

WHI Biomarker Data

Jingmin Liu, WHI CCC
Fred Hutch, Seattle WA
September 9, 2016



Outline

- Source of biomarker data
- Where to find biomarker data
- Specimen data files
- Analytic considerations on specimen data files
- Example
- CVD Biomarker Subset

Source of biomarker data

- WHI holds a large repository of biological specimens collected from WHI participants at various times over the past 20 and more years.
- Some biological specimens were assayed and biomarker data are available through:
 - Core studies
 - Ancillary studies
 - BAA studies

Source of biomarker data

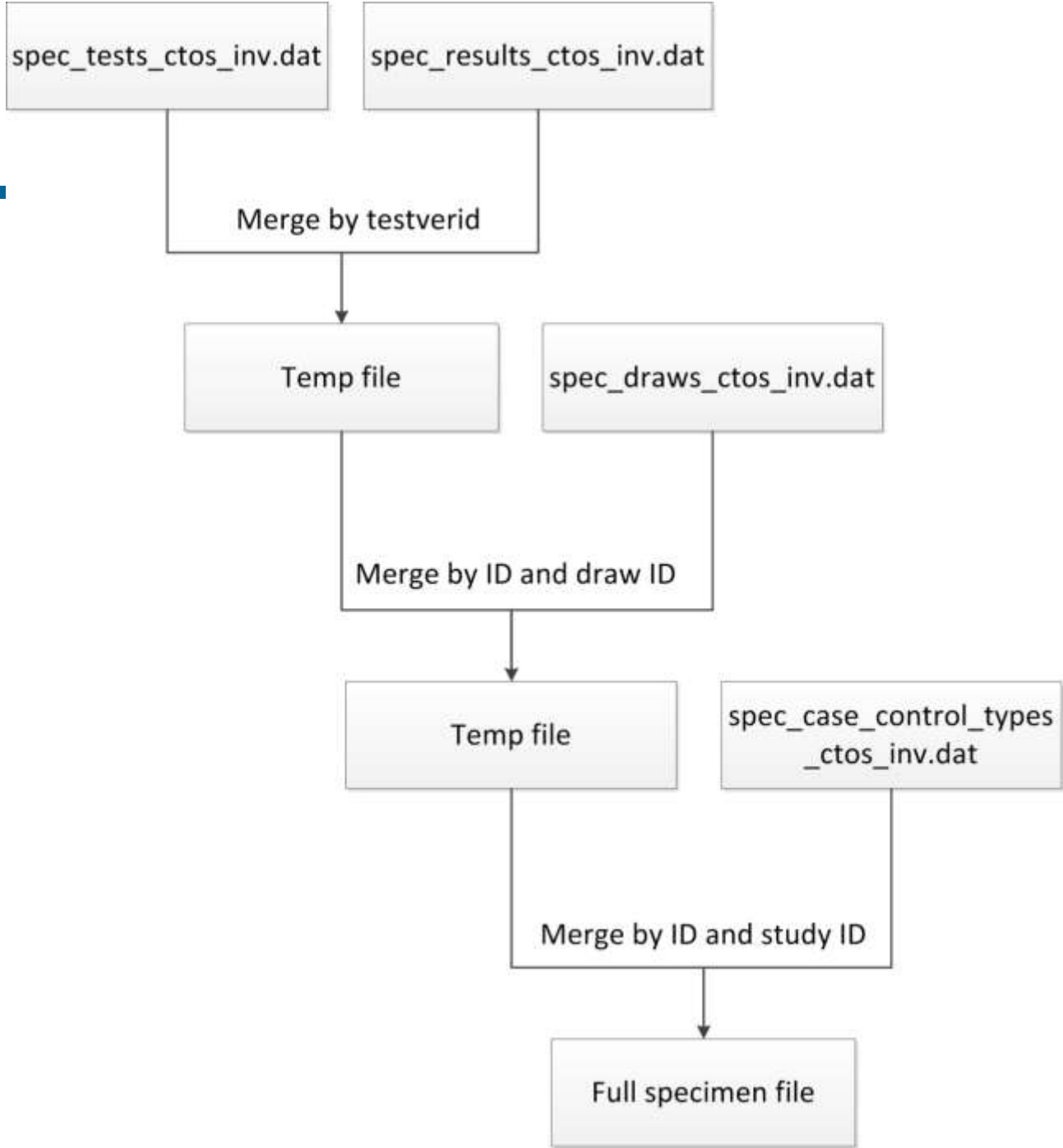
- Biomarker data derived from WHI core funds are available as soon as they are ready
- Biomarker data derived from ancillary or BAA studies will be available after study's first publication or > 1 year past end of funding period
- If the study is not yet complete, ask PI for permission to use its data

How to find biomarker data

Ways to look up biomarker data	Where	Purpose
Online look-up	www.whi.org – researchers – Data – Dataset Documentation – Specimen Results – Specimen Results Descriptions	Quick way to look up information on studies, tests, samples etc.
Query builder	www.whi.org – researchers – Data – query builder	A more custom look- up; availability of existing biomarkers and bio specimens
Full specimen data set	www.whi.org – researchers – Data – download	More appropriate for data analysis

Specimen data files

- 4 separate data files with biomarker data
 - spec_case_control_types_ctos_inv.dat
(info on study, case/control status)
 - spec_draws_ctos_inv.dat
(draw information)
 - spec_results_ctos_inv.dat
(test results)
 - spec_tests_ctos_inv.dat
(assay information)
- SAS load files, pdf data dictionaries for these data sets are also provided



Full Specimen data set

The resulting file has one record/row per participant per test per visit per study

ID	DRAWVY	DRAWVTYP	TESTNAME	TESTVAL	TESTUNITS	LABNAME	STUDYID
18	0	Screening	Estradiol, free	0.43	pg/dL	Hormone Lab	BA7
18	0	Screening	SHBG	30.20	nmol/L	Hormone Lab	BA7
18	0	Screening	HCT	38.30	%	WHI Labs	W1
18	0	Screening	Hemoglobin	13.40	GM/dL	WHI Labs	W1
18	0	Screening	Insulin	16.0	uIU/mL	PPD	W14
18	0	Screening	Insulin	101.0	pmol/L	UMMC	W58
18	1	Annual visit	Insulin	7.8	uIU/mL	PPD	W14
18	1	Annual visit	Glucose	109.0	mg/dL	PPD	W14
18	1	Annual visit	ATIII	90.0	%	Vermont	W6
18	1	Annual visit	CRP	3.79	mg/L	Vermont	W6

Dissecting Full Specimen data set - by test

- As of Dec 2015, a total of 343 different tests were done on different subsets of WHI participants
- Almost all ($\sim 99\%$) participants had baseline data on four biomarkers: Hemoglobin, HCT, PLT, and WBC
- $\sim 23\%$ of all WHI participants had baseline lipids data (e.g. LDLC, HDLC)

Dissecting Full Specimen data set - by visit

- Data from 5 different time points during main trial period
- Data from WHI Extension II through LLS (non-routine visits)
- Baseline had all 343 biomarkers
- Some biomarkers had longitudinal measures

TESTNAME	BL	Y1	Y3	Y6	Y9
CRP	46496(28.7%)	1132(0.7%)	3567(2.2%)	-	-
Carotene, alpha	7446(4.6%)	4028(2.5%)	3542(2.2%)	3281(2.0%)	-
Cholesterol	38668(23.9%)	5071(3.1%)	5772(3.6%)	3295(2.0%)	-
D-dimers	6380(3.9%)	1154(0.7%)	-	-	-
Estradiol, free	11462(7.1%)	1597(1.0%)	302(0.2%)		
Fibrinogen	13941(8.6%)	4910(3.0%)	3666(2.27%)	3245(2.0%)	-
Glucose	37455(23.2%)	5593(3.5%)	6949(4.29%)	3298(2.0%)	-
HCT	160130(98.9%)	6562(4.1%)	76236(47.1%)	159(0.1%)	5(0.0%)

Dissecting Full Specimen data set - by study

- 70 studies contributed (sample sizes: 58 – 160,130)
- Number of tests/markers per study: 1 – 64
- Some participants had multiple test results of the same biomarker from different studies
- Labs and/or assay methods may vary from one study to another
- Each study designed for specific goal
- Sample not representative of the WHI population

Tips on analyzing biomarker data

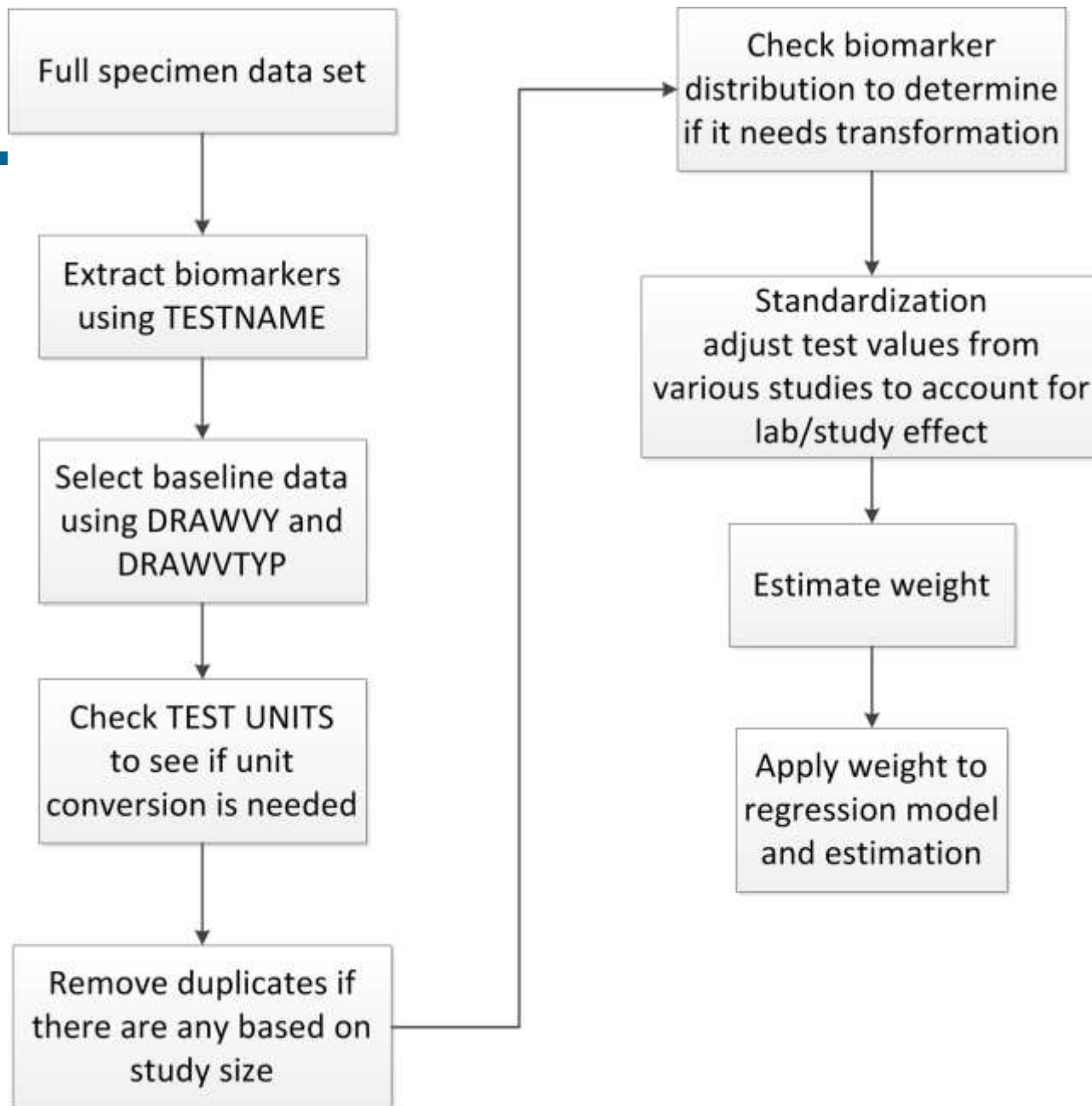
- Select the biomarker using TESTNAME
- Select the time point of specimen draw (e.g. DRAWVY=0 & DRAWVTYP=1 for baseline)
- Make sure test units are the same before combining test results from different studies
- Have a set rule to remove duplicate measures (based on study size, testing lab, testing date, etc.)
- When pooling test results from several studies, consider standardizing
- Core/Ancillary/BAA studies often oversample cases, consider weighting

Weighting

- Is weighting necessary for your analysis?
- Which weighting method?(frequency, IPW etc.)
- What variables to use to calculate weights
- Use sandwich estimators to get correct SE

Example - overview

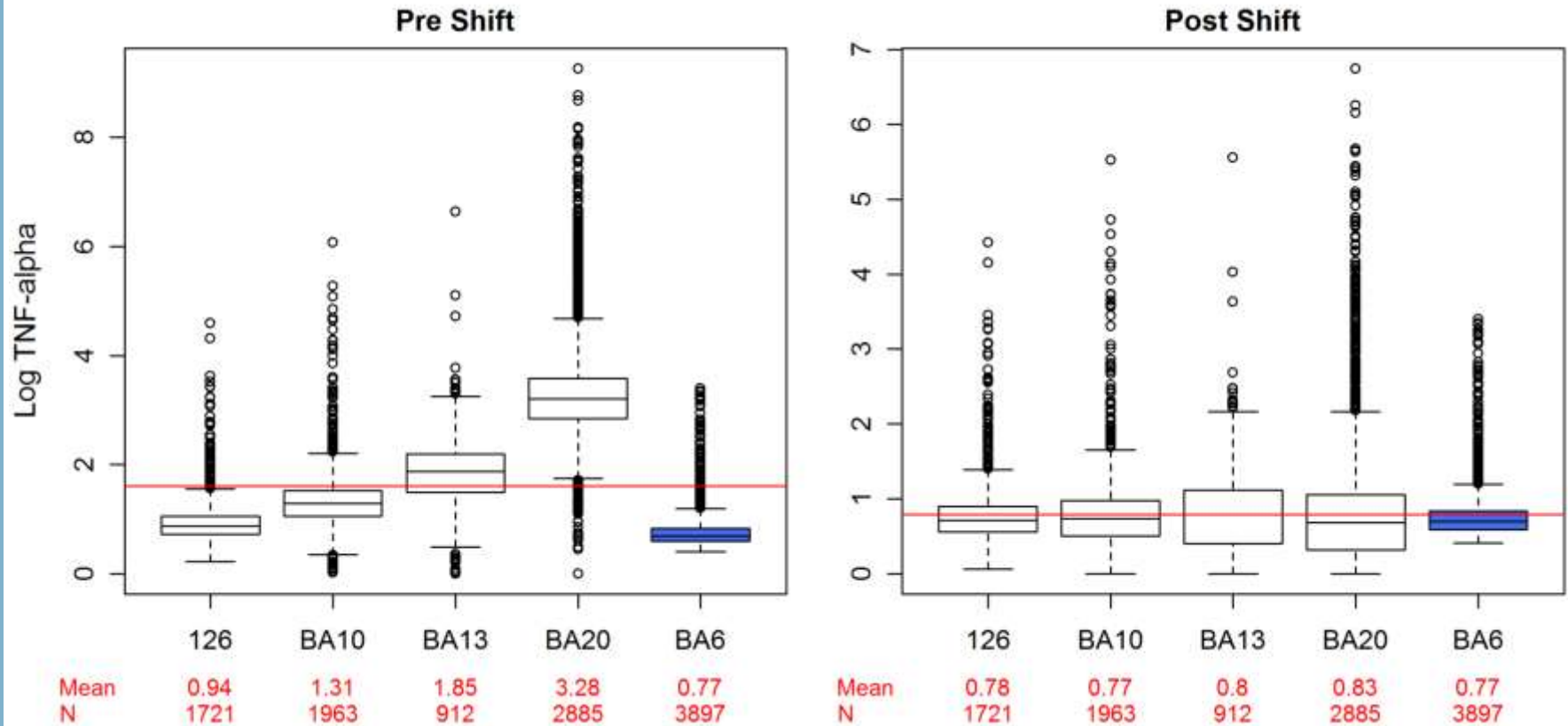
- Goal of the analysis is to look at the effect of physical activity on various biomarkers
- A total of 22 markers from 42 different WHI ancillary, core, or BAA studies
- For each marker, use all available data. Sample sizes range from 299 to 46,496
- Number of studies contributed data ranges from 1 to 22



Example - standardization

- Fit a model with the biomarker as a function of study + various demographics (same model for each marker)
- Pull the beta coefficient for each study from output, then shift the test values by corresponding beta coefficient
- As a test, rerun the same model with the adjusted marker values. If done correctly, the beta coefficients for study should now all be zero

Example - standardization



Example - weighting

- Use Inverse probability weighting
- Model the probability of being in a sample (in this case having data for a biomarker) and then invert the predicted probabilities for each participant to create the weights
- IPW models should include all of the following from each study involved:
 - Outcome variables
 - Matching variables
 - Exclusion criteria variables

Example - weighting

Demographics											
		Age		Race/Ethnicity				WHI Components			
		White	Black	Hispanic	Other	HT	DM	CAD	OS		
Biomarker	n	Mean	SD	%	%	%	%	%	%	%	%
All	161808	63.2	7.2	83	9	4	4	17	30	22	58
Cholesterol	38668	64.5	7.3	59	25	11	5	40	26	30	46
CRP	46496	64.6	7.3	65	22	9	4	33	23	26	54
Fibrinogen	11199	65.7	7.1	72	15	7	7	37	27	28	48
Glucose	37455	64.1	7.3	57	27	11	5	41	26	31	45
HDL-C	37421	64.6	7.3	58	26	11	5	41	27	31	44
Insulin	37017	64.0	7.3	58	26	11	5	42	31	33	40
LDL-C	31061	64.2	7.3	53	30	13	5	49	32	37	33
TNF-alpha	11378	65.2	7.3	78	17	3	2	5	7	5	89
Triglycerides	35389	64.4	7.3	57	27	12	5	44	29	33	40

Example - weighting

Weighted Demographics											
		Age		Race/Ethnicity				WHI Components			
		Mean	SD	White	Black	Hispanic	Other	HT	DM	CAD	OS
Biomarker	n			%	%	%	%	%	%	%	%
All	161808	63.2	7.2	83	9	4	4	17	30	22	58
Cholesterol	38668	62.8	14.6	79	10	4	6	18	27	21	61
CRP	46496	62.9	13.4	80	10	4	6	18	30	22	58
Fibrinogen	11199	63.5	26.7	80	10	4	5	18	32	24	55
Glucose	37455	62.5	14.7	79	10	4	6	18	26	21	62
HDL-C	37421	62.9	14.8	79	10	5	6	18	29	22	59
Insulin	37017	62.6	15.0	80	10	4	6	18	33	24	55
LDL-C	31061	62.4	15.8	77	11	5	8	19	32	23	54
TNF-alpha	11378	63.7	24.5	80	10	6	4	18	23	17	72
Triglycerides	35389	62.6	15.3	79	10	4	6	18	27	21	61

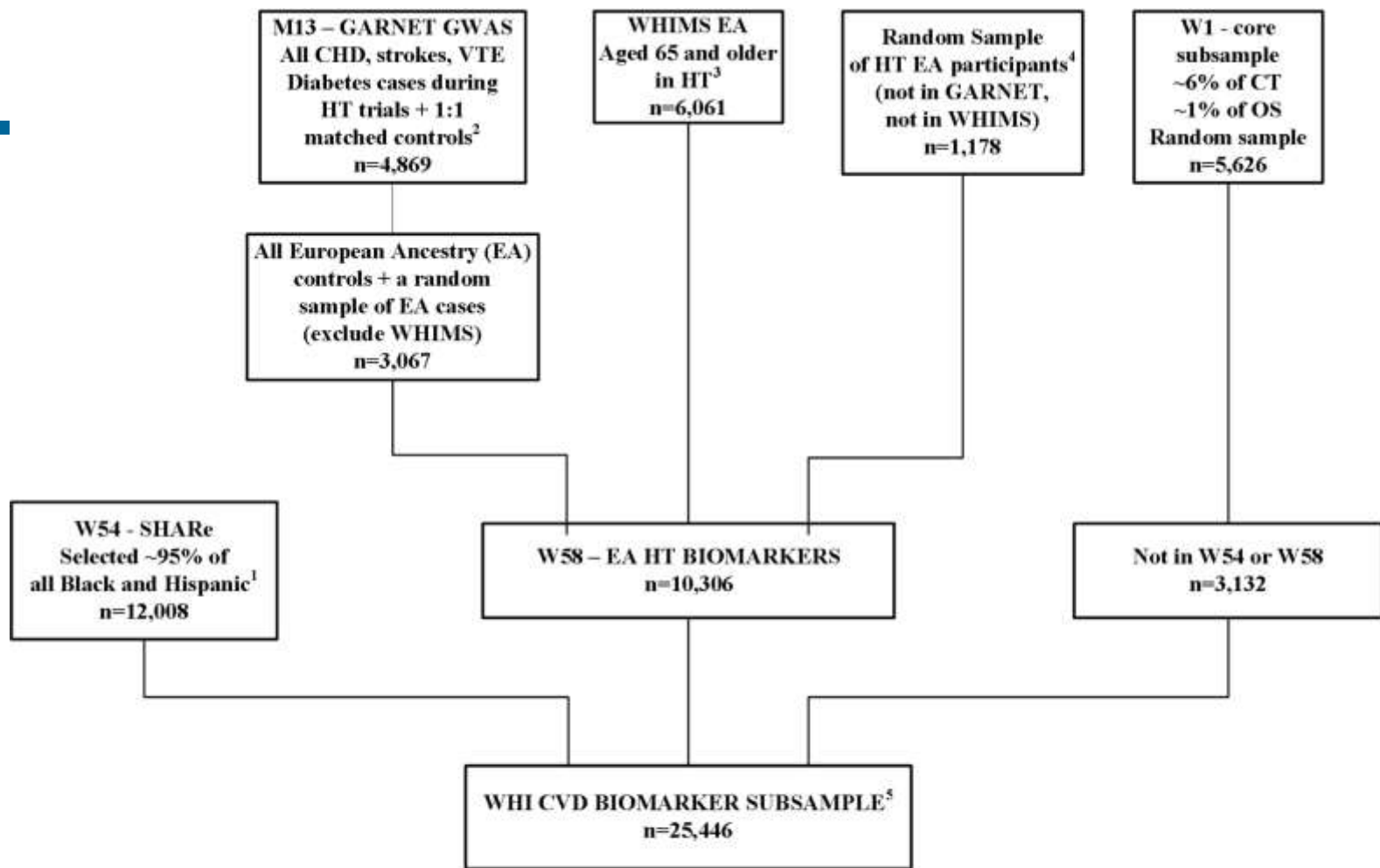
Example - estimates

Model: $\log(\text{biomarker}) = \text{physical activity (continuous)}$

Biomarker	n	Unweighted		Weighted	
		Estimate (95% CI)	p	Estimate (95% CI)	p
Cholesterol	38668	0.00%(-0.14%,0.15%)	0.992	-0.14%(-0.35%,0.08%)	0.212
CRP	46496	-8.57%(-9.06%,-8.07%)	<.001	-8.77%(-9.43%,-8.11%)	<.001
Fibrinogen	11199	-1.17%(-1.54%,-0.81%)	<.001	-1.29%(-1.72%,-0.86%)	<.001
Glucose	37455	-1.65%(-1.81%,-1.48%)	<.001	-1.60%(-1.80%,-1.39%)	<.001
HDL-C	37421	2.63%(2.41%,2.85%)	<.001	2.73%(2.38%,3.08%)	<.001
Insulin	37017	-8.49%(-8.93%,-8.05%)	<.001	-8.71%(-9.30%,-8.11%)	<.001
LDL-C	31061	-0.39%(-0.63%,-0.15%)	0.001	-0.67%(-1.16%,-0.18%)	0.008
TNF-alpha	11378	-0.96%(-1.67%,-0.25%)	0.008	-0.24%(-1.33%,0.86%)	0.667
Triglycerides	35389	-3.20%(-3.59%,-2.80%)	<.001	-3.84%(-4.42%,-3.25%)	<.001

CVD Biomarker Subset

- A large cohort of ~25,000 participants had baseline CVD biomarkers
- Contributed by 3 WHI core studies: W1, W54, and W58
- Biomarkers are HDL, LDL, total cholesterol, triglycerides, glucose, insulin, CRP, creatinine
- Over 87% also had GWAS data



1. dbGaP eligible. Some dbGaP eligible participants not selected due to funding constraints

2. Matching factors were age, race/ethnicity, hysterectomy status, enrollment date, length of follow-up

3. Must also be willing to participate WHIMS activities and sign WHIMS study consent

4. Selection was stratified by age, year of enrollment, medical history at baseline to provide a more balanced cross-section of HT participants

5. 87.7% (n=22,311) of this subsample had GWAS data.

CVD Biomarker Subset

	CVD Biomarker cohort	WHI
N	25,446	161,808
Age		
50-59	7922 (31.13%)	53559 (33.1%)
60-69	11377 (44.71%)	72589 (44.86%)
70-79	6147 (24.16)	35660 (22.04%)
Race/ethnicity		
White	12132 (47.68%)	133541 (82.53%)
Black	8836 (34.72%)	14618 (9.03%)
Hispanic	3841 (15.09%)	6484 (4.01%)
American Indian	133 (0.52%)	713 (0.44%)
Asian/Pacific Islander	411 (1.62%)	4190 (2.59%)
Unknown	93 (0.37%)	2262 (1.4%)

Online resources

- [Approved and Funded Ancillary Studies](#)
- [BAA list](#)
- [WHI Specimen Test Results \(Read Me\)](#)
- [Standardization of assays from multiple labs for a given analyte](#)
- [Data Preparation and Use of WHI datasets](#)
- More questions? WHI Helpdesk helpdesk@whi.org

Acknowledgment

Joe Larson

Mary Pettinger