January 20th 2016
TOPMED – sample selection

Sample: 5929 cases and 5171 controls. The cases had Ischemic or Hemorrhagic strokes or VTE. Controls were DBGaP eligible, with DNA already extracted and needed to have at least 5 biomarkers (LDL, HDL, Triglycerides, Glucose, Insulin, CRP and/or Creatinine). Controls were matched based on age at baseline and HRT participation.

Steps for selecting cases:

We looked at the 3 outcomes by race and whether DNA was available or needed to be extracted. We took all samples in green. Four of the red cells were randomly selected as backups and the rest were not selected. In the blue cell, 203 were randomly selected as samples and the remainder were used as backups.

VTE			
	Available	Needs buffy coat	total
Missing	**1**	**1**	2
American Indian or Alaskan Native	**5**	**2**	7
Asian or Pacific Islander	**4**	**0**	4
Black or African-American	**197**	**1**	198
Hispanic/Latino	**30**	**1**	31
White	**796**	**75**	871
Other	**1**	**5**	6
Total	1034	85	1119

Hemorrahgic Stroke No VTE			
	Available	Needs buffy coat	total
Missing	**1**	**3**	4
American Indian or Alaskan Native	**5**	**1**	6

Asian or Pacific Islander	**24**	**6**	30
Black or African-American	**69**	**3**	72
Hispanic/Latino	**23**	**0**	23
White	**544**	**178**	722
Other	**2**	**4**	6
Total	668	195	863

Hemorrahgic Stroke and VTE			
	Available	Needs buffy coat	total
Missing	**0**	**0**	0
American Indian or Alaskan Native	**1**	**0**	1
Asian or Pacific Islander	**0**	**0**	0
Black or African-American	**7**	**1**	8
Hispanic/Latino	**0**	**0**	0
White	**13**	**0**	13
Other	**0**	**0**	0
Total	21	1	22

Ischemic Stroke No VTE			
	Available	Needs buffy coat	total
Missing	**7**	**7**	14
American Indian or Alaskan Native	**11**	**1**	12
Asian or Pacific Islander	**63**	**5**	68

January 20th 2016
TOPMED – sample selection

Black or African-American	**430**	**7**	437
Hispanic/Latino	**99**	**0**	99
White	**3034**	**472**	3506
Other	**17**	**31**	48
Total	3661	523	4184

Ischemic Stroke and VTE			
	Available	Needs buffy coat	total
Missing	**2**	**0**	2
American Indian or Alaskan Native	**0**	**0**	0
Asian or Pacific Islander	**0**	**0**	0
Black or African-American	**13**	**0**	13
Hispanic/Latino	**2**	**0**	2
White	**65**	**2**	67
Other	**0**	**1**	1
Total	82	3	85

Steps for selecting controls:

Subjects eligible for controls were DBGaP eligible, no history of stroke or VTE and needed to have DNA available. The controls were selected to match the number of cases when they were stratified by age and HRT participation. Controls were selected within cells of age strata and race to match cases. Preference was given to samples with the greatest number of biomarkers; most cells had enough subjects with all 7 of the biomarkers (LDL, HDL, INSU, GLUC, CREA, CRP and TRI). If there were not enough with 7 biomarkers, we looked at cells with fewer biomarkers until enough subjects were selected. We also gave preference to subjects in W1 and W6 core studies. Finally, when cells were big enough that controls were selected randomly, we selected for the type of extraction giving preference to prime 5 extraction. The order of the preference for selecting controls was, biomarker number, participation core studies and then extraction method.

Distribution of cases by age decade at baseline and race for non HRT participants

	Missing	Am Indian/Alaska Nat	Asian	Black	Hispanic	White	Other	Total
<=60	0	4	17	184	42	394	8	649
61-70	5	3	35	233	50	1471	19	1816
71-80	5	4	31	90	23	1091	18	1262
>80								
Total	10	11	83	507	115	2956	45	3727

Distribution of eligible controls with >=5 biomarkers available by age decade at baseline and race for non HRT participants

	Missing	Am Indian/Alaska Nat	Asian	Black	Hispanic	White	Other	Total
<=60		205	82	2898	1383	343		5105
61-70		129	226	2352	959	1096*		4762
71-80		47	118	695	206	834*		1900
>80								
Total		381	426	5945	2548	2467		11767

*For these two cells, the totals are less than the target, we selected the remaining from those with 4 biomarkers (164 in age 61-70 category and 101 in the 71-80 category)

Distribution of eligible controls with 7 biomarkers available by age decade at baseline and race for non HRT participants

	Missing	Am Indian/Alaska Nat	Asian	Black	Hispanic	White	Other	Total
<=60		201	1	2758	1288	1		4249
61-70		125	2	2218	875	3		3223
71-80		46	0	635	191	4		876
>80								

January 20th 2016
TOPMED – sample selection

Total		372	3	5611	2354	8		8348

The number of controls selected by age decade and race for non HRT participants. These were selected from the yellow cells in the above tables (except for the White and Asians which did not have enough subjects and for which we had to select subjects with 4 biomarkers). The controls were selected randomly within each cell.

		Am Indian/Alaska Nat	Asian	Black	Hispanic	White		Total
<=60		4	17	184	42	339		586
61-70		3	35	233	50	1265	*	1586
71-80		4	31	90	23	939	*	1087
>80								
Total		11	83	507	115	2543		3259

Distribution of cases by age decade at baseline and race for HRT participants

	Missing	Am Indian/Alaska Nat	Asian	Black	Hispanic	White	Other	Total
<=60	3	7	5	69	16	371	3	474
61-70	7	5	9	112	22	948	5	1108
71-80	2	3	5	40	2	560	8	620
>80								
Total	12	15	19	221	40	1879	16	2202

Distribution of eligible controls with >=5 biomarkers available by age decade at baseline and race for HRT participants

	Missing	Am Indian/Alaska Nat	Asian	Black	Hispanic	White	Other	Total
<=60		43	59	850	609	1511		3072
61-70		30	54	653	336	4885		5958

71-80		9	41	180	65	2588		2883
>80		0	0	0	0	1		1
Total		82	154	1683	1010	8985		11914

Distribution of eligible controls with 7 biomarkers available by age decade at baseline and race for HRT participants

	Missing	Am Indian/Alaska Nat	Asian	Black	Hispanic	White	Other	Total
<=60		41	0	799	558	1417		2815
61-70		30	0	600	300	4675		5605
71-80		9	0	160	58	2508		2735
>80								
Total		80	0	1559	916	8600		11155

The number of controls selected by age decade and race for non HRT participants. These were selected from the yellow cells in the above tables. The controls were selected randomly within each cell.

		Am Indian/Alaska Nat	Asian	Black	Hispanic	White		Total
<=60		7	5	69	16	319		416
61-70		5	9	112	22	816		964
71-80		3	5	40	2	482		532
>80								
Total		15	19	221	40	1617		1912

In addition, 282 control backups were selected from the whites who were in HRT. We randomly selected 94 from each of the age categories from those that were not already selected with preference given to those who were in the core studies and who had prime 5 extraction.