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 Principle 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

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 ($\overline{w}^*$) is calculated using the formula:

$$\overline{w}^* = \frac{\sum_{i=1}^{n_R} BMXWT_i \times WTMEC_i}{\sum_{i=1}^{n_R} WTMEC_i} \tag{1}$$

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

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

Step 3) Consistent with AC 120-27E, Appendix 2, sample data with BMXWT more than twice σ^* either greater than or less than $\overline{w}^*$ 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:

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

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

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

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:

$$e = \frac{1.96 \times \sigma \times 100}{W \times \sqrt{n_f}} \tag{5}$$

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

$$N_A = N_F + N_M + N_X \tag{6}$$

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_C, 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:

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

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:

$$M:F = N_M : N_F \tag{8}$$

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:

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

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

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

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

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

<u>Part 3, Survey-derived Average Crewmember Body Weights from NHANES Data</u>

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

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

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.

<div align="center">DISCUSSION</div>

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 [wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx].
- A technical Tutorial for data analysis which discusses background concepts and includes practical examples is at [wwwn.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
        select(
          read.xport(sprintf('BMX_%s.XPT' , LETTERS[cn])),
          # only needed columns from the BMX file here
          'SEQN','BMXWT','BMIWT'
        ),
        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.

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 origin[2]:									
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]	--- ---	--- ---	--- ---	--- ---	184.1 (2.4)	187.0 (2.4)	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 origin[2]:									
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)	160.1 (1.6)	162.4 (1.3)	166.3 (2.1)	169.0 (1.6)
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.
[1]Includes other races not shown separately.
[2]Age adjusted to the projected 2000 U.S. census population using age groups 20–39, 40–59, and 60 and over.
[3]Includes Mexican-American persons.

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

Page 6

National Health Statistics Reports ∎

```
# 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 <u>METHOD DEFINITION</u>, 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(ww)/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 <u>METHOD DEFINITION</u>.

In Part 2 of <u>METHOD DEFINITION</u>, 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,            ) %>% summarise( Nt=sum(WTMEC4YR) ),
           by=c('cycle4')
                )                                                           %>%
         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 core, which is freely available for a variety of platforms from [www.r-project.org]. Use is simplified by installing a graphical development interface, such as RStudio Desktop (available as of writing at no cost from [rstudio.com]), and pasting the entire code into it directly. The code may be executed line by line or by selecting "Run All".

Packages required for installation are: dplyr, foreign, survey, and tidyr.

Other modifications are generally not required except to assure the path at setwd(…) exists and is accessible for writing. System specific details such as what directories may be accessible to the user are not addressed here.

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

```
#---------------------------------------------------------------------

# Script for DISCUSSION section '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 auth0rs, 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 '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)

# 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 '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
      select(
        read.xport(sprintf('BMX_%s.XPT' , LETTERS[cn])),
        # only needed columns from the BMX file here
        'SEQN','BMXWT','BMIWT'
      ),
      by = 'SEQN' )
    )
  }
  return(allData)
}

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

# 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
)

# based heavily on https://wwwn.cdc.gov/nchs/nhanes/tutorials/samplecode.aspx
#  and https://wwwn.cdc.gov/nchs/data/tutorials/DB303_Fig1_R.R
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) )
}

# 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

# 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 ) )

# 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(ww)/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%'))

# 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%'))

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

fmtWTLB( NN )

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) )

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

# 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%'))

# 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%'))

# 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 )) )

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 )) )

#---------------------------------------------------------------------
```

END