)
# (C) 2020 Society of Aircraft Performance and Operations Engineers (SAPOE)
# SAPOE is a member-based organization promoting the safety and efficiency of
# flight through knowledge of aircraft performance and engineering principles.
# Except for brief quotations with appropriate citation, copies may be made
# and distributed only of the complete document including cover page and
# this disclaimer.
# SAPOE, the authors, and reviewers assume no liability whatsoever and make
# no warranty of any kind. Anyone who uses this report, or the source code
# and data contained in it, is solely responsible for their own operations
# and for any outcome of such use.
# The 'dpRr' library simplifies and improves readability, especially with
# 'mutate()' to add calculated columns to a data table,
# groupBy() to define categories within the data, and
# summarise() (note spelling) to calculate summary statistics
library(dpRr)

# Define a helper to keep cycle number and year conventions consistent
cyleYears <- function( cycleYear, cycleNumbers )
  yr <- 2009+2*(cycleYear-1)
  sprintf('%s-%s',yr-(cycleYear-1),yr)
}

# Define a helper to download needed data files to the working directory
downloadDataCycles <- function( cycleNumbers )
  {
    for( cn in cycleNumbers ){
      files <- sprintf('DBM_%s.XPT', BMK_%s.XPT ) %>% sprintf(LETTERS[cn])
      for( f in files ){
        download.file(
          sprintf('https://www.cdc.gov/nchs/ohanes/%s/%s',cyleYears(2,cn),f),
          file.path(
            '.',mode='w'
          )
        )
      }
    }
  }

# Change the working directory to one the current user can save files locally
setwd( 'C:/NAMES/' )

# Download XPT files for the cycles of interest
# Do this just once (ever) and then comment out this line
downloadDataCycles(3:18)

# The 'foreign' library reads SAS Transport(XPT) files
library(foreign)

# Define a helper to load the needed data columns files in the working directory
loadDataCycles <- function( cycleNumbers )
  {
    allData <- data.frame(NULL)
    for( cn in cycleNumbers ){
      allData <- bind_rows(allData,left_join(
        subset(
          read.xport(sprintf('DBM_%s.XPT', LETTERS[cn])),
          only needed columns from the DBM file here
          'SOMSTR', 'SNWPSU', 'NAGEGR', 'NAGEEXP',
          'SOMSTR', 'SNWPSU', 'NAGEGR'
        ),
        allData,by='SEQN'
      ))
    }
  }

# Load the data into memory once each session
loadData <- loadDataCycles(3:18)

# Create a copy with calculated columns added (preserve rawData)
# This is the same as the survey design parameters are lost.
allData <- mutate(rawData,
  one = 1,
  # combine 2-year cycles as recommended by Analytical Guidelines
```

```
cycle4 = 2*ceiling(SDSRVR/2),
WMEC4YR = WMEC2YR/2,
# an alias to ease comparing 4- and 2-year cycles
cycle2 = SDSRVR,
# use with WMEC2YR

# make it easy to work in US units
WTLB = BMDT / 0.45359237, # kg to lb

# use cut(right=FALSE) to build age intervals that include the lower
# and exclude the upper boundary

# for FAA, infants have not yet reached 2nd birthday.
# Children have not yet reached 13th birthday.
# Use cut(right=FALSE) to build age intervals that include the lower
# and exclude the upper boundary
age = cut(right=FALSE, labels=c('Inf', 'Infant', 'Child', 'Adult'),
  range = c(0, 2, 13, 18), right=FALSE, include.lowest=TRUE)
# for comparison to CDC/NHIS, Adults are age 20 and over
age = cut(right=FALSE, labels=c('Inf', '20', 'Inf'),
  range = c(0, 20, 18), right=FALSE, include.lowest=TRUE)

# Define a function to filter the data, like 'inAnalysis' in tutorials
filterData <- function( data, filter )
  {
    allData <- mutate(
      NAgeGR = ifelse(NAgeGR < 13, 'Infant', 'Adult'),
      NAgeEXP = ifelse(NAgeEXP < 13, 'Infant', 'Adult'),
      NAgeGR2 = ifelse(NAgeGR < 13, 'Infant', 'Adult'),
      NAgeEXP2 = ifelse(NAgeEXP < 13, 'Infant', 'Adult')
    )
    # for RIDEEXP, 1-19 is 3rd through 19th birthday, 20-24 is 20th through 24th birthday, 25-29 is 25th through 29th birthday, 30-34 is 30th through 34th birthday, 35-39 is 35th through 39th birthday, 40-44 is 40th through 44th birthday, 45-49 is 45th through 49th birthday, 50-54 is 50th through 54th birthday, 55-59 is 55th through 59th birthday, 60-64 is 60th through 64th birthday, 65-69 is 65th through 69th birthday, 70-74 is 70th through 74th birthday, 75-79 is 75th through 79th birthday, 80-84 is 80th through 84th birthday, 85-89 is 85th through 89th birthday, 90-94 is 90th through 94th birthday, 95-99 is 95th through 99th birthday, 100-104 is 100th through 104th birthday, 105-109 is 105th through 109th birthday, 110-114 is 110th through 114th birthday, 115-119 is 115th through 119th birthday, 120-124 is 120th through 124th birthday, 125-129 is 125th through 129th birthday, 130-134 is 130th through 134th birthday, 135-139 is 135th through 139th birthday, 140-144 is 140th through 144th birthday, 145-149 is 145th through 149th birthday, 150-154 is 150th through 154th birthday, 155-159 is 155th through 159th birthday, 160-164 is 160th through 164th birthday, 165-169 is 165th through 169th birthday, 170-174 is 170th through 174th birthday, 175-179 is 175th through 179th birthday, 180-184 is 180th through 184th birthday, 185-189 is 185th through 189th birthday, 190-194 is 190th through 194th birthday, 195-199 is 195th through 199th birthday, 200-204 is 200th through 204th birthday, 205-209 is 205th through 209th birthday, 210-214 is 210th through 214th birthday, 215-219 is 215th through 219th birthday, 220-224 is 220th through 224th birthday, 225-229 is 225th through 229th birthday, 230-234 is 230th through 234th birthday, 235-239 is 235th through 239th birthday, 240-244 is 240th through 244th birthday, 245-249 is 245th through 249th birthday, 250-254 is 250th through 254th birthday, 255-259 is 255th through 259th birthday, 260-264 is 260th through 264th birthday, 265-269 is 265th through 269th birthday, 270-274 is 270th through 274th birthday, 275-279 is 275th through 279th birthday, 280-284 is 280th through 284th birthday, 285-289 is 285th through 289th birthday, 290-294 is 290th through 294th birthday, 295-299 is 295th through 299th birthday, 300-304 is 300th through 304th birthday, 305-309 is 305th through 309th birthday, 310-314 is 310th through 314th birthday, 315-319 is 315th through 319th birthday, 320-324 is 320th through 324th birthday, 325-329 is 325th through 329th birthday, 330-334 is 330th through 334th birthday, 335-339 is 335th through 339th birthday, 340-344 is 340th through 344th birthday, 345-349 is 345th through 349th birthday, 350-354 is 350th through 354th birthday, 355-359 is 355th through 359th birthday, 360-364 is 360th through 364th birthday, 365-369 is 365th through 369th birthday, 370-374 is 370th through 374th birthday, 375-379 is 375th through 379th birthday, 380-384 is 380th through 384th birthday, 385-389 is 385th through 389th birthday, 390-394 is 390th through 394th birthday, 395-399 is 395th through 399th birthday, 400-404 is 400th through 404th birthday, 405-409 is 405th through 409th birthday, 410-414 is 410th through 414th birthday, 415-419 is 415th through 419th birthday, 420-424 is 420th through 424th birthday, 425-429 is 425th through 429th birthday, 430-434 is 430th through 434th birthday, 435-439 is 435th through 439th birthday, 440-444 is 440th through 444th birthday, 445-449 is 445th through 449th birthday, 450-454 is 450th through 454th birthday, 455-459 is 455th through 459th birthday, 460-464 is 460th through 464th birthday, 465-469 is 465th through 469th birthday, 470-474 is 470th through 474th birthday, 475-479 is 475th through 479th birthday, 480-484 is 480th through 484th birthday, 485-489 is 485th through 489th birthday, 490-494 is 490th through 494th birthday, 495-499 is 495th through 499th birthday, 500-504 is 500th through 504th birthday, 505-509 is 505th through 509th birthday, 510-514 is 510th through 514th birthday, 515-519 is 515th through 519th birthday, 520-524 is 520th through 524th birthday, 525-529 is 525th through 529th birthday, 530-534 is 530th through 534th birthday, 535-539 is 535th through 539th birthday, 540-544 is 540th through 544th birthday, 545-549 is 545th through 549th birthday, 550-554 is 550th through 554th birthday, 555-559 is 555th through 559th birthday, 560-564 is 560th through 564th birthday, 565-569 is 565th through 569th birthday, 570-574 is 570th through 574th birthday, 575-579 is 575th through 579th birthday, 580-584 is 580th through 584th birthday, 585-589 is 585th through 589th birthday, 590-594 is 590th through 594th birthday, 595-599 is 595th through 599th birthday, 600-604 is 600th through 604th birthday, 605-609 is 605th through 609th birthday, 610-614 is 610th through 614th birthday, 615-619 is 615th through 619th birthday, 620-624 is 620th through 624th birthday, 625-629 is 625th through 629th birthday, 630-634 is 630th through 634th birthday, 635-639 is 635th through 639th birthday, 640-644 is 640th through 644th birthday, 645-649 is 645th through 649th birthday, 650-654 is 650th through 654th birthday, 655-659 is 655th through 659th birthday, 660-664 is 660th through 664th birthday, 665-669 is 665th through 669th birthday, 670-674 is 670th through 674th birthday, 675-679 is 675th through 679th birthday, 680-684 is 680th through 684th birthday, 685-689 is 685th through 689th birthday, 690-694 is 690th through 694th birthday, 695-699 is 695th through 699th birthday, 700-704 is 700th through 704th birthday, 705-709 is 705th through 709th birthday, 710-714 is 710th through 714th birthday, 715-719 is 715th through 719th birthday, 720-724 is 720th through 724th birthday, 725-729 is 725th through 729th birthday, 730-734 is 730th through 734th birthday, 735-739 is 735th through 739th birthday, 740-744 is 740th through 744th birthday, 745-749 is 745th through 749th birthday, 750-754 is 750th through 754th birthday, 755-759 is 755th through 759th birthday, 760-764 is 760th through 764th birthday, 765-769 is 765th through 769th birthday, 770-774 is 770th through 774th birthday, 775-779 is 775th through 779th birthday, 780-784 is 780th through 784th birthday, 785-789 is 785th through 789th birthday, 790-794 is 790th through 794th birthday, 795-799 is 795th through 799th birthday, 800-804 is 800th through 804th birthday, 805-809 is 805th through 809th birthday, 810-814 is 810th through 814th birthday, 815-819 is 815th through 819th birthday, 820-824 is 820th through 824th birthday, 825-829 is 825th through 829th birthday, 830-834 is 830th through 834th birthday, 835-839 is 835th through 839th birthday, 840-844 is 840th through 844th birthday, 845-849 is 845th through 849th birthday, 850-854 is 850th through 854th birthday, 855-859 is 855th through 859th birthday, 860-864 is 860th through 864th birthday, 865-869 is 865th through 869th birthday, 870-874 is 870th through 874th birthday, 875-879 is 875th through 879th birthday, 880-884 is 880th through 884th birthday, 885-889 is 885th through 889th birthday, 890-894 is 890th through 894th birthday, 895-899 is 895th through 899th birthday, 900-904 is 900th through 904th birthday, 905-909 is 905th through 909th birthday, 910-914 is 910th through 914th birthday, 915-919 is 915th through 919th birthday, 920-924 is 920th through 924th birthday, 925-929 is 925th through 929th birthday, 930-934 is 930th through 934th birthday, 935-939 is 935th through 939th birthday, 940-944 is 940th through 944th birthday, 945-949 is 945th through 949th birthday, 950-954 is 950th through 954th birthday, 955-959 is 955th through 959th birthday, 960-964 is 960th through 964th birthday, 965-969 is 965th through 969th birthday, 970-974 is 970th through 974th birthday, 975-979 is 975th through 979th birthday, 980-984 is 980th through 984th birthday, 985-989 is 985th through 989th birthday, 990-994 is 990th through 994th birthday, 995-999 is 995th through 999th birthday, 1000-1004 is 1000th through 1004th birthday, 1005-1009 is 1005th through 1009th birthday, 1010-1014 is 1010th through 1014th birthday, 1015-1019 is 1015th through 1019th birthday, 1020-1024 is 1020th through 1024th birthday, 1025-1029 is 1025th through 1029th birthday, 1030-1034 is 1030th through 1034th birthday, 1035-1039 is 1035th through 1039th birthday, 1040-1044 is 1040th through 1044th birthday, 1045-1049 is 1045th through 1049th birthday, 1050-1054 is 1050th through 1054th birthday, 1055-1059 is 1055th through 1059th birthday, 1060-1064 is 1060th through 1064th birthday, 1065-1069 is 1065th through 1069th birthday, 1070-1074 is 1070th through 1074th birthday, 1075-1079 is 1075th through 1079th birthday, 1080-1084 is 1080th through 1084th birthday, 1085-1089 is 1085th through 1089th birthday, 1090-1094 is 1090th through 1094th birthday, 1095-1099 is 1095th through 1099th birthday, 1100-1104 is 1100th through 1104th birthday, 1105-1109 is 1105th through 1109th birthday, 1110-1114 is 1110th through 1114th birthday, 1115-1119 is 1115th through 1119th birthday, 1120-1124 is 1120th through 1124th birthday, 1125-1129 is 1125th through 1129th birthday, 1130-1134 is 1130th through 1134th birthday, 1135-1139 is 1135th through 1139th birthday, 1140-1144 is 1140th through 1144th birthday, 1145-1149 is 1145th through 1149th birthday, 1150-1154 is 1150th through 1154th birthday, 1155-1159 is 1155th through 1159th birthday, 1160-1164 is 1160th through 1164th birthday, 1165-1169 is 1165th through 1169th birthday, 1170-1174 is 1170th through 1174th birthday, 1175-1179 is 1175th through 1179th birthday, 1180-1184 is 1180th through 1184th birthday, 1185-1189 is 1185th through 1189th birthday, 1190-1194 is 1190th through 1194th birthday, 1195-1199 is 1195th through 1199th birthday, 1200-1204 is 1200th through 1204th birthday, 1205-1209 is 1205th through 1209th birthday, 1210-1214 is 1210th through 1214th birthday, 1215-1219 is 1215th through 1219th birthday, 1220-1224 is 1220th through 1224th birthday, 1225-1229 is 1225th through 1229th birthday, 1230-1234 is 1230th through 1234th birthday, 1235-1239 is 1235th through 1239th birthday, 1240-1244 is 1240th through 1244th birthday, 1245-1249 is 1245th through 1249th birthday, 1250-1254 is 1250th through 1254th birthday, 1255-1259 is 1255th through 1259th birthday, 1260-1264 is 1260th through 1264th birthday, 1265-1269 is 1265th through 1269th birthday, 1270-1274 is 1270th through 1274th birthday, 1275-1279 is 1275th through 1279th birthday, 1280-1284 is 1280th through 1284th birthday, 1285-1289 is 1285th through 1289th birthday, 1290-1294 is 1290th through 1294th birthday, 1295-1299 is 1295th through 1299th birthday, 1300-1304 is 1300th through 1304th birthday, 1305-1309 is 1305th through 1309th birthday, 1310-1314 is 1310th through 1314th birthday, 1315-1319 is 1315th through 1319th birthday, 1320-1324 is 1320th through 1324th birthday, 1325-1329 is 1325th through 1329th birthday, 1330-1334 is 1330th through 1334th birthday, 1335-1339 is 1335th through 1339th birthday, 1340-1344 is 1340th through 1344th birthday, 1345-1349 is 1345th through 1349th birthday, 1350-1354 is 1350th through 1354th birthday, 1355-1359 is 1355th through 1359th birthday, 1360-1364 is 1360th through 1364th birthday, 1365-1369 is 1365th through 1369th birthday, 1370-1374 is 1370th through 1374th birthday, 1375-1379 is 1375th through 1379th birthday, 1380-1384 is 1380th through 1384th birthday, 1385-1389 is 1385th through 1389th birthday, 1390-1394 is 1390th through 1394th birthday, 1395-1399 is 1395th through 1399th birthday, 1400-1404 is 1400th through 1404th birthday, 1405-1409 is 1405th through 1409th birthday, 1410-1414 is 1410th through 1414th birthday, 1415-1419 is 1415th through 1419th birthday, 1420-1424 is 1420th through 1424th birthday, 1425-1429 is 1425th through 1429th birthday, 1430-1434 is 1430th through 1434th birthday, 1435-1439 is 1435th through 1439th birthday, 1440-1444 is 1440th through 1444th birthday, 1445-1449 is 1445th through 1449th birthday, 1450-1454 is 1450th through 1454th birthday, 1455-1459 is 1455th through 1459th birthday, 1460-1464 is 1460th through 1464th birthday, 1465-1469 is 1465th through 1469th birthday, 1470-1474 is 1470th through 1474th birthday, 1475-1479 is 1475th through 1479th birthday, 1480-1484 is 1480th through 1484th birthday, 1485-1489 is 1485th through 1489th birthday, 1490-1494 is 1490th through 1494th birthday, 1495-1499 is 1495th through 1499th birthday, 1500-1504 is 1500th through 1504th birthday, 1505-1509 is 1505th through 1509th birthday, 1510-1514 is 1510th through 1514th birthday, 1515-1519 is 1515th through 1519th birthday, 1520-1524 is 1520th through 1524th birthday, 1525-1529 is 1525th through 1529th birthday, 1530-1534 is 1530th through 1534th birthday, 1535-1539 is 1535th through 1539th birthday, 1540-1544 is 1540th through 1544th birthday, 1545-1549 is 1545th through 1549th birthday, 1550-1554 is 1550th through 1554th birthday, 1555-1559 is 1555th through 1559th birthday, 1560-1564 is 1560th through 1564th birthday, 1565-1569 is 1565th through 1569th birthday, 1570-1574 is 1570th through 1574th birthday, 1575-1579 is 1575th through 1579th birthday, 1580-1584 is 1580th through 1584th birthday, 1585-1589 is 1585th through 1589th birthday, 1590-1594 is 1590th through 1594th birthday, 1595-1599 is 1595th through 1599th birthday, 1600-1604 is 1600th through 1604th birthday, 1605-1609 is 1605th through 1609th birthday, 1610-1614 is 1610th through 1614th birthday, 1615-1619 is 1615th through 1619th birthday, 1620-1624 is 1620th through 1624th birthday, 1625-1629 is 1625th through 1629th birthday, 1630-1634 is 1630th through 1634th birthday, 1635-1639 is 1635th through 1639th birthday, 1640-1644 is 1640th through 1644th birthday, 1645-1649 is 1645th through 1649th birthday, 1650-1654 is 1650th through 1654th birthday, 1655-1659 is 1655th through 1659th birthday, 1660-1664 is 1660th through 1664th birthday, 1665-1669 is 1665th through 1669th birthday, 1670-1674 is 1670th through 1674th birthday, 1675-1679 is 1675th through 1679th birthday, 1680-1684 is 1680th through 1684th birthday, 1685-1689 is 1685th through 1689th birthday, 1690-1694 is 1690th through 1694th birthday, 1695-1699 is 1695th through 1699th birthday, 1700-1704 is 1700th through 1704th birthday, 1705-1709 is 1705th through 1709th birthday, 1710-1714 is 1710th through 1714th birthday, 1715-1719 is 1715th through 1719th birthday, 1720-1724 is 1720th through 1724th birthday, 1725-1729 is 1725th through 1729th birthday, 1730-1734 is 1730th through 1734th birthday, 1735-1739 is 1735th through 1739th birthday, 1740-1744 is 1740th through 1744th birthday, 1745-1749 is 1745th through 1749th birthday, 1750-1754 is 1750th through 1754th birthday, 1755-1759 is 1755th through 1759th birthday, 1760-1764 is 1760th through 1764th birthday, 1765-1769 is 1765th through 1769th birthday, 1770-1774 is 1770th through 1774th birthday, 1775-1779 is 1775th through 1779th birthday, 1780-1784 is 1780th through 1784th birthday, 1785-1789 is 1785th through 1789th birthday, 1790-1794 is 1790th through 1794th birthday, 1795-1799 is 1795th through 1799th birthday, 1800-1804 is 1800th through 1804th birthday, 1805-1809 is 1805th through 1809th birthday, 1810-1814 is 1810th through 1814th birthday, 1815-1819 is 1815th through 1819th birthday, 1820-1824 is 1820th through 1824th birthday, 1825-1829 is 1825th through 1829th birthday, 1830-1834 is 1830th through 1834th birthday, 1835-1839 is 1835th through 1839th birthday, 1840-1844 is 1840th through 1844th birthday, 1845-1849 is 1845th through 1849th birthday, 1850-1854 is 1850th through 1854th birthday, 1855-1859 is 1855th through 1859th birthday, 1860-1864 is 1860th through 1864th birthday, 1865-1869 is 1865th through 1869th birthday, 1870-1874 is 1870th through 1874th birthday, 1875-1879 is 1875th through 1879th birthday, 1880-1884 is 1880th through 1884th birthday, 1885-1889 is 1885th through 1889th birthday, 1890-1894 is 1890th through 1894th birthday, 1895-1899 is 1895th through 1899th birthday, 1900-1904 is 1900th through 1904th birthday, 1905-1909 is 1905th through 1909th birthday, 1910-1914 is 1910th through 1914th birthday, 1915-1919 is 1915th through 1919th birthday, 1920-1924 is 1920th through 1924th birthday, 1925-1929 is 1925th through 1929th birthday, 1930-1934 is 1930th through 1934th birthday, 1935-1939 is 1935th through 1939th birthday, 1940-1944 is 1940th through 1944th birthday, 1945-1949 is 1945th through 1949th birthday, 1950-1954 is 1950th through 1954th birthday, 1955-1959 is 1955th through 1959th birthday, 1960-1964 is 1960th through 1964th birthday, 1965-1969 is 1965th through 1969th birthday, 1970-1974 is 1970th through 1974th birthday, 1975-1979 is 1975th through 1979th birthday, 1980-1984 is 1980th through 1984th birthday, 1985-1989 is 1985th through 1989th birthday, 1990-1994 is 1990th through 1994th birthday, 1995-1999 is 1995th through 1999th birthday, 2000-2004 is 2000th through 2004th birthday, 2005-2009 is 2005th through 2009th birthday, 2010-2014 is 2010th through 2014th birthday, 2015-2019 is 2015th through 2019th birthday, 2020-2024 is 2020th through 2024th birthday, 2025-2029 is 2025th through 2029th birthday, 2030-2034 is 2030th through 2034th birthday, 2035-2039 is 2035th through 2039th birthday, 2040-2044 is 2040th through 2044th birthday, 2045-2049 is 2045th through 2049th birthday, 2050-2054 is 2050th through 2054th birthday, 2055-2059 is 2055th through 2059th birthday, 2060-2064 is 2060th through 2064th birthday, 2065-2069 is 2065th through 2069th birthday, 2070-2074 is 2070th through 2074th birthday, 2075-2079 is 2075th through 2079th birthday, 2080-2084 is 2080th through 2084th birthday, 2085-2089 is 2085th through 2089th birthday, 2090-2094 is 2090th through 2094th birthday, 2095-2099 is 2095th through 2099th birthday, 2100-2104 is 2100th through 2104th birthday, 2105-2109 is 2105th through 2109th birthday, 2110-2114 is 2110th through 2114th birthday, 2115-2119 is 2115th through 2119th birthday, 2120-2124 is 2120th through 2124th birthday, 2125-2129 is 2125th through 2129th birthday, 2130-2134 is 2130th through 2134th birthday, 2135-2139 is 2135th through 2139th birthday, 2140-2144 is 2140th through 2144th birthday, 2145-2149 is 2145th through 2149th birthday, 2150-2154 is 2150th through 2154th birthday, 2155-2159 is 2155th through 2159th birthday, 2160-2164 is 2160th through 2164th birthday, 2165-2169 is 2165th through 2169th birthday, 2170-2174 is 2170th through 2174th birthday, 2175-2179 is 2175th through 2179th birthday, 2180-2184 is 2180th through 2184th birthday, 2185-2189 is 2185th through 2189th birthday, 2190-2194 is 2190th through 2194th birthday, 2195-2199 is 2195th through 2199th birthday, 2200-2204 is 2200th through 2204th birthday, 2205-2209 is 2205th through 2209th birthday, 2210-2214 is 2210th through 2214th birthday, 2215-2219 is 2215th through 2219th birthday, 2220-2224 is 2220th through 2224th birthday, 222
```