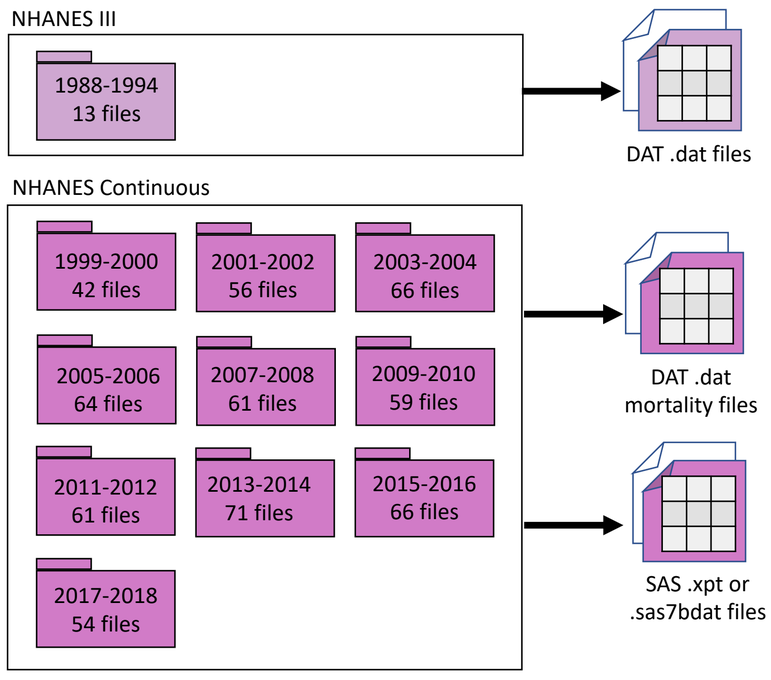
A) Tables (example subset) of file names of NHANES datasets

Variable name	Description	File name	Study period
BMXBMI	Body Mass Index replicate 1 (kg/m**2)	EXAM	1988-1994 Examination
BMXBMI	Body Mass Index replicate 2 (kg/m**2)	EXAMSE	1988-1994 Examination - Second Exam
BMXBMI	Body Mass Index (kg/m**2)	BMX	1999-2000 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_B	2001-2002 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_C	2003-2004 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_D	2005-2006 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_E	2007-2008 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_F	2009-2010 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_G	2011-2012 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_H	2013-2014 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_I	2015-2016 Examination
BMXBMI	Body Mass Index (kg/m**2)	BMX_J	2017-2018 Examination

B) Files by Study Period

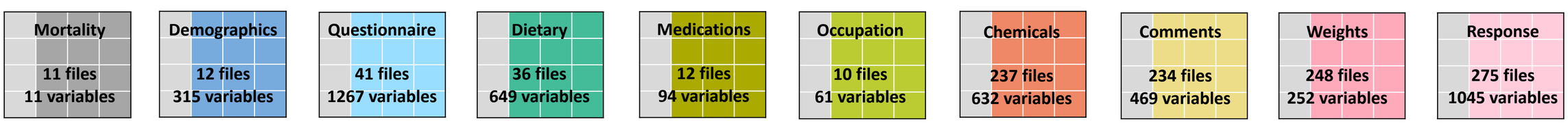


C)

Compile unclean modules

Format DAT files into *Nxp* format
Read in SAS with R package nhanesA
Create new identifier "SEQN_new"
Encode study period ("SDDSRVYR") for NHANES III (1988-1994) as -1
Merge individual datasets with "SEQN", "SEQN_new", and "SDDSRVYR"

D) Unclean NHANES Modules



E)

Tabulate inconsistencies in each module

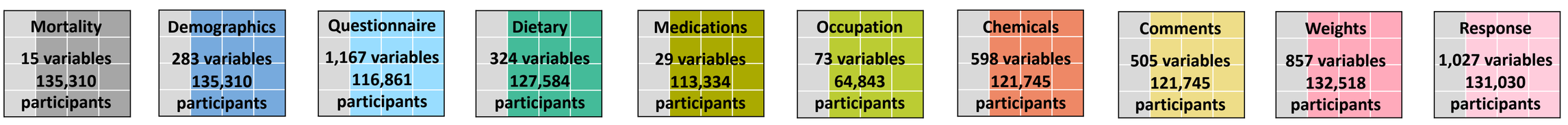
Determine values indicating "Blank but applicable"
Identify changes in variable nomenclature
Identify changes in measurement units
Determine changes in levels for the same category
Determine variables with multiple replicates
Include additional variables
Form consolidated survey weights specific for each chemical biomarker

F) Table (example subset) documenting inconsistencies

Original variable name	Harmonized variable name	Description	Study period	Value indicating "Blank but applicable"	Change in variable name	Change in measurement units	Conversion factor	Statistic on replicates	Replicates used for calculation
HFA6XCR	DMDBORN4	In what country {were you/was SP} born?	1988-1994	8	1				
DMDBORN	DMDBORN4	In what country {were you/was SP} born?	1999-2000		1				
DMDBORN	DMDBORN4	In what country {were you/was SP} born?	2001-2002		1				
DMDBORN	DMDBORN4	In what country {were you/was SP} born?	2003-2004		1				
DMDBORN	DMDBORN4	In what country {were you/was SP} born?	2005-2006		1				
DMDBORN2	DMDBORN4	In what country {were you/was SP} born?	2007-2008		1				
DMDBORN2	DMDBORN4	In what country {were you/was SP} born?	2009-2010		1				
DMDBORN4	DMDBORN4	In what country {were you/was SP} born?	2011-2012						
DMDBORN4	DMDBORN4	In what country {were you/was SP} born?	2013-2014						
DMDBORN4	DMDBORN4	In what country {were you/was SP} born?	2015-2016						
DMDBORN4	DMDBORN4	In what country {were you/was SP} born?	2017-2018						
HFA8R	DMDEDUC2	Education level - Adults 20+	1988-1994	88	1				
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	1999-2000						
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	2001-2002						
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	2003-2004						
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	2005-2006						
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	2007-2008						
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	2009-2010						
LBXVCF	LBXVCF	Blood Chloroform (pg/mL)	2011-2012						
LBXVCF	LBXVCF	Blood Chloroform (ng/mL)	2013-2014			1	*1000		
LBXVCF	LBXVCF	Blood Chloroform (ng/mL)	2015-2016			1	*1000		
LBXVCF	LBXVCF	Blood Chloroform (ng/mL)	2017-2018			1	*1000		
COP	LBXCOT1	Serum cotinine replicate 1 (ng/mL)	1988-1994	88888	1				
COP	LBXCOT2	Serum cotinine replicate 2 (ng/mL)	1988-1994	88888	1				
COR	LBXCOT3	Serum cotinine replicate 3 (ng/mL)	1988-1994	88888	1				
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	1988-1994					rowMeans	LBXCOT1, LBXCOT2, LBXCOT3
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	1999-2000						
LB2COT	LBXCOT1	Serum cotinine replicate 1 (ng/mL)	2001-2002		1				
LBXCOT	LBXCOT2	Serum cotinine replicate 2 (ng/mL)	2001-2002		1				
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2001-2002					rowMeans	LBXCOT1, LBXCOT2
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2003-2004						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2005-2006						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2007-2008						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2009-2010						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2011-2012						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2013-2014						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2015-2016						
LBXCOT	LBXCOT	Serum cotinine (ng/mL)	2017-2018						

G) Write coding pipeline to clean each module

H) Cleaned NHANES Modules



I) Form data dictionary

J) Data Dictionary (example subset)

Variable name	Description	Module	Category	Number of participants	Study Periods	Units	CAS Number	Comment code	Chemical family	Chemical family shorten
ELIGSTAT	Eligibility Status for Mortality Follow-up	Mortality	Mortality	135310	1988-1994, 1999-2018					
DRXTSFAT	Total saturated fatty acids (gm)	Dietary	Total Nutrient Intakes	117180	1988-1994, 1999-2018					
RIAGENDR	Gender of the participant.	Demographics	Demographics	135310	1988-1994, 1999-2018					
BMXBMI	Body Mass Index (kg/m**2)	Response	Body Measures	115629	1988-1994, 1999-2018					
RXDDRUG	GENERIC DRUG NAME	Medications	Prescription Medications	51352	1988-1994, 1999-2018					
MCQ220	Ever told you had cancer or malignancy	Questionnaire	Medical Conditions	75127	1988-1994, 1999-2018					
URXTRS	Urinary Triclosan (ng/mL)	Chemicals	Personal Care and Consumer Product Chemicals and Metabolites	18244	2003-2016	(ng/mL)	3380-34-5	URDTRSLC	Personal Care & Consumer Product Compounds	PCCPCs
URDTRSLC	Comments code for Urinary Triclosan	Comments	Personal Care and Consumer Product Chemicals and Metabolites	18244	2003-2016					
WT_URXTRS	Survey weights for Urinary Triclosan	Weights	Survey Weights - Chemicals	19307	2003-2016					
VNCURRJOB	Harmonized job code and description for current job	Occupation	Occupation	26523	1999-2014					