

Supplemental Appendix V
SAS Code for Importing, Harmonizing, and Merging Data

```

/**** SAS Code for DLW Database Preparation ****/
/** Indiana University, Biostatistics Consulting Center **/
/** Last revised 10/13/2022 **/

%let mypath=C:\Users\lgolzarr\Indiana University\0365-BCC -
Documents\Projects\Allison, David\DRI Energy;
*%let mypath=C:\Users\sd3\Indiana University\0365-BCC - Documents\Projects\Allison,
David\DRI Energy;

libname clean "&mypath.\Data\Clean data for analysis";
libname mysolnas "&mypath.\Data\NHLBI Solnas\HCHS_SOLNAS\SOLNAS_data";

/*****
*****/
/***** IAEA *****/
*****/
/*****
*****/

/*****/
/* Reading IAEA data n=7710 with empty rows, n=7696 actual data*/
PROC IMPORT OUT= IAEARaw
            DATAFILE= "&mypath.\Data\IAEA\IAEA DLW database 3.6.1
abbreviated for DRI group (allison).xlsx"
            DBMS=xlsx REPLACE;
            SHEET="Sheet1";
            GETNAMES=YES;
RUN;
/*NOTE: I had to modify the imported Excel file by:
-remove spaces from numeric variables in excel since it was reading columns as
characters and
it was creating problems when transforming it to numeric.
-Also deleted empty rows at the end with spaces and empty column with spaces
-Rename bmi to uppercase BMI */

*From Note on Log: 'Variable Name Change. BEE (MJ/day) -> VAR24';
data iaearaw; set iaearaw;
rename VAR24 = BEEmjday
        LBM=FFM;
run;

/*****/
/* Reading data for categories by income of countries
https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-
and-lending-groups */
PROC IMPORT OUT= class
            DATAFILE= "&mypath.\Data\IAEA\CLASS 2022-06-13.xlsx"

```

```
        DBMS=xlsx REPLACE;
        SHEET="List of economies";
    GETNAMES=YES;
RUN;
```

```
* Keep only relevant variables;
data class; set class;
keep Economy Code Income_group;
rename Code=ISO;
label Code= ;
run;
```

```
/* Reading data for publication descriptions to remove */
PROC IMPORT OUT= publication
            DATAFILE= "&mypath.\Data\IAEA\IAEA publications description
IU 060722 - LGA notes.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="IAEA Database";
            GETNAMES=YES;
RUN;
```

```
* Remove duplicate IDs, I checked that the duplicated ones all stay or all are
excluded; * Has to be 129 studies;
proc sort data=publication out=publication nodupkey; by stud_n; run;
```

```
* Keep only relevant variables;
data publication; set publication;
keep stud_n Study_Burden Medical_Conditions Remove;
run;
```

```
/* Reading data for kids BMI to use growth charts for BMI percentiles */
PROC IMPORT OUT=bmigirls
            DATAFILE= "&mypath.\Data\BMI kids\BMIcharts.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="Females";
            GETNAMES=YES;
RUN;
```

```
PROC IMPORT OUT=bmiboys
            DATAFILE= "&mypath.\Data\BMI kids\BMIcharts.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="Males";
            GETNAMES=YES;
RUN;
```

```
/* Reading data for infants weight-for-length. Length is in cm.*/
```

```

PROC IMPORT OUT=wlgirls
            DATAFILE= "&mypath.\Data\BMI kids\Weightlength_WHO.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="Girls";
            GETNAMES=YES;
RUN;

```

```

PROC IMPORT OUT=wlboys
            DATAFILE= "&mypath.\Data\BMI kids\Weightlength_WHO.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="Boys";
            GETNAMES=YES;
RUN;

```

```

/*****/
/* Clean IAEA by steps */
*Meeting Notes/Working Group 1 Agenda 060922.docx;
*Meeting Notes/INCLUSION EXCLUSION 053022.docx;
*Data/IAEA/IAEA_International Database instructions v4.docx;
/*****/

```

* Step 1: Use the codes under the column, ISO, in the DLW datasheet to exclude all studies being done in non-high-income countries.

- Get countries:
<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
 (link to download data doesn't work, but it is on 'databank.worldbank.org/data/download/site-content/CLASS.xlsx' includes 3 letter code)
- I am using the excel file from the link to subset only high-income countries, saved on IAEA folder as 'CLASS 2022-06-13.xlsx'
- if needed 3 letter codes:
http://www.nationsonline.org/oneworld/country_code_list.htm;

```

* Sort data sets by ISO code to merge;
proc sort data=class; by ISO; run;
proc sort data=iaearaw; by ISO; run;

```

```

* Merge data;
data iaea;
    merge iaearaw(in=x) class(in=y);
    by ISO;
    if x; *Only keep the countries in the iaea data. Drop the rest;
run; *n=7696;

```

```
* Only keep High-Income countries; *n=6989;
data iaea2; set iaea;
      where Income_group="High income";
run;
```

* Step 2: Then use the codes under the column, Health, in the DLW datasheet to exclude subjects with diseases that might affect TEE. Under the column, healthy subjects were labeled as "H". Any subjects with a code beginning with a D such as D1 or D15 should be excluded.;

```
* See which categories they have;
proc freq data=iaea2; tables Health; run;
* Categories H, Cystic Fybrois, D1, D10, D17, D4, D5, D6, D9, NA, missing; *NOTE:
What to do with NA, Missing and Cystic Fybrois;
```

```
* Create new variable with only first letter of Health;
data iaea2; set iaea2;
Health2 = substr(Health,1,1);
run;
/*proc freq data=iaea; tables Health2; run;*/
```

```
* Remove if new variable Health2=D; *n=6744;
data iaea3; set iaea2;
where Health2 ne 'D'; *Removes 245;
run;
/*
proc freq data=iaea3; tables Health2; run;
*/
```

* Step 3: Then use the codes under the column, ath, in the DLW datasheet to exclude professional athletes (PA);

```
* See which categories they have;
proc freq data=iaea3; tables ath; run;
* Categories AA, AANCT, NA, NotA, NotAUA, PA, PANCT, missing;
```

```
* Remove PA categories; *n=6706;
data iaea4; set iaea3;
where ath ne 'PA'; *Removes 38;
run;
/*proc freq data=iaea; tables ath; run;*/
```

* Step 4: Remove study from Wong ;

* Note WG1 originally said to remove Wong study #129 in meeting agenda, but 129 is Brage and has no reason to be excluded. WG1 confirmed in meeting 6/16/22 to remove study #103 from Wong and leave #129 Brage included.

* Remove PWong #103;

data iaea5; set iaea4;

if stud_n = 103 and scan(stud_ID,1," ")="Wong" then delete;

run;*n=6550;

* Step 5: Finally, exclude those studies highlighted in yellow in the IAEA publications description IU 060722.xlsx spreadsheet.;

* - LGA went to the file and identified the ID number of studies highlighted in yellow and created the variable 'remove' to filter.

Study ids and remove identifiers are in file "IAEA publications description IU 060722 - LGA notes.xlsx";

* Read in as 'publication';

* From Bill email on 7/15/22 "Therefore, you should exclude the Yamada (2018) and the Persson (2000) data from further analyses.";

* Sort data sets by study_n;

proc sort data=publication; by stud_n; run;

proc sort data=iaea5; by stud_n; run;

* Merge data;

data iaea6;

merge iaea5(in=x) publication(in=y);

by stud_n;

if x;

if remove=1 then delete; * These are the ones highlighted in yellow to

remove;

run;*n=6033;

* Step extra 1: Get sample sizes for Pregnant and Lactating. They should be treated separately than grouping with other Female.

*PP pre-puberty

RANR reproductive age (post puberty) but not reproducing,

P pregnant with trimester if known (ie p1, p2, p3) if not just P

L Lactating

PM post menopausal. ;

/* Nancy said via email 8/16/22:

For the IAEA data:

Most 2019 [Stud_n=120]: use 14.5 wk for P2 and 36 wk for P3.
Matsiko [Stud_n=125]: use 12 wk postpartum for these lactating women.
Most 2020 [Stud_n=120]: use 25 wk postpartum for these lactating women.
*/

```
* See which categories they have;  
proc freq data=iaea6; tables rep_statF; run;  
*Categories L,NA,P2,P3,PM,PP,RANR;
```

```
data iaea6;  
  length Strata P_stage $25. Weeks 8.;  
  set iaea6;  
  if rep_statF='L' then Lactating='Yes';  
    else Lactating='No';  
  if rep_statF='P2' or rep_statF='P3' then Pregnant='Yes';  
    else Pregnant='No';  
  *There was one person in P1 but dropped in filters above;  
  if rep_statF='P2' then do; P_stage = "2nd Trimester"; Weeks=14.5; end;  
  if rep_statF='P3' then do; P_stage = "3rd Trimester"; Weeks=36; end;  
  if rep_statF='L' and stud_n=120 then do; P_stage= "Lactating 4-6 mo";  
Weeks=25; end; *Most 2020;  
  if rep_statF='L' and stud_n=125 then do; P_stage= "Lactating 1-3 mo";  
Weeks=12; end; *Motsiko;
```

```
  if rep_statF in ('L','P p1','P2','P3') then Strata='Preg/Lac';  
  else do;  
    if 0<Age<3.0 then Strata='Infant';  
    if 3.0<=Age<19 and Sex='F' then Strata='Child Girls 3-18';  
    if 3.0<=Age<19 and Sex='M' then Strata='Child Boys 3-18';  
    if 19<=Age and Sex='F' then Strata='Adult Women 19+';  
    if 19<=Age and Sex='M' then Strata='Adult Men 19+';  
  end;
```

```
run;  
proc freq data=iaea6; tables Strata P_stage; run;  
*161 are missing Strata because missing Age and Sex;  
proc means data=iaea6(where=(Strata='Preg/Lac'));  
  var weeks;  
run; *174 are in 'Preg/Lac' and all have data for weeks;
```

* Step extra 2: The BMI categories for kids <18 should be based on percentiles <5th, 5th to 85th, 85th to 95th, >95th ;

```
*Calculate age month, round to 0.5;  
data iaea6; set iaea6;  
  age_mo = round(Age*12,0.5);  
  int_mo = int(age_mo);  
  if age_mo=int_mo then age_mo5=age_mo+0.5;  
  else age_mo5=age_mo;
```

```

    *age_mo5 = round(Age*12)+0.5;
    drop age_mo int_mo;
run;

*Calculate length, round to 0.5;
data iaea6; set iaea6;
    ht_mo = round(ht,0.1);
    *int_mo = int(ht_mo);
    *if ht_mo=int_mo then ht_5=ht_mo+0.5;
    *else ht_5=ht_mo;
    *drop ht_mo int_mo;
run;

```

*I'll subset the data by girls, boys and adults so it is easier to merge data.
From n=6,033;

```

data babygirls; set iaea6; if .<Age<2 & Sex='F'; run; *n=229;
data babyboys; set iaea6; if .<Age<2 & Sex='M'; run; *n=190;
data girls; set iaea6; if 2<=Age<19 & Sex='F'; run; *n=452;
data boys; set iaea6; if 2<=Age<19 & Sex='M'; run; *n=428;
data adults; set iaea6; if Age=. or Age>=19; run; *n=4734;
    *Note we are including people missing Age as adults;

```

```

*Sort to merge;
proc sort data=girls; by age_mo5; run;
proc sort data=boys; by age_mo5; run;
proc sort data=bmigirls; by Age_months; run;
proc sort data=bmiboys; by Age_months; run;

```

```

*Calculate percentile;
data boys2;
    merge boys (in=x rename=(age_mo5=age_months)) bmiboys;
    by Age_months;
    if x;
    length Percentile_group $25;
    if 0<=bmi<=Pct3 then do; Percentile =((BMI-0)/(Pct3 -0))*(3 -0)+0 ; if
Percentile<0 or Percentile>3 then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct3 <=bmi<=Pct5 then do; Percentile =((BMI-Pct3)/(Pct5 -Pct3))*(5 -3
)+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=bmi<=Pct10 then do; Percentile =((BMI-Pct5)/(Pct10-Pct5))*(10-5
)+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=bmi<=Pct25 then do; Percentile
=((BMI-Pct10)/(Pct25-Pct10))*(25-10)+10; if Percentile<10 or Percentile>25 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=bmi<=Pct50 then do; Percentile
=((BMI-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then put
"WARNING: " pat_person_nbr= Percentile= ; end;

```



```

        if Pct50<=bmi<=Pct75 then do; Percentile
= ((BMI-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct75<=bmi<=Pct85 then do; Percentile
= ((BMI-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct85<=bmi<=Pct90 then do; Percentile
= ((BMI-Pct85)/(Pct90-Pct85))*(90-85)+85; if Percentile<85 or Percentile>90 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct90<=bmi<=Pct95 then do; Percentile
= ((BMI-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct95<=bmi<=Pct97 then do; Percentile
= ((BMI-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if .<Pct97<=bmi<=50 then do; Percentile = ((BMI-Pct97)/(
100-Pct97))*(100-97)+97; if Percentile<97 or Percentile>100 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
        if Percentile<5 then Percentile_group='<5th';
        if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
        if Percentile>=85 & Percentile<=95 then Percentile_group='85th to 95th';
        if Percentile>95 then Percentile_group='>95th';
        if Percentile=. then Percentile_group='';
        drop Pct3 Pct5 Pct10 Pct25 Pct50 Pct75 Pct85 Pct90 Pct95 Pct97
pat_person_nbr;
run;

```

```

data girls2;
    merge girls (in=x rename=(age_mo5=age_months)) bmigirls;
    by Age_months;
    if x;
    length Percentile_group $25;
    if 0<=bmi<=Pct3 then do; Percentile = ((BMI-0)/(Pct3 -0 ))*(3 -0 )+0 ; if
Percentile<0 or Percentile>3 then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct3 <=bmi<=Pct5 then do; Percentile = ((BMI-Pct3)/(Pct5 -Pct3 ))*(5 -3
)+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=bmi<=Pct10 then do; Percentile = ((BMI-Pct5)/(Pct10-Pct5 ))*(10-5
)+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=bmi<=Pct25 then do; Percentile
= ((BMI-Pct10)/(Pct25-Pct10))*(25-10)+10; if Percentile<10 or Percentile>25 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=bmi<=Pct50 then do; Percentile
= ((BMI-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct50<=bmi<=Pct75 then do; Percentile
= ((BMI-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then put
"WARNING: " pat_person_nbr= Percentile= ; end;

```

```

        if Pct75<=bmi<=Pct85 then do; Percentile
= ((BMI-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct85<=bmi<=Pct90 then do; Percentile
= ((BMI-Pct85)/(Pct90-Pct85))*(90-85)+85; if Percentile<85 or Percentile>90 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct90<=bmi<=Pct95 then do; Percentile
= ((BMI-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct95<=bmi<=Pct97 then do; Percentile
= ((BMI-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if .<Pct97<=bmi<=50 then do; Percentile = ((BMI-Pct97)/(
100-Pct97))*(100-97)+97; if Percentile<97 or Percentile>100 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
        if Percentile<5 then Percentile_group='<5th';
        if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
        if Percentile>=85 & Percentile<=95 then Percentile_group='85th to 95th';
        if Percentile>95 then Percentile_group='>95th';
        if Percentile=. then Percentile_group='';
        drop Pct3 Pct5 Pct10 Pct25 Pct50 Pct75 Pct85 Pct90 Pct95 Pct97
pat_person_nbr;
run;

```

* Infants percentile;

*Sort to merge;

```

proc sort data=babygirls; by ht_mo; run;
proc sort data=babyboys; by ht_mo; run;
proc sort data=wlgirls; by Length; run;
proc sort data=wlboys; by Length; run;

```

*Calculate percentile;

```

data babyboys2;
    merge babyboys (in=x rename=(ht_mo=Length)) wlboys;
    by Length;
    if x;
    length Percentile_group $25;
    length Percentile_infant $25;
    if 0<=wt_init<=Pct01 then do; Percentile = ((wt_init-0 )/(Pct01 -0 ))*(0.1 -0
)+0 ; if Percentile<0 or Percentile>0.1 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct01 <=wt_init<=Pct1 then do; Percentile = ((wt_init-Pct01 )/(Pct1 -Pct01
))*(1 -0.1 )+0.1 ; if Percentile<0.1 or Percentile>1 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
    if Pct1 <=wt_init<=Pct3 then do; Percentile = ((wt_init-Pct1 )/(Pct3 -Pct1
))*(3 -1 )+1 ; if Percentile<1 or Percentile>3 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct3 <=wt_init<=Pct5 then do; Percentile = ((wt_init-Pct3 )/(Pct5 -Pct3

```

```

))*(5 -3 )+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=wt_init<=Pct10 then do; Percentile =((wt_init-Pct5 )/(Pct10-Pct5
))*(10-5 )+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=wt_init<=Pct15 then do; Percentile
=((wt_init-Pct10)/(Pct15-Pct10))*(15-10)+10; if Percentile<10 or Percentile>15 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct15<=wt_init<=Pct25 then do; Percentile
=((wt_init-Pct15)/(Pct25-Pct15))*(25-15)+15; if Percentile<15 or Percentile>25 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=wt_init<=Pct50 then do; Percentile
=((wt_init-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct50<=wt_init<=Pct75 then do; Percentile
=((wt_init-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct75<=wt_init<=Pct85 then do; Percentile
=((wt_init-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct85<=wt_init<=Pct90 then do; percentile
=((wt_init-Pct85)/(Pct90-Pct85))*(90-85)+85; if percentile<85 or percentile>90 then
put "WARNING: " pat_person_nbr= percentile= ; end;
    if Pct90<=wt_init<=Pct95 then do; Percentile
=((wt_init-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct95<=wt_init<=Pct97 then do; Percentile
=((wt_init-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct97<=wt_init<=Pct99 then do; Percentile
=((wt_init-Pct97)/(Pct99-Pct97))*(99-97)+97; if Percentile<97 or Percentile>99 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct99<=wt_init<=Pct999 then do; Percentile
=((wt_init-Pct99)/(Pct999-Pct99))*(99.9-99)+99; if Percentile<99 or Percentile>99.9
then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if .<Pct999<=wt_init<=50 then do; Percentile =((wt_init-Pct999)/(
100-Pct999))*(100-99.9)+99.9;if Percentile<99.9 or Percentile>100 then put "WARNING:
" pat_person_nbr= Percentile= ; end;
    if Percentile<5 then Percentile_group='<5th';
    if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
    if Percentile>=85 & Percentile<=95 then Percentile_group='85th to 95th';
    if Percentile>95 then Percentile_group='>95th';
    if Percentile=. then Percentile_group='';
    if Percentile<3 then Percentile_infant='<3th';
    if Percentile>=3 & Percentile<=97 then Percentile_infant='3th to 97th';
    if Percentile>97 then Percentile_infant='>97th';
    if Percentile=. then Percentile_infant='';
    drop L M S Pct01 Pct1 Pct3 Pct5 Pct10 Pct15 Pct25 Pct50 Pct75 Pct85 Pct90
Pct95 Pct97 Pct99 Pct999 pat_person_nbr; * ;
run;

```

```

data babygirls2;
    merge babygirls (in=x rename=(ht_mo=Length)) wlgirls;
    by Length;
    if x;
    length Percentile_group $25;
    length Percentile_infant $25;
    if 0<=wt_init<=Pct01 then do; Percentile =((wt_init-0 )/(Pct01 -0 ))*(0.1 -0
)+0 ; if Percentile<0 or Percentile>0.1 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct01 <=wt_init<=Pct1 then do; Percentile =((wt_init-Pct01 )/(Pct1 -Pct01
))* (1 -0.1 )+0.1 ; if Percentile<0.1 or Percentile>1 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
    if Pct1 <=wt_init<=Pct3 then do; Percentile =((wt_init-Pct1 )/(Pct3 -Pct1
))* (3 -1 )+1 ; if Percentile<1 or Percentile>3 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct3 <=wt_init<=Pct5 then do; Percentile =((wt_init-Pct3 )/(Pct5 -Pct3
))* (5 -3 )+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=wt_init<=Pct10 then do; Percentile =((wt_init-Pct5 )/(Pct10-Pct5
))* (10-5 )+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=wt_init<=Pct15 then do; Percentile
=((wt_init-Pct10)/(Pct15-Pct10))*(15-10)+10; if Percentile<10 or Percentile>15 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct15<=wt_init<=Pct25 then do; Percentile
=((wt_init-Pct15)/(Pct25-Pct15))*(25-15)+15; if Percentile<15 or Percentile>25 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=wt_init<=Pct50 then do; Percentile
=((wt_init-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct50<=wt_init<=Pct75 then do; Percentile
=((wt_init-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct75<=wt_init<=Pct85 then do; Percentile
=((wt_init-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct85<=wt_init<=Pct90 then do; percentile
=((wt_init-Pct85)/(Pct90-Pct85))*(90-85)+85; if percentile<85 or percentile>90 then
put "WARNING: " pat_person_nbr= percentile= ; end;
    if Pct90<=wt_init<=Pct95 then do; Percentile
=((wt_init-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct95<=wt_init<=Pct97 then do; Percentile
=((wt_init-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct97<=wt_init<=Pct99 then do; Percentile
=((wt_init-Pct97)/(Pct99-Pct97))*(99-97)+97; if Percentile<97 or Percentile>99 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct99<=wt_init<=Pct999 then do; Percentile

```

```

=((wt_init-Pct99)/(Pct999-Pct99))*(99.9-99)+99; if Percentile<99 or Percentile>99.9
then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if .<Pct999<=wt_init<=50 then do; Percentile =((wt_init-Pct999)/(
100-Pct999))*(100-99.9)+99.9;if Percentile<99.9 or Percentile>100 then put "WARNING:
" pat_person_nbr= Percentile= ; end;
    if Percentile<5 then Percentile_group='<5th';
    if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
    if Percentile>=85 & Percentile<=95 then Percentile_group='85th to 95th';
    if Percentile>95 then Percentile_group='>95th';
    if Percentile=. then Percentile_group='';
    if Percentile<3 then Percentile_infant='<3th';
    if Percentile>=3 & Percentile<=97 then Percentile_infant='3th to 97th';
    if Percentile>97 then Percentile_infant='>97th';
    if Percentile=. then Percentile_infant='';
    drop L M S Pct01 Pct1 Pct3 Pct5 Pct10 Pct15 Pct25 Pct50 Pct75 Pct85 Pct90
Pct95 Pct97 Pct99 Pct999 pat_person_nbr; *;
run;

```

```

** Merge them back together; *n=6033;
data iaea7; set babygirls2 babyboys2 girls2 boys2 adults; run;

```

```

** Create Age in months and rename rounded age in months;
data iaea7; set iaea7; drop Age_months;run;
data iaea7; set iaea7; Age_months = Age*12;run;

```

* Step extra 3: In the exclusion of "IAEA Publications description..." check column E for "Use". See orange highlights in column E in "IAEA Publications description IU 060722-LGA". I made a flag for Remove=2.

There are 6 studies to only use the Healthy (H) participants. The non-H should already be filtered out.

Only use US in #118 (Nigeria may already be filtered out?)

Only use Netherlands in #125 (Rwanda may be already filtered out?)

"Follow-up measurements on same subject". Not sure what to do with these??

**NOTE from LGA, these are pregnant and lactating women, would be separated.

"This set of study contained repeated studies. Please use just the first study";

```

data iaea8; set iaea7;
    if (stud_n=27 or stud_n=28 or stud_n=84 or stud_n=86 or stud_n=89 or
stud_n=91) & Health ne 'H' then remove01=1;
    *6 from stud_n=84 would be remove because their Health is missing;
    if stud_n=118 & ISO ne 'USA' then remove01=1;
    *none are removed at this step, probably deleted before;
    if stud_n=125 & ISO ne 'NLD' then remove01=1;
    *none are removed at this step, probably deleted before;
run;

```

```

data iaea9;set iaea8;
    if remove01 = 1 then delete;
    drop remove01;
run; *n=6027;

** Remove duplicated from study #128;
** From the 'Use' column in the file "IAEA publications description IU 060722 - LGA
notes",
"This set of study contained repeated studies. Please use just the first study in
the analysis.",
therefore we are only keeping the first observation by ID not not have repeated
observations;

*Subset study 128;
data stud128; set iaea9; where stud_n=128; run; *n=167;
data allstud; set iaea9; where stud_n ne 128; run; *n=5860;

*Sort by ind_id and year to choose first;
proc sort data=stud128; by ind_id year; run;

*Choose first to keep;
data stud128; set stud128;
by ind_id;
if first.ind_id then keep=1;
run;

*Only keep first;
data stud128; set stud128; where keep=1; drop keep; run; *n=106;

** Merge all back together;
data iaea10; set allstud stud128; run; *n=5966;

****
* Step extra 4: For Medical condition, Amy Luke said to recode the "1/3" to be "1";

/*proc freq data=iaea; table medical_conditions; run;*/
data iaea10; set iaea10;
if medical_conditions='1/3' then medical_conditions='1';
run; *n=5966;

****
* Remove Butte (2016): row number 5337, IND_ID 4873, Age = 25, Sex = F with missing
BMI value.
The subject had a genetic disease (Neonatal Progeroid Syndrome).;

data iaea11; set iaea10;

```

```
if stud_n=101 & IND_ID=4873 & DBID=5341 &
  Age=25 & Sex=F and bmi=. then delete;
run; *n=5966;
*This study was already removed because this subject was removed in the
step where health started with D were removed;
```

```
* Create age category variable;
* ... But we also have Strata variable coded above ~line 235
* These are the groupings in the Life Stages document provided by WG1;
data iaea11; set iaea11;
  length Age_cat $25;
  if .<Age<1 then Age_cat="Infants";
  if 1<=Age<9 then Age_cat="Children";
  if 9<=Age<19 then Age_cat="Teenagers";
  if Age>=19 then Age_cat="Adults";
  if Age = . then Age_cat="";
  if Pregnant="Yes" then Age_cat="Pregnant";
  if Lactating="Yes" then Age_cat="Lactating";
run; *n=5966;

proc freq data=iaea11; table Age_cat strata; run;
```

```
* Change TEE and BEE from MJ/day to Kcal/day;
data iaea12; set iaea11;
  label TEE='TEE MJ/day';
  rename TEE=TEE_MJ;
run;
data iaea12; set iaea12;
  TEE = (TEE_MJ*1000)/4.184; *TEE is now in kcal/day;
  BEE = (BEEmjday*1000)/4.184; *BEE is now in kcal/day;
  Weight = wt_init; *weight in kg.;
  Height = ht; *height in cm ;
  PAL = TEE/BEE;
  label TEE= 'TEE in kcal/day'
        BEE= 'BEE in kcal/day'
        PAL= 'PAL in kcal/day';
run;
```

```
* Create PAL category;
/** These cutoffs were used based on IOM 2005 report;
data iaea12; set iaea12;
  length PALCAT $8;
  if PAL<1 then PALCAT="PAL<1";
  if 1<=PAL<1.4 then PALCAT="S";
  if 1.4<=PAL<1.6 then PALCAT="LA";
  if 1.6<=PAL<1.9 then PALCAT="A";
```

```

    if 1.9<=PAL<2.5 then PALCAT="VA";
    if 2.5<=PAL then PALCAT="PAL>2.5";
    if PAL=. then PALCAT="";
run;*/
** These are the cutoffs defined by percentiles that WG1 emailed week of 8/17/22;
data iaea12; set iaea12;
    length PALCAT $8;
    if .<age<9.0 then do; *3.0<=age<9.0;
        if .<PAL<1 then PALCAT='PAL<1';
        if 1.0=<PAL<1.31 then PALCAT='S';
        if 1.31=<PAL<1.44 then PALCAT='LA';
        if 1.44=<PAL<1.59 then PALCAT='A';
        if 1.59=<PAL<2.5 then PALCAT='VA';
        if 2.5<=PAL then PALCAT='PAL>2.5';
    end;
    if 9.0=<age<14.0 then do;
        if .<PAL<1 then PALCAT='PAL<1';
        if 1.0=<PAL<1.44 then PALCAT='S';
        if 1.44=<PAL<1.59 then PALCAT='LA';
        if 1.59=<PAL<1.77 then PALCAT='A';
        if 1.77=<PAL<2.5 then PALCAT='VA';
        if 2.5<=PAL then PALCAT='PAL>2.5';
    end;
    if 14.0=<age<19.0 then do;
        if .<PAL<1 then PALCAT='PAL<1';
        if 1.0=<PAL<1.56 then PALCAT='S';
        if 1.56=<PAL<1.73 then PALCAT='LA';
        if 1.73=<PAL<1.92 then PALCAT='A';
        if 1.92=<PAL<2.5 then PALCAT='VA';
        if 2.5<=PAL then PALCAT='PAL>2.5';
    end;
    if 19.0=<age then do;
    *Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
        if .<PAL<1 then PALCAT='PAL<1';
        if 1.0=<PAL<1.53 then PALCAT='S';
        if 1.53=<PAL<1.68 then PALCAT='LA';
        if 1.68=<PAL<1.85 then PALCAT='A';
        if 1.85=<PAL<2.5 then PALCAT='VA';
        if 2.5<=PAL then PALCAT='PAL>2.5';
    end;
run;

proc freq data= iaea12; tables PALCAT; run;

* Calculate a BMR_kjd_Schofield . Divide Height by 100 because data in cm and put
in m for equations.;
* Where BEE and PAL are missing, these estimated measures of BMR
will be used as a "proxy" for BMR to use in multiple imputation.;
data iaea13; set iaea12;

```



```

    if .<Age<3 & Sex="M" then BMR_Mjd_Schofield = 0.0007*Weight +
6.349*Height/100 - 2.584;
    if 3<=Age<10 & Sex="M" then BMR_Mjd_Schofield = 0.082*Weight +
0.545*Height/100 + 1.736;
    if 10<=Age<18 & Sex="M" then BMR_Mjd_Schofield = 0.068*Weight +
0.574*Height/100 + 2.157;
    if 18<=Age<30 & Sex="M" then BMR_Mjd_Schofield = 0.063*Weight -
0.042*Height/100 + 2.953;
    if 30<=Age<60 & Sex="M" then BMR_Mjd_Schofield = 0.048*Weight -
0.011*Height/100 + 3.670;
    if 60<=Age & Sex="M" then BMR_Mjd_Schofield = 0.038*Weight +
4.068*Height/100 - 3.491;

    if Age<3 & Sex="F" then BMR_Mjd_Schofield = 0.068*Weight + 4.281*Height/100
- 1.730;
    if 3<=Age<10 & Sex="F" then BMR_Mjd_Schofield = 0.071*Weight +
0.677*Height/100 + 1.553;
    if 10<=Age<18 & Sex="F" then BMR_Mjd_Schofield = 0.035*Weight +
1.948*Height/100 + 0.837;
    if 18<=Age<30 & Sex="F" then BMR_Mjd_Schofield = 0.057*Weight +
1.184*Height/100 + 0.411;
    if 30<=Age<60 & Sex="F" then BMR_Mjd_Schofield = 0.034*Weight +
0.006*Height/100 + 3.530;
    if 60<=Age & Sex="F" then BMR_Mjd_Schofield = 0.033*Weight +
1.917*Height/100 + 0.074;

*Convert to kcal from MJ (BMR_kcal_Schofield);
BMR_kcal_Schofield = (BMR_Mjd_Schofield*1000)/4.184;

*Calculate PAL_est=TEE_kcal/BMR_kcal_Schofield;
PAL_est=TEE/BMR_kcal_Schofield;
label PAL_est= 'PAL estimated from BMR Schofield, in kcal/day';

* Create PAL_est category;
/*if PAL_est<1 then PALCAT_est="PAL<1";
if 1<=PAL_est<1.4 then PALCAT_est="S";
if 1.4<=PAL_est<1.6 then PALCAT_est="LA";
if 1.6<=PAL_est<1.9 then PALCAT_est="A";
if 1.9<=PAL_est<2.5 then PALCAT_est="VA";
if 2.5<=PAL_est then PALCAT_est="PAL>2.5";
if PAL_est=. then PALCAT_est="";*/

if .<age<9.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.31 then PALCAT_est='S';
    if 1.31<=PAL_est<1.44 then PALCAT_est='LA';
    if 1.44<=PAL_est<1.59 then PALCAT_est='A';
    if 1.59<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';

```

```

end;
if 9.0=<age<14.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.44 then PALCAT_est='S';
    if 1.44=<PAL_est<1.59 then PALCAT_est='LA';
    if 1.59=<PAL_est<1.77 then PALCAT_est='A';
    if 1.77=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 14.0=<age<19.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.56 then PALCAT_est='S';
    if 1.56=<PAL_est<1.73 then PALCAT_est='LA';
    if 1.73=<PAL_est<1.92 then PALCAT_est='A';
    if 1.92=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 19.0=<age then do;
    *Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.53 then PALCAT_est='S';
    if 1.53=<PAL_est<1.68 then PALCAT_est='LA';
    if 1.68=<PAL_est<1.85 then PALCAT_est='A';
    if 1.85=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

** Combine estimate and observed PAL;
if PAL ne . then PAL_obs_est=PAL;
if PAL = . then PAL_obs_est=PAL_est;
if PALCAT ne '' then PALCAT_obs_est=PALCAT;
if PALCAT = '' then PALCAT_obs_est=PALCAT_est;

run; *n=5966;

*Calculate BMI categories and Life stages, ethnicity;
data iaea13; set iaea13; *n=5966;
    length BMICat $25;
    length Life_Stage $25;
    length Ethnicity $50;

    if BMI=. then BMI=Weight/((Height/100)**2); /*Calculate BMI for those*/

    if .<BMI<18.5 then BMICat="< 18.5 kg/m2";
    if 18.5<=BMI<=24.9999 then BMICat=">= 18.5 & <25 kg/m2";
    if 24.9999<BMI<30 then BMICat=">= 25 & <30 kg/m2";
    if 30<=BMI<35 then BMICat=">= 30 & <35 kg/m2";
    if 35<=BMI<40 then BMICat=">= 35 & <40 kg/m2";

```

```

if 40<=BMI then BMIcat=">= 40 kg/m2";
if BMI=. then BMIcat="";

if .<Age<0.5833 then Life_stage="0-6 months";
if 0.5833<=Age<1 then Life_stage="7-11 months";
if 1<=Age<4 then Life_stage="1-3 years";
if 4<=Age<9 then Life_stage="4-8 years";
if 9<=Age<14 then Life_stage="9-13 years";
if 14<=Age<19 then Life_stage="14-18 years";
if 19<=Age<31 then Life_stage="19-30 years";
if 31<=Age<51 then Life_stage="31-50 years";
if 51<=Age<71 then Life_stage="51-70 years";
if Age>=71 then Life_stage=">=71 years";
if Age=. then Life_stage="";

if eth1="A" then Ethnicity="African";
if eth1="AA" then Ethnicity="African American";
if eth1="AS" then Ethnicity="Asian";
if eth1="C" then Ethnicity="Caucasian";
if eth1="H" then Ethnicity="Hispanic";
if eth1="O" then Ethnicity="Other";
if eth1="NA" then Ethnicity="Unknown or data not available";

run;

/*data trial;set iaea2; where DBID=2028;run;*/

proc freq data=iaea13; tables eth1; run;

*Calculate FM and FM_percent;
data iaea13; set iaea13; *n=5966;
    FM_pct=_FM;
    *FM2=Weight*FM_pct;
    FM=(FM_pct/100*FFM)/(1-FM_pct/100);
    *FMcheck=Weight-FFM;
    *FM_pctcheck=FMcheck/(FMcheck+FFM);
run;

data iaea14; set iaea13;

    /*Remove people without Age or sex*/
    if Age ne . and Sex ne "";

    /*Change format of numeric for TEE*/
    format TEE 15.6;

    /*Create unique ID variable from Study ID and IND ID*/
    SID = CATX("-", stud_n, IND_ID);

```

```

run; *n=5805;

/*Create clean dataset , removing inelligible PAL values

In infants (0 to 11.9 mo) AND CHILDREN 1-3 YEAR
    If PAL (observed) is <1 (n=9) or >2.5 (n=0), then drop that infant
    If PAL estimated from Schofield is <1 (n=34) or >2.5 (n=0) then drop that
infant.

For children (4+) & adults,
    if observed PAL >2.5 (n=54) then drop that person.
    If observed PAL < 1 (n=9) then make PAL missing and use multiple imputation
        MI will bound PAL to be set 1 to 2.5
    if unobserved PAL, but PAL estimated from Schofield >2.5 (n=~30?) then drop
these people
    These are people with very high TEE for their age & wt
    if unobserved PAL, but PAL estimated from Schofield <1 (n=~45), leave them
be because PAL will be imputed.
*/
proc freq data=iaea14;
    table PALCAT;
run;

data iaea_final; set iaea14;
    if .<Age<4 & (.<PAL<1 or PAL>2.5) then Not_include=1;
    if .<Age<4 & (.<PAL_est<1 or PAL_est>2.5) then Not_include=1;
    if Age>=4 & PAL>2.5 then Not_include=1;
    if Age>=4 & .<PAL<1 then do; PAL=. ; PALCAT=""; end; *Make missing, to be
imputed;
    if Age>=4 & missing(PAL) & PAL_est>2.5 then Not_include=1;
run;*n=5805;

/*Clean dataset*/
data iaea_clean; set iaea_final;
    if Not_include not in (1,2);
    drop Not_include;
run;*n=5717;

proc freq data=iaea_clean;
    table Strata PALCAT PALCAT_est;
run;
*47 are missing PALCAT_est because either missing BMR (missing height) or TEE;

/*****/
/* SAVE IAEA dataset (before filtering PAL) */

* As SAS data;
data "&mypath.\Data\Clean data for analysis\IAEA_&SYSDATE9..sas7bdat";
    set iaea_final;

```

```

run;

* As csv data;
proc export data=iaea_final
  outfile="&mypath.\Data\Clean data for analysis\IAEA_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;

/* SAVE IAEA clean analysis dataset (after filtering PAL)*/

* As SAS data;
data "&mypath.\Data\Clean data for analysis\IAEA_clean_&SYSDATE9..sas7bdat";
  set iaea_clean;
run;

* As csv data;
proc export data=iaea_clean
  outfile="&mypath.\Data\Clean data for analysis\IAEA_clean_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;

/*****
*****/
/***** IOM *****/
*****/
/*****
*****/

/*****/
/* TABLE I-1 Infants and Very Young Children (0 Through 2 Years of Age) Within the
3rd to 97th Percentile for Body Mass Index (BMI)*/
* n=320;
* Age(y) Height(m) Weight(kg) BEEo(kcal/d) BEEp(kcal/d) TEE(kcal/d) TEE/kg BMRo/kg
BMp/kg ;
PROC IMPORT OUT= T1
          DATAFILE=
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
          DBMS=xlsx REPLACE;
          SHEET="Table I-1";
          GETNAMES=YES;
RUN;

*Set percentile group;
data t1; set t1;

```

```

length Percentile_group2 $25;
Percentile_group2='3rd to 97th';
PALo=TEE/BEEo; *kcal/d; *Note that we shouldnt use PAL for infants but just
in case.;
run;

/* TABLE I-2 Normal Weight Children, 3 Through 18 Years of Age with Body Mass
Index (BMI) ? 85th Percentile */
* n=525;
* Note that this PALCAT is what they used in 2005, but not what we will use for 2022
so drop it;
PROC IMPORT OUT= T2
                DATAFILE=
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
                DBMS=xlsx REPLACE;
                SHEET="Table I-2";
                GETNAMES=YES;
RUN;

*Set percentile group;
data t2; set t2;
        drop PALCAT; *This was from 2005 IOM report. We are re-classifying below;
        length Percentile_group2 $25;
        Percentile_group2='<85th';
run;

/* TABLE I-3 Normal Weight Adults with Body Mass Index (BMI) from 18.5 up to 25
kg/m2 */
* n=407;
* Note that this PALCAT is what they used in 2005, but not what we will use for 2022
so drop it;
PROC IMPORT OUT= T3
                DATAFILE=
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
                DBMS=xlsx REPLACE;
                SHEET="Table I-3";
                GETNAMES=YES;
RUN;

data t3; set t3;
        drop PALCAT; *This was from 2005 IOM report. We are re-classifying below;
run;

/* TABLE I-4 Pregnant Women with Prepregnancy Body Mass Index (BMI) from 18.5 up to
25 kg/m2 */
* n=22;
* Gestation(wk), Age(y), Height(m), weight(kg), BMI, BEEo(kcal/d), TEE(kcal/d),
PALo, PALCAT;
* Note that this PALCAT is what they used in 2005, but not what we will use for 2022
so drop it;

```

```

PROC IMPORT OUT= T4
              DATAFILE=
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
              DBMS=xlsx REPLACE;
              SHEET="Table I-4";
              GETNAMES=YES;
RUN;

data t4;
  length SID Strata P_stage $25;
  set t4;
  drop PALCAT; *This was from 2005 IOM report. We are re-classifying below;
  Sex="F";
  Pregnant="Yes";
  Strata='Preg/Lac';
  if Gestation="6" then do; P_stage="1st Trimester"; weeks=6; end;
  if Gestation="8-10" then do; P_stage="1st Trimester"; weeks=9; end;
  if Gestation="12" then do; P_stage="1st Trimester"; weeks=12; end;
  if Gestation="16-18" then do; P_stage="2nd Trimester"; weeks=17; end;
  if Gestation="18" then do; P_stage="2nd Trimester"; weeks=18; end;
  if Gestation="24" then do; P_stage="2nd Trimester"; weeks=24; end;
  if Gestation="24-26" then do; P_stage="2nd Trimester"; weeks=25; end;
  if Gestation="30" then do; P_stage="3rd Trimester"; weeks=30; end;
  if Gestation="34-36" then do; P_stage="3rd Trimester"; weeks=35; end;
  if Gestation="36" then do; P_stage="3rd Trimester"; weeks=36; end;
  SID="IOM_I4_"||left(trim(_N_));
run;

proc freq data=t4; tables strata Gestation*P_stage/nopercent norow nocol;run;

/* TABLE I-5 Lactating Women with Prepregnancy Body Mass Index (BMI) from 18.5 up
to 25 kg/m2 */
* n=35;
PROC IMPORT OUT= T5
              DATAFILE=
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
              DBMS=xlsx REPLACE;
              SHEET="Table I-5";
              GETNAMES=YES;
RUN;

data t5;
  length SID Strata P_stage $25 ;
  set t5;
  *drop PALCAT; *There was no PALCAT for Lactating women;
  Sex="F";
  Strata='Preg/Lac';
  Lactating="Yes";
  if LactationMonth="1" then do; P_stage="Lactating 1-3 mo"; Weeks=4.3; end;
  if LactationMonth="1-1.5" then do; P_stage="Lactating 1-3 mo"; Weeks=5.4;

```

```

end;
    if LactationMonth="2" then do; P_stage="Lactating 1-3 mo"; Weeks=8.7; end;
    if LactationMonth="3" then do; P_stage="Lactating 1-3 mo"; Weeks=13; end;
    if LactationMonth="3-6.5" then do; P_stage="Lactating 4-6 mo"; Weeks=20.6;
end;
    if LactationMonth="6" then do; P_stage="Lactating 4-6 mo"; Weeks=26; end;
    SID="IOM_I5_"||left(trim(_N_));
run;

proc freq data=t5; tables Strata LactationMonth;run;

/* BONUS DATA !! */
/* WG1 sent MORE Data for Pregnant and Lactating week of August 17 */
/* Nancy made a sheet for Preg Lac data from IOM2005 which were not previously
included because the tables in the IOM report
    only included healthy weights, so there are more here */
/* She also CNRC pregnancy data*/

PROC IMPORT OUT=PregLacUWOW
            DATAFILE= "&mypath.\Data\Pregnancy
Data\10_DLW_Database_IOM2005 Preg Lac Weeks 2022-08-20.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="PregLacUWOW";
            GETNAMES=YES;
RUN;
data PregLacUwow;
    length SID Strata P_stage $25.;
    Sex="F";
    Strata='Preg/Lac';
    set PreglacUwow (rename=(BMRo_kcal_d=BEEo TEE_kcal_d=TEE)
                    drop=BMRo_MJ_d BMRe_MJ_d BMRe_kcal_d
BMRe_Henry TEE_MJ_d--BMRe_kg PALe);

SID=trim(left(substr(Author,1,8)))||trim(left(Study))||"_"||trim(left(put(SubID,4.)
));
run;

PROC IMPORT OUT=PregLac
            DATAFILE= "&mypath.\Data\Pregnancy
Data\10_DLW_Database_IOM2005 Preg Lac Weeks 2022-08-20.xlsx"
            DBMS=xlsx REPLACE;
            SHEET="PregLac";
            GETNAMES=YES;
RUN;
data Preglac;
    length SID $25.;
    Sex="F";
    Strata='Preg/Lac';
    set Preglac (rename=(BMRo_kcal_d=BEEo TEE_kcal_d=TEE)
                drop=BMRo_MJ_d BMRe_MJ_d BMRe_kcal_d

```



```

BMRe_Henry TEE_MJ_d--BMRe_kg PALe);
    if nmiss(Age, Height, Weight, BMI, BEEo, TEE, PALo)=7 then delete;

SID=trim(left(substr(Author,1,8)))||trim(left(Study))||"_"||trim(left(put(SubID,4.)
));
run;

data PLcombo;
    set t4 t5 PregLacUwow PregLac ;
    if Pregnant="Yes" then Code="P"; *For t4 and t5;
    if Lactating="Yes" then Code="L";
    *Status and Code are almost identical but not quite;
*   if missing(Status) then Status=Code;
*   if missing(Code) then Code=Status;
    Beeround=round(BEEo,1);
    TEERound=round(TEE,1);
    Ageround=round(Age,1);
    BMIround=round(BMI,.1);
    format BEEo TEE Age Beeround TEERound Ageround 8.2;
    drop Activity Occupation Notes Time;
run;*n=421;

proc sort data=PLcombo out=PLcombo2 nodupkey;
    by TEERound BEEround Ageround BMIround; *39 duplicates were removed as
identical;
run; *n=382;
proc freq data=PLcombo2;
    table status*code/norow nocol nopercnt;
run;

data PLcombo3;
    set PLcombo2;
    Sex="F";
    Strata='Preg/Lac';
    if Status="P" or Code="P" then Pregnant="Yes";
    if Status="L" or Code="L" then Lactating="Yes";
    if Status ne "P" and Code ne "P" then Pregnant="No";
    if Status ne "L" and Code ne "L" then Lactating="No";
    if (Status="P" or Code="P") and 1<=Weeks<14 then P_Stage="1st Trimester";
    if (Status="P" or Code="P") and 14<=Weeks<28 then P_Stage="2nd Trimester";
    if (Status="P" or Code="P") and 28<=Weeks<=40 then P_Stage="3rd Trimester";

    if (Status="L" or Code="L") and 1<=Weeks<14 then P_Stage="Lactating 1-3 mo";
    if (Status="L" or Code="L") and 14<=Weeks<=27 then P_Stage="Lactating 4-6
mo";
    if Status="N" then P_Stage="NPNL";
    if Status="PL" then P_Stage="NPNL"; *Nancy said via email 8/20 that these
are NPNL;
    if missing(Status) and Code="N" then P_Stage="NPNL";
    if nmiss(Age, Height, Weight, BMI, BEEo, TEE, PALo)=7 then delete;

```

```
keep SID Strata Sex Age Height Weight BMI BEEo TEE PALo Weeks Status Code
Pregnant Lactating P_stage;
run; *n=382;
```

```
proc freq data=PLcombo3; table Strata code status P_Stage; run;
proc means data=PLcombo3 maxdec=2;
var Age height weight BEEo TEE PALo;
run;
```

```
proc freq data=PLcombo3;
table P_stage*(pregnant lactating)/norow nocol nopercnt;
run;
```

```
/******
*****/
```

```
/* TABLE I-6 Overweight/Obese Children, 3 Through 18 Years of Age, with Body Mass
Index (BMI) > 85th Percentile */
```

```
* n=319;
```

```
PROC IMPORT OUT= T6
```

```
DATAFILE=
```

```
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
```

```
DBMS=xlsx REPLACE;
```

```
SHEET="Table I-6";
```

```
GETNAMES=YES;
```

```
RUN;
```

```
*Set percentile group;
```

```
data t6; set t6;
```

```
drop PALCAT; *This was from 2005 IOM report. We are re-classifying below;
Percentile_group2='>85th';
```

```
run;
```

```
/* TABLE I-7 Overweight/Obese Adults with Body Mass Index (BMI) > 25 kg/m2 */
/*Age (y) ,Height (m),Weight (kg) , BMI, BEEo, (kcal/d) */
```

```
* n=360;
```

```
PROC IMPORT OUT= T7
```

```
DATAFILE=
```

```
"&mypath.\Data\IOM\IOM2005_AppendixI_Tables_JM_LGA.xlsx"
```

```
DBMS=xlsx REPLACE;
```

```
SHEET="Table I-7";
```

```
GETNAMES=YES;
```

```
RUN;
```

```
data t7; set t7;
```

```
drop PALCAT; *This was from 2005 IOM report. We are re-classifying below;
```

```
run;
```

```
* Step extra : The BMI categories for kids <18 should be based on percentiles <5th,
```

```
5th to 85th, 85th to 95th, >95th ;
data kids; set t1 t2 t6; run; *n=1164;
```

```
*Calculate age month, round to 0.5;
data kids; set kids;
    age_mo = round(Age*12,0.5);
    int_mo = int(age_mo);
    if age_mo=int_mo then age_mo5=age_mo+0.5;
    else age_mo5=age_mo;
    *age_mo5 = round(Age*12)+0.5;
    drop age_mo int_mo;
run;
```

```
*Calculate length, round to 0.5;
data kids; set kids;
    ht_mo = round((Height*100),0.1); *Height was in m. This puts in cm to match
length in growth tables;
    *int_mo = int(ht_mo);
    *if ht_mo=int_mo then ht_5=ht_mo+0.5;
    *else ht_5=ht_mo;
    *drop ht_mo int_mo;
run;*1164;
```

```
*I'll subset the data by girls, boys and babies;
data babygirlsiom; set kids; where Age ne . & Age<2 & Sex='F'; run; *n=182;
data babyboysiom; set kids; where Age ne . & Age<2 & Sex='M'; run; *n=107;
data girlsiom; set kids; where Age ne . & 2<=Age<19 & Sex='F'; run; *n=570;
data boysiom; set kids; where Age ne . & 2<=Age<19 & Sex='M'; run; *n=305;
```

```
*Sort to merge;
proc sort data=girlsiom; by age_mo5; run;
proc sort data=boysiom; by age_mo5; run;
proc sort data=bmigirls; by Age_months; run;
proc sort data=bmiboys; by Age_months; run;
```

```
*Calculate percentile;
data boysiom2;
    merge boysiom (in=x rename=(age_mo5=age_months)) bmiboys;
    by Age_months;
    if x;
    length Percentile_group $25;
    if 0<=bmi<=Pct3 then do; Percentile =((BMI-0)/(Pct3 -0))*(3 -0)+0 ; if
Percentile<0 or Percentile>3 then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct3 <=bmi<=Pct5 then do; Percentile =((BMI-Pct3)/(Pct5 -Pct3))*(5 -3
)+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=bmi<=Pct10 then do; Percentile =((BMI-Pct5)/(Pct10-Pct5))*(10-5
)+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
```

```

Percentile= ; end;
    if Pct10<=bmi<=Pct25 then do; Percentile
=((BMI-Pct10)/(Pct25-Pct10))*(25-10)+10; if Percentile<10 or Percentile>25 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=bmi<=Pct50 then do; Percentile
=((BMI-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct50<=bmi<=Pct75 then do; Percentile
=((BMI-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct75<=bmi<=Pct85 then do; Percentile
=((BMI-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct85<=bmi<=Pct90 then do; Percentile
=((BMI-Pct85)/(Pct90-Pct85))*(90-85)+85; if Percentile<85 or Percentile>90 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct90<=bmi<=Pct95 then do; Percentile
=((BMI-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct95<=bmi<=Pct97 then do; Percentile
=((BMI-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
    if .<Pct97<=bmi<=50 then do; Percentile =((BMI-Pct97)/(
100-Pct97))*(100-97)+97;if Percentile<97 or Percentile>100 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
    if Percentile<5 then Percentile_group='<5th';
    if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
    if Percentile>=85 & Percentile<95 then Percentile_group='85th to 95th';
    if Percentile>=95 then Percentile_group='>95th';
    if Percentile=. then Percentile_group='';
drop Pct3 Pct5 Pct10 Pct25 Pct50 Pct75 Pct85 Pct90 Pct95 Pct97 pat_person_nbr;
run;

```

```

data girlsiom2;
    merge girlsiom (in=x rename=(age_mo5=age_months)) bmigirls;
    by Age_months;
    if x;
    length Percentile_group $25;
    if 0<=bmi<=Pct3 then do; Percentile =((BMI-0)/(Pct3 -0))*(3 -0)+0 ; if
Percentile<0 or Percentile>3 then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct3 <=bmi<=Pct5 then do; Percentile =((BMI-Pct3)/(Pct5 -Pct3))*(5 -3
)+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=bmi<=Pct10 then do; Percentile =((BMI-Pct5)/(Pct10-Pct5))*(10-5
)+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=bmi<=Pct25 then do; Percentile
=((BMI-Pct10)/(Pct25-Pct10))*(25-10)+10; if Percentile<10 or Percentile>25 then put
"WARNING: " pat_person_nbr= Percentile= ; end;

```

```

        if Pct25<=bmi<=Pct50 then do; Percentile
= ((BMI-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct50<=bmi<=Pct75 then do; Percentile
= ((BMI-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct75<=bmi<=Pct85 then do; Percentile
= ((BMI-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct85<=bmi<=Pct90 then do; Percentile
= ((BMI-Pct85)/(Pct90-Pct85))*(90-85)+85; if Percentile<85 or Percentile>90 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct90<=bmi<=Pct95 then do; Percentile
= ((BMI-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if Pct95<=bmi<=Pct97 then do; Percentile
= ((BMI-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then put
"WARNING: " pat_person_nbr= Percentile= ; end;
        if .<Pct97<=bmi<=50 then do; Percentile = ((BMI-Pct97)/(
100-Pct97))*(100-97)+97; if Percentile<97 or Percentile>100 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
        if Percentile<5 then Percentile_group='<5th';
        if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
        if Percentile>=85 & Percentile<95 then Percentile_group='85th to 95th';
        if Percentile>=95 then Percentile_group='>95th';
        if Percentile=. then Percentile_group='';
drop Pct3 Pct5 Pct10 Pct25 Pct50 Pct75 Pct85 Pct90 Pct95 Pct97 pat_person_nbr;
run;

```

* Infants percentile;

*Sort to merge;

```

proc sort data=babygirlsiom; by ht_mo; run;
proc sort data=babyboysiom; by ht_mo; run;
proc sort data=wlgirls; by Length; run;
proc sort data=wlboys; by Length; run;

```

*Calculate percentile; *ht_mo is in cm, and length is in cm;

```

data babyboysiom2;
    merge babyboysiom (in=x rename=(ht_mo=Length)) wlboys;
    by Length;
    if x;
    length Percentile_group $25;
    length Percentile_infant $25;
    if 0<=Weight<=Pct01 then do; Percentile = ((Weight-0 )/(Pct01 -0 ))*(0.1 -0
)+0 ; if Percentile<0 or Percentile>0.1 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct01 <=Weight<=Pct1 then do; Percentile = ((Weight-Pct01 )/(Pct1 -Pct01
))*(1 -0.1 )+0.1 ; if Percentile<0.1 or Percentile>1 then put "WARNING: "

```

```

pat_person_nbr= Percentile= ; end;
    if Pct1 <=Weight<=Pct3 then do; Percentile =((Weight-Pct1 )/(Pct3 -Pct1
))*(3 -1 )+1 ; if Percentile<1 or Percentile>3 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct3 <=Weight<=Pct5 then do; Percentile =((Weight-Pct3 )/(Pct5 -Pct3
))*(5 -3 )+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=Weight<=Pct10 then do; Percentile =((Weight-Pct5 )/(Pct10-Pct5
))*(10-5 )+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=Weight<=Pct15 then do; Percentile
=((Weight-Pct10)/(Pct15-Pct10))*(15-10)+10; if Percentile<10 or Percentile>15 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct15<=Weight<=Pct25 then do; Percentile
=((Weight-Pct15)/(Pct25-Pct15))*(25-15)+15; if Percentile<15 or Percentile>25 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=Weight<=Pct50 then do; Percentile
=((Weight-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct50<=Weight<=Pct75 then do; Percentile
=((Weight-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct75<=Weight<=Pct85 then do; Percentile
=((Weight-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct85<=Weight<=Pct90 then do; percentile
=((Weight-Pct85)/(Pct90-Pct85))*(90-85)+85; if percentile<85 or percentile>90 then
put "WARNING: " pat_person_nbr= percentile= ; end;
    if Pct90<=Weight<=Pct95 then do; Percentile
=((Weight-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct95<=Weight<=Pct97 then do; Percentile
=((Weight-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct97<=Weight<=Pct99 then do; Percentile
=((Weight-Pct97)/(Pct99-Pct97))*(99-97)+97; if Percentile<97 or Percentile>99 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct99<=Weight<=Pct999 then do; Percentile
=((Weight-Pct99)/(Pct999-Pct99))*(99.9-99)+99; if Percentile<99 or Percentile>99.9
then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if .<Pct999<=Weight<=50 then do; Percentile =((Weight-Pct999)/(
100-Pct999))*(100-99.9)+99.9;if Percentile<99.9 or Percentile>100 then put "WARNING:
" pat_person_nbr= Percentile= ; end;
    if Percentile<5 then Percentile_group='<5th';
    if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
    if Percentile>=85 & Percentile<=95 then Percentile_group='85th to 95th';
    if Percentile>95 then Percentile_group='>95th';
    if Percentile=. then Percentile_group='';
    if Percentile<3 then Percentile_infant='<3th';
    if Percentile>=3 & Percentile<=97 then Percentile_infant='3th to 97th';

```

```

    if Percentile>97 then Percentile_infant='>97th';
    if Percentile=. then Percentile_infant='';
    drop L M S Pct01 Pct1 Pct3 Pct5 Pct10 Pct15 Pct25 Pct50 Pct75 Pct85 Pct90
Pct95 Pct97 Pct99 Pct999 pat_person_nbr; * ;
run;

```

```

data babygirls2; *ht_mo is in cm, so length is in cm;
    merge babygirls1 (in=x rename=(ht_mo=Length)) wlgirls;
    by Length;
    if x;
    length Percentile_group $25;
    length Percentile_infant $25;
    if 0<=Weight<=Pct01 then do; Percentile =((Weight-0)/(Pct01 -0))*(0.1 -0
)+0 ; if Percentile<0 or Percentile>0.1 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct01 <=Weight<=Pct1 then do; Percentile =((Weight-Pct01)/(Pct1 -Pct01
))*(1 -0.1)+0.1 ; if Percentile<0.1 or Percentile>1 then put "WARNING: "
pat_person_nbr= Percentile= ; end;
    if Pct1 <=Weight<=Pct3 then do; Percentile =((Weight-Pct1)/(Pct3 -Pct1
))*(3 -1)+1 ; if Percentile<1 or Percentile>3 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct3 <=Weight<=Pct5 then do; Percentile =((Weight-Pct3)/(Pct5 -Pct3
))*(5 -3)+3 ; if Percentile<3 or Percentile>5 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct5 <=Weight<=Pct10 then do; Percentile =((Weight-Pct5)/(Pct10-Pct5
))*(10-5)+5 ; if Percentile<5 or Percentile>10 then put "WARNING: " pat_person_nbr=
Percentile= ; end;
    if Pct10<=Weight<=Pct15 then do; Percentile
=((Weight-Pct10)/(Pct15-Pct10))*(15-10)+10; if Percentile<10 or Percentile>15 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct15<=Weight<=Pct25 then do; Percentile
=((Weight-Pct15)/(Pct25-Pct15))*(25-15)+15; if Percentile<15 or Percentile>25 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct25<=Weight<=Pct50 then do; Percentile
=((Weight-Pct25)/(Pct50-Pct25))*(50-25)+25; if Percentile<25 or Percentile>50 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct50<=Weight<=Pct75 then do; Percentile
=((Weight-Pct50)/(Pct75-Pct50))*(75-50)+50; if Percentile<50 or Percentile>75 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct75<=Weight<=Pct85 then do; Percentile
=((Weight-Pct75)/(Pct85-Pct75))*(85-75)+75; if Percentile<75 or Percentile>85 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct85<=Weight<=Pct90 then do; percentile
=((Weight-Pct85)/(Pct90-Pct85))*(90-85)+85; if percentile<85 or percentile>90 then
put "WARNING: " pat_person_nbr= percentile= ; end;
    if Pct90<=Weight<=Pct95 then do; Percentile
=((Weight-Pct90)/(Pct95-Pct90))*(95-90)+90; if Percentile<90 or Percentile>95 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct95<=Weight<=Pct97 then do; Percentile

```

```

=((Weight-Pct95)/(Pct97-Pct95))*(97-95)+95; if Percentile<95 or Percentile>97 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct97<=Weight<=Pct99 then do; Percentile
=((Weight-Pct97)/(Pct99-Pct97))*(99-97)+97; if Percentile<97 or Percentile>99 then
put "WARNING: " pat_person_nbr= Percentile= ; end;
    if Pct99<=Weight<=Pct999 then do; Percentile
=((Weight-Pct99)/(Pct999-Pct99))*(99.9-99)+99; if Percentile<99 or Percentile>99.9
then put "WARNING: " pat_person_nbr= Percentile= ; end;
    if .<Pct999<=Weight<=50 then do; Percentile =((Weight-Pct999)/(
100-Pct999))*(100-99.9)+99.9;if Percentile<99.9 or Percentile>100 then put "WARNING:
" pat_person_nbr= Percentile= ; end;
    if Percentile<5 then Percentile_group='<5th';
    if Percentile>=5 & Percentile<85 then Percentile_group='5th to 85th';
    if Percentile>=85 & Percentile<=95 then Percentile_group='85th to 95th';
    if Percentile>95 then Percentile_group='>95th';
    if Percentile=. then Percentile_group='';
    if Percentile<3 then Percentile_infant='<3th';
    if Percentile>=3 & Percentile<=97 then Percentile_infant='3th to 97th';
    if Percentile>97 then Percentile_infant='>97th';
    if Percentile=. then Percentile_infant='';
    drop L M S Pct01 Pct1 Pct3 Pct5 Pct10 Pct15 Pct25 Pct50 Pct75 Pct85 Pct90
Pct95 Pct97 Pct99 Pct999 pat_person_nbr; *;
run;

```

```

** Merge them back together; *n=1164;
data kids2; set babygirlsiom2 babyboysiom2 girlsiom2 boysiom2; run;

```

```

/*****
/**** MERGE ALL IOM DATA ****

```

```

* Merging all tables together;
* t3 is healthy adults, t7 is overweight adults;
data iom1;
    set t3 t7 kids2 PLcombo3;
run;*n=2313;

```

```

* Create age category variable according to Life Stages document as WG requested;
* But note that 2005 IOM models use Infants 0-2.9, and Children 3.0 to 18.9;
data iom2; set iom1;

```

```

    length Age_cat $25 ;
    if Age<1 then Age_cat="Infants";
    if 1<=Age<9 then Age_cat="Children";
    if 9<=Age<19 then Age_cat="Teenagers";
    if Age>=19 then Age_cat="Adults";
    if Pregnant="Yes" then Age_cat="Pregnant";
    if Lactating="Yes" then Age_cat="Lactating";

    if Strata ne 'Preg/Lac' /*These were coded above*/

```



```

then do;
    if 0<Age<3.0 then Strata='Infant';
    if 3.0<=Age<19 and Sex='F' then Strata='Child Girls 3-18';
    if 3.0<=Age<19 and Sex='M' then Strata='Child Boys 3-18';
    if 19<=Age and Sex='F' then Strata='Adult Women 19+';
    if 19<=Age and Sex='M' then Strata='Adult Men 19+';
end;

BEE = BEEo;
PAL = PALo;
run;*n=2313;
proc freq data=iom2; tables Age_cat Strata; run;

* Create PAL category;
*These are new PAL cutoffs WG1 sent the week of 8/17;
data iom2; set iom2;
if .<age<9.0 then do; *3.0<=age<9.0;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0<=PAL<1.31 then PALCAT='S';
    if 1.31<=PAL<1.44 then PALCAT='LA';
    if 1.44<=PAL<1.59 then PALCAT='A';
    if 1.59<=PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;
if 9.0<=age<14.0 then do;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0<=PAL<1.44 then PALCAT='S';
    if 1.44<=PAL<1.59 then PALCAT='LA';
    if 1.59<=PAL<1.77 then PALCAT='A';
    if 1.77<=PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;
if 14.0<=age<19.0 then do;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0<=PAL<1.56 then PALCAT='S';
    if 1.56<=PAL<1.73 then PALCAT='LA';
    if 1.73<=PAL<1.92 then PALCAT='A';
    if 1.92<=PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;
if 19.0<=age then do;
*Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0<=PAL<1.53 then PALCAT='S';
    if 1.53<=PAL<1.68 then PALCAT='LA';
    if 1.68<=PAL<1.85 then PALCAT='A';
    if 1.85<=PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;

```

```
rename height=height_m;
run;
```

```
* Calculate a variable BMR_kjd_Schofield as an estimate for BEE;
* Note that final equations don't use this estimate;
* This is just an estimate to look for extreme low and high PAL to filter data;
* Equations from Nancy email 7/24. NOTE: BMR is in MJ/d. Weight is in kg. Height var
is in m - convert to cm to match iom and solnas;
data iom2; set iom2;
```

```
length PALCAT_est $25;
```

```
height=height_m*100; *convert height to cm to match iaea and solnas;
```

```
if .<Age<3 & Sex="M" then BMR_Mjd_Schofield = 0.0007*Weight +
6.349*Height/100 - 2.584;
if 3<=Age<10 & Sex="M" then BMR_Mjd_Schofield = 0.082*Weight +
0.545*Height/100 + 1.736;
if 10<=Age<18 & Sex="M" then BMR_Mjd_Schofield = 0.068*Weight +
0.574*Height/100 + 2.157;
if 18<=Age<30 & Sex="M" then BMR_Mjd_Schofield = 0.063*Weight -
0.042*Height/100 + 2.953;
if 30<=Age<60 & Sex="M" then BMR_Mjd_Schofield = 0.048*Weight -
0.011*Height/100 + 3.670;
if 60<=Age & Sex="M" then BMR_Mjd_Schofield = 0.038*Weight +
4.068*Height/100 - 3.491;
```

```
if .<Age<3 & Sex="F" then BMR_Mjd_Schofield = 0.068*Weight +
4.281*Height/100 - 1.730;
if 3<=Age<10 & Sex="F" then BMR_Mjd_Schofield = 0.071*Weight +
0.677*Height/100 + 1.553;
if 10<=Age<18 & Sex="F" then BMR_Mjd_Schofield = 0.035*Weight +
1.948*Height/100 + 0.837;
if 18<=Age<30 & Sex="F" then BMR_Mjd_Schofield = 0.057*Weight +
1.184*Height/100 + 0.411;
if 30<=Age<60 & Sex="F" then BMR_Mjd_Schofield = 0.034*Weight +
0.006*Height/100 + 3.530;
if 60<=Age & Sex="F" then BMR_Mjd_Schofield = 0.033*Weight +
1.917*Height/100 + 0.074;
```

```
*Convert to kcal from MJ (BMR_kcal_Schofield);
BMR_kcal_Schofield = (BMR_Mjd_Schofield*1000)/4.184;
```

```
*Calculate PAL_est=TEE_kcal/BMR_kcal_Schofield;
PAL_est=TEE/BMR_kcal_Schofield;
```

```
* Create PAL_est category;
/*if PAL_est<1 then PALCAT_est="PAL<1";
if 1<=PAL_est<1.4 then PALCAT_est="S";
if 1.4<=PAL_est<1.6 then PALCAT_est="LA";
```

```

if 1.6<=PAL_est<1.9 then PALCAT_est="A";
if 1.9<=PAL_est<2.5 then PALCAT_est="VA";
if 2.5<=PAL_est then PALCAT_est="PAL>2.5";
if PAL_est=. then PALCAT_est="";*/

```

```

if .<age<9.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.31 then PALCAT_est='S';
    if 1.31<=PAL_est<1.44 then PALCAT_est='LA';
    if 1.44<=PAL_est<1.59 then PALCAT_est='A';
    if 1.59<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

```

```

if 9.0<=age<14.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.44 then PALCAT_est='S';
    if 1.44<=PAL_est<1.59 then PALCAT_est='LA';
    if 1.59<=PAL_est<1.77 then PALCAT_est='A';
    if 1.77<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

```

```

if 14.0<=age<19.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.56 then PALCAT_est='S';
    if 1.56<=PAL_est<1.73 then PALCAT_est='LA';
    if 1.73<=PAL_est<1.92 then PALCAT_est='A';
    if 1.92<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

```

```

if 19.0<=age then do;
*Note that these are based on percentiles of 19 to 70.99, but 71+ use these

```

too;

```

    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.53 then PALCAT_est='S';
    if 1.53<=PAL_est<1.68 then PALCAT_est='LA';
    if 1.68<=PAL_est<1.85 then PALCAT_est='A';
    if 1.85<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

```

```

end;

```

```

** Combine estimate and observed PAL;
if PAL ne . then PAL_obs_est=PAL;
if PAL = . then PAL_obs_est=PAL_est;
if PALCAT ne '' then PALCAT_obs_est=PALCAT;
if PALCAT = '' then PALCAT_obs_est=PALCAT_est;

```

```

run; *n=2313;

```

```

*Calculate BMI categories and Life stages;
data iom3; set iom2;
    length BMICat $25  Life_Stage $25  Ethnicity $50;

    if .<BMI<18.5 then BMICat="< 18.5 kg/m2";
    if 18.5<=BMI<25 then BMICat=">= 18.5 & <25 kg/m2";
    if 25<=BMI<30 then BMICat=">= 25 & <30 kg/m2";
    if 30<=BMI<35 then BMICat=">= 30 & <35 kg/m2";
    if 35<=BMI<40 then BMICat=">= 35 & <40 kg/m2";
    if 40<=BMI then BMICat=">= 40 kg/m2";
    if BMI=. then BMICat="";

    if .<Age<0.5833 then Life_stage="0-6 months";
    if 0.5833<=Age<1 then Life_stage="7-11 months";
    if 1<=Age<4 then Life_stage="1-3 years";
    if 4<=Age<9 then Life_stage="4-8 years";
    if 9<=Age<14 then Life_stage="9-13 years";
    if 14<=Age<19 then Life_stage="14-18 years";
    if 19<=Age<31 then Life_stage="19-30 years";
    if 31<=Age<51 then Life_stage="31-50 years";
    if 51<=Age<71 then Life_stage="51-70 years";
    if Age>=71 then Life_stage=">=71 years";
    if Age=. then Life_stage="";

    Ethnicity="Unknown or data not available"; *Because there is no ethnicity in
this data;

run; *n=2313;

** Set as "No" for lactating and pregnant if not already filled in;
data iom3;set iom3;
if Pregnant='' then Pregnant='No';
if Lactating='' then Lactating='No';
run;

** Create Age in months and rename rounded age in months;
data iom3; set iom3; drop Age_months; run;
data iom3; set iom3; Age_months = Age*12;run;

** Create Subject ID (if not already done in Preg/Lac data);
data iom3; set iom3;
    if missing (SID) then SID = CATX("-", 'IOM', _N_);
;run;

/*Remove people without Age or sex*/
data iom4; set iom3; *n=2535;
    where Age ne . and Sex ne "";

    /*Change format of numeric for TEE*/

```

```

format TEE 15.6; run;
run;

*NOTE: Remove invalid PAL per discussion with WG1;
/*Create clean dataset

In infants (0 to 11.9 mo) AND CHILDREN 1-3 YEAR
If PAL (observed) is <1 (n=9) or >2.5 (n=0), then drop that infant
If PAL estimated from Schofield is <1 (n=34) or >2.5 (n=0) then drop that infant.

For children (4+) & adults,
if observed PAL >2.5 (n=54) then drop that person.
If observed PAL < 1 (n=9) then make PAL missing and use multiple imputation
MI will bound PAL to be set 1 to 2.5
if unobserved PAL, but PAL estimated from Schofield >2.5 (n=~30?) then drop these
people
These are people with very high TEE for their age & wt
if unobserved PAL, but PAL estimated from Schofield <1 (n=~45), leave them be
because PAL will be imputed.
*/

data iom_final; set iom4;
    if .<Age<4 & PAL ne . & (PAL<1 or PAL>2.5) then Not_include=1;
    if .<Age<4 & PAL_est ne . & (PAL_est<1 or PAL_est>2.5) then Not_include=1;
    if Age>=4 & PAL ne . & PAL>2.5 then Not_include=1;
    if Age>=4 & PAL ne . & PAL<1 then do; PAL=. ; PALCAT=""; end;
    if Age>=4 & PAL = . & PAL_est>2.5 then Not_include=1;
    *if Age ne . & Age>=4 & PAL = . & PAL_est<1 ;
run;*n=2313;

proc freq data=iom_final; tables strata age_cat PALCAT; run;

/*Clean dataset*/
data iom_clean; set iom_final;
where Not_include ne 1;
drop Not_include;
run;
*n=2283;
proc freq data=iom_clean; tables strata age_cat PALCAT; run;

/*****/
/* SAVE IOM dataset */

* As SAS data;
data "&mypath.\Data\Clean data for analysis\IOM_&SYSDATE9..sas7bdat";
    set iom_final;
run;

```

```

* As csv data;
proc export data=iom_final
  outfile="&mypath.\Data\Clean data for analysis\IOM_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;

/* SAVE IOM clean dataset */

* As SAS data;
data "&mypath.\Data\Clean data for analysis\IOM_clean_&SYSDATE9..sas7bdat";
  set iom_clean;
run;

* As csv data;
proc export data=iom_clean
  outfile="&mypath.\Data\Clean data for analysis\IOM_clean_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;

/*****
*****/
/*****                               CNRC Pregnancy
*****/
/*****
*****/
* Age (y), Weight (kg), Height (cm), BMR and TEE kcal/d ;

PROC IMPORT OUT=CNRC0
              DATAFILE= "&mypath.\Data\Pregnancy Data\CNRC Pregnancy w
weeks 2022-08-20.xlsx"
              DBMS=xlSX REPLACE;
  GETNAMES=YES;
RUN;
data CNRC1;
  length Strata $25.;
  set CNRC0 (rename=(BMRkcald_rcal=BEE TEE_DLW=TEE Maternal_Age=Age
Status=Status_n)
           drop= A VAR13 ID2);
  Sex="F";
  Strata='Preg/Lac';
  label BEE= TEE= Age= Status_n= ;
  if missing(Weeks) and SS=1 then Weeks=0;
  if missing(Weeks) and SS=2 then Weeks=9;
  if missing(Weeks) and SS=3 then Weeks=22;
  if missing(Weeks) and SS=4 then Weeks=36;
  if missing(Weeks) and SS=5 then Weeks=2; *But this is weeks Postpartum;
  if missing(Weeks) and SS=6 then Weeks=6; *But this is weeks Postpartum;

```

```

if missing(Weeks) and SS=7 then Weeks=27; *But this is weeks Postpartum;
if SS=1 then Code="N";
if SS in (2,3,4) then Code="P";
if SS in (5,6,7) and status_n=4 then Code="PF"; *Postpartum formula;
if SS in (5,6,7) and status_n=5 or PRESENTLY_BREASTFEEDING_=1 then Code="L";
* Postpartum breastfeeding;
if missing(TEE) then delete; * We dont necessarily have to delete if we want
to impute these?;
run; *n=222;
data CNRC1;
length SID dataset $12. P_stage $20.;
set CNRC1;
Age_cat="Pregnant";
if Code="P" then do; Pregnant="Yes"; Lactating="No "; end;
if Code="L" then do; Lactating="Yes"; Pregnant="No "; end;
if Code="P" and 1<=Weeks<14 then P_Stage="1st Trimester";
if Code="P" and 14<=Weeks<28 then P_Stage="2nd Trimester";
if Code="P" and 28<=Weeks<=40 then P_Stage="3rd Trimester";
if Code="L" and 1<=Weeks<14 then P_Stage="Lactating 1-3 mo";
if Code="L" and 14<=Weeks<=27 then P_Stage="Lactating 4-6 mo";
if Code="N" then do; P_Stage="NPNL"; Pregnant="No "; Lactating="No "; end;
if nmiss(Age, Height, Weight, BMI, BEE, TEE, PAL)=7 then delete;

SID="CNRC_"||trim(left(ID));
dataset="CNRC";
keep SID dataset Strata Sex Age Age_Cat Height Weight BMI BEE TEE Weeks Code
Pregnant Lactating P_stage Maternal_Race;
run; *n=222;

data CNRC2;
set CNRC1 ;

length PALCAT_est $8;
PAL=TEE/BEE;

if .<age<9.0 then do; *3.0<=age<9.0;
if .<PAL<1 then PALCAT='PAL<1';
if 1.0=<PAL<1.31 then PALCAT='S';
if 1.31=<PAL<1.44 then PALCAT='LA';
if 1.44=<PAL<1.59 then PALCAT='A';
if 1.59=<PAL<2.5 then PALCAT='VA';
if 2.5<=PAL then PALCAT='PAL>2.5';
end;
if 9.0=<age<14.0 then do;
if .<PAL<1 then PALCAT='PAL<1';
if 1.0=<PAL<1.44 then PALCAT='S';
if 1.44=<PAL<1.59 then PALCAT='LA';
if 1.59=<PAL<1.77 then PALCAT='A';
if 1.77=<PAL<2.5 then PALCAT='VA';
if 2.5<=PAL then PALCAT='PAL>2.5';

```

```

end;
if 14.0=<age<19.0 then do;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0=<PAL<1.56 then PALCAT='S';
    if 1.56=<PAL<1.73 then PALCAT='LA';
    if 1.73=<PAL<1.92 then PALCAT='A';
    if 1.92=<PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;
if 19.0=<age then do;
*Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0=<PAL<1.53 then PALCAT='S';
    if 1.53=<PAL<1.68 then PALCAT='LA';
    if 1.68=<PAL<1.85 then PALCAT='A';
    if 1.85=<PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;

* Height is in cm in the dataset, so divide by 100 for meters ;
* See equations in email from Nancy 7/24;
if .<Age<3 & Sex="M" then BMR_Mjd_Schofield = 0.0007*Weight +
6.349*Height/100 - 2.584;
if 3<=Age<10 & Sex="M" then BMR_Mjd_Schofield = 0.082*Weight +
0.545*Height/100 + 1.736;
if 10<=Age<18 & Sex="M" then BMR_Mjd_Schofield = 0.068*Weight +
0.574*Height/100 + 2.157;
if 18<=Age<30 & Sex="M" then BMR_Mjd_Schofield = 0.063*Weight -
0.042*Height/100 + 2.953;
if 30<=Age<60 & Sex="M" then BMR_Mjd_Schofield = 0.048*Weight -
0.011*Height/100 + 3.670;
if 60<=Age & Sex="M" then BMR_Mjd_Schofield = 0.038*Weight +
4.068*Height/100 - 3.491;

if .<Age<3 & Sex="F" then BMR_Mjd_Schofield = 0.068*Weight +
4.281*Height/100 - 1.730;
if 3<=Age<10 & Sex="F" then BMR_Mjd_Schofield = 0.071*Weight +
0.677*Height/100 + 1.553;
if 10<=Age<18 & Sex="F" then BMR_Mjd_Schofield = 0.035*Weight +
1.948*Height/100 + 0.837;
if 18<=Age<30 & Sex="F" then BMR_Mjd_Schofield = 0.057*Weight +
1.184*Height/100 + 0.411;
if 30<=Age<60 & Sex="F" then BMR_Mjd_Schofield = 0.034*Weight +
0.006*Height/100 + 3.530;
if 60<=Age & Sex="F" then BMR_Mjd_Schofield = 0.033*Weight +
1.917*Height/100 + 0.074;

*Convert to kcal from MJ (BMR_kcal_Schofield);
BMR_kcal_Schofield = (BMR_Mjd_Schofield*1000)/4.184;

```



```

*Calculate PAL_est=TEE_kcal/BMR_kcal_Schofield;
PAL_est=TEE/BMR_kcal_Schofield;

* Create PAL_est category;
if .<age<9.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.31 then PALCAT_est='S';
    if 1.31=<PAL_est<1.44 then PALCAT_est='LA';
    if 1.44=<PAL_est<1.59 then PALCAT_est='A';
    if 1.59=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 9.0=<age<14.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.44 then PALCAT_est='S';
    if 1.44=<PAL_est<1.59 then PALCAT_est='LA';
    if 1.59=<PAL_est<1.77 then PALCAT_est='A';
    if 1.77=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 14.0=<age<19.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.56 then PALCAT_est='S';
    if 1.56=<PAL_est<1.73 then PALCAT_est='LA';
    if 1.73=<PAL_est<1.92 then PALCAT_est='A';
    if 1.92=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 19.0=<age then do;
*Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0=<PAL_est<1.53 then PALCAT_est='S';
    if 1.53=<PAL_est<1.68 then PALCAT_est='LA';
    if 1.68=<PAL_est<1.85 then PALCAT_est='A';
    if 1.85=<PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

** Create a variable that combines estimate and observed PAL where
available;
if PAL ne . then PAL_obs_est=PAL;
if PAL = . then PAL_obs_est=PAL_est;
if PALCAT ne '' then PALCAT_obs_est=PALCAT;
if PALCAT = '' then PALCAT_obs_est=PALCAT_est;

run; *n=222 ;

```

```

*Calculate BMI categories and Life stages;
data CNRC3; set CNRC2;
    length BMICat $25 Life_Stage $25 Ethnicity $25;

    if .<BMI<18.5 then BMICat="< 18.5 kg/m2";
    if 18.5<=BMI<25 then BMICat=">= 18.5 & <25 kg/m2";
    if 25<=BMI<30 then BMICat=">= 25 & <30 kg/m2";
    if 30<=BMI<35 then BMICat=">= 30 & <35 kg/m2";
    if 35<=BMI<40 then BMICat=">= 35 & <40 kg/m2";
    if 40<=BMI then BMICat=">= 40 kg/m2";
    if BMI=. then BMICat="";

    if .<Age<0.5833 then Life_stage="0-6 months";
    if 0.5833<=Age<1 then Life_stage="7-11 months";
    if 1<=Age<4 then Life_stage="1-3 years";
    if 4<=Age<9 then Life_stage="4-8 years";
    if 9<=Age<14 then Life_stage="9-13 years";
    if 14<=Age<19 then Life_stage="14-18 years";
    if 19<=Age<31 then Life_stage="19-30 years";
    if 31<=Age<51 then Life_stage="31-50 years";
    if 51<=Age<71 then Life_stage="51-70 years";
    if Age>=71 then Life_stage=">=71 years";
    if Age=. then Life_stage="";

    if Maternal_race=1 then Ethnicity="Caucasian";
    if Maternal_race=2 then Ethnicity="African American";
    if Maternal_race=3 then Ethnicity="Hispanic";
    if Maternal_race=4 then Ethnicity="Other";
run; *222;

data CNRC_final; set CNRC3;
    if .<Age<4 & PAL ne . & (PAL<1 or PAL>2.5) then Not_include=1;
    if .<Age<4 & PAL_est ne . & (PAL_est<1 or PAL_est>2.5) then Not_include=1;
    if Age>=4 & PAL ne . & PAL>2.5 then Not_include=1;
    if Age>=4 & PAL ne . & PAL<1 then do; PAL=. ; PALCAT=""; end;
    if Age>=4 & PAL = . & PAL_est>2.5 then Not_include=1;
    *if Age ne . & Age>=4 & PAL = . & PAL_est<1 ;

run;*n=222;

proc freq data=CNRC_final; tables strata age_cat PALCAT; run;

/*Clean dataset*/
data CNRC_clean; set CNRC_final;
where Not_include ne 1;
drop Not_include;
run;
*n=220;
proc freq data=CNRC_clean; tables strata age_cat PALCAT; run;

```

```
/* *****  
/* SAVE CNRC dataset */
```

```
* As SAS data;  
data "&mypath.\Data\Clean data for analysis\CNRC_&SYSDATE9..sas7bdat";  
  set CNRC_final;  
run;
```

```
* As csv data;  
proc export data=CNRC_final  
  outfile="&mypath.\Data\Clean data for analysis\CNRC_&SYSDATE9..csv"  
  dbms=csv REPLACE;  
run;
```

```
* As SAS data;  
data "&mypath.\Data\Clean data for analysis\CNRC_clean_&SYSDATE9..sas7bdat";  
  set CNRC_clean;  
run;
```

```
* As csv data;  
proc export data=CNRC_clean  
  outfile="&mypath.\Data\Clean data for analysis\CNRC_clean_&SYSDATE9..csv"  
  dbms=csv REPLACE;  
run;
```

```
/* *****  
*****/  
/* ***** SOLNAS *****  
*****/  
/* *****  
*****/
```

```
/* *****  
/* DLW data*/  
data dlw;  
set mysolnas.dlwa_lad1;  
run;
```

```
/* VSEA data (height weight)*/  
data vsea;  
set mysolnas.vsea_lad1;  
run;
```

```

/* BIEA data (gender)*/
data biea;
set mysolnas.biea_lad1;
run;

/* CSEA data (Age wieght)*/
data csea;
set mysolnas.csea_lad1;
run;

/* PA DERV data (Physical Activity)*/
data pal;
set mysolnas.pa_derv_solnas_lad1;
run;

*****
** DLW **
* Keep only urine data for TEE;
* Create new variable with only first letter of Sample collection - DLWA4;
data dlw; set dlw;
Sample = substr(DLWA4,1,1);
run;
/*proc freq data=dlw; tables Sample; run;*/

* Keep only urine where TEE is not empty (TEE is DLWA33) & SOLNAS Main Study
(SOLNAS_STUDY); *n=382;
data dlw; set dlw;
where Sample='U' & DLWA33 ne . & SOLNAS_STUDY='M';
run;
/*proc sort data=dlw nodupkey; by PID; run; *There are no more duplicated IDs; */

*****
* Rename variables to use an donly keep TEE and BMI;
data dlw1; set dlw; *n=382;
TEE = DLWA33;
BMI = DLWA34;
FFM = DLWA35;
FM = DLWA36;
FM_pct = DLWA37;
*FM2=(FM_pct*FFM)/(1-FM_pct);
keep PID TEE BMI FFM FM FM_pct FM2;
run;

*****
** VSEA **
* Keep only data from Main study for Height and Weight;
* Height is in cm and weight is in kg in solnas;
data vsea1; set vsea; *n=393;
where SOLNAS_STUDY='M';

```

```
Height = VSEA3A; * height in cm ;
Weight = VSEA3B; * weight in kg;
keep PID Height Weight;
run;
/*proc sort data=vsea1 nodupkey;    by PID; run; *There are no more duplicated IDs;
*/
```

```
*****
** BIEA **
* Keep only data from Main study for Gender;
data biea1; set biea; *n=393;
where SOLNAS_STUDY='M';
if BIEA1=1 then Sex = 'M';
if BIEA1=2 then Sex = 'F';
keep PID Sex;
run;
/*proc sort data=biea1 nodupkey;    by PID; run; *There are no more duplicated IDs;
*/
```

```
*****
** CSEA **
* Keep only data from Main study for Age and Colorimeter weight;
data csea1; set csea; *n=385;
where SOLNAS_STUDY='M';
Weight_calorim = CSEA2;
Age = CSEA3;
EE_mean_kcald= CSEA4D1;
EE_SD_kcald= CSEA4D2;
EE_CV_kcald= CSEA4D3;
keep PID Weight_calorim Age EE_mean_kcald EE_SD_kcald EE_CV_kcald;
run;
/*proc sort data=csea1 nodupkey;    by PID; run; *There are no more duplicated IDs;
*/
```

```
*****
** PAL **
* Keep only data from Main study for Age and Colorimeter weight;
data pal1; set pal; *n=69;
keep PID SED_DAY LIGHT_DAY MOD_DAY VIG_DAY;
run;
```

```
*****
** Sort to merge;
proc sort data= biea1; by PID; run;
proc sort data=vsea1; by PID; run;
proc sort data=csea1; by PID; run;
proc sort data=dlw1; by PID; run;
```

```

proc sort data=pal1; by PID; run;

* Merge;
data solnas; *n=393;
  length SID Strata Age_cat PALCAT P_Stage $25;
  merge biea1 vsea1 csea1 dlw1 pal1;
  by PID;
  Age_cat="Adults";
    if 0<Age<3.0 then Strata='Infant';
    if 3.0<=Age<19 and Sex='F' then Strata='Child Girls 3-18';
    if 3.0<=Age<19 and Sex='M' then Strata='Child Boys 3-18';
    if 19<=Age and Sex='F' then Strata='Adult Women 19+';
    if 19<=Age and Sex='M' then Strata='Adult Men 19+';

  BEE = EE_mean_kcald;
  PAL = TEE/BEE;
run;

proc freq data=solnas; table Strata; run;
*11 are missing Strata because missing age;

* Create PAL category;
/** These cutoffs were used based on IOM 2005 report;
data iaea11; set iaea11;
  length PALCAT $25;
  if PAL<1 then PALCAT="PAL<1";
  if 1<=PAL<1.4 then PALCAT="S";
  if 1.4<=PAL<1.6 then PALCAT="LA";
  if 1.6<=PAL<1.9 then PALCAT="A";
  if 1.9<=PAL<2.5 then PALCAT="VA";
  if 2.5<=PAL then PALCAT="PAL>2.5";
  if PAL=. then PALCAT="";
run;*/
data solnas; set solnas;
if .<age<9.0 then do; *3.0<=age<9.0;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0<=PAL<1.31 then PALCAT='S';
    if 1.31<=PAL<1.44 then PALCAT='LA';
    if 1.44<=PAL<1.59 then PALCAT='A';
    if 1.59<=PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
  end;
if 9.0<=age<14.0 then do;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0<=PAL<1.44 then PALCAT='S';
    if 1.44<=PAL<1.59 then PALCAT='LA';
    if 1.59<=PAL<1.77 then PALCAT='A';
    if 1.77<=PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
  end;
end;

```

```

if 14.0=<age<19.0 then do;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0=<PAL<1.56 then PALCAT='S';
    if 1.56=<PAL<1.73 then PALCAT='LA';
    if 1.73=<PAL<1.92 then PALCAT='A';
    if 1.92=<PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;
if 19.0=<age then do;
*Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
    if .<PAL<1 then PALCAT='PAL<1';
    if 1.0=<PAL<1.53 then PALCAT='S';
    if 1.53=<PAL<1.68 then PALCAT='LA';
    if 1.68=<PAL<1.85 then PALCAT='A';
    if 1.85=<PAL<2.5 then PALCAT='VA';
    if 2.5<=PAL then PALCAT='PAL>2.5';
end;
run;

```

```

* Calculate a variable BMR_kjd_Schofield;
data solnas; set solnas;

```

```

if .<Age<3 & Sex="M" then BMR_Mjd_Schofield = 0.0007*Weight + 6.349*Height/100 -
2.584;
if 3<=Age<10 & Sex="M" then BMR_Mjd_Schofield = 0.082*Weight + 0.545*Height/100 +
1.736;
if 10<=Age<18 & Sex="M" then BMR_Mjd_Schofield = 0.068*Weight + 0.574*Height/100 +
2.157;
if 18<=Age<30 & Sex="M" then BMR_Mjd_Schofield = 0.063*Weight - 0.042*Height/100 +
2.953;
if 30<=Age<60 & Sex="M" then BMR_Mjd_Schofield = 0.048*Weight - 0.011*Height/100 +
3.670;
if 60<=Age & Sex="M" then BMR_Mjd_Schofield = 0.038*Weight + 4.068*Height/100 -
3.491;

```

```

if .<Age<3 & Sex="F" then BMR_Mjd_Schofield = 0.068*Weight + 4.281*Height/100 -
1.730;
if 3<=Age<10 & Sex="F" then BMR_Mjd_Schofield = 0.071*Weight + 0.677*Height/100 +
1.553;
if 10<=Age<18 & Sex="F" then BMR_Mjd_Schofield = 0.035*Weight + 1.948*Height/100 +
0.837;
if 18<=Age<30 & Sex="F" then BMR_Mjd_Schofield = 0.057*Weight + 1.184*Height/100 +
0.411;
if 30<=Age<60 & Sex="F" then BMR_Mjd_Schofield = 0.034*Weight + 0.006*Height/100 +
3.530;
if 60<=Age & Sex="F" then BMR_Mjd_Schofield = 0.033*Weight + 1.917*Height/100 +
0.074;

```

```

*Convert to kcal from MJ (BMR_kcal_Schofield);
BMR_kcal_Schofield = (BMR_Mjd_Schofield*1000)/4.184;

*Calculate PAL_est=TEE_kcal/BMR_kcal_Schofield;
PAL_est=TEE/BMR_kcal_Schofield;

* Create PAL_est category;
/*if PAL_est<1 then PALCAT_est="PAL<1";
if 1<=PAL_est<1.4 then PALCAT_est="S";
if 1.4<=PAL_est<1.6 then PALCAT_est="LA";
if 1.6<=PAL_est<1.9 then PALCAT_est="A";
if 1.9<=PAL_est<2.5 then PALCAT_est="VA";
if 2.5<=PAL_est then PALCAT_est="PAL>2.5";
if PAL_est=. then PALCAT_est="";*/

if .<age<9.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.31 then PALCAT_est='S';
    if 1.31<=PAL_est<1.44 then PALCAT_est='LA';
    if 1.44<=PAL_est<1.59 then PALCAT_est='A';
    if 1.59<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 9.0<=age<14.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.44 then PALCAT_est='S';
    if 1.44<=PAL_est<1.59 then PALCAT_est='LA';
    if 1.59<=PAL_est<1.77 then PALCAT_est='A';
    if 1.77<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 14.0<=age<19.0 then do;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.56 then PALCAT_est='S';
    if 1.56<=PAL_est<1.73 then PALCAT_est='LA';
    if 1.73<=PAL_est<1.92 then PALCAT_est='A';
    if 1.92<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;
if 19.0<=age then do;
*Note that these are based on percentiles of 19 to 70.99, but 71+ use these
too;
    if .<PAL_est<1 then PALCAT_est='PAL<1';
    if 1.0<=PAL_est<1.53 then PALCAT_est='S';
    if 1.53<=PAL_est<1.68 then PALCAT_est='LA';
    if 1.68<=PAL_est<1.85 then PALCAT_est='A';
    if 1.85<=PAL_est<2.5 then PALCAT_est='VA';
    if 2.5<=PAL_est then PALCAT_est='PAL>2.5';
end;

```



```

** Combine estimate and observed PAL;
if PAL ne . then PAL_obs_est=PAL;
if PAL = . then PAL_obs_est=PAL_est;
if PALCAT ne '' then PALCAT_obs_est=PALCAT;
if PALCAT = '' then PALCAT_obs_est=PALCAT_est;

run; *n=393;

*Calculate BMI categories and Life stages;
data solnas; set solnas; *n=393;
length BMICat $25 Life_Stage $25 Ethnicity $50;

if BMI=. then BMI=Weight/((Height/100)**2); /*Calculate BMI for those not on the DLW
dataset*/

if BMI<18.5 then BMICat="< 18.5 kg/m2";
if 18.5<=BMI<25 then BMICat=">= 18.5 & <25 kg/m2";
if 25<=BMI<30 then BMICat=">= 25 & <30 kg/m2";
if 30<=BMI<35 then BMICat=">= 30 & <35 kg/m2";
if 35<=BMI<40 then BMICat=">= 35 & <40 kg/m2";
if 40<=BMI then BMICat=">= 40 kg/m2";
if BMI=. then BMICat="";

if Age<0.5833 then Life_stage="0-6 months";
if 0.5833<=Age<1 then Life_stage="7-11 months";
if 1<=Age<4 then Life_stage="1-3 years";
if 4<=Age<9 then Life_stage="4-8 years";
if 9<=Age<14 then Life_stage="9-13 years";
if 14<=Age<19 then Life_stage="14-18 years";
if 19<=Age<31 then Life_stage="19-30 years";
if 31<=Age<51 then Life_stage="31-50 years";
if 51<=Age<71 then Life_stage="51-70 years";
if Age>=71 then Life_stage=">=71 years";
if Age=. then Life_stage="";

Ethnicity="Hispanic";
Pregnant="No";
Lactating="No";

run;

** Create Age in months;
data solnas; set solnas; Age_months = Age*12;run;

```

```

** Create ID;
data solnas; set solnas; SID=PID;run;

/*Remove people without Age or sex*/
data solnas; set solnas;
    where Age ne . and Sex ne "";
run; *n=382;
proc freq data=solnas; table strata; run;
*n=382 with no missing Strata;

/*Calculate BMI for those not on the DLW dataset*/
data solnas; set solnas;
if BMI=. then BMI=Weight/((Height/100)**2);
run; *n=382;

/*Change format of numeric for TEE*/
data solnas; set solnas; format TEE 15.6; run;

/*proc freq data=solnas; tables LIGHT_DAY;RUN;*/

*NOTE: Remove invalid data from the committe;
/*Create clean dataset

In infants (0 to 11.9 mo) AND CHILDREN 1-3 YEAR
If PAL (observed) is <1 (n=9) or >2.5 (n=0), then drop that infant
If PAL estimated from Schofield is <1 (n=34) or >2.5 (n=0) then drop that infant.

For children (4+) & adults,
if observed PAL >2.5 (n=54) then drop that person.
If observed PAL < 1 (n=9) then make PAL missing and use multiple imputation
MI will bound PAL to be set 1 to 2.5
if unobserved PAL, but PAL estimated from Schofield >2.5 (n=~30?) then drop these
people
These are people with very high TEE for their age & wt
if unobserved PAL, but PAL estimated from Schofield <1 (n=~45), leave them be
because PAL will be imputed.
*/

data solnas_final; set solnas;
if .<Age<4 & (.<PAL<1 or PAL>2.5) then Not_include=1;
if .<Age<4 & (.<PAL_est<1 or PAL_est>2.5) then Not_include=1;
if Age>=4 & PAL>2.5 then Not_include=1;
if Age>=4 & .<PAL<1 then do; PAL=. ; PALCAT=""; end;
if Age>=4 & PAL = . & PAL_est>2.5 then Not_include=1;
*if Age ne . & Age>=4 & PAL = . & PAL_est<1 ;
run;*n=382;

/*Clean dataset*/
data solnas_clean; set solnas_final;
where Not_include ne 1;

```

```
drop Not_include;
run;*n=380;
```

```
/* *****
/* SAVE SOLNAS dataset */
```

```
* As SAS data;
data "&mypath.\Data\Clean data for analysis\SOLNAS_&SYSDATE9..sas7bdat";
  set solnas_final;
run;
```

```
* As csv data;
proc export data=solnas_final
  outfile="&mypath.\Data\Clean data for analysis\SOLNAS_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;
```

```
* As SAS data;
data "&mypath.\Data\Clean data for analysis\SOLNAS_clean_&SYSDATE9..sas7bdat";
  set solnas_clean;
run;
```

```
* As csv data;
proc export data=solnas_clean
  outfile="&mypath.\Data\Clean data for analysis\SOLNAS_clean_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;
```

```
/* *****
*****/
/* ***** MERGE ALL *****
*****/
/* *****
*****/
```

```
* This merges each "final" dataset before they are cleaned/filtered for PAL ;
```

```
data iaea_merge; set iaea_final; *n=5805;
length dataset $25;
dataset="iaea";
keep dataset SID Strata Age_cat Age Age_months Life_Stage Ethnicity Sex BMI BMIcat
Height Weight TEE BEE Percentile Percentile_group Percentile_infant
```

```
Lactating Pregnant P_stage weeks PAL PALCAT PAL_obs_est PALCAT_obs_est PAL_est
PALCAT_est BMR_kcal_Schofield FFM FM FM_pct;
run;
```

```
data iom_merge; set iom_final; *n=2313;
length dataset $25;
dataset="iom";
keep dataset SID Strata Age_cat Age Age_months Life_Stage Ethnicity Sex BMI BMIcat
Height Weight TEE BEE Percentile Percentile_group Percentile_infant
Lactating Pregnant P_stage weeks PAL PALCAT PAL_obs_est PALCAT_obs_est PAL_est
PALCAT_est BMR_kcal_Schofield ;
run;
```

```
data cnrc_merge; set cnrc_final; *n=222;
length dataset $25;
dataset="cnrc";
keep dataset SID Strata Age_cat Age Life_Stage Ethnicity Sex BMI BMIcat Height
Weight TEE BEE
Lactating Pregnant P_stage weeks PAL PALCAT PAL_obs_est PALCAT_obs_est PAL_est
PALCAT_est BMR_kcal_Schofield ;
run;
```

```
data solnas_merge; set solnas_final; *n=382;
length dataset $25;
dataset="solnas";
keep dataset SID Strata Age_cat Age Age_months Life_Stage Ethnicity Sex BMI BMIcat
Height Weight TEE BEE Lactating Pregnant
PAL PALCAT PAL_obs_est PALCAT_obs_est PAL_est PALCAT_est BMR_kcal_Schofield FFM FM
FM_pct;
run;
```

```
* Merge all datasets;
data all; *n=8722;
    length dataset $8.;
    set iaea_merge iom_merge cnrc_merge solnas_merge ;
run;
```

```
*NOTE: Consider invalid data from the committe;
/*Create clean dataset
```

```
In infants (0 to 11.9 mo) AND CHILDREN 1-3 YEAR
If PAL (observed) is <1 (n=9) or >2.5 (n=0), then drop that infant (A&B)
If PAL estimated from Schofield is <1 (n=34) or >2.5 (n=0) then drop that infant
(C&D).
```

```
For children (4+) & adults,
    if observed PAL >2.5 (n=54) then drop that person. (E)
    If observed PAL < 1 (n=9) then make PAL missing and use multiple imputation
```

(X)

MI will bound PAL to be set 1 to 2.5

if unobserved PAL, but PAL estimated from Schofield >2.5 (n=~30?) then drop these people(F)

These are people with very high TEE for their age & wt

if unobserved PAL, but PAL estimated from Schofield <1 (n=~45), leave them be because PAL will be imputed.(Y)

*/

```
data all_final; set all; *n=8722;
  if .<Age<4 & (.<PAL<1) then Not_include="A";
  if .<Age<4 & (PAL>2.5) then Not_include="B";
  if .<Age<4 & (.<PAL_est<1 ) then Not_include="C";
  if .<Age<4 & (PAL_est>2.5) then Not_include="D";
  if Age>=4 & PAL>2.5 then Not_include="E";
  if Age>=4 & .<PAL<1 then do; PAL=. ; PALCAT=""; Not_include="X"; end;
  if Age>=4 & PAL = . & PAL_est>2.5 then Not_include="F";
  if Age>=4 & PAL = . & .<PAL_est<1 then Not_include="Y";
run;

proc freq data=all_final; where Not_include ne "" ;
  table Life_stage*Not_include / norow nocol nopercnt;run;
/*proc freq data=all_clean; where Not_include=1; table Life_stage*PALCAT_est/ norow
nocol nopercnt;run;*/

proc freq data=all_final; table strata; run;
proc freq data=all_final (where=(strata="Preg/Lac"));
  table P_stage*(pregnant lactating) PALCAT Not_include/norow nocol nopercnt;
run;

/*Clean dataset*/
data all_clean; set all_final;
if Not_include in ("A","B","C","D","E","F") then delete ;
drop Not_include;
run;*n=8600;

proc freq data=all_clean; table strata P_stage; run;
proc freq data=all_clean; table dataset*strata/norow nocol nopercnt ; run;
proc means data=all_clean (where=(strata="Preg/Lac"));
  var weeks;
run;
data spl; set all_clean (where=(strata="Preg/Lac")); *n=766 Preg/Lac;
run;
proc means data=spl;
  var weeks tee;
run; *766 all have data for weeks;
proc freq data=spl;
  table dataset*P_stage*(pregnant lactating)/norow nocol nopercnt;
run;
```

```

/*****/
/* SAVE ALL dataset */

* As SAS data;
data "&mypath.\Data\Clean data for analysis\ALLDATA_&SYSDATE9..sas7bdat";
  set all_final;
run;

* As csv data;
proc export data=all_final
  outfile="&mypath.\Data\Clean data for analysis\ALLDATA_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;

/* SAVE ALL clean dataset */

* As SAS data;
data "&mypath.\Data\Clean data for analysis\ALLDATA_clean_&SYSDATE9..sas7bdat";
  set all_clean;
run;

* As csv data;
proc export data=all_clean
  outfile="&mypath.\Data\Clean data for analysis\ALLDATA_clean_&SYSDATE9..csv"
  dbms=csv REPLACE;
run;

/**
If you need to read it back in to check anything without rerunning it all:
data all_clean; set clean.alldata_clean_&SYSDATE9; run;
data all_final; set clean.alldata_&SYSDATE9; run;

**/

```