

python assembling framingham databases

October 10, 2017

```
In [7]: import os
import pandas as pd
```

```
In [8]: os.chdir('P:\\Framingham hip\\Original datasets of interest')
```

```
In [9]: os.getcwd()
```

```
Out[9]: 'P:\\\\Framingham hip\\\\Original datasets of interest'
```

```
In [10]: dforiginal=pd.read_csv('Original_dataset_bmi_diab.csv')
dforiginal.head
```

```
Out[10]: <bound method NDFrame.head of
          PID  idtype  sex  att1  age1  examyr1  att2 da
0      10000142      0    1    1   45   1949      1  924   1951  48
1      10008641      0    1    1   36   1952      1  707   1954  38
2      10009718      0    1    1   34   1952      1  763   1954  36
3      10013675      0    2    1   47   1949      1  800   1951  50
4      10014385      0    2    1   57   1949      1  906   1951  59
5      10015218      0    2    1   45   1949      0
6      10018470      0    2    1   33   1951      1  729   1953  35
7      10018519      0    2    1   53   1949      1  766   1951  55
8      10021384      0    1    1   33   1950      1  734   1952  35
9      10024359      0    2    1   43   1951      1  686   1952  45
10     10028367      0    2    1   33   1951      0
11     10028966      0    2    1   33   1950      1  706   1952  35
12     10033400      0    2    1   56   1951      0
13     10033434      0    1    1   49   1952      1  699   1954  51
14     10034205      0    1    1   32   1949      1  787   1951  35
15     10035343      0    2    1   36   1951      1  744   1953  38
16     10037287      0    1    1   46   1949      1  637   1951  48
17     10037330      0    1    1   51   1952      1  664   1954  53
18     10037336      0    1    1   36   1951      1  645   1953  37
19     10038395      0    2    1   43   1951      1  671   1953  45
20     10040425      0    1    1   47   1951      1  666   1952  49
21     10042702      0    2    1   44   1951      1  676   1953  45
22     10046366      0    2    1   34   1952      1  717   1954  36
23     10046544      0    2    1   40   1950      1  722   1952  42
24     10048014      0    1    1   62   1949      1  782   1951  64
```

25	10052831	0	1	1	55	1951	1	670	1952	57
26	10055229	0	2	1	50	1949	1	913	1951	52
27	10055254	0	2	1	49	1951	1	746	1953	51
28	10056052	0	2	1	39	1951	1	700	1952	41
29	10058420	0	1	1	31	1948	1	971	1951	34
...
5049	19955175	0	2	1	54	1949	1	780	1951	56
5050	19956122	0	2	1	43	1952	0			
5051	19962079	0	1	1	38	1951	1	833	1953	41
5052	19962312	0	1	1	40	1950	0			
5053	19964296	0	2	1	35	1952	1	636	1954	37
5054	19966273	0	2	1	48	1948	1	953	1951	51
5055	19966326	0	2	1	45	1951	1	682	1952	47
5056	19967802	0	1	1	51	1951	1	663	1953	53
5057	19968397	0	2	1	37	1949	1	938	1952	39
5058	19970781	0	2	1	46	1949	1	778	1951	48
5059	19970948	0	1	1	48	1951	1	658	1952	50
5060	19971112	0	1	1	50	1951	1	893	1953	53
5061	19974096	0	1	1	60	1950	0			
5062	19975520	0	2	1	41	1948	1	907	1951	44
5063	19976090	0	1	1	47	1951	1	660	1952	48
5064	19976145	0	2	1	40	1948	1	1192	1952	43
5065	19976158	0	1	1	43	1950	1	841	1952	45
5066	19976680	0	2	1	42	1950	1	812	1952	44
5067	19979016	0	1	1	56	1951	1	666	1953	58
5068	19981203	0	1	1	40	1949	0			
5069	19981956	0	2	1	42	1949	1	933	1951	45
5070	19982857	0	2	1	35	1949	1	826	1951	37
5071	19983897	0	2	1	36	1951	1	746	1953	38
5072	19986805	0	2	1	46	1949	1	856	1952	48
5073	19989455	0	2	1	38	1952	0			
5074	19995267	0	1	1	49	1949	0			
5075	19995599	0	1	1	57	1952	1	697	1954	59
5076	19996887	0	2	1	40	1951	1	749	1953	42
5077	19997237	0	1	1	49	1951	1	667	1952	51
5078	19997933	0	1	1	34	1951	1	696	1953	36

	...	att26	date26	age26	examyr26	BMI16	BMI21	\
0	...	0						
1	...	0				23.8455366367454	19.119470432968	
2	...	0				23.0725131866375	21.9096844993141	
3	...	0				20.2596338158504		
4	...	0						
5	...	0						
6	...	0				24.4082088164176	23.671395479114	
7	...	0						
8	...	0						
9	...	0						

10	...	0						
11	...	0				25.070566983276	26.1233136094675	
12	...	0						
13	...	0						
14	...	0				32.906552094522	30.5020661157025	
15	...	1	18041	85	2001	30.9222047674339	30.4384268902776	
16	...	0				26.2175863877396		
17	...	0				29.6464100346021		
18	...	0				28.922969476675	29.1015063252964	
19	...	0				26.4217909204498		
20	...	0						
21	...	0						
22	...	1	17658	83	2000	34.4779420481295	35.5317259750017	
23	...	0				16.8787515006002		
24	...	0						
25	...	0						
26	...	0						
27	...	0						
28	...	0				21.7954325259516		
29	...	1	18947	83	2000	26.5883659665361	28.0812672176309	
...
5049	...	0				28.4823793164626		
5050	...	0						
5051	...	0				23.7974214876033		
5052	...	0				26.1496539792388	28.4871938018273	
5053	...	1	17745	84	2001	27.2494797086368	26.0470137442732	
5054	...	0						
5055	...	0				22.483642578125		
5056	...	0				25.0354671698844		
5057	...	0					23.513427734375	
5058	...	0						
5059	...	0						
5060	...	0						
5061	...	0						
5062	...	0						
5063	...	0						
5064	...	0						
5065	...	0						
5066	...	0					23.6052614184678	
5067	...	0						
5068	...	0				24.9872925619835	22.514072520799	
5069	...	0						
5070	...	0				21.3748430167228	20.6127446897126	
5071	...	1	17854	85	2000	24.7971781305115		
5072	...	0				25.1055862111724		
5073	...	0						
5074	...	0						
5075	...	0						

5076	...	0			24.8166219583816	26.123965742488
5077	...	0				
5078	...	1	18292	84	2001	21.3992865636147

	bmi26	Diab_16	Diab_21	Diab_26
0				
1		0	0	
2		0	1	
3		0		
4				
5				
6		1	1	
7				
8				
9		0		
10				
11		0	0	
12				
13				
14			1	
15		0	0	
16		0		
17		0		
18		0	0	
19		0		
20				
21				
22	33.6725197388001	0	0	0
23		0		
24				
25				
26				
27				
28		0		
29	27.9198806244261	0	0	0
...
5049		1		
5050				
5051		1		
5052		1	1	
5053	24.1999863397309	0	0	0
5054				
5055		1		
5056		1		
5057			1	
5058				
5059				
5060				

```

5061
5062
5063
5064
5065
5066          0
5067
5068          0      0
5069
5070          0      0
5071          0
5072          0
5073          1
5074
5075
5076          0      0
5077
5078 23.18923611111111 0      0      0

```

```
[5079 rows x 28 columns]>
```

```
In [11]: dflex17=pd.read_csv('lex0_17.csv')
dflex17.head
```

```
Out[11]: <bound method NDFrame.head of
```

	FJ4	FJ5	FJ8	FJ9	FJ10	FJ11	FJ12	FJ13	FJ14	FJ15	...
0	110	66	0	0	0	0	0	0	0	0	...
1	132	80	0	0	0	0	0	0	0	0	...
2	108	60	0	0	0	0	0	0	0	1	...
3	112	72	0	0	0	0	0	0	0	0	...
4	128	80	0	0	0	0	0	0	0	1	...
5	144	80	0	0	0	0	0	0	0	1	...
6	130	70	0	2	1	0	1	0	0	0	...
7	122	72	0	0	0	0	0	0	0	1	...
8	118	82	1	0	0	0	0	0	0	1	...
9	120	66	0	0	0	0	0	0	0	0	...
10	120	64	0	0	0	0	0	0	0	0	...
11	130	78	0	0	0	0	1	0	0	0	...
12	144	86	0	0	0	0	0	0	0	0	...
13	172	66	0	0	0	0	0	0	0	0	...
14	114	60	0	0	0	0	0	0	0	0	...
15	148	68	0	0	0	0	0	0	0	1	...
16	138	84	0	0	0	0	0	0	0	0	...
17	144	78	1	0	0	0	0	0	0	0	...
18	130	90	0	0	0	0	0	0	0	0	...
19	150	74	1	1	1	0	0	0	0	1	...
20	148	82	0	0	0	0	0	0	0	0	...
21	116	84	0	0	0	0	0	0	0	1	...
22	126	64	0	0	0	0	0	0	0	0	...

23	134	80	0	0	0	0	0	0	0	0	...	193	506	141
24	138	78	0	0	0	0	0	0	0	1	...	193	320	94
25	138	86	0	0	0	0	0	0	0	0	...	353	115	711
26	162	86	0	0	0	0	0	0	0	0	...	270	490	133
27	116	86	0	0	0	0	0	0	0	0	...			
28	152	82	0	0	0	0	0	0	0	1	...	241	432	104
29	148	66	0	0	0	0	0	0	0	0	...	156	390	78
...
1999	130	78	0	0	0	0	0	0	0	0	...	260	691	256
2000	200	90	0	0	0	0	1	0	0	1	...	277	419	161
2001	132	80	0	0	0	0	0	0	0	1	...	314	644	270
2002	134	74	0	0	0	0	0	0	0	0	...	131	229	36
2003	126	60	0	0	0	0	0	0	0	0	...	243	274	215
2004	154	88	0	0	0	0	1	0	0	1	...	287	728	277
2005	140	74	0	0	0	0	0	0	0	0	...	254	298	85
2006	140	96	0	0	0	0	0	0	0	0	...	362	805	410
2007	160	84	0	0	1	0	1	0	0	1	...	217	376	105
2008	150	86	0	0	0	0	0	0	0	1	...	127	230	38
2009	142	92	0	0	0	0	0	0	0	1	...	167	204	135
2010	132	60	1	1	0	0	1	1	0	0	...			
2011	150	70	0	0	0	0	0	0	0	1	...			
2012	120	70	2	2	2	0	0	0	0	0	...	252	624	225
2013	136	78	0	0	0	0	0	0	0	0	...	173	373	128
2014	134	78	0	0	0	0	1	0	0	1	...	277	607	221
2015	114	76	0	0	1	0	1	1	0	0	...	375	726	194
2016	110	70	0	0	0	0	0	0	0	0	...	424	111	230
2017	138	70	0	0	0	0	0	0	0	0	...	199	238	60
2018	144	76	1	0	0	0	1	0	0	1	...	371	890	167
2019	170	70	1	1	1	0	0	0	0	1	...			
2020	142	70	0	0	0	0	1	0	0	1	...	323	607	127
2021	160	90	0	0	0	0	0	0	0	0	...	322	553	283
2022	142	78	0	1	0	0	1	0	0	1	...	209	505	156
2023	138	80	0	0	0	0	1	0	0	0	...	153	468	81
2024	120	78	1	0	0	0	0	1	0	0	...	369	894	220
2025	128	58	0	0	0	0	0	0	0	0	...	225	340	191
2026	134	82	0	0	0	0	0	0	0	0	...	235	573	171
2027	136	84	0	0	0	0	0	0	0	1	...	278	559	219
2028	130	76	0	0	0	0	0	0	0	0	...	418	105	298

	FJ375	FJ376	FJ377	FJ380	FJ381	PID	IDTYPE
0	320	76	25	47	72	10008641	0
1	542	449	76	41	72	10009718	0
2						10013675	0
3						10014385	0
4	762	330	59	50	94	10018470	0
5	415	322	51	44	74	10028966	0
6	494	255	68	44	110	10034205	0
7	440	152	43	48	84	10035343	0

8	374	289	145	45	87	10037287	0
9	394	488	120	47	81	10037330	0
10	825	545	77	50	67	10037336	0
11	413	301	69	40	97	10046366	0
12	766	213	34	53	71	10058420	0
13	440	178	42	43	107	10060287	0
14	464	177	25	49	84	10063541	0
15	322	144	25	46	99	10072107	0
16	528	196	43	50	88	10075161	0
17	484	138	43	51	101	10077053	0
18	120	60	17	53	84	10086413	0
19	278	101	42	50	105	10088892	0
20	783	340	94	45	83	10107826	0
21	440	491	102	47	84	10113174	0
22	742	261	84	48	83	10114362	0
23	489	211	51	45	77	10118794	0
24	304	135	34	50	67	10128761	0
25	105	72	320	47	84	10129815	0
26	404	163	60			10132062	0
27						10137050	0
28	381	161	34	45	95	10152890	0
29	381	119	25	43	100	10154847	0
...
1999	649	363	126	46	91	19882074	0
2000	394	240	51	47	83	19883575	0
2001	610	407	110	50	84	19892253	0
2002	51	42	17	51	92	19892541	0
2003	145	197	86	43	80	19899790	0
2004	678	618	68	56	97	19903851	0
2005	255	128	34	51	154	19911069	0
2006	779	610	203	50	99	19914392	0
2007	376	180	26	42	71	19917074	0
2008	94	60	17	54	80	19923378	0
2009	196	187	60	48	93	19928195	0
2010				43	115	19928813	0
2011				46	117	19932675	0
2012	624	363	76	49	86	19934475	0
2013	339	152	59	45	121	19949586	0
2014	582	422	59	48	112	19951222	0
2015	709	406	69	47	70	19952723	0
2016	111	536	43	48	85	19953146	0
2017	145	77	26	50	98	19953696	0
2018	881	274	60	50	96	19954168	0
2019				42	77	19955175	0
2020	464	185	51	53	99	19962312	0
2021	553	417	102	45	67	19964296	0
2022	496	274	60	42	77	19968397	0
2023	460	119	17	46	93	19976680	0

2024	723	323	60	47	88	19981203	0
2025	340	255	60	45	88	19982857	0
2026	565	261	51	45	75	19983897	0
2027	525	305	76	49	78	19996887	0
2028	105	496	120	52	65	19997933	0

[2029 rows x 333 columns]>

In [12]: dflex17['PID']

Out[12]:

0	10008641
1	10009718
2	10013675
3	10014385
4	10018470
5	10028966
6	10034205
7	10035343
8	10037287
9	10037330
10	10037336
11	10046366
12	10058420
13	10060287
14	10063541
15	10072107
16	10075161
17	10077053
18	10086413
19	10088892
20	10107826
21	10113174
22	10114362
23	10118794
24	10128761
25	10129815
26	10132062
27	10137050
28	10152890
29	10154847
	...
1999	19882074
2000	19883575
2001	19892253
2002	19892541
2003	19899790
2004	19903851
2005	19911069

2006	19914392
2007	19917074
2008	19923378
2009	19928195
2010	19928813
2011	19932675
2012	19934475
2013	19949586
2014	19951222
2015	19952723
2016	19953146
2017	19953696
2018	19954168
2019	19955175
2020	19962312
2021	19964296
2022	19968397
2023	19976680
2024	19981203
2025	19982857
2026	19983897
2027	19996887
2028	19997933

Name: PID, Length: 2029, dtype: int64

```
In [13]: dfmerged = pd.merge(dforiginal, dflex17, on='PID', how='outer')
dfmerged.head
```

```
Out[13]: <bound method NDFrame.head of
```

	PID	idtype	sex	att1	age1	examy1	att2 da
0	10000142	0	1	1	45	1949	1 924 1951 48
1	10008641	0	1	1	36	1952	1 707 1954 38
2	10009718	0	1	1	34	1952	1 763 1954 36
3	10013675	0	2	1	47	1949	1 800 1951 50
4	10014385	0	2	1	57	1949	1 906 1951 59
5	10015218	0	2	1	45	1949	0
6	10018470	0	2	1	33	1951	1 729 1953 35
7	10018519	0	2	1	53	1949	1 766 1951 55
8	10021384	0	1	1	33	1950	1 734 1952 35
9	10024359	0	2	1	43	1951	1 686 1952 45
10	10028367	0	2	1	33	1951	0
11	10028966	0	2	1	33	1950	1 706 1952 35
12	10033400	0	2	1	56	1951	0
13	10033434	0	1	1	49	1952	1 699 1954 51
14	10034205	0	1	1	32	1949	1 787 1951 35
15	10035343	0	2	1	36	1951	1 744 1953 38
16	10037287	0	1	1	46	1949	1 637 1951 48
17	10037330	0	1	1	51	1952	1 664 1954 53
18	10037336	0	1	1	36	1951	1 645 1953 37

19	10038395	0	2	1	43	1951	1	671	1953	45
20	10040425	0	1	1	47	1951	1	666	1952	49
21	10042702	0	2	1	44	1951	1	676	1953	45
22	10046366	0	2	1	34	1952	1	717	1954	36
23	10046544	0	2	1	40	1950	1	722	1952	42
24	10048014	0	1	1	62	1949	1	782	1951	64
25	10052831	0	1	1	55	1951	1	670	1952	57
26	10055229	0	2	1	50	1949	1	913	1951	52
27	10055254	0	2	1	49	1951	1	746	1953	51
28	10056052	0	2	1	39	1951	1	700	1952	41
29	10058420	0	1	1	31	1948	1	971	1951	34
...
5049	19955175	0	2	1	54	1949	1	780	1951	56
5050	19956122	0	2	1	43	1952	0			
5051	19962079	0	1	1	38	1951	1	833	1953	41
5052	19962312	0	1	1	40	1950	0			
5053	19964296	0	2	1	35	1952	1	636	1954	37
5054	19966273	0	2	1	48	1948	1	953	1951	51
5055	19966326	0	2	1	45	1951	1	682	1952	47
5056	19967802	0	1	1	51	1951	1	663	1953	53
5057	19968397	0	2	1	37	1949	1	938	1952	39
5058	19970781	0	2	1	46	1949	1	778	1951	48
5059	19970948	0	1	1	48	1951	1	658	1952	50
5060	19971112	0	1	1	50	1951	1	893	1953	53
5061	19974096	0	1	1	60	1950	0			
5062	19975520	0	2	1	41	1948	1	907	1951	44
5063	19976090	0	1	1	47	1951	1	660	1952	48
5064	19976145	0	2	1	40	1948	1	1192	1952	43
5065	19976158	0	1	1	43	1950	1	841	1952	45
5066	19976680	0	2	1	42	1950	1	812	1952	44
5067	19979016	0	1	1	56	1951	1	666	1953	58
5068	19981203	0	1	1	40	1949	0			
5069	19981956	0	2	1	42	1949	1	933	1951	45
5070	19982857	0	2	1	35	1949	1	826	1951	37
5071	19983897	0	2	1	36	1951	1	746	1953	38
5072	19986805	0	2	1	46	1949	1	856	1952	48
5073	19989455	0	2	1	38	1952	0			
5074	19995267	0	1	1	49	1949	0			
5075	19995599	0	1	1	57	1952	1	697	1954	59
5076	19996887	0	2	1	40	1951	1	749	1953	42
5077	19997237	0	1	1	49	1951	1	667	1952	51
5078	19997933	0	1	1	34	1951	1	696	1953	36

	...	FJ371	FJ372	FJ373	FJ374	FJ375	FJ376	FJ377	FJ380	FJ381	IDTYPE
0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	...	189	269	422	71	320	76	25	47	72	0.0
2	...	325	401	542	235	542	449	76	41	72	0.0
3	...										0.0

4	...										0.0
5	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	...	257	308	779	195	762	330	59	50	94	0.0
7	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	...	228	278	424	170	415	322	51	44	74	0.0
12	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	...	237	288	587	189	494	255	68	44	110	0.0
15	...	169	203	466	129	440	152	43	48	84	0.0
16	...	280	343	374	260	374	289	145	45	87	0.0
17	...	234	277	488	306	394	488	120	47	81	0.0
18	...	348	424	834	273	825	545	77	50	67	0.0
19	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	...	231	283	430	182	413	301	69	40	97	0.0
23	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
27	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	...	246	297	800	141	766	213	34	53	71	0.0
...
5049	...								42	77	0.0
5050	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5051	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5052	...	239	323	607	127	464	185	51	53	99	0.0
5053	...	276	322	553	283	553	417	102	45	67	0.0
5054	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5055	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5056	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5057	...	179	209	505	156	496	274	60	42	77	0.0
5058	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5059	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5060	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5061	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5062	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5063	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5064	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5065	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5066	...	127	153	468	81	460	119	17	46	93	0.0
5067	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5068	...	305	369	894	220	723	323	60	47	88	0.0
5069	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5070	...	191	225	340	191	340	255	60	45	88	0.0
5071	...	197	235	573	171	565	261	51	45	75	0.0
5072	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5073	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5074	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5075	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5076	...	241	278	559	219	525	305	76	49	78	0.0
5077	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5078	...	354	418	105	298	105	496	120	52	65	0.0

[5079 rows x 360 columns]>

```
In [14]: dfmerged.to_csv('inception_lex017_merged')
```

```
In [15]: dflex16=pd.read_csv('lex0_16.csv')
         dflex16['FI43']
```

C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:2717: DtypeWarning:
interactivity=interactivity, compiler=compiler, result=result)

```
Out[15]: 0      3
         1      3
         2      3
         3      3
         4      3
         5      3
         6      3
         7      3
         8      3
         9      3
        10      3
        11      3
        12      3
        13      3
        14      3
        15      3
        16      3
        17      3
        18      3
        19      3
        20      3
        21      3
        22      3
        23      3
        24      3
        25      3
        26      3
        27      3
```

```

28      3
29      3
...
2206    3
2207    4
2208    3
2209    3
2210    5
2211    3
2212    3
2213    3
2214    3
2215    3
2216    4
2217    3
2218    3
2219    3
2220    3
2221    3
2222    3
2223    3
2224    4
2225    3
2226    3
2227    3
2228    4
2229    3
2230    3
2231    3
2232    3
2233    3
2234    3
2235    3
Name: FI43, Length: 2236, dtype: object

```

```

In [16]: dflex16_short =dflex16[['FI43','PID']]
         dflex16_short.head

```

```

Out[16]: <bound method NDFrame.head of          FI43          PID
0         3  10008641
1         3  10009718
2         3  10013675
3         3  10018470
4         3  10024359
5         3  10028966
6         3  10034205
7         3  10035343
8         3  10037287

```

9	3	10037330
10	3	10037336
11	3	10038395
12	3	10046366
13	3	10046544
14	3	10056052
15	3	10058420
16	3	10060287
17	3	10063541
18	3	10072107
19	3	10073680
20	3	10075161
21	3	10077053
22	3	10086413
23	3	10088892
24	3	10095046
25	3	10095159
26	3	10107826
27	3	10114362
28	3	10118794
29	3	10128761
...
2206	3	19867629
2207	4	19878064
2208	3	19882074
2209	3	19883575
2210	5	19892253
2211	3	19892541
2212	3	19899790
2213	3	19903851
2214	3	19911069
2215	3	19914392
2216	4	19917074
2217	3	19923378
2218	3	19928195
2219	3	19951222
2220	3	19952723
2221	3	19953146
2222	3	19953696
2223	3	19954168
2224	4	19955175
2225	3	19962079
2226	3	19962312
2227	3	19964296
2228	4	19966326
2229	3	19967802
2230	3	19981203
2231	3	19982857

```

2232    3  19983897
2233    3  19986805
2234    3  19996887
2235    3  19997933

```

```
[2236 rows x 2 columns]>
```

```

In [17]: df_orig_17=pd.read_csv('inception_lex017_merged.csv')
df_orig_17.head

```

```

Out[17]: <bound method NDFrame.head of
          0      1      2      3      4      5      6      7      8      9      10      11      12      13      14      15      16      17      18      19      20      21      22      23      24      25      26      27      28      29      ...
          0  10000142    0    1    1    45    1949    1    924
          1    10008641    0    1    1    36    1952    1    707
          2    10009718    0    1    1    34    1952    1    763
          3    10013675    0    2    1    47    1949    1    800
          4    10014385    0    2    1    57    1949    1    906
          5    10015218    0    2    1    45    1949    0
          6    10018470    0    2    1    33    1951    1    729
          7    10018519    0    2    1    53    1949    1    766
          8    10021384    0    1    1    33    1950    1    734
          9    10024359    0    2    1    43    1951    1    686
         10    10028367    0    2    1    33    1951    0
         11    10028966    0    2    1    33    1950    1    706
         12    10033400    0    2    1    56    1951    0
         13    10033434    0    1    1    49    1952    1    699
         14    10034205    0    1    1    32    1949    1    787
         15    10035343    0    2    1    36    1951    1    744
         16    10037287    0    1    1    46    1949    1    637
         17    10037330    0    1    1    51    1952    1    664
         18    10037336    0    1    1    36    1951    1    645
         19    10038395    0    2    1    43    1951    1    671
         20    10040425    0    1    1    47    1951    1    666
         21    10042702    0    2    1    44    1951    1    676
         22    10046366    0    2    1    34    1952    1    717
         23    10046544    0    2    1    40    1950    1    722
         24    10048014    0    1    1    62    1949    1    782
         25    10052831    0    1    1    55    1951    1    670
         26    10055229    0    2    1    50    1949    1    913
         27    10055254    0    2    1    49    1951    1    746
         28    10056052    0    2    1    39    1951    1    700
         29    10058420    0    1    1    31    1948    1    971
         ...      ...      ...      ...      ...      ...      ...      ...
        5049    19955175    0    2    1    54    1949    1    780
        5050    19956122    0    2    1    43    1952    0
        5051    19962079    0    1    1    38    1951    1    833
        5052    19962312    0    1    1    40    1950    0
        5053    19964296    0    2    1    35    1952    1    636
        5054    19966273    0    2    1    48    1948    1    953

```

5055	5055	19966326	0	2	1	45	1951	1	682
5056	5056	19967802	0	1	1	51	1951	1	663
5057	5057	19968397	0	2	1	37	1949	1	938
5058	5058	19970781	0	2	1	46	1949	1	778
5059	5059	19970948	0	1	1	48	1951	1	658
5060	5060	19971112	0	1	1	50	1951	1	893
5061	5061	19974096	0	1	1	60	1950	0	
5062	5062	19975520	0	2	1	41	1948	1	907
5063	5063	19976090	0	1	1	47	1951	1	660
5064	5064	19976145	0	2	1	40	1948	1	1192
5065	5065	19976158	0	1	1	43	1950	1	841
5066	5066	19976680	0	2	1	42	1950	1	812
5067	5067	19979016	0	1	1	56	1951	1	666
5068	5068	19981203	0	1	1	40	1949	0	
5069	5069	19981956	0	2	1	42	1949	1	933
5070	5070	19982857	0	2	1	35	1949	1	826
5071	5071	19983897	0	2	1	36	1951	1	746
5072	5072	19986805	0	2	1	46	1949	1	856
5073	5073	19989455	0	2	1	38	1952	0	
5074	5074	19995267	0	1	1	49	1949	0	
5075	5075	19995599	0	1	1	57	1952	1	697
5076	5076	19996887	0	2	1	40	1951	1	749
5077	5077	19997237	0	1	1	49	1951	1	667
5078	5078	19997933	0	1	1	34	1951	1	696

	examy2	...	bmi26	Diab_16	Diab_21	Diab_26	\
0	1951	...					
1	1954	...		0		0	
2	1954	...		0		1	
3	1951	...		0			
4	1951	...					
5		...					
6	1953	...		1		1	
7	1951	...					
8	1952	...					
9	1952	...		0			
10		...					
11	1952	...		0		0	
12		...					
13	1954	...					
14	1951	...				1	
15	1953	...		0		0	
16	1951	...		0			
17	1954	...		0			
18	1953	...		0		0	
19	1953	...		0			
20	1952	...					
21	1953	...					

22	1954	...	33.67251974	0	0	0
23	1952	...		0		
24	1951	...				
25	1952	...				
26	1951	...				
27	1953	...				
28	1952	...		0		
29	1951	...	27.91988062	0	0	0
...
5049	1951	...		1		
5050		...				
5051	1953	...		1		
5052		...		1	1	
5053	1954	...	24.19998634	0	0	0
5054	1951	...				
5055	1952	...		1		
5056	1953	...		1		
5057	1952	...			1	
5058	1951	...				
5059	1952	...				
5060	1953	...				
5061		...				
5062	1951	...				
5063	1952	...				
5064	1952	...				
5065	1952	...				
5066	1952	...			0	
5067	1953	...				
5068		...		0	0	
5069	1951	...				
5070	1951	...		0	0	
5071	1953	...		0		
5072	1952	...		0		
5073		...			1	
5074		...				
5075	1954	...				
5076	1953	...		0	0	
5077	1952	...				
5078	1953	...	23.18923611	0	0	0

	Estrogen_1980	Steroid_1980	Smoke_1980	Beer1980	Wine_1980	Cocktail_1980
0	NaN	NaN	NaN	NaN	NaN	NaN
1	0	1	0	0	7	7
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0		0	0	0
5	NaN	NaN	NaN	NaN	NaN	NaN
6	0	0	30	0	0	0

7	NaN	NaN	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN	NaN	NaN
11	1	0	88	0	4	4
12	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN	NaN
14	0	0	88	0	0	1
15	0	0	88	0	0	0
16	0	0	88	0	0	0
17	0	0	88	0	7	0
18	0	0	20	1	1	0
19	NaN	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN	NaN
22	0	0	88	0	3	1
23	NaN	NaN	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN	NaN
29	0	0	0	0	0	1
...
5049	0	0	88	0	0	0
5050	NaN	NaN	NaN	NaN	NaN	NaN
5051	NaN	NaN	NaN	NaN	NaN	NaN
5052	0	0	0	2	0	1
5053	0	0	88	0	1	0
5054	NaN	NaN	NaN	NaN	NaN	NaN
5055	NaN	NaN	NaN	NaN	NaN	NaN
5056	NaN	NaN	NaN	NaN	NaN	NaN
5057	0	0	88	5	2	14
5058	NaN	NaN	NaN	NaN	NaN	NaN
5059	NaN	NaN	NaN	NaN	NaN	NaN
5060	NaN	NaN	NaN	NaN	NaN	NaN
5061	NaN	NaN	NaN	NaN	NaN	NaN
5062	NaN	NaN	NaN	NaN	NaN	NaN
5063	NaN	NaN	NaN	NaN	NaN	NaN
5064	NaN	NaN	NaN	NaN	NaN	NaN
5065	NaN	NaN	NaN	NaN	NaN	NaN
5066	0	0	0	0	0	8
5067	NaN	NaN	NaN	NaN	NaN	NaN
5068	0	0	88	0	0	0
5069	NaN	NaN	NaN	NaN	NaN	NaN
5070	0	0	0	1	7	2
5071	0	0	88	0	0	4
5072	NaN	NaN	NaN	NaN	NaN	NaN

5073	NaN	NaN	NaN	NaN	NaN	NaN
5074	NaN	NaN	NaN	NaN	NaN	NaN
5075	NaN	NaN	NaN	NaN	NaN	NaN
5076	0	0	0	0	2	0
5077	NaN	NaN	NaN	NaN	NaN	NaN
5078	0	0	88	1	21	1

[5079 rows x 35 columns]>

```
In [18]: dfmerged = pd.merge(df_orig_17, dflex16_short, on='PID', how='outer')
```

```
In [19]: dfmerged.head
```

```
Out[19]: <bound method NDFrame.head of      Unnamed: 0      PID idtype sex att1 age1 exam
0          0  10000142      0      1      1    45   1949      1   924
1          1  10008641      0      1      1    36   1952      1   707
2          2  10009718      0      1      1    34   1952      1   763
3          3  10013675      0      2      1    47   1949      1   800
4          4  10014385      0      2      1    57   1949      1   906
5          5  10015218      0      2      1    45   1949      0
6          6  10018470      0      2      1    33   1951      1   729
7          7  10018519      0      2      1    53   1949      1   766
8          8  10021384      0      1      1    33   1950      1   734
9          9  10024359      0      2      1    43   1951      1   686
10         10  10028367      0      2      1    33   1951      0
11         11  10028966      0      2      1    33   1950      1   706
12         12  10033400      0      2      1    56   1951      0
13         13  10033434      0      1      1    49   1952      1   699
14         14  10034205      0      1      1    32   1949      1   787
15         15  10035343      0      2      1    36   1951      1   744
16         16  10037287      0      1      1    46   1949      1   637
17         17  10037330      0      1      1    51   1952      1   664
18         18  10037336      0      1      1    36   1951      1   645
19         19  10038395      0      2      1    43   1951      1   671
20         20  10040425      0      1      1    47   1951      1   666
21         21  10042702      0      2      1    44   1951      1   676
22         22  10046366      0      2      1    34   1952      1   717
23         23  10046544      0      2      1    40   1950      1   722
24         24  10048014      0      1      1    62   1949      1   782
25         25  10052831      0      1      1    55   1951      1   670
26         26  10055229      0      2      1    50   1949      1   913
27         27  10055254      0      2      1    49   1951      1   746
28         28  10056052      0      2      1    39   1951      1   700
29         29  10058420      0      1      1    31   1948      1   971
...         ...         ...         ...         ...         ...         ...         ...         ...
5049        5049  19955175      0      2      1    54   1949      1   780
5050        5050  19956122      0      2      1    43   1952      0
5051        5051  19962079      0      1      1    38   1951      1   833
```

5052	5052	19962312	0	1	1	40	1950	0	
5053	5053	19964296	0	2	1	35	1952	1	636
5054	5054	19966273	0	2	1	48	1948	1	953
5055	5055	19966326	0	2	1	45	1951	1	682
5056	5056	19967802	0	1	1	51	1951	1	663
5057	5057	19968397	0	2	1	37	1949	1	938
5058	5058	19970781	0	2	1	46	1949	1	778
5059	5059	19970948	0	1	1	48	1951	1	658
5060	5060	19971112	0	1	1	50	1951	1	893
5061	5061	19974096	0	1	1	60	1950	0	
5062	5062	19975520	0	2	1	41	1948	1	907
5063	5063	19976090	0	1	1	47	1951	1	660
5064	5064	19976145	0	2	1	40	1948	1	1192
5065	5065	19976158	0	1	1	43	1950	1	841
5066	5066	19976680	0	2	1	42	1950	1	812
5067	5067	19979016	0	1	1	56	1951	1	666
5068	5068	19981203	0	1	1	40	1949	0	
5069	5069	19981956	0	2	1	42	1949	1	933
5070	5070	19982857	0	2	1	35	1949	1	826
5071	5071	19983897	0	2	1	36	1951	1	746
5072	5072	19986805	0	2	1	46	1949	1	856
5073	5073	19989455	0	2	1	38	1952	0	
5074	5074	19995267	0	1	1	49	1949	0	
5075	5075	19995599	0	1	1	57	1952	1	697
5076	5076	19996887	0	2	1	40	1951	1	749
5077	5077	19997237	0	1	1	49	1951	1	667
5078	5078	19997933	0	1	1	34	1951	1	696

	examyr2	...	Diab_16	Diab_21	Diab_26	Estrogen_1980	Steroid_1980	\
0	1951	...				NaN	NaN	
1	1954	...	0	0		0	1	
2	1954	...	0	1		0	0	
3	1951	...	0			0	0	
4	1951	...				0	0	
5		...				NaN	NaN	
6	1953	...	1	1		0	0	
7	1951	...				NaN	NaN	
8	1952	...				NaN	NaN	
9	1952	...	0			NaN	NaN	
10		...				NaN	NaN	
11	1952	...	0	0		1	0	
12		...				NaN	NaN	
13	1954	...				NaN	NaN	
14	1951	...		1		0	0	
15	1953	...	0	0		0	0	
16	1951	...	0			0	0	
17	1954	...	0			0	0	
18	1953	...	0	0		0	0	

19	1953	...	0			NaN	NaN
20	1952	...				NaN	NaN
21	1953	...				NaN	NaN
22	1954	...	0	0	0	0	0
23	1952	...	0			NaN	NaN
24	1951	...				NaN	NaN
25	1952	...				NaN	NaN
26	1951	...				NaN	NaN
27	1953	...				NaN	NaN
28	1952	...	0			NaN	NaN
29	1951	...	0	0	0	0	0
...
5049	1951	...	1			0	0
5050		...				NaN	NaN
5051	1953	...	1			NaN	NaN
5052		...	1	1		0	0
5053	1954	...	0	0	0	0	0
5054	1951	...				NaN	NaN
5055	1952	...	1			NaN	NaN
5056	1953	...	1			NaN	NaN
5057	1952	...		1		0	0
5058	1951	...				NaN	NaN
5059	1952	...				NaN	NaN
5060	1953	...				NaN	NaN
5061		...				NaN	NaN
5062	1951	...				NaN	NaN
5063	1952	...				NaN	NaN
5064	1952	...				NaN	NaN
5065	1952	...				NaN	NaN
5066	1952	...		0		0	0
5067	1953	...				NaN	NaN
5068		...	0	0		0	0
5069	1951	...				NaN	NaN
5070	1951	...	0	0		0	0
5071	1953	...	0			0	0
5072	1952	...	0			NaN	NaN
5073		...		1		NaN	NaN
5074		...				NaN	NaN
5075	1954	...				NaN	NaN
5076	1953	...	0	0		0	0
5077	1952	...				NaN	NaN
5078	1953	...	0	0	0	0	0

	Smoke_1980	Beer1980	Wine_1980	Cocktail_1980	FI43
0	NaN	NaN	NaN	NaN	NaN
1	0	0	7	7	3
2	0	0	0	0	3
3	0	0	0	0	3

4		0	0	0	NaN
5	NaN	NaN	NaN	NaN	NaN
6	30	0	0	0	3
7	NaN	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	3
10	NaN	NaN	NaN	NaN	NaN
11	88	0	4	4	3
12	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN
14	88	0	0	1	3
15	88	0	0	0	3
16	88	0	0	0	3
17	88	0	7	0	3
18	20	1	1	0	3
19	NaN	NaN	NaN	NaN	3
20	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN
22	88	0	3	1	3
23	NaN	NaN	NaN	NaN	3
24	NaN	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	3
29	0	0	0	1	3
...
5049	88	0	0	0	4
5050	NaN	NaN	NaN	NaN	NaN
5051	NaN	NaN	NaN	NaN	3
5052	0	2	0	1	3
5053	88	0	1	0	3
5054	NaN	NaN	NaN	NaN	NaN
5055	NaN	NaN	NaN	NaN	4
5056	NaN	NaN	NaN	NaN	3
5057	88	5	2	14	NaN
5058	NaN	NaN	NaN	NaN	NaN
5059	NaN	NaN	NaN	NaN	NaN
5060	NaN	NaN	NaN	NaN	NaN
5061	NaN	NaN	NaN	NaN	NaN
5062	NaN	NaN	NaN	NaN	NaN
5063	NaN	NaN	NaN	NaN	NaN
5064	NaN	NaN	NaN	NaN	NaN
5065	NaN	NaN	NaN	NaN	NaN
5066	0	0	0	8	NaN
5067	NaN	NaN	NaN	NaN	NaN
5068	88	0	0	0	3
5069	NaN	NaN	NaN	NaN	NaN

5070	0	1	7	2	3
5071	88	0	0	4	3
5072	NaN	NaN	NaN	NaN	3
5073	NaN	NaN	NaN	NaN	NaN
5074	NaN	NaN	NaN	NaN	NaN
5075	NaN	NaN	NaN	NaN	NaN
5076	0	0	2	0	3
5077	NaN	NaN	NaN	NaN	NaN
5078	88	1	21	1	3

[5079 rows x 36 columns]>

```
In [20]: dfmerged.to_csv('original_lex17_bblock_bmi')
```

```
In [21]: dfmeno=pd.read_csv('vr_meno.csv')
dfmerged_meno = pd.merge(dfmerged, dfmeno, on='PID', how='outer')
dfmerged_meno.head
```

```
Out[21]: <bound method NDFrame.head of
0      0  10000142      0  1  1  45  1949  1  924
1      1  10008641      0  1  1  36  1952  1  707
2      2  10009718      0  1  1  34  1952  1  763
3      3  10013675      0  2  1  47  1949  1  800
4      4  10014385      0  2  1  57  1949  1  906
5      5  10015218      0  2  1  45  1949  0
6      6  10018470      0  2  1  33  1951  1  729
7      7  10018519      0  2  1  53  1949  1  766
8      8  10021384      0  1  1  33  1950  1  734
9      9  10024359      0  2  1  43  1951  1  686
10     10 10028367      0  2  1  33  1951  0
11     11 10028966      0  2  1  33  1950  1  706
12     12 10033400      0  2  1  56  1951  0
13     13 10033434      0  1  1  49  1952  1  699
14     14 10034205      0  1  1  32  1949  1  787
15     15 10035343      0  2  1  36  1951  1  744
16     16 10037287      0  1  1  46  1949  1  637
17     17 10037330      0  1  1  51  1952  1  664
18     18 10037336      0  1  1  36  1951  1  645
19     19 10038395      0  2  1  43  1951  1  671
20     20 10040425      0  1  1  47  1951  1  666
21     21 10042702      0  2  1  44  1951  1  676
22     22 10046366      0  2  1  34  1952  1  717
23     23 10046544      0  2  1  40  1950  1  722
24     24 10048014      0  1  1  62  1949  1  782
25     25 10052831      0  1  1  55  1951  1  670
26     26 10055229      0  2  1  50  1949  1  913
27     27 10055254      0  2  1  49  1951  1  746
28     28 10056052      0  2  1  39  1951  1  700
```

29	29	10058420	0	1	1	31	1948	1	971
...
5049	5049	19955175	0	2	1	54	1949	1	780
5050	5050	19956122	0	2	1	43	1952	0	
5051	5051	19962079	0	1	1	38	1951	1	833
5052	5052	19962312	0	1	1	40	1950	0	
5053	5053	19964296	0	2	1	35	1952	1	636
5054	5054	19966273	0	2	1	48	1948	1	953
5055	5055	19966326	0	2	1	45	1951	1	682
5056	5056	19967802	0	1	1	51	1951	1	663
5057	5057	19968397	0	2	1	37	1949	1	938
5058	5058	19970781	0	2	1	46	1949	1	778
5059	5059	19970948	0	1	1	48	1951	1	658
5060	5060	19971112	0	1	1	50	1951	1	893
5061	5061	19974096	0	1	1	60	1950	0	
5062	5062	19975520	0	2	1	41	1948	1	907
5063	5063	19976090	0	1	1	47	1951	1	660
5064	5064	19976145	0	2	1	40	1948	1	1192
5065	5065	19976158	0	1	1	43	1950	1	841
5066	5066	19976680	0	2	1	42	1950	1	812
5067	5067	19979016	0	1	1	56	1951	1	666
5068	5068	19981203	0	1	1	40	1949	0	
5069	5069	19981956	0	2	1	42	1949	1	933
5070	5070	19982857	0	2	1	35	1949	1	826
5071	5071	19983897	0	2	1	36	1951	1	746
5072	5072	19986805	0	2	1	46	1949	1	856
5073	5073	19989455	0	2	1	38	1952	0	
5074	5074	19995267	0	1	1	49	1949	0	
5075	5075	19995599	0	1	1	57	1952	1	697
5076	5076	19996887	0	2	1	40	1951	1	749
5077	5077	19997237	0	1	1	49	1951	1	667
5078	5078	19997933	0	1	1	34	1951	1	696

	examyr2	...	Smoke_1980	Beer1980	Wine_1980	Cocktail_1980	FI43	AM2	\
0	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	
1	1954	...	0	0	7	7	3	NaN	
2	1954	...	0	0	0	0	3	NaN	
3	1951	...	0	0	0	0	3	51	
4	1951	...		0	0	0	NaN	45	
5		...	NaN	NaN	NaN	NaN	NaN		
6	1953	...	30	0	0	0	3	50	
7	1951	...	NaN	NaN	NaN	NaN	NaN	48	
8	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	
9	1952	...	NaN	NaN	NaN	NaN	3	48	
10		...	NaN	NaN	NaN	NaN	NaN	32	
11	1952	...	88	0	4	4	3	55	
12		...	NaN	NaN	NaN	NaN	NaN	45	
13	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	

14	1951	...	88	0	0	1	3	NaN
15	1953	...	88	0	0	0	3	52
16	1951	...	88	0	0	0	3	NaN
17	1954	...	88	0	7	0	3	NaN
18	1953	...	20	1	1	0	3	NaN
19	1953	...	NaN	NaN	NaN	NaN	3	49
20	1952	...	NaN	NaN	NaN	NaN	NaN	NaN
21	1953	...	NaN	NaN	NaN	NaN	NaN	43
22	1954	...	88	0	3	1	3	41
23	1952	...	NaN	NaN	NaN	NaN	3	49
24	1951	...	NaN	NaN	NaN	NaN	NaN	NaN
25	1952	...	NaN	NaN	NaN	NaN	NaN	NaN
26	1951	...	NaN	NaN	NaN	NaN	NaN	43
27	1953	...	NaN	NaN	NaN	NaN	NaN	26
28	1952	...	NaN	NaN	NaN	NaN	3	51
29	1951	...	0	0	0	1	3	NaN
...
5049	1951	...	88	0	0	0	4	47
5050		...	NaN	NaN	NaN	NaN	NaN	
5051	1953	...	NaN	NaN	NaN	NaN	3	NaN
5052		...	0	2	0	1	3	NaN
5053	1954	...	88	0	1	0	3	47
5054	1951	...	NaN	NaN	NaN	NaN	NaN	47
5055	1952	...	NaN	NaN	NaN	NaN	4	52
5056	1953	...	NaN	NaN	NaN	NaN	3	NaN
5057	1952	...	88	5	2	14	NaN	48
5058	1951	...	NaN	NaN	NaN	NaN	NaN	52
5059	1952	...	NaN	NaN	NaN	NaN	NaN	NaN
5060	1953	...	NaN	NaN	NaN	NaN	NaN	NaN
5061		...	NaN	NaN	NaN	NaN	NaN	NaN
5062	1951	...	NaN	NaN	NaN	NaN	NaN	28
5063	1952	...	NaN	NaN	NaN	NaN	NaN	NaN
5064	1952	...	NaN	NaN	NaN	NaN	NaN	51
5065	1952	...	NaN	NaN	NaN	NaN	NaN	NaN
5066	1952	...	0	0	0	8	NaN	55
5067	1953	...	NaN	NaN	NaN	NaN	NaN	NaN
5068		...	88	0	0	0	3	NaN
5069	1951	...	NaN	NaN	NaN	NaN	NaN	47
5070	1951	...	0	1	7	2	3	44
5071	1953	...	88	0	0	4	3	48
5072	1952	...	NaN	NaN	NaN	NaN	3	49
5073		...	NaN	NaN	NaN	NaN	NaN	43
5074		...	NaN	NaN	NaN	NaN	NaN	NaN
5075	1954	...	NaN	NaN	NaN	NaN	NaN	NaN
5076	1953	...	0	0	2	0	3	52
5077	1952	...	NaN	NaN	NaN	NaN	NaN	NaN
5078	1953	...	88	1	21	1	3	NaN

	AM3	AM4	AM5	IDTYPE
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN
3	0	0	3	0.0
4	0	0	1	0.0
5				0.0
6	0	0	10	0.0
7	0	0	1	0.0
8	NaN	NaN	NaN	NaN
9	0	0	4	0.0
10	1	1	1	0.0
11	1	2	12	0.0
12	0	0	1	0.0
13	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN
15	0	0	9	0.0
16	NaN	NaN	NaN	NaN
17	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	NaN
19	1	2	4	0.0
20	NaN	NaN	NaN	NaN
21	1		1	0.0
22	1	0	5	0.0
23	0	0	6	0.0
24	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN
26	1	2	1	0.0
27	1	2	1	0.0
28	0	0	7	0.0
29	NaN	NaN	NaN	NaN
...
5049	0	2	1	0.0
5050				0.0
5051	NaN	NaN	NaN	NaN
5052	NaN	NaN	NaN	NaN
5053	1	2	7	0.0
5054	0	0	1	0.0
5055	0	0	5	0.0
5056	NaN	NaN	NaN	NaN
5057	0	0	7	0.0
5058	0	0	4	0.0
5059	NaN	NaN	NaN	NaN
5060	NaN	NaN	NaN	NaN
5061	NaN	NaN	NaN	NaN
5062	1	2	1	0.0
5063	NaN	NaN	NaN	NaN
5064	0	0	6	0.0

5065	NaN	NaN	NaN	NaN
5066	0	0	8	0.0
5067	NaN	NaN	NaN	NaN
5068	NaN	NaN	NaN	NaN
5069	1	2	3	0.0
5070	0	0	6	0.0
5071	0	0	7	0.0
5072	1	2	3	0.0
5073	1	1	4	0.0
5074	NaN	NaN	NaN	NaN
5075	NaN	NaN	NaN	NaN
5076	0	0	7	0.0
5077	NaN	NaN	NaN	NaN
5078	NaN	NaN	NaN	NaN

[5079 rows x 41 columns]>

// At this point, I have assembled the bmi, vr diabetes, vr meno, of lex 17 data plus beta blocker into one file

Now I need to assemble lex 21 and 26 data. I'll make this easier, but creating dataframes of just the columns i want. '

```
In [22]: dfmerged_meno.to_csv('Original_lex17_bBlock_bmi_meo_diab.csv')
```

bring up lex 21 and select just he columns we need.

```
In [23]: df_lex21=pd.read_csv('lex0_21.csv')
```

```
In [24]: df_lex21_short=df_lex21[['FN185','FN125','FN175','FN178','FN181','FN124','FN99','PID']]
```

```
In [25]: dfmerged2= pd.merge(dfmerged_meno, df_lex21_short, on='PID', how='outer')
```

```
In [26]: dfmerged2.head
```

```
Out[26]: <bound method NDFrame.head of      Unnamed: 0      PID  idtype  sex  att1  age1  exam
0      0  10000142      0      1      1   45   1949      1   924
1      1  10008641      0      1      1   36   1952      1   707
2      2  10009718      0      1      1   34   1952      1   763
3      3  10013675      0      2      1   47   1949      1   800
4      4  10014385      0      2      1   57   1949      1   906
5      5  10015218      0      2      1   45   1949      0
6      6  10018470      0      2      1   33   1951      1   729
7      7  10018519      0      2      1   53   1949      1   766
8      8  10021384      0      1      1   33   1950      1   734
9      9  10024359      0      2      1   43   1951      1   686
10     10  10028367      0      2      1   33   1951      0
11     11  10028966      0      2      1   33   1950      1   706
12     12  10033400      0      2      1   56   1951      0
13     13  10033434      0      1      1   49   1952      1   699
```

14	14	10034205	0	1	1	32	1949	1	787
15	15	10035343	0	2	1	36	1951	1	744
16	16	10037287	0	1	1	46	1949	1	637
17	17	10037330	0	1	1	51	1952	1	664
18	18	10037336	0	1	1	36	1951	1	645
19	19	10038395	0	2	1	43	1951	1	671
20	20	10040425	0	1	1	47	1951	1	666
21	21	10042702	0	2	1	44	1951	1	676
22	22	10046366	0	2	1	34	1952	1	717
23	23	10046544	0	2	1	40	1950	1	722
24	24	10048014	0	1	1	62	1949	1	782
25	25	10052831	0	1	1	55	1951	1	670
26	26	10055229	0	2	1	50	1949	1	913
27	27	10055254	0	2	1	49	1951	1	746
28	28	10056052	0	2	1	39	1951	1	700
29	29	10058420	0	1	1	31	1948	1	971
...
5049	5049	19955175	0	2	1	54	1949	1	780
5050	5050	19956122	0	2	1	43	1952	0	
5051	5051	19962079	0	1	1	38	1951	1	833
5052	5052	19962312	0	1	1	40	1950	0	
5053	5053	19964296	0	2	1	35	1952	1	636
5054	5054	19966273	0	2	1	48	1948	1	953
5055	5055	19966326	0	2	1	45	1951	1	682
5056	5056	19967802	0	1	1	51	1951	1	663
5057	5057	19968397	0	2	1	37	1949	1	938
5058	5058	19970781	0	2	1	46	1949	1	778
5059	5059	19970948	0	1	1	48	1951	1	658
5060	5060	19971112	0	1	1	50	1951	1	893
5061	5061	19974096	0	1	1	60	1950	0	
5062	5062	19975520	0	2	1	41	1948	1	907
5063	5063	19976090	0	1	1	47	1951	1	660
5064	5064	19976145	0	2	1	40	1948	1	1192
5065	5065	19976158	0	1	1	43	1950	1	841
5066	5066	19976680	0	2	1	42	1950	1	812
5067	5067	19979016	0	1	1	56	1951	1	666
5068	5068	19981203	0	1	1	40	1949	0	
5069	5069	19981956	0	2	1	42	1949	1	933
5070	5070	19982857	0	2	1	35	1949	1	826
5071	5071	19983897	0	2	1	36	1951	1	746
5072	5072	19986805	0	2	1	46	1949	1	856
5073	5073	19989455	0	2	1	38	1952	0	
5074	5074	19995267	0	1	1	49	1949	0	
5075	5075	19995599	0	1	1	57	1952	1	697
5076	5076	19996887	0	2	1	40	1951	1	749
5077	5077	19997237	0	1	1	49	1951	1	667
5078	5078	19997933	0	1	1	34	1951	1	696

	examyr2	...	AM4	AM5	IDTYPE	FN185	FN125	FN175	FN178	FN181	FN124	FN99
0	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1954	...	NaN	NaN	NaN	0.0	0.0	0	0	0	0	0.0
2	1954	...	NaN	NaN	NaN	0.0	0.0	0	0	0	0	0.0
3	1951	...	0	3	0.0	0.0	0.0	0	0	0	0	0.0
4	1951	...	0	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5		...			0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	1953	...	0	10	0.0	0.0	0.0	0	0	0	0	0.0
7	1951	...	0	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	1952	...	0	4	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10		...	1	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	1952	...	2	12	0.0	0.0	0.0	0	1	8	1	1.0
12		...	0	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	1951	...	NaN	NaN	NaN	0.0	0.0	0	0	0	0	0.0
15	1953	...	0	9	0.0	0.0	0.0	0	0	0	0	0.0
16	1951	...	NaN	NaN	NaN	0.0	0.0	0	0	0	0	0.0
17	1954	...	NaN	NaN	NaN	0.0	0.0	0	0	0	0	0.0
18	1953	...	NaN	NaN	NaN	1.0	0.0	0	0	0	0	0.0
19	1953	...	2	4	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	1953	...		1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	1954	...	0	5	0.0	0.0	0.0	1	2	1	0	0.0
23	1952	...	0	6	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26	1951	...	2	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
27	1953	...	2	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	1952	...	0	7	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	1951	...	NaN	NaN	NaN	0.0	0.0	0	4	1	0	0.0
...
5049	1951	...	2	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5050		...			0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5051	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5052		...	NaN	NaN	NaN	0.0	0.0	1	1	1	0	0.0
5053	1954	...	2	7	0.0	0.0	0.0	0	0	0	0	1.0
5054	1951	...	0	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5055	1952	...	0	5	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5056	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5057	1952	...	0	7	0.0	0.0	0.0	0	0	7	0	0.0
5058	1951	...	0	4	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5059	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5060	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5061		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5062	1951	...	2	1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5063	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5064	1952	...	0	6	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5065	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5066	1952	...	0	8	0.0	0.0	0.0	0	0	7	0	1.0
5067	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5068		...	NaN	NaN	NaN	0.0	0.0	0	0	0	0	0.0
5069	1951	...	2	3	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5070	1951	...	0	6	0.0	0.0	0.0	0	1	6	0	0.0
5071	1953	...	0	7	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5072	1952	...	2	3	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5073		...	1	4	0.0	0.0	0.0	0	0	0	0	0.0
5074		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5075	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5076	1953	...	0	7	0.0	0.0	0.0	0	1	0	0	0.0
5077	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5078	1953	...	NaN	NaN	NaN	0.0	0.0	2	7	3	0	0.0

[5079 rows x 48 columns]>

In [27]: dfmerged2.to_csv('Original_lex17_bBlock_bmi_meo_diab_1990.csv')

In [28]: df_lex26=pd.read_csv('lex0_26.csv')

In [29]: df_lex26_short=df_lex26[['FS347','FS291','FS345', 'FS339', 'FS336', 'FS333', 'FS329'],'F'

In [30]: dfmerged3=pd.merge(dfmerged2, df_lex26_short, on='PID', how='outer')
dfmerged3.to_csv('Original_lex17_bBlock_bmi_meo_diab_1990_2000.csv')
dfmerged3.head

Out[30]: <bound method NDFrame.head of Unnamed: 0 PID idtype sex att1 age1 exam

0	0	10000142	0	1	1	45	1949	1	924	
1	1	10008641	0	1	1	36	1952	1	707	
2	2	10009718	0	1	1	34	1952	1	763	
3	3	10013675	0	2	1	47	1949	1	800	
4	4	10014385	0	2	1	57	1949	1	906	
5	5	10015218	0	2	1	45	1949	0		
6	6	10018470	0	2	1	33	1951	1	729	
7	7	10018519	0	2	1	53	1949	1	766	
8	8	10021384	0	1	1	33	1950	1	734	
9	9	10024359	0	2	1	43	1951	1	686	
10	10	10028367	0	2	1	33	1951	0		
11	11	10028966	0	2	1	33	1950	1	706	
12	12	10033400	0	2	1	56	1951	0		
13	13	10033434	0	1	1	49	1952	1	699	
14	14	10034205	0	1	1	32	1949	1	787	
15	15	10035343	0	2	1	36	1951	1	744	
16	16	10037287	0	1	1	46	1949	1	637	
17	17	10037330	0	1	1	51	1952	1	664	
18	18	10037336	0	1	1	36	1951	1	645	
19	19	10038395	0	2	1	43	1951	1	671	
20	20	10040425	0	1	1	47	1951	1	666	

21	21	10042702	0	2	1	44	1951	1	676
22	22	10046366	0	2	1	34	1952	1	717
23	23	10046544	0	2	1	40	1950	1	722
24	24	10048014	0	1	1	62	1949	1	782
25	25	10052831	0	1	1	55	1951	1	670
26	26	10055229	0	2	1	50	1949	1	913
27	27	10055254	0	2	1	49	1951	1	746
28	28	10056052	0	2	1	39	1951	1	700
29	29	10058420	0	1	1	31	1948	1	971
...
5049	5049	19955175	0	2	1	54	1949	1	780
5050	5050	19956122	0	2	1	43	1952	0	
5051	5051	19962079	0	1	1	38	1951	1	833
5052	5052	19962312	0	1	1	40	1950	0	
5053	5053	19964296	0	2	1	35	1952	1	636
5054	5054	19966273	0	2	1	48	1948	1	953
5055	5055	19966326	0	2	1	45	1951	1	682
5056	5056	19967802	0	1	1	51	1951	1	663
5057	5057	19968397	0	2	1	37	1949	1	938
5058	5058	19970781	0	2	1	46	1949	1	778
5059	5059	19970948	0	1	1	48	1951	1	658
5060	5060	19971112	0	1	1	50	1951	1	893
5061	5061	19974096	0	1	1	60	1950	0	
5062	5062	19975520	0	2	1	41	1948	1	907
5063	5063	19976090	0	1	1	47	1951	1	660
5064	5064	19976145	0	2	1	40	1948	1	1192
5065	5065	19976158	0	1	1	43	1950	1	841
5066	5066	19976680	0	2	1	42	1950	1	812
5067	5067	19979016	0	1	1	56	1951	1	666
5068	5068	19981203	0	1	1	40	1949	0	
5069	5069	19981956	0	2	1	42	1949	1	933
5070	5070	19982857	0	2	1	35	1949	1	826
5071	5071	19983897	0	2	1	36	1951	1	746
5072	5072	19986805	0	2	1	46	1949	1	856
5073	5073	19989455	0	2	1	38	1952	0	
5074	5074	19995267	0	1	1	49	1949	0	
5075	5075	19995599	0	1	1	57	1952	1	697
5076	5076	19996887	0	2	1	40	1951	1	749
5077	5077	19997237	0	1	1	49	1951	1	667
5078	5078	19997933	0	1	1	34	1951	1	696

	examyr2	...	FS347	FS291	FS345	FS339	FS336	FS333	FS329	FS314	FS307	\
0	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

6	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	1953	...	0	0	0	0	0	0	0	0.0	1
16	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	1954	...	0	0	0	0	1	0	0	0.0	0
23	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
27	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	1951	...	0	0	0	0	0	0	0	8.0	0
...
5049	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5050		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5051	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5052		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5053	1954	...	0	0	0	0	0	0	0	0.0	0
5054	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5055	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5056	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5057	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5058	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5059	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5060	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5061		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5062	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5063	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5064	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5065	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5066	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5067	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5068		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5069	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5070	1951	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5071	1953	...	0	0	0	0	0	0	0	0.0	0

5072	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5073		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5074		...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5075	1954	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5076	1953	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5077	1952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5078	1953	...	0	0	3	0	2	2	0	8.0	0	

FS245

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	NaN
10	NaN
11	NaN
12	NaN
13	NaN
14	NaN
15	1
16	NaN
17	NaN
18	NaN
19	NaN
20	NaN
21	NaN
22	0
23	NaN
24	NaN
25	NaN
26	NaN
27	NaN
28	NaN
29	0
...	...
5049	NaN
5050	NaN
5051	NaN
5052	NaN
5053	1
5054	NaN
5055	NaN
5056	NaN

```

5057    NaN
5058    NaN
5059    NaN
5060    NaN
5061    NaN
5062    NaN
5063    NaN
5064    NaN
5065    NaN
5066    NaN
5067    NaN
5068    NaN
5069    NaN
5070    NaN
5071      0
5072    NaN
5073    NaN
5074    NaN
5075    NaN
5076    NaN
5077    NaN
5078      0

```

```
[5079 rows x 58 columns]>
```

And, we're done. Renamed the columns using excel'
Moving on to Offsprings. Start by loading vr_dates_yr file

```
In [31]: os.chdir('P:\Framingham hip\Framingham Offspring Cohort datasets of Interest')
```

```
In [32]: df_offspring=pd.read_csv('vr_dates.csv')
df_offspring.head
```

```
Out[32]: <bound method NDFrame.head of          idtype      PID  sex  att1  age1  examyr1  att2 dat
0          1    2924    2    1    36    1972     1  2949   44    1980
1          1    3297    1    1    31    1971     1  2947   39    1979
2          7    4040    2    1    43    1997     0
3          1    4061    1    1    31    1972     1  2984   39    1980
4          7    4301    2    1    51    1996     0
5          1    5350    2    1    28    1975     0
6          7    6719    2    1    44    1996     1  1870   49    2001
7          1    7248    2    1    41    1973     0
8          1    8059    2    1    38    1973     1  3405   47    1982
9          7    8459    2    1    51    1996     0
10         1    9458    2    1    36    1972     1  2953   44    1980
11         7    9912    2    1    59    1996     1  1199   62    1999
12         7   10791    1    1    46    1994     1  1825   51    1999
13         7   11016    2    1    59    1994     1  1739   63    1999
14         7   11651    2    1    50    1996     1  1355   54    1999
```

15	1	14463	2	1	34	1974	1	2676	41	1982
16	1	14764	2	1	44	1972	0			
17	1	15059	2	1	24	1971	1	2971	32	1980
18	7	15398	2	1	53	1995	0			
19	7	18098	1	1	44	1996	1	1092	47	1999
20	7	18103	2	1	41	1997	1	1015	44	1999
21	1	18844	2	1	19	1973	1	2942	27	1981
22	7	19353	2	1	52	1995	0			
23	7	19892	1	1	69	1998	1	665	70	2000
24	1	20206	2	1	37	1975	1	2648	44	1982
25	1	22267	1	1	30	1972	1	2926	38	1980
26	7	22567	1	1	64	1998	1	735	66	2000
27	1	22996	2	1	61	1973	0			
28	7	24738	1	1	57	1996	1	1605	61	2000
29	7	25240	2	1	53	1997	0			
...
5477	1	9939621	1	1	45	1975	1	2569	52	1982
5478	1	9939888	1	1	47	1974	1	2679	55	1982
5479	1	9940525	2	1	58	1973	0			
5480	1	9940870	1	1	33	1973	1	2885	41	1981
5481	1	9941870	2	1	29	1974	1	2695	37	1982
5482	1	9943257	1	1	35	1972	0			
5483	1	9945338	2	1	38	1972	1	3014	46	1980
5484	1	9947765	2	1	26	1974	1	2815	34	1981
5485	1	9947948	2	1	46	1972	1	2948	54	1980
5486	1	9948725	2	1	41	1973	1	2803	49	1981
5487	1	9951579	1	1	43	1974	1	2802	51	1981
5488	1	9957447	1	1	45	1972	0			
5489	1	9958150	2	1	47	1973	1	2590	54	1980
5490	1	9965528	1	1	25	1971	1	3110	34	1980
5491	1	9967264	2	1	35	1974	0			
5492	1	9973742	1	1	22	1973	0			
5493	1	9975906	2	1	28	1973	1	2800	35	1981
5494	1	9977699	1	1	27	1972	1	2962	35	1980
5495	1	9980291	2	1	27	1971	1	2962	35	1979
5496	1	9982799	2	1	28	1973	1	2775	36	1981
5497	1	9982997	2	1	28	1972	1	3452	38	1981
5498	1	9985882	1	1	35	1973	1	2897	43	1981
5499	1	9986747	1	1	48	1973	1	2773	56	1981
5500	1	9987740	1	1	17	1972	1	2963	25	1980
5501	1	9988354	2	1	32	1972	1	2992	41	1980
5502	1	9988615	2	1	38	1975	0			
5503	1	9989304	1	1	47	1972	1	2993	55	1980
5504	1	9990078	1	1	49	1974	0			
5505	1	9995187	2	1	34	1972	1	2957	42	1980
5506	1	9995880	1	1	53	1973	1	2889	61	1981

att4 date4 age4 examyr4 att7 date7 age7 site7 examyr7

0	1	5778	52	1987	1	9787	63		1998
1	0				0				
2	0								
3	1	5783	47	1988	1	9596	57		1998
4	0								
5	1	5665	44	1990	1	9399	54	2	2001
6	1	6410	61	2014					
7	1	5685	57	1989	1	9705	68		2000
8	1	5948	54	1989	1	10254	66	2	2001
9	0								
10	1	5790	52	1988	1	9781	62		1999
11	0								
12	1	6550	64	2012					
13	0								
14	0								
15	1	5641	49	1990	1	9456	60	0	2000
16	0				0				
17	0				1	10490	53	0	2000
18	0								
19	1	5789	60	2012					
20	1	5655	56	2012					
21	1	5950	35	1989	0				
22	0								
23	1	5105	83	2012					
24	1	5636	53	1990	1	9300	63	0	2000
25	1	5782	45	1988	1	9842	57		1999
26	0								
27	0				0				
28	0								
29	0								
...
5477	1	5630	61	1990	1	9278	71	0	2000
5478	1	5645	63	1990	0				
5479	0				0				
5480	1	5726	49	1989	1	9373	59		1999
5481	1	5657	45	1990	1	9052	54		1999
5482	1	5795	51	1988	1	10252	63	0	2001
5483	1	5786	53	1988	1	10085	65		1999
5484	1	5629	42	1989	1	9307	52		1999
5485	1	5770	62	1988	1	9971	73		1999
5486	1	5646	57	1989	1	9496	67		1999
5487	1	5728	59	1989	0				
5488	1	6021	62	1989	1	10272	73	2	2001
5489	1	5848	63	1989	1	9458	73		1999
5490	1	6077	42	1988	1	10094	53		1999
5491	1	5342	50	1989	1	9223	61		1999
5492	1	5737	38	1989	0				
5493	1	5690	43	1989	1	9432	53		1999

5494	1	5800	43	1987	1	9907	54	1999
5495	1	5845	43	1987	1	10215	55	1999
5496	1	5635	43	1989	1	9236	53	1999
5497	1	5897	45	1988	1	10055	56	1999
5498	1	5719	51	1989	1	9556	61	1999
5499	1	5632	64	1989	1	9235	73	1999
5500	1	5811	33	1988	1	9946	44	1999
5501	1	5772	48	1988	1	10073	60	2000
5502	1	5368	53	1989	1	8903	63	1999
5503	1	5790	63	1988	0			
5504	0				0			
5505	1	6545	52	1990	0			
5506	0				0			

[5507 rows x 19 columns]>

```
In [33]: df_lex1_2 = pd.read_csv('lex1_2v2.csv')
df_lex1_2.head
```

```
Out[33]: <bound method NDFrame.head of
0      2  24.0  23.0  128.0  78.0  106.0  72.0  100.0  70.0  NaN  ...
1      1  12.0  13.0  148.0  94.0  154.0  102.0  156.0  96.0  NaN  ...
2      1  27.0  28.0  142.0  100.0  134.0  90.0  130.0  88.0  4.0  ...
3      2  30.0  29.0  148.0  88.0  142.0  84.0  134.0  78.0  3.0  ...
4      2  12.0  11.0  106.0  62.0  126.0  68.0  120.0  76.0  4.0  ...
5      2  30.0  31.0  138.0  100.0  136.0  110.0  134.0  74.0  12.0  ...
6      2   5.0   6.0  106.0  60.0  104.0  56.0  106.0  60.0  NaN  ...
7      2  15.0  13.0  116.0  72.0  124.0  74.0  122.0  72.0  8.0  ...
8      2  30.0  27.0  120.0  74.0  116.0  70.0  120.0  74.0  9.0  ...
9      1  17.0  16.0  124.0  88.0  124.0  86.0  124.0  86.0  7.0  ...
10     2  36.0  36.0  144.0  90.0  136.0  90.0  124.0  93.0  5.0  ...
11     2  23.0  21.0  104.0  74.0  106.0  70.0  100.0  68.0  8.0  ...
12     1  15.0  13.0  130.0  86.0  136.0  84.0  124.0  84.0  9.0  ...
13     1  21.0  20.0  148.0  100.0  150.0  90.0  120.0  96.0  4.0  ...
14     2  12.0  11.0  124.0  76.0  108.0  74.0  118.0  76.0  7.0  ...
15     1  10.0   9.0  112.0  82.0  120.0  78.0  120.0  80.0  6.0  ...
16     2   5.0   6.0  102.0  74.0   90.0  70.0   92.0  72.0  5.0  ...
17     1   6.0   7.0  130.0  96.0  108.0  82.0  114.0  80.0  36.0  ...
18     2  25.0  24.0  106.0  62.0  112.0  66.0  122.0  70.0  2.0  ...
19     1   8.0   7.0  116.0  70.0  122.0  70.0  124.0  70.0  19.0  ...
20     1   4.0   5.0  152.0  92.0  158.0  86.0  148.0  84.0  7.0  ...
21     2  18.0  17.0  110.0  70.0  122.0  82.0  114.0  80.0  3.0  ...
22     2  13.0  14.0  108.0  80.0  134.0  80.0  122.0  78.0  3.0  ...
23     1   9.0   8.0  106.0  76.0  112.0  70.0  114.0  70.0  11.0  ...
24     2  54.0  51.0  120.0  80.0  138.0  80.0  140.0  82.0  3.0  ...
25     2  16.0  17.0  128.0  82.0  112.0  68.0  118.0  76.0  5.0  ...
26     2  36.0  37.0  110.0  62.0  112.0  74.0   98.0  66.0  7.0  ...
27     1  26.0  30.0  120.0  76.0  118.0  76.0  122.0  76.0  14.0  ...
```

28	1	24.0	24.0	164.0	88.0	160.0	82.0	154.0	78.0	4.0	...
29	1	7.0	9.0	154.0	100.0	184.0	110.0	150.0	88.0	20.0	...
...
3745	1	20.0	20.0	120.0	82.0	136.0	86.0	124.0	86.0	12.0	...
3746	1	14.0	15.0	140.0	92.0	142.0	90.0	126.0	86.0	8.0	...
3747	2	19.0	18.0	110.0	78.0	108.0	68.0	106.0	70.0	3.0	...
3748	2	19.0	18.0	106.0	58.0	114.0	60.0	114.0	58.0	11.0	...
3749	2	49.0	46.0	92.0	74.0	100.0	70.0	100.0	70.0	4.0	...
3750	2	7.0	8.0	102.0	80.0	104.0	80.0	120.0	86.0	5.0	...
3751	1	7.0	7.0	148.0	94.0	138.0	92.0	140.0	92.0	3.0	...
3752	1	19.0	20.0	130.0	88.0	138.0	92.0	144.0	90.0	3.0	...
3753	1	14.0	15.0	156.0	92.0	142.0	80.0	152.0	86.0	12.0	...
3754	1	9.0	8.0	126.0	80.0	120.0	80.0	124.0	76.0	4.0	...
3755	2	18.0	20.0	104.0	70.0	104.0	70.0	110.0	70.0	21.0	...
3756	2	28.0	24.0	110.0	78.0	104.0	68.0	102.0	76.0	8.0	...
3757	2	27.0	27.0	92.0	70.0	104.0	70.0	104.0	66.0	4.0	...
3758	2	13.0	16.0	124.0	78.0	112.0	72.0	126.0	72.0	38.0	...
3759	2	49.0	50.0	142.0	100.0	156.0	100.0	152.0	96.0	8.0	...
3760	1	17.0	16.0	142.0	90.0	148.0	90.0	170.0	94.0	3.0	...
3761	2	14.0	12.0	112.0	74.0	110.0	78.0	114.0	76.0	3.0	...
3762	1	4.0	4.0	100.0	74.0	114.0	68.0	104.0	68.0	6.0	...
3763	2	15.0	15.0	102.0	72.0	96.0	74.0	100.0	70.0	12.0	...
3764	1	4.0	3.0	140.0	102.0	150.0	98.0	148.0	100.0	26.0	...
3765	2	19.0	18.0	98.0	70.0	114.0	86.0	106.0	80.0	NaN	...
3766	2	14.0	13.0	120.0	70.0	114.0	68.0	110.0	62.0	23.0	...
3767	2	34.0	32.0	114.0	74.0	114.0	80.0	116.0	80.0	3.0	...
3768	1	23.0	24.0	166.0	94.0	168.0	88.0	152.0	96.0	13.0	...
3769	1	9.0	10.0	134.0	84.0	130.0	84.0	124.0	88.0	6.0	...
3770	1	4.0	4.0	96.0	64.0	110.0	64.0	100.0	66.0	5.0	...
3771	2	9.0	8.0	94.0	60.0	104.0	64.0	98.0	60.0	3.0	...
3772	1	7.0	5.0	122.0	90.0	120.0	82.0	130.0	84.0	4.0	...
3773	2	12.0	11.0	128.0	82.0	132.0	80.0	136.0	82.0	4.0	...
3774	1	5.0	5.0	128.0	88.0	130.0	80.0	130.0	78.0	6.0	...

	B731	B732	B733	B734	AGE2	WGTGP2	YRMAGP2	pid	educ_2	IDTYPE
0	NaN	NaN	NaN	NaN	44	7	4.0	2924	2.0	1
1	NaN	NaN	NaN	NaN	39	18	11.0	3297	2.0	1
2	NaN	NaN	NaN	NaN	39	24	3.0	4061	4.0	1
3	NaN	NaN	NaN	NaN	47	9	6.0	8059	3.0	1
4	NaN	NaN	NaN	NaN	44	0	5.0	9458	2.0	1
5	NaN	NaN	NaN	NaN	41	9	5.0	14463	2.0	1
6	NaN	NaN	NaN	NaN	32	1	3.0	15059	4.0	1
7	NaN	NaN	NaN	NaN	27	2	1.0	18844	2.0	1
8	NaN	NaN	NaN	NaN	44	10	5.0	20206	3.0	1
9	NaN	NaN	NaN	NaN	38	25	2.0	22267	2.0	1
10	NaN	NaN	NaN	NaN	48	12	6.0	27195	3.0	1
11	NaN	NaN	NaN	NaN	34	5	3.0	27363	3.0	1
12	NaN	NaN	NaN	NaN	45	14	4.0	32688	3.0	1

13	NaN	NaN	NaN	NaN	41	14	2.0	33693	4.0	1
14	NaN	NaN	NaN	NaN	41	6	3.0	35867	3.0	1
15	NaN	NaN	NaN	NaN	40	15	5.0	38579	2.0	1
16	NaN	NaN	NaN	NaN	35	2	11.0	43540	2.0	1
17	NaN	NaN	NaN	NaN	47	11	6.0	46959	2.0	1
18	NaN	NaN	NaN	NaN	31	5	2.0	47291	4.0	1
19	NaN	NaN	NaN	NaN	27	15	2.0	49050	2.0	1
20	NaN	NaN	NaN	NaN	64	12	6.0	50122	2.0	1
21	NaN	NaN	NaN	NaN	45	5	6.0	52532	3.0	1
22	NaN	NaN	NaN	NaN	43	3	5.0	55025	3.0	1
23	NaN	NaN	NaN	NaN	21	9	11.0	55986	3.0	1
24	1.0	NaN	NaN	NaN	55	20	7.0	58198	2.0	1
25	NaN	NaN	NaN	NaN	27	5	11.0	65987	3.0	1
26	NaN	NaN	NaN	NaN	40	10	3.0	73470	1.0	1
27	NaN	NaN	NaN	NaN	42	23	3.0	78430	4.0	1
28	NaN	NaN	NaN	NaN	57	17	1.0	80711	3.0	1
29	NaN	NaN	NaN	NaN	56	13	7.0	83839	2.0	1
...
3745	NaN	NaN	NaN	NaN	59	15	7.0	9925701	3.0	1
3746	NaN	NaN	NaN	NaN	35	25	3.0	9926341	3.0	1
3747	NaN	NaN	NaN	NaN	41	2	11.0	9928033	2.0	1
3748	NaN	NaN	NaN	NaN	54	4	7.0	9930609	1.0	1
3749	NaN	NaN	NaN	NaN	51	11	11.0	9934124	1.0	1
3750	NaN	NaN	NaN	NaN	39	8	4.0	9936752	3.0	1
3751	NaN	NaN	NaN	NaN	54	17	5.0	9937420	4.0	1
3752	4.0	NaN	NaN	NaN	52	20	5.0	9939621	4.0	1
3753	NaN	NaN	NaN	NaN	55	14	7.0	9939888	2.0	1
3754	NaN	NaN	NaN	NaN	41	20	4.0	9940870	4.0	1
3755	NaN	NaN	NaN	NaN	37	8	4.0	9941870	2.0	1
3756	NaN	NaN	NaN	NaN	46	8	11.0	9945338	4.0	1
3757	NaN	NaN	NaN	NaN	34	4	3.0	9947765	2.0	1
3758	NaN	NaN	NaN	NaN	54	7	7.0	9947948	3.0	1
3759	NaN	NaN	NaN	NaN	49	22	4.0	9948725	2.0	1
3760	NaN	NaN	NaN	NaN	51	15	4.0	9951579	3.0	1
3761	NaN	NaN	NaN	NaN	54	1	7.0	9958150	1.0	1
3762	NaN	NaN	NaN	NaN	34	12	2.0	9965528	4.0	1
3763	NaN	NaN	NaN	NaN	35	9	3.0	9975906	3.0	1
3764	NaN	NaN	NaN	NaN	35	7	3.0	9977699	4.0	1
3765	NaN	NaN	NaN	NaN	35	7	NaN	9980291	3.0	1
3766	NaN	NaN	NaN	NaN	36	7	3.0	9982799	2.0	1
3767	NaN	NaN	NaN	NaN	38	6	3.0	9982997	3.0	1
3768	NaN	NaN	NaN	NaN	43	15	4.0	9985882	3.0	1
3769	NaN	NaN	NaN	NaN	56	22	7.0	9986747	2.0	1
3770	NaN	NaN	NaN	NaN	25	15	11.0	9987740	4.0	1
3771	NaN	NaN	NaN	NaN	41	5	11.0	9988354	2.0	1
3772	NaN	NaN	NaN	NaN	55	19	7.0	9989304	2.0	1
3773	NaN	NaN	NaN	NaN	42	7	4.0	9995187	2.0	1
3774	NaN	NaN	NaN	NaN	61	3	6.0	9995880	1.0	1

```
[3775 rows x 701 columns]>
```

```
In [34]: df_bmi1=pd.read_csv('BMIex2.csv')
```

```
In [35]: df_bmi1.rename(columns={'pid':'PID'}, inplace=True)
```

```
In [36]: df_bmi1.head
```

```
Out[36]: <bound method NDFrame.head of
```

			EXAM	BMI	PID	IDTYPE
0	2	22.648360726409	2924	1		
1	2	26.476583565412	3297	1		
2	2	31.644675765722	4061	1		
3	2	23.8403797120094	8059	1		
4	2	18.033647098607	9458	1		
5	2	26.1290845162335	14463	1		
6	2	16.8390869326117	15059	1		
7	2	21.8407503749796	18844	1		
8	2	24.5982740486191	20206	1		
9	2	30.7171810534853	22267	1		
10	2	24.7054403665501	27195	1		
11	2	19.9757795965296	27363	1		
12	2	26.1038615193737	32688	1		
13	2	26.9008802480415	33693	1		
14	2	20.6560330476033	35867	1		
15	2	26.2468085489804	38579	1		
16	2	19.415361411368	43540	1		
17	2	23.9285397260926	46959	1		
18	2	20.6048996082367	47291	1		
19	2	29.2896093604687	49050	1		
20	2	21.1241741927928	50122	1		
21	2	21.7914268028083	52532	1		
22	2	20.8069483593952	55025	1		
23	2	21.908598060903	55986	1		
24	2	39.0729020469819	58198	1		
25	2	18.9563756271643	65987	1		
26	2	30.894086758447	73470	1		
27	2	31.9788701861141	78430	1		
28	2	30.8222051990979	80711	1		
29	2	22.0660163542549	83839	1		
...		
3745	2	25.9451210943944	9925701	1		
3746	2	31.3163891633906	9926341	1		
3747	2	18.5614384900645	9928033	1		
3748	2	21.5567483002744	9930609	1		
3749	2	28.3049477681863	9934124	1		
3750	2	21.9370273451291	9936752	1		
3751	2	24.9058224757989	9937420	1		
3752	2	33.2062431702988	9939621	1		

3753	2	24.7590287241066	9939888	1
3754	2	25.2201015724164	9940870	1
3755	2	20.3639031845538	9941870	1
3756	2	22.9889375794377	9945338	1
3757	2	22.1190982559337	9947765	1
3758	2	22.1374954468659	9947948	1
3759	2	37.3791070162786	9948725	1
3760	2	23.007858515717	9951579	1
3761	2	21.6828688505022	9958150	1
3762	2	23.6824651849996	9965528	1
3763	2	21.908598060903	9975906	1
3764	2	19.9113720026748	9977699	1
3765	2	22.648360726409	9980291	1
3766	2	21.4615813847012	9982799	1
3767	2	24.1810480933512	9982997	1
3768	2	30.5782924491019	9985882	1
3769	2	28.6416958471687	9986747	1
3770	2	22.1631195636387	9987740	1
3771	2	21.7914268028083	9988354	1
3772	2	32.6953778907558	9989304	1
3773	2	21.6266704722759	9995187	1
3774	2	18.7317269165788	9995880	1

[3775 rows x 4 columns]>

```
In [37]: df_offspring_bmi = pd.merge(df_offspring, df_bmi1, on='PID', how='outer')
df_offspring_bmi.head
```

```
Out[37]: <bound method NDFrame.head of          idtype      PID  sex  att1  age1  examyr1  att2 dat
0           1    2924    2     1    36   1972     1  2949   44   1980
1           1    3297    1     1    31   1971     1  2947   39   1979
2           7    4040    2     1    43   1997     0
3           1    4061    1     1    31   1972     1  2984   39   1980
4           7    4301    2     1    51   1996     0
5           1    5350    2     1    28   1975     0
6           7    6719    2     1    44   1996     1  1870   49   2001
7           1    7248    2     1    41   1973     0
8           1    8059    2     1    38   1973     1  3405   47   1982
9           7    8459    2     1    51   1996     0
10          1    9458    2     1    36   1972     1  2953   44   1980
11          7    9912    2     1    59   1996     1  1199   62   1999
12          7   10791    1     1    46   1994     1  1825   51   1999
13          7   11016    2     1    59   1994     1  1739   63   1999
14          7   11651    2     1    50   1996     1  1355   54   1999
15          1   14463    2     1    34   1974     1  2676   41   1982
16          1   14764    2     1    44   1972     0
17          1   15059    2     1    24   1971     1  2971   32   1980
18          7   15398    2     1    53   1995     0
```

19	7	18098	1	1	44	1996	1	1092	47	1999
20	7	18103	2	1	41	1997	1	1015	44	1999
21	1	18844	2	1	19	1973	1	2942	27	1981
22	7	19353	2	1	52	1995	0			
23	7	19892	1	1	69	1998	1	665	70	2000
24	1	20206	2	1	37	1975	1	2648	44	1982
25	1	22267	1	1	30	1972	1	2926	38	1980
26	7	22567	1	1	64	1998	1	735	66	2000
27	1	22996	2	1	61	1973	0			
28	7	24738	1	1	57	1996	1	1605	61	2000
29	7	25240	2	1	53	1997	0			
...
5477	1	9939621	1	1	45	1975	1	2569	52	1982
5478	1	9939888	1	1	47	1974	1	2679	55	1982
5479	1	9940525	2	1	58	1973	0			
5480	1	9940870	1	1	33	1973	1	2885	41	1981
5481	1	9941870	2	1	29	1974	1	2695	37	1982
5482	1	9943257	1	1	35	1972	0			
5483	1	9945338	2	1	38	1972	1	3014	46	1980
5484	1	9947765	2	1	26	1974	1	2815	34	1981
5485	1	9947948	2	1	46	1972	1	2948	54	1980
5486	1	9948725	2	1	41	1973	1	2803	49	1981
5487	1	9951579	1	1	43	1974	1	2802	51	1981
5488	1	9957447	1	1	45	1972	0			
5489	1	9958150	2	1	47	1973	1	2590	54	1980
5490	1	9965528	1	1	25	1971	1	3110	34	1980
5491	1	9967264	2	1	35	1974	0			
5492	1	9973742	1	1	22	1973	0			
5493	1	9975906	2	1	28	1973	1	2800	35	1981
5494	1	9977699	1	1	27	1972	1	2962	35	1980
5495	1	9980291	2	1	27	1971	1	2962	35	1979
5496	1	9982799	2	1	28	1973	1	2775	36	1981
5497	1	9982997	2	1	28	1972	1	3452	38	1981
5498	1	9985882	1	1	35	1973	1	2897	43	1981
5499	1	9986747	1	1	48	1973	1	2773	56	1981
5500	1	9987740	1	1	17	1972	1	2963	25	1980
5501	1	9988354	2	1	32	1972	1	2992	41	1980
5502	1	9988615	2	1	38	1975	0			
5503	1	9989304	1	1	47	1972	1	2993	55	1980
5504	1	9990078	1	1	49	1974	0			
5505	1	9995187	2	1	34	1972	1	2957	42	1980
5506	1	9995880	1	1	53	1973	1	2889	61	1981

	...	age4	examyr4	att7	date7	age7	site7	examyr7	EXAM	\
0	...	52	1987	1	9787	63		1998	2.0	
1	...			0					2.0	
2	...								NaN	
3	...	47	1988	1	9596	57		1998	2.0	

4	...								NaN
5	...	44	1990	1	9399	54	2	2001	NaN
6	...	61	2014						NaN
7	...	57	1989	1	9705	68		2000	NaN
8	...	54	1989	1	10254	66	2	2001	2.0
9	...								NaN
10	...	52	1988	1	9781	62		1999	2.0
11	...								NaN
12	...	64	2012						NaN
13	...								NaN
14	...								NaN
15	...	49	1990	1	9456	60	0	2000	2.0
16	...			0					NaN
17	...			1	10490	53	0	2000	2.0
18	...								NaN
19	...	60	2012						NaN
20	...	56	2012						NaN
21	...	35	1989	0					2.0
22	...								NaN
23	...	83	2012						NaN
24	...	53	1990	1	9300	63	0	2000	2.0
25	...	45	1988	1	9842	57		1999	2.0
26	...								NaN
27	...			0					NaN
28	...								NaN
29	...								NaN
...
5477	...	61	1990	1	9278	71	0	2000	2.0
5478	...	63	1990	0					2.0
5479	...			0					NaN
5480	...	49	1989	1	9373	59		1999	2.0
5481	...	45	1990	1	9052	54		1999	2.0
5482	...	51	1988	1	10252	63	0	2001	NaN
5483	...	53	1988	1	10085	65		1999	2.0
5484	...	42	1989	1	9307	52		1999	2.0
5485	...	62	1988	1	9971	73		1999	2.0
5486	...	57	1989	1	9496	67		1999	2.0
5487	...	59	1989	0					2.0
5488	...	62	1989	1	10272	73	2	2001	NaN
5489	...	63	1989	1	9458	73		1999	2.0
5490	...	42	1988	1	10094	53		1999	2.0
5491	...	50	1989	1	9223	61		1999	NaN
5492	...	38	1989	0					NaN
5493	...	43	1989	1	9432	53		1999	2.0
5494	...	43	1987	1	9907	54		1999	2.0
5495	...	43	1987	1	10215	55		1999	2.0
5496	...	43	1989	1	9236	53		1999	2.0
5497	...	45	1988	1	10055	56		1999	2.0

5498	...	51	1989	1	9556	61	1999	2.0
5499	...	64	1989	1	9235	73	1999	2.0
5500	...	33	1988	1	9946	44	1999	2.0
5501	...	48	1988	1	10073	60	2000	2.0
5502	...	53	1989	1	8903	63	1999	NaN
5503	...	63	1988	0				2.0
5504	...			0				NaN
5505	...	52	1990	0				2.0
5506	...			0				2.0

	BMI	IDTYPE
0	22.648360726409	1.0
1	26.476583565412	1.0
2	NaN	NaN
3	31.644675765722	1.0
4	NaN	NaN
5	NaN	NaN
6	NaN	NaN
7	NaN	NaN
8	23.8403797120094	1.0
9	NaN	NaN
10	18.033647098607	1.0
11	NaN	NaN
12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	26.1290845162335	1.0
16	NaN	NaN
17	16.8390869326117	1.0
18	NaN	NaN
19	NaN	NaN
20	NaN	NaN
21	21.8407503749796	1.0
22	NaN	NaN
23	NaN	NaN
24	24.5982740486191	1.0
25	30.7171810534853	1.0
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...
5477	33.2062431702988	1.0
5478	24.7590287241066	1.0
5479	NaN	NaN
5480	25.2201015724164	1.0
5481	20.3639031845538	1.0
5482	NaN	NaN

5483	22.9889375794377	1.0
5484	22.1190982559337	1.0
5485	22.1374954468659	1.0
5486	37.3791070162786	1.0
5487	23.007858515717	1.0
5488	NaN	NaN
5489	21.6828688505022	1.0
5490	23.6824651849996	1.0
5491	NaN	NaN
5492	NaN	NaN
5493	21.908598060903	1.0
5494	19.9113720026748	1.0
5495	22.648360726409	1.0
5496	21.4615813847012	1.0
5497	24.1810480933512	1.0
5498	30.5782924491019	1.0
5499	28.6416958471687	1.0
5500	22.1631195636387	1.0
5501	21.7914268028083	1.0
5502	NaN	NaN
5503	32.6953778907558	1.0
5504	NaN	NaN
5505	21.6266704722759	1.0
5506	18.7317269165788	1.0

[5507 rows x 22 columns]>

In [38]: df_offspring_bmi.drop('IDTYPE', axis=1, inplace=True)

In [39]: df_offspring_bmi.drop('EXAM', axis=1, inplace=True)

In []:

In [40]: df_offspring_bmi.head

Out[40]: <bound method NDFrame.head of

					idtype	PID	sex	att1	age1	examy1	att2	dat
0	1	2924	2	1	36	1972	1	2949	44	1980		
1	1	3297	1	1	31	1971	1	2947	39	1979		
2	7	4040	2	1	43	1997	0					
3	1	4061	1	1	31	1972	1	2984	39	1980		
4	7	4301	2	1	51	1996	0					
5	1	5350	2	1	28	1975	0					
6	7	6719	2	1	44	1996	1	1870	49	2001		
7	1	7248	2	1	41	1973	0					
8	1	8059	2	1	38	1973	1	3405	47	1982		
9	7	8459	2	1	51	1996	0					
10	1	9458	2	1	36	1972	1	2953	44	1980		
11	7	9912	2	1	59	1996	1	1199	62	1999		
12	7	10791	1	1	46	1994	1	1825	51	1999		

13	7	11016	2	1	59	1994	1	1739	63	1999
14	7	11651	2	1	50	1996	1	1355	54	1999
15	1	14463	2	1	34	1974	1	2676	41	1982
16	1	14764	2	1	44	1972	0			
17	1	15059	2	1	24	1971	1	2971	32	1980
18	7	15398	2	1	53	1995	0			
19	7	18098	1	1	44	1996	1	1092	47	1999
20	7	18103	2	1	41	1997	1	1015	44	1999
21	1	18844	2	1	19	1973	1	2942	27	1981
22	7	19353	2	1	52	1995	0			
23	7	19892	1	1	69	1998	1	665	70	2000
24	1	20206	2	1	37	1975	1	2648	44	1982
25	1	22267	1	1	30	1972	1	2926	38	1980
26	7	22567	1	1	64	1998	1	735	66	2000
27	1	22996	2	1	61	1973	0			
28	7	24738	1	1	57	1996	1	1605	61	2000
29	7	25240	2	1	53	1997	0			
...
5477	1	9939621	1	1	45	1975	1	2569	52	1982
5478	1	9939888	1	1	47	1974	1	2679	55	1982
5479	1	9940525	2	1	58	1973	0			
5480	1	9940870	1	1	33	1973	1	2885	41	1981
5481	1	9941870	2	1	29	1974	1	2695	37	1982
5482	1	9943257	1	1	35	1972	0			
5483	1	9945338	2	1	38	1972	1	3014	46	1980
5484	1	9947765	2	1	26	1974	1	2815	34	1981
5485	1	9947948	2	1	46	1972	1	2948	54	1980
5486	1	9948725	2	1	41	1973	1	2803	49	1981
5487	1	9951579	1	1	43	1974	1	2802	51	1981
5488	1	9957447	1	1	45	1972	0			
5489	1	9958150	2	1	47	1973	1	2590	54	1980
5490	1	9965528	1	1	25	1971	1	3110	34	1980
5491	1	9967264	2	1	35	1974	0			
5492	1	9973742	1	1	22	1973	0			
5493	1	9975906	2	1	28	1973	1	2800	35	1981
5494	1	9977699	1	1	27	1972	1	2962	35	1980
5495	1	9980291	2	1	27	1971	1	2962	35	1979
5496	1	9982799	2	1	28	1973	1	2775	36	1981
5497	1	9982997	2	1	28	1972	1	3452	38	1981
5498	1	9985882	1	1	35	1973	1	2897	43	1981
5499	1	9986747	1	1	48	1973	1	2773	56	1981
5500	1	9987740	1	1	17	1972	1	2963	25	1980
5501	1	9988354	2	1	32	1972	1	2992	41	1980
5502	1	9988615	2	1	38	1975	0			
5503	1	9989304	1	1	47	1972	1	2993	55	1980
5504	1	9990078	1	1	49	1974	0			
5505	1	9995187	2	1	34	1972	1	2957	42	1980
5506	1	9995880	1	1	53	1973	1	2889	61	1981

	att4	date4	age4	examyr4	att7	date7	age7	site7	examyr7	BMI
0	1	5778	52	1987	1	9787	63		1998	22.648360726409
1	0				0					26.476583565412
2	0									NaN
3	1	5783	47	1988	1	9596	57		1998	31.644675765722
4	0									NaN
5	1	5665	44	1990	1	9399	54	2	2001	NaN
6	1	6410	61	2014						NaN
7	1	5685	57	1989	1	9705	68		2000	NaN
8	1	5948	54	1989	1	10254	66	2	2001	23.8403797120094
9	0									NaN
10	1	5790	52	1988	1	9781	62		1999	18.033647098607
11	0									NaN
12	1	6550	64	2012						NaN
13	0									NaN
14	0									NaN
15	1	5641	49	1990	1	9456	60	0	2000	26.1290845162335
16	0				0					NaN
17	0				1	10490	53	0	2000	16.8390869326117
18	0									NaN
19	1	5789	60	2012						NaN
20	1	5655	56	2012						NaN
21	1	5950	35	1989	0					21.8407503749796
22	0									NaN
23	1	5105	83	2012						NaN
24	1	5636	53	1990	1	9300	63	0	2000	24.5982740486191
25	1	5782	45	1988	1	9842	57		1999	30.7171810534853
26	0									NaN
27	0				0					NaN
28	0									NaN
29	0									NaN
...
5477	1	5630	61	1990	1	9278	71	0	2000	33.2062431702988
5478	1	5645	63	1990	0					24.7590287241066
5479	0				0					NaN
5480	1	5726	49	1989	1	9373	59		1999	25.2201015724164
5481	1	5657	45	1990	1	9052	54		1999	20.3639031845538
5482	1	5795	51	1988	1	10252	63	0	2001	NaN
5483	1	5786	53	1988	1	10085	65		1999	22.9889375794377
5484	1	5629	42	1989	1	9307	52		1999	22.1190982559337
5485	1	5770	62	1988	1	9971	73		1999	22.1374954468659
5486	1	5646	57	1989	1	9496	67		1999	37.3791070162786
5487	1	5728	59	1989	0					23.007858515717
5488	1	6021	62	1989	1	10272	73	2	2001	NaN
5489	1	5848	63	1989	1	9458	73		1999	21.6828688505022
5490	1	6077	42	1988	1	10094	53		1999	23.6824651849996
5491	1	5342	50	1989	1	9223	61		1999	NaN

5492	1	5737	38	1989	0					NaN
5493	1	5690	43	1989	1	9432	53	1999	21.908598060903	
5494	1	5800	43	1987	1	9907	54	1999	19.9113720026748	
5495	1	5845	43	1987	1	10215	55	1999	22.648360726409	
5496	1	5635	43	1989	1	9236	53	1999	21.4615813847012	
5497	1	5897	45	1988	1	10055	56	1999	24.1810480933512	
5498	1	5719	51	1989	1	9556	61	1999	30.5782924491019	
5499	1	5632	64	1989	1	9235	73	1999	28.6416958471687	
5500	1	5811	33	1988	1	9946	44	1999	22.1631195636387	
5501	1	5772	48	1988	1	10073	60	2000	21.7914268028083	
5502	1	5368	53	1989	1	8903	63	1999	NaN	
5503	1	5790	63	1988	0				32.6953778907558	
5504	0				0				NaN	
5505	1	6545	52	1990	0				21.6266704722759	
5506	0				0				18.7317269165788	

[5507 rows x 20 columns]>

```
In [41]: df_offspring_bmi.rename(columns={'BMI':'BMI_1980'}, inplace=True)
df_offspring_bmi.head
```

```
Out[41]: <bound method NDFrame.head of          idtype      PID  sex  att1  age1  examyr1  att2 dat
0          1    2924    2    1    36    1972    1  2949   44   1980
1          1    3297    1    1    31    1971    1  2947   39   1979
2          7    4040    2    1    43    1997    0
3          1    4061    1    1    31    1972    1  2984   39   1980
4          7    4301    2    1    51    1996    0
5          1    5350    2    1    28    1975    0
6          7    6719    2    1    44    1996    1  1870   49   2001
7          1    7248    2    1    41    1973    0
8          1    8059    2    1    38    1973    1  3405   47   1982
9          7    8459    2    1    51    1996    0
10         1    9458    2    1    36    1972    1  2953   44   1980
11         7    9912    2    1    59    1996    1  1199   62   1999
12         7   10791    1    1    46    1994    1  1825   51   1999
13         7   11016    2    1    59    1994    1  1739   63   1999
14         7   11651    2    1    50    1996    1  1355   54   1999
15         1   14463    2    1    34    1974    1  2676   41   1982
16         1   14764    2    1    44    1972    0
17         1   15059    2    1    24    1971    1  2971   32   1980
18         7   15398    2    1    53    1995    0
19         7   18098    1    1    44    1996    1  1092   47   1999
20         7   18103    2    1    41    1997    1  1015   44   1999
21         1   18844    2    1    19    1973    1  2942   27   1981
22         7   19353    2    1    52    1995    0
23         7   19892    1    1    69    1998    1   665   70   2000
24         1   20206    2    1    37    1975    1  2648   44   1982
25         1   22267    1    1    30    1972    1  2926   38   1980
```


26	7	22567	1	1	64	1998	1	735	66	2000
27	1	22996	2	1	61	1973	0			
28	7	24738	1	1	57	1996	1	1605	61	2000
29	7	25240	2	1	53	1997	0			
...
5477	1	9939621	1	1	45	1975	1	2569	52	1982
5478	1	9939888	1	1	47	1974	1	2679	55	1982
5479	1	9940525	2	1	58	1973	0			
5480	1	9940870	1	1	33	1973	1	2885	41	1981
5481	1	9941870	2	1	29	1974	1	2695	37	1982
5482	1	9943257	1	1	35	1972	0			
5483	1	9945338	2	1	38	1972	1	3014	46	1980
5484	1	9947765	2	1	26	1974	1	2815	34	1981
5485	1	9947948	2	1	46	1972	1	2948	54	1980
5486	1	9948725	2	1	41	1973	1	2803	49	1981
5487	1	9951579	1	1	43	1974	1	2802	51	1981
5488	1	9957447	1	1	45	1972	0			
5489	1	9958150	2	1	47	1973	1	2590	54	1980
5490	1	9965528	1	1	25	1971	1	3110	34	1980
5491	1	9967264	2	1	35	1974	0			
5492	1	9973742	1	1	22	1973	0			
5493	1	9975906	2	1	28	1973	1	2800	35	1981
5494	1	9977699	1	1	27	1972	1	2962	35	1980
5495	1	9980291	2	1	27	1971	1	2962	35	1979
5496	1	9982799	2	1	28	1973	1	2775	36	1981
5497	1	9982997	2	1	28	1972	1	3452	38	1981
5498	1	9985882	1	1	35	1973	1	2897	43	1981
5499	1	9986747	1	1	48	1973	1	2773	56	1981
5500	1	9987740	1	1	17	1972	1	2963	25	1980
5501	1	9988354	2	1	32	1972	1	2992	41	1980
5502	1	9988615	2	1	38	1975	0			
5503	1	9989304	1	1	47	1972	1	2993	55	1980
5504	1	9990078	1	1	49	1974	0			
5505	1	9995187	2	1	34	1972	1	2957	42	1980
5506	1	9995880	1	1	53	1973	1	2889	61	1981

	att4	date4	age4	examyr4	att7	date7	age7	site7	examyr7	BMI_1980
0	1	5778	52	1987	1	9787	63		1998	22.648360726409
1	0				0					26.476583565412
2	0									NaN
3	1	5783	47	1988	1	9596	57		1998	31.644675765722
4	0									NaN
5	1	5665	44	1990	1	9399	54	2	2001	NaN
6	1	6410	61	2014						NaN
7	1	5685	57	1989	1	9705	68		2000	NaN
8	1	5948	54	1989	1	10254	66	2	2001	23.8403797120094
9	0									NaN
10	1	5790	52	1988	1	9781	62		1999	18.033647098607

11	0											NaN
12	1	6550	64	2012								NaN
13	0											NaN
14	0											NaN
15	1	5641	49	1990	1	9456	60	0	2000	26.1290845162335		
16	0				0							NaN
17	0				1	10490	53	0	2000	16.8390869326117		
18	0											NaN
19	1	5789	60	2012								NaN
20	1	5655	56	2012								NaN
21	1	5950	35	1989	0					21.8407503749796		
22	0											NaN
23	1	5105	83	2012								NaN
24	1	5636	53	1990	1	9300	63	0	2000	24.5982740486191		
25	1	5782	45	1988	1	9842	57		1999	30.7171810534853		
26	0											NaN
27	0				0							NaN
28	0											NaN
29	0											NaN
...
5477	1	5630	61	1990	1	9278	71	0	2000	33.2062431702988		
5478	1	5645	63	1990	0					24.7590287241066		
5479	0				0							NaN
5480	1	5726	49	1989	1	9373	59		1999	25.2201015724164		
5481	1	5657	45	1990	1	9052	54		1999	20.3639031845538		
5482	1	5795	51	1988	1	10252	63	0	2001			NaN
5483	1	5786	53	1988	1	10085	65		1999	22.9889375794377		
5484	1	5629	42	1989	1	9307	52		1999	22.1190982559337		
5485	1	5770	62	1988	1	9971	73		1999	22.1374954468659		
5486	1	5646	57	1989	1	9496	67		1999	37.3791070162786		
5487	1	5728	59	1989	0					23.007858515717		
5488	1	6021	62	1989	1	10272	73	2	2001			NaN
5489	1	5848	63	1989	1	9458	73		1999	21.6828688505022		
5490	1	6077	42	1988	1	10094	53		1999	23.6824651849996		
5491	1	5342	50	1989	1	9223	61		1999			NaN
5492	1	5737	38	1989	0							NaN
5493	1	5690	43	1989	1	9432	53		1999	21.908598060903		
5494	1	5800	43	1987	1	9907	54		1999	19.9113720026748		
5495	1	5845	43	1987	1	10215	55		1999	22.648360726409		
5496	1	5635	43	1989	1	9236	53		1999	21.4615813847012		
5497	1	5897	45	1988	1	10055	56		1999	24.1810480933512		
5498	1	5719	51	1989	1	9556	61		1999	30.5782924491019		
5499	1	5632	64	1989	1	9235	73		1999	28.6416958471687		
5500	1	5811	33	1988	1	9946	44		1999	22.1631195636387		
5501	1	5772	48	1988	1	10073	60		2000	21.7914268028083		
5502	1	5368	53	1989	1	8903	63		1999			NaN
5503	1	5790	63	1988	0					32.6953778907558		
5504	0				0							NaN

5505	1	6545	52	1990	0	21.6266704722759
5506	0				0	18.7317269165788

[5507 rows x 20 columns]>

BMI for 1980 added. now add all the 1980 variables.

```
In [42]: df_lex1_2_short = df_lex1_2[['pid', 'B87', 'B117', 'B118', 'B119', 'B81', 'B54',]]
```

```
In [43]: df_lex1_2_short
```

```
Out[43]:
```

	pid	B87	B117	B118	B119	B81	B54
0	2924	0.0	0.0	1.0	0.0	0.0	3
1	3297	1.0	36.0	0.0	0.0	8.0	3
2	4061	1.0	18.0	0.0	0.0	8.0	4
3	8059	2.0	0.0	0.0	0.0	0.0	3
4	9458	0.0	0.0	1.0	2.0	0.0	3
5	14463	3.0	0.0	0.0	7.0	0.0	3
6	15059	0.0	0.0	1.0	0.0	0.0	3
7	18844	1.0	1.0	1.0	0.0	0.0	3
8	20206	3.0	1.0	0.0	4.0	0.0	3
9	22267	3.0	0.0	0.0	0.0	8.0	3
10	27195	0.0	0.0	2.0	2.0	0.0	3
11	27363	3.0	0.0	1.0	0.0	0.0	3
12	32688	3.0	10.0	5.0	1.0	8.0	3
13	33693	3.0	1.0	0.0	0.0	8.0	3
14	35867	0.0	0.0	0.0	0.0	0.0	3
15	38579	0.0	0.0	2.0	0.0	8.0	3
16	43540	1.0	0.0	0.0	0.0	0.0	3
17	46959	3.0	3.0	0.0	1.0	8.0	3
18	47291	1.0	0.0	0.0	0.0	0.0	3
19	49050	3.0	6.0	0.0	28.0	8.0	3
20	50122	1.0	0.0	28.0	0.0	8.0	3
21	52532	1.0	0.0	0.0	0.0	0.0	3
22	55025	1.0	0.0	1.0	1.0	0.0	3
23	55986	3.0	14.0	0.0	0.0	8.0	3
24	58198	0.0	0.0	1.0	2.0	0.0	3
25	65987	3.0	0.0	1.0	1.0	0.0	3
26	73470	3.0	0.0	4.0	0.0	0.0	3
27	78430	3.0	6.0	0.0	0.0	8.0	3
28	80711	3.0	0.0	0.0	1.0	8.0	3
29	83839	3.0	1.0	1.0	1.0	8.0	3
...
3745	9925701	3.0	1.0	0.0	0.0	8.0	3
3746	9926341	1.0	3.0	1.0	1.0	8.0	3
3747	9928033	0.0	0.0	1.0	0.0	0.0	3
3748	9930609	3.0	0.0	0.0	0.0	0.0	3
3749	9934124	0.0	0.0	0.0	0.0	0.0	3
3750	9936752	2.0	0.0	2.0	4.0	0.0	3

3751	9937420	1.0	2.0	0.0	2.0	8.0	3
3752	9939621	3.0	12.0	1.0	3.0	8.0	3
3753	9939888	3.0	24.0	0.0	32.0	8.0	3
3754	9940870	0.0	0.0	7.0	7.0	8.0	3
3755	9941870	3.0	0.0	28.0	0.0	0.0	3
3756	9945338	3.0	1.0	0.0	1.0	0.0	3
3757	9947765	0.0	0.0	2.0	0.0	0.0	3
3758	9947948	3.0	0.0	0.0	0.0	0.0	3
3759	9948725	0.0	0.0	14.0	3.0	0.0	3
3760	9951579	1.0	0.0	0.0	0.0	8.0	3
3761	9958150	0.0	0.0	0.0	0.0	2.0	3
3762	9965528	0.0	1.0	1.0	2.0	8.0	3
3763	9975906	3.0	0.0	0.0	0.0	0.0	3
3764	9977699	3.0	14.0	21.0	0.0	8.0	3
3765	9980291	3.0	0.0	5.0	2.0	0.0	3
3766	9982799	3.0	0.0	1.0	1.0	0.0	3
3767	9982997	0.0	0.0	1.0	0.0	0.0	3
3768	9985882	3.0	0.0	0.0	2.0	8.0	3
3769	9986747	1.0	0.0	0.0	7.0	8.0	3
3770	9987740	0.0	7.0	1.0	0.0	8.0	3
3771	9988354	0.0	0.0	2.0	2.0	0.0	3
3772	9989304	0.0	0.0	0.0	0.0	8.0	3
3773	9995187	1.0	0.0	0.0	0.0	0.0	3
3774	9995880	3.0	35.0	0.0	0.0	8.0	3

[3775 rows x 7 columns]

In [44]: df_lex1_2_short.rename(columns={'pid':'PID'}, inplace=True)

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:2844: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/indexing.html#](http://pandas.pydata.org/pandas-docs/stable/indexing.html#**kwargs)
**kwargs)

In [45]: df_lex1_2_short

Out[45]:

	PID	B87	B117	B118	B119	B81	B54
0	2924	0.0	0.0	1.0	0.0	0.0	3
1	3297	1.0	36.0	0.0	0.0	8.0	3
2	4061	1.0	18.0	0.0	0.0	8.0	4
3	8059	2.0	0.0	0.0	0.0	0.0	3
4	9458	0.0	0.0	1.0	2.0	0.0	3
5	14463	3.0	0.0	0.0	7.0	0.0	3
6	15059	0.0	0.0	1.0	0.0	0.0	3
7	18844	1.0	1.0	1.0	0.0	0.0	3
8	20206	3.0	1.0	0.0	4.0	0.0	3
9	22267	3.0	0.0	0.0	0.0	8.0	3

10	27195	0.0	0.0	2.0	2.0	0.0	3
11	27363	3.0	0.0	1.0	0.0	0.0	3
12	32688	3.0	10.0	5.0	1.0	8.0	3
13	33693	3.0	1.0	0.0	0.0	8.0	3
14	35867	0.0	0.0	0.0	0.0	0.0	3
15	38579	0.0	0.0	2.0	0.0	8.0	3
16	43540	1.0	0.0	0.0	0.0	0.0	3
17	46959	3.0	3.0	0.0	1.0	8.0	3
18	47291	1.0	0.0	0.0	0.0	0.0	3
19	49050	3.0	6.0	0.0	28.0	8.0	3
20	50122	1.0	0.0	28.0	0.0	8.0	3
21	52532	1.0	0.0	0.0	0.0	0.0	3
22	55025	1.0	0.0	1.0	1.0	0.0	3
23	55986	3.0	14.0	0.0	0.0	8.0	3
24	58198	0.0	0.0	1.0	2.0	0.0	3
25	65987	3.0	0.0	1.0	1.0	0.0	3
26	73470	3.0	0.0	4.0	0.0	0.0	3
27	78430	3.0	6.0	0.0	0.0	8.0	3
28	80711	3.0	0.0	0.0	1.0	8.0	3
29	83839	3.0	1.0	1.0	1.0	8.0	3
...
3745	9925701	3.0	1.0	0.0	0.0	8.0	3
3746	9926341	1.0	3.0	1.0	1.0	8.0	3
3747	9928033	0.0	0.0	1.0	0.0	0.0	3
3748	9930609	3.0	0.0	0.0	0.0	0.0	3
3749	9934124	0.0	0.0	0.0	0.0	0.0	3
3750	9936752	2.0	0.0	2.0	4.0	0.0	3
3751	9937420	1.0	2.0	0.0	2.0	8.0	3
3752	9939621	3.0	12.0	1.0	3.0	8.0	3
3753	9939888	3.0	24.0	0.0	32.0	8.0	3
3754	9940870	0.0	0.0	7.0	7.0	8.0	3
3755	9941870	3.0	0.0	28.0	0.0	0.0	3
3756	9945338	3.0	1.0	0.0	1.0	0.0	3
3757	9947765	0.0	0.0	2.0	0.0	0.0	3
3758	9947948	3.0	0.0	0.0	0.0	0.0	3
3759	9948725	0.0	0.0	14.0	3.0	0.0	3
3760	9951579	1.0	0.0	0.0	0.0	8.0	3
3761	9958150	0.0	0.0	0.0	0.0	2.0	3
3762	9965528	0.0	1.0	1.0	2.0	8.0	3
3763	9975906	3.0	0.0	0.0	0.0	0.0	3
3764	9977699	3.0	14.0	21.0	0.0	8.0	3
3765	9980291	3.0	0.0	5.0	2.0	0.0	3
3766	9982799	3.0	0.0	1.0	1.0	0.0	3
3767	9982997	0.0	0.0	1.0	0.0	0.0	3
3768	9985882	3.0	0.0	0.0	2.0	8.0	3
3769	9986747	1.0	0.0	0.0	7.0	8.0	3
3770	9987740	0.0	7.0	1.0	0.0	8.0	3
3771	9988354	0.0	0.0	2.0	2.0	0.0	3

```

3772 9989304 0.0 0.0 0.0 0.0 8.0 3
3773 9995187 1.0 0.0 0.0 0.0 0.0 3
3774 9995880 3.0 35.0 0.0 0.0 8.0 3

```

[3775 rows x 7 columns]

```
In [46]: newdf = pd.merge(df_offspring_bmi,df_lex1_2_short,on='PID', how='outer')
```

```
In [47]: newdf
```

```

Out[47]:
   idtype  PID sex att1 age1 examyr1 att2 date2 age2 examyr2 ... \
0         1  2924  2    1   36   1972    1  2949   44   1980 ...
1         1  3297  1    1   31   1971    1  2947   39   1979 ...
2         7  4040  2    1   43   1997    0                ...
3         1  4061  1    1   31   1972    1  2984   39   1980 ...
4         7  4301  2    1   51   1996    0                ...
5         1  5350  2    1   28   1975    0                ...
6         7  6719  2    1   44   1996    1  1870   49   2001 ...
7         1  7248  2    1   41   1973    0                ...
8         1  8059  2    1   38   1973    1  3405   47   1982 ...
9         7  8459  2    1   51   1996    0                ...
10        1  9458  2    1   36   1972    1  2953   44   1980 ...
11        7  9912  2    1   59   1996    1  1199   62   1999 ...
12        7 10791  1    1   46   1994    1  1825   51   1999 ...
13        7 11016  2    1   59   1994    1  1739   63   1999 ...
14        7 11651  2    1   50   1996    1  1355   54   1999 ...
15        1 14463  2    1   34   1974    1  2676   41   1982 ...
16        1 14764  2    1   44   1972    0                ...
17        1 15059  2    1   24   1971    1  2971   32   1980 ...
18        7 15398  2    1   53   1995    0                ...
19        7 18098  1    1   44   1996    1  1092   47   1999 ...
20        7 18103  2    1   41   1997    1  1015   44   1999 ...
21        1 18844  2    1   19   1973    1  2942   27   1981 ...
22        7 19353  2    1   52   1995    0                ...
23        7 19892  1    1   69   1998    1   665   70   2000 ...
24        1 20206  2    1   37   1975    1  2648   44   1982 ...
25        1 22267  1    1   30   1972    1  2926   38   1980 ...
26        7 22567  1    1   64   1998    1   735   66   2000 ...
27        1 22996  2    1   61   1973    0                ...
28        7 24738  1    1   57   1996    1  1605   61   2000 ...
29        7 25240  2    1   53   1997    0                ...
...      ...    ...    ...    ...    ...    ...    ...    ...
5477     1 9939621  1    1   45   1975    1  2569   52   1982 ...
5478     1 9939888  1    1   47   1974    1  2679   55   1982 ...
5479     1 9940525  2    1   58   1973    0                ...
5480     1 9940870  1    1   33   1973    1  2885   41   1981 ...
5481     1 9941870  2    1   29   1974    1  2695   37   1982 ...
5482     1 9943257  1    1   35   1972    0                ...

```

5483	1	9945338	2	1	38	1972	1	3014	46	1980 ...
5484	1	9947765	2	1	26	1974	1	2815	34	1981 ...
5485	1	9947948	2	1	46	1972	1	2948	54	1980 ...
5486	1	9948725	2	1	41	1973	1	2803	49	1981 ...
5487	1	9951579	1	1	43	1974	1	2802	51	1981 ...
5488	1	9957447	1	1	45	1972	0			...
5489	1	9958150	2	1	47	1973	1	2590	54	1980 ...
5490	1	9965528	1	1	25	1971	1	3110	34	1980 ...
5491	1	9967264	2	1	35	1974	0			...
5492	1	9973742	1	1	22	1973	0			...
5493	1	9975906	2	1	28	1973	1	2800	35	1981 ...
5494	1	9977699	1	1	27	1972	1	2962	35	1980 ...
5495	1	9980291	2	1	27	1971	1	2962	35	1979 ...
5496	1	9982799	2	1	28	1973	1	2775	36	1981 ...
5497	1	9982997	2	1	28	1972	1	3452	38	1981 ...
5498	1	9985882	1	1	35	1973	1	2897	43	1981 ...
5499	1	9986747	1	1	48	1973	1	2773	56	1981 ...
5500	1	9987740	1	1	17	1972	1	2963	25	1980 ...
5501	1	9988354	2	1	32	1972	1	2992	41	1980 ...
5502	1	9988615	2	1	38	1975	0			...
5503	1	9989304	1	1	47	1972	1	2993	55	1980 ...
5504	1	9990078	1	1	49	1974	0			...
5505	1	9995187	2	1	34	1972	1	2957	42	1980 ...
5506	1	9995880	1	1	53	1973	1	2889	61	1981 ...

	age7	site7	examyr7	BMI_1980	B87	B117	B118	B119	B81	B54
0	63		1998	22.648360726409	0.0	0.0	1.0	0.0	0.0	3.0
1				26.476583565412	1.0	36.0	0.0	0.0	8.0	3.0
2				NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	57		1998	31.644675765722	1.0	18.0	0.0	0.0	8.0	4.0
4				NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	54	2	2001	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6				NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	68		2000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	66	2	2001	23.8403797120094	2.0	0.0	0.0	0.0	0.0	3.0
9				NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	62		1999	18.033647098607	0.0	0.0	1.0	2.0	0.0	3.0
11				NaN	NaN	NaN	NaN	NaN	NaN	NaN
12				NaN	NaN	NaN	NaN	NaN	NaN	NaN
13				NaN	NaN	NaN	NaN	NaN	NaN	NaN
14				NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	60	0	2000	26.1290845162335	3.0	0.0	0.0	7.0	0.0	3.0
16				NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	53	0	2000	16.8390869326117	0.0	0.0	1.0	0.0	0.0	3.0
18				NaN	NaN	NaN	NaN	NaN	NaN	NaN
19				NaN	NaN	NaN	NaN	NaN	NaN	NaN
20				NaN	NaN	NaN	NaN	NaN	NaN	NaN
21				21.8407503749796	1.0	1.0	1.0	0.0	0.0	3.0

22					NaN	NaN	NaN	NaN	NaN	NaN	NaN
23					NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	63	0	2000	24.5982740486191	3.0	1.0	0.0	4.0	0.0	3.0	
25	57		1999	30.7171810534853	3.0	0.0	0.0	0.0	8.0	3.0	
26					NaN	NaN	NaN	NaN	NaN	NaN	NaN
27					NaN	NaN	NaN	NaN	NaN	NaN	NaN
28					NaN	NaN	NaN	NaN	NaN	NaN	NaN
29					NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
5477	71	0	2000	33.2062431702988	3.0	12.0	1.0	3.0	8.0	3.0	
5478				24.7590287241066	3.0	24.0	0.0	32.0	8.0	3.0	
5479					NaN	NaN	NaN	NaN	NaN	NaN	NaN
5480	59		1999	25.2201015724164	0.0	0.0	7.0	7.0	8.0	3.0	
5481	54		1999	20.3639031845538	3.0	0.0	28.0	0.0	0.0	3.0	
5482	63	0	2001		NaN	NaN	NaN	NaN	NaN	NaN	NaN
5483	65		1999	22.9889375794377	3.0	1.0	0.0	1.0	0.0	3.0	
5484	52		1999	22.1190982559337	0.0	0.0	2.0	0.0	0.0	3.0	
5485	73		1999	22.1374954468659	3.0	0.0	0.0	0.0	0.0	3.0	
5486	67		1999	37.3791070162786	0.0	0.0	14.0	3.0	0.0	3.0	
5487				23.007858515717	1.0	0.0	0.0	0.0	8.0	3.0	
5488	73	2	2001		NaN	NaN	NaN	NaN	NaN	NaN	NaN
5489	73		1999	21.6828688505022	0.0	0.0	0.0	0.0	2.0	3.0	
5490	53		1999	23.6824651849996	0.0	1.0	1.0	2.0	8.0	3.0	
5491	61		1999		NaN	NaN	NaN	NaN	NaN	NaN	NaN
5492					NaN	NaN	NaN	NaN	NaN	NaN	NaN
5493	53		1999	21.908598060903	3.0	0.0	0.0	0.0	0.0	3.0	
5494	54		1999	19.9113720026748	3.0	14.0	21.0	0.0	8.0	3.0	
5495	55		1999	22.648360726409	3.0	0.0	5.0	2.0	0.0	3.0	
5496	53		1999	21.4615813847012	3.0	0.0	1.0	1.0	0.0	3.0	
5497	56		1999	24.1810480933512	0.0	0.0	1.0	0.0	0.0	3.0	
5498	61		1999	30.5782924491019	3.0	0.0	0.0	2.0	8.0	3.0	
5499	73		1999	28.6416958471687	1.0	0.0	0.0	7.0	8.0	3.0	
5500	44		1999	22.1631195636387	0.0	7.0	1.0	0.0	8.0	3.0	
5501	60		2000	21.7914268028083	0.0	0.0	2.0	2.0	0.0	3.0	
5502	63		1999		NaN	NaN	NaN	NaN	NaN	NaN	NaN
5503				32.6953778907558	0.0	0.0	0.0	0.0	8.0	3.0	
5504					NaN	NaN	NaN	NaN	NaN	NaN	NaN
5505				21.6266704722759	1.0	0.0	0.0	0.0	0.0	3.0	
5506				18.7317269165788	3.0	35.0	0.0	0.0	8.0	3.0	

[5507 rows x 26 columns]

```
In [48]: df_lex1_4= pd.read_csv('lex1_4.csv') #1990 dataset.
```

```
In [49]: df_lex_1_4short=df_lex1_4[['pid', 'D443', 'D093', 'D014', 'D082', 'D085', 'D088', 'D061', 'D0
```

```
In [50]: df_lex_1_4short.rename(columns={'pid': 'PID'}, inplace=True)
```

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:2844: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/indexing.html#](http://pandas.pydata.org/pandas-docs/stable/indexing.html#**kwargs)
**kwargs)

```
In [51]: newdf2 = pd.merge(newdf, df_lex_1_4short, on='PID', how='outer')
```

```
In [52]: list(newdf2.columns.values)
```

```
Out[52]: ['idtype',  
          'PID',  
          'sex',  
          'att1',  
          'age1',  
          'examy1',  
          'att2',  
          'date2',  
          'age2',  
          'examy2',  
          'att4',  
          'date4',  
          'age4',  
          'examy4',  
          'att7',  
          'date7',  
          'age7',  
          'site7',  
          'examy7',  
          'BMI_1980',  
          'B87',  
          'B117',  
          'B118',  
          'B119',  
          'B81',  
          'B54',  
          'D443',  
          'D093',  
          'D014',  
          'D082',  
          'D085',  
          'D088',  
          'D061',  
          'D063',  
          'D014']
```

```
In [53]: newdf2.to_csv('offspring_bmi_1980_1990') #now to get 2000 risk factors.
```

```
In [54]: df_lex1_7=pd.read_csv('lex1_7.csv') #2000 dataset
```

```
In [55]: df_lex1_7_short = df_lex1_7[['pid', 'G453', 'G116', 'G680', 'G104', 'G107', 'G110', 'G113',
```

```
In [56]: df_lex1_7_short.rename(columns={'pid': 'PID'}, inplace=True)
```

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:2844: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/indexing.html#](http://pandas.pydata.org/pandas-docs/stable/indexing.html#**kwargs)
**kwargs)

```
In [57]: newdf3=pd.merge(newdf2,df_lex_1_4short,on='PID', how='outer')
```

```
In [58]: list(newdf3.columns.values)
```

```
Out[58]: ['idtype',  
          'PID',  
          'sex',  
          'att1',  
          'age1',  
          'examy1',  
          'att2',  
          'date2',  
          'age2',  
          'examy2',  
          'att4',  
          'date4',  
          'age4',  
          'examy4',  
          'att7',  
          'date7',  
          'age7',  
          'site7',  
          'examy7',  
          'BMI_1980',  
          'B87',  
          'B117',  
          'B118',  
          'B119',  
          'B81',  
          'B54',  
          'D443_x',  
          'D093_x',  
          'D014_x',  
          'D082_x',  
          'D085_x',  
          'D088_x',  
          'D061_x',  
          'D063_x',
```

```

'D014_x',
'D443_y',
'D093_y',
'D014_y',
'D082_y',
'D085_y',
'D088_y',
'D061_y',
'D063_y',
'D014_y']

```

```
In [59]: newdf3.to_csv('offspring_bmi_1980_1990_2000')
```

now I have all the data from the individual exam datasets. need to add diabetes and menopause

```
In [60]: diab_set=pd.read_csv('ldia1_7.csv')
```

```
In [61]: diab_set.rename(columns={'pid':'PID'},inplace=True)
```

```
In [62]: diab_set_short = diab_set[['PID','DIAB2','DIAB4','DIAB7']]
```

```
In [63]: newdf4=pd.merge(newdf3,diab_set_short, on='PID', how='outer')
```

```
In [64]: list(newdf4.columns.values) #getting a strange duplicaiton of the D variables. not sur
```

```

Out[64]: ['idtype',
          'PID',
          'sex',
          'att1',
          'age1',
          'examy1',
          'att2',
          'date2',
          'age2',
          'examy2',
          'att4',
          'date4',
          'age4',
          'examy4',
          'att7',
          'date7',
          'age7',
          'site7',
          'examy7',
          'BMI_1980',
          'B87',
          'B117',
          'B118',
          'B119',

```

```

'B81',
'B54',
'D443_x',
'D093_x',
'D014_x',
'D082_x',
'D085_x',
'D088_x',
'D061_x',
'D063_x',
'D014_x',
'D443_y',
'D093_y',
'D014_y',
'D082_y',
'D085_y',
'D088_y',
'D061_y',
'D063_y',
'D014_y',
'DIAB2',
'DIAB4',
'DIAB7']

```

```
In [65]: Menopause_set = pd.read_csv('vr_meno_ex07_1_0152d.csv')
```

```
In [66]: Menopause_set_short=Menopause_set[['PID', 'STOP_AGE', 'STOP2', 'STOP4', 'STOP7']]
```

```
In [67]: newdf5=pd.merge(newdf4,Menopause_set_short, on='PID', how='outer')
```

```
In [68]: newdf5.to_csv('offspring_bmi_1980_1990_2000_Diabetes_menopause')
```

```
In [69]: vr_death= pd.read_csv('vr_death_offspring.csv')
```

```
In [70]: newdf6=pd.merge(newdf5,vr_death, on='PID', how='outer')
```

```
In [71]: newdf6.to_csv('offspring_bmi_1980_1990_2000_Diabetes_menopause_death.csv')
```

```
In [72]: df_lex1_7=pd.read_csv('lex1_7.csv')
```

```
In [73]: df_lex1_7_short = df_lex1_7[['pid', 'G453', 'G116', 'G680', 'G104', 'G107', 'G110', 'G113',
```

```
In [74]: df_lex1_7_short.rename(columns={'pid': 'PID'}, inplace=True)
```

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:2844: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/indexing.html#](http://pandas.pydata.org/pandas-docs/stable/indexing.html#**kwargs)
**kwargs)

```
newdf7=pd.merge(newdf7,df_lex1_7_short, on='PID', how='outer')
```

```
In [75]: newdf7=pd.merge(newdf6,df_lex1_7_short, on='PID', how='outer')
```

```
In [76]: newdf7.to_csv('offspring_full_2000_redone.csv')
```

Stopping here. Need to rename the 2000's to the proper headings. then calculate alcohol..do in python. also calculate estrogen in python '

```
In [77]: list(newdf7.columns.values)
```

```
Out[77]: ['idtype',  
          'PID',  
          'sex',  
          'att1',  
          'age1',  
          'examy1',  
          'att2',  
          'date2',  
          'age2',  
          'examy2',  
          'att4',  
          'date4',  
          'age4',  
          'examy4',  
          'att7',  
          'date7',  
          'age7',  
          'site7',  
          'examy7',  
          'BMI_1980',  
          'B87',  
          'B117',  
          'B118',  
          'B119',  
          'B81',  
          'B54',  
          'D443_x',  
          'D093_x',  
          'D014_x',  
          'D082_x',  
          'D085_x',  
          'D088_x',  
          'D061_x',  
          'D063_x',  
          'D014_x',  
          'D443_y',  
          'D093_y',  
          'D014_y',
```

```

'D082_y',
'D085_y',
'D088_y',
'D061_y',
'D063_y',
'D014_y',
'DIAB2',
'DIAB4',
'DIAB7',
'STOP_AGE',
'STOP2',
'STOP4',
'STOP7',
'datedth',
'lastatt',
'lastcon',
'lastsoe',
'G453',
'G116',
'G680',
'G104',
'G107',
'G110',
'G113',
'G088',
'G075',
'G017',
'G018']

```

```
In [78]: df_offspring=pd.read_csv('offspring_full_2000_redone_columns renamed.csv')
```

```
In [79]: df_offspring
```

```
Out[79]:
```

	Unnamed: 0	idtype	PID	sex	att1	age1	examy1	att2	date2	age2	\
0	0	1	2924	2	1	36	1972	1	2949	44	
1	1	1	3297	1	1	31	1971	1	2947	39	
2	2	7	4040	2	1	43	1997	0			
3	3	1	4061	1	1	31	1972	1	2984	39	
4	4	7	4301	2	1	51	1996	0			
5	5	1	5350	2	1	28	1975	0			
6	6	7	6719	2	1	44	1996	1	1870	49	
7	7	1	7248	2	1	41	1973	0			
8	8	1	8059	2	1	38	1973	1	3405	47	
9	9	7	8459	2	1	51	1996	0			
10	10	1	9458	2	1	36	1972	1	2953	44	
11	11	7	9912	2	1	59	1996	1	1199	62	
12	12	7	10791	1	1	46	1994	1	1825	51	
13	13	7	11016	2	1	59	1994	1	1739	63	

14	14	7	11651	2	1	50	1996	1	1355	54
15	15	1	14463	2	1	34	1974	1	2676	41
16	16	1	14764	2	1	44	1972	0		
17	17	1	15059	2	1	24	1971	1	2971	32
18	18	7	15398	2	1	53	1995	0		
19	19	7	18098	1	1	44	1996	1	1092	47
20	20	7	18103	2	1	41	1997	1	1015	44
21	21	1	18844	2	1	19	1973	1	2942	27
22	22	7	19353	2	1	52	1995	0		
23	23	7	19892	1	1	69	1998	1	665	70
24	24	1	20206	2	1	37	1975	1	2648	44
25	25	1	22267	1	1	30	1972	1	2926	38
26	26	7	22567	1	1	64	1998	1	735	66
27	27	1	22996	2	1	61	1973	0		
28	28	7	24738	1	1	57	1996	1	1605	61
29	29	7	25240	2	1	53	1997	0		
...
5477	5477	1	9939621	1	1	45	1975	1	2569	52
5478	5478	1	9939888	1	1	47	1974	1	2679	55
5479	5479	1	9940525	2	1	58	1973	0		
5480	5480	1	9940870	1	1	33	1973	1	2885	41
5481	5481	1	9941870	2	1	29	1974	1	2695	37
5482	5482	1	9943257	1	1	35	1972	0		
5483	5483	1	9945338	2	1	38	1972	1	3014	46
5484	5484	1	9947765	2	1	26	1974	1	2815	34
5485	5485	1	9947948	2	1	46	1972	1	2948	54
5486	5486	1	9948725	2	1	41	1973	1	2803	49
5487	5487	1	9951579	1	1	43	1974	1	2802	51
5488	5488	1	9957447	1	1	45	1972	0		
5489	5489	1	9958150	2	1	47	1973	1	2590	54
5490	5490	1	9965528	1	1	25	1971	1	3110	34
5491	5491	1	9967264	2	1	35	1974	0		
5492	5492	1	9973742	1	1	22	1973	0		
5493	5493	1	9975906	2	1	28	1973	1	2800	35
5494	5494	1	9977699	1	1	27	1972	1	2962	35
5495	5495	1	9980291	2	1	27	1971	1	2962	35
5496	5496	1	9982799	2	1	28	1973	1	2775	36
5497	5497	1	9982997	2	1	28	1972	1	3452	38
5498	5498	1	9985882	1	1	35	1973	1	2897	43
5499	5499	1	9986747	1	1	48	1973	1	2773	56
5500	5500	1	9987740	1	1	17	1972	1	2963	25
5501	5501	1	9988354	2	1	32	1972	1	2992	41
5502	5502	1	9988615	2	1	38	1975	0		
5503	5503	1	9989304	1	1	47	1972	1	2993	55
5504	5504	1	9990078	1	1	49	1974	0		
5505	5505	1	9995187	2	1	34	1972	1	2957	42
5506	5506	1	9995880	1	1	53	1973	1	2889	61

		G453	Smoke_2000	Steroid_2000	Beer_2000	Whitewine_2000	\
0	...	NaN	0.0	NaN	0.0	0.0	
1	...	NaN	NaN	NaN	NaN	NaN	
2	...	NaN	NaN	NaN	NaN	NaN	
3	...	465.0	0.0	NaN	28.0	0.0	
4	...	NaN	NaN	NaN	NaN	NaN	
5	...	NaN	0.0	NaN	0.0	1.0	
6	...	NaN	NaN	NaN	NaN	NaN	
7	...	517.0	0.0	NaN	0.0	0.0	
8	...	NaN	1.0	NaN	0.0	1.0	
9	...	NaN	NaN	NaN	NaN	NaN	
10	...	701.0	0.0	NaN	0.0	3.0	
11	...	NaN	NaN	NaN	NaN	NaN	
12	...	NaN	NaN	NaN	NaN	NaN	
13	...	NaN	NaN	NaN	NaN	NaN	
14	...	NaN	NaN	NaN	NaN	NaN	
15	...	593.0	0.0	1.0	0.0	0.0	
16	...	NaN	NaN	NaN	NaN	NaN	
17	...	NaN	0.0	NaN	0.0	0.0	
18	...	NaN	NaN	NaN	NaN	NaN	
19	...	NaN	NaN	NaN	NaN	NaN	
20	...	NaN	NaN	NaN	NaN	NaN	
21	...	NaN	NaN	NaN	NaN	NaN	
22	...	NaN	NaN	NaN	NaN	NaN	
23	...	NaN	NaN	NaN	NaN	NaN	
24	...	585.0	0.0	NaN	0.0	6.0	
25	...	444.0	0.0	NaN	2.0	0.0	
26	...	NaN	NaN	NaN	NaN	NaN	
27	...	NaN	NaN	NaN	NaN	NaN	
28	...	NaN	NaN	NaN	NaN	NaN	
29	...	NaN	NaN	NaN	NaN	NaN	
...	
5477	...	480.0	0.0	NaN	1.0	1.0	
5478	...	NaN	NaN	NaN	NaN	NaN	
5479	...	NaN	NaN	NaN	NaN	NaN	
5480	...	NaN	0.0	NaN	2.0	4.0	
5481	...	605.0	1.0	NaN	0.0	14.0	
5482	...	404.0	0.0	NaN	7.0	1.0	
5483	...	582.0	0.0	NaN	0.0	0.0	
5484	...	733.0	0.0	NaN	0.0	1.0	
5485	...	601.0	1.0	NaN	0.0	0.0	
5486	...	430.0	0.0	NaN	0.0	2.0	
5487	...	NaN	NaN	NaN	NaN	NaN	
5488	...	NaN	0.0	NaN	0.0	0.0	
5489	...	621.0	0.0	NaN	0.0	0.0	
5490	...	477.0	0.0	NaN	1.0	0.0	
5491	...	462.0	0.0	NaN	0.0	0.0	
5492	...	NaN	NaN	NaN	NaN	NaN	

5493	...	611.0	1.0	NaN	0.0	0.0
5494	...	561.0	0.0	NaN	10.0	0.0
5495	...	637.0	0.0	NaN	0.0	0.0
5496	...	490.0	0.0	NaN	0.0	0.0
5497	...	630.0	0.0	NaN	0.0	0.0
5498	...	367.0	0.0	NaN	0.0	0.0
5499	...	419.0	0.0	NaN	0.0	0.0
5500	...	458.0	0.0	NaN	5.0	0.0
5501	...	501.0	0.0	NaN	0.0	0.0
5502	...	572.0	0.0	NaN	0.0	4.0
5503	...	NaN	NaN	NaN	NaN	NaN
5504	...	NaN	NaN	NaN	NaN	NaN
5505	...	NaN	NaN	NaN	NaN	NaN
5506	...	NaN	NaN	NaN	NaN	NaN

	RedWine_2000	Cocktail_2000	Estrogen_2000	Bisphosphonates_2000	Beta_2000
0	0.0	0.0	0.0	0.0	0.0
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	0.0	0.0	8.0	0.0	1.0
4	NaN	NaN	NaN	NaN	NaN
5	0.0	1.0	2.0	0.0	1.0
6	NaN	NaN	NaN	NaN	NaN
7	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0
9	NaN	NaN	NaN	NaN	NaN
10	0.0	0.0	1.0	0.0	0.0
11	NaN	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN	NaN
15	0.0	0.0	1.0	0.0	0.0
16	NaN	NaN	NaN	NaN	NaN
17	0.0	0.0	0.0	0.0	0.0
18	NaN	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN	NaN
24	0.0	0.0	1.0	1.0	0.0
25	0.0	0.0	8.0	0.0	0.0
26	NaN	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN
29	NaN	NaN	NaN	NaN	NaN
...
5477	3.0	14.0	8.0	0.0	0.0

5478	NaN	NaN	NaN	NaN	NaN
5479	NaN	NaN	NaN	NaN	NaN
5480	0.0	0.0	8.0	0.0	0.0
5481	0.0	0.0	1.0	0.0	0.0
5482	4.0	1.0	8.0	0.0	0.0
5483	0.0	0.0	1.0	0.0	1.0
5484	0.0	0.0	1.0	0.0	0.0
5485	0.0	0.0	0.0	0.0	1.0
5486	1.0	0.0	0.0	0.0	0.0
5487	NaN	NaN	NaN	NaN	NaN
5488	0.0	0.0	8.0	0.0	0.0
5489	0.0	0.0	1.0	0.0	0.0
5490	2.0	2.0	8.0	0.0	0.0
5491	0.0	0.0	0.0	0.0	1.0
5492	NaN	NaN	NaN	NaN	NaN
5493	0.0	0.0	0.0	0.0	0.0
5494	14.0	0.0	8.0	0.0	1.0
5495	0.0	0.0	1.0	0.0	0.0
5496	0.0	0.0	1.0	0.0	0.0
5497	0.0	0.0	0.0	0.0	0.0
5498	0.0	0.0	8.0	0.0	1.0
5499	7.0	7.0	8.0	0.0	0.0
5500	1.0	0.0	8.0	0.0	0.0
5501	0.0	0.0	1.0	0.0	0.0
5502	0.0	8.0	1.0	0.0	0.0
5503	NaN	NaN	NaN	NaN	NaN
5504	NaN	NaN	NaN	NaN	NaN
5505	NaN	NaN	NaN	NaN	NaN
5506	NaN	NaN	NaN	NaN	NaN

[5507 rows x 57 columns]

```
In [80]: df_lex1_7=pd.read_csv('lex1_7.csv')
df_lex1_7_short = df_lex1_7[['pid', 'G709']]
df_lex1_7_short.rename(columns={'pid': 'PID'}, inplace=True)
```

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:2844: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/indexing.html#](http://pandas.pydata.org/pandas-docs/stable/indexing.html#**kwargs)
**kwargs)

```
In [81]: df_lex1_7_short
```

```
Out[81]:
```

	PID	G709
0	2924	25.32
1	4061	33.33
2	5350	NaN

3	7248	32.03
4	8059	NaN
5	9458	20.08
6	14463	31.12
7	15059	22.96
8	20206	26.25
9	22267	30.15
10	27195	24.03
11	27363	21.49
12	32688	25.72
13	33693	29.14
14	35867	22.46
15	38579	25.45
16	44583	29.54
17	46959	28.84
18	47291	21.99
19	49050	31.55
20	52532	24.52
21	55025	24.19
22	55986	23.42
23	65987	21.80
24	70080	30.92
25	78430	28.82
26	85697	28.41
27	88392	29.93
28	89464	36.65
29	89854	31.31
...
3462	9926341	34.96
3463	9928033	20.61
3464	9929361	23.00
3465	9930609	24.40
3466	9934124	26.45
3467	9935469	33.76
3468	9936752	24.32
3469	9937420	23.84
3470	9939621	30.99
3471	9940870	28.85
3472	9941870	23.36
3473	9943257	26.63
3474	9945338	27.57
3475	9947765	19.67
3476	9947948	20.18
3477	9948725	33.96
3478	9957447	NaN
3479	9958150	24.91
3480	9965528	24.15
3481	9967264	35.95

```

3482 9975906 19.91
3483 9977699 23.35
3484 9980291 23.39
3485 9982799 26.69
3486 9982997 24.37
3487 9985882 37.82
3488 9986747 28.80
3489 9987740 25.31
3490 9988354 26.95
3491 9988615 25.74

```

```
[3492 rows x 2 columns]
```

```

In [83]: df_offspring2=pd.merge(df_offspring,df_lex1_7_short, on='PID', how='outer')
df_offspring2.rename(columns={'G709':'BMI_2000'},inplace=True)
df_offspring2.to_csv('offspring_full_2000_redone_columns_bmi.csv')
df_offspring2

```

```

Out[83]:
   Unnamed: 0  idtype  PID  sex  att1  age1  examyr1  att2  date2  age2  \
0           0        1   2924    2    1    36    1972    1   2949    44
1           1        1   3297    1    1    31    1971    1   2947    39
2           2        7   4040    2    1    43    1997    0
3           3        1   4061    1    1    31    1972    1   2984    39
4           4        7   4301    2    1    51    1996    0
5           5        1   5350    2    1    28    1975    0
6           6        7   6719    2    1    44    1996    1   1870    49
7           7        1   7248    2    1    41    1973    0
8           8        1   8059    2    1    38    1973    1   3405    47
9           9        7   8459    2    1    51    1996    0
10          10        1   9458    2    1    36    1972    1   2953    44
11          11        7   9912    2    1    59    1996    1   1199    62
12          12        7  10791    1    1    46    1994    1   1825    51
13          13        7  11016    2    1    59    1994    1   1739    63
14          14        7  11651    2    1    50    1996    1   1355    54
15          15        1  14463    2    1    34    1974    1   2676    41
16          16        1  14764    2    1    44    1972    0
17          17        1  15059    2    1    24    1971    1   2971    32
18          18        7  15398    2    1    53    1995    0
19          19        7  18098    1    1    44    1996    1   1092    47
20          20        7  18103    2    1    41    1997    1   1015    44
21          21        1  18844    2    1    19    1973    1   2942    27
22          22        7  19353    2    1    52    1995    0
23          23        7  19892    1    1    69    1998    1    665    70
24          24        1  20206    2    1    37    1975    1   2648    44
25          25        1  22267    1    1    30    1972    1   2926    38
26          26        7  22567    1    1    64    1998    1    735    66
27          27        1  22996    2    1    61    1973    0
28          28        7  24738    1    1    57    1996    1   1605    61

```

29	29	7	25240	2	1	53	1997	0		
...
5477	5477	1	9939621	1	1	45	1975	1	2569	52
5478	5478	1	9939888	1	1	47	1974	1	2679	55
5479	5479	1	9940525	2	1	58	1973	0		
5480	5480	1	9940870	1	1	33	1973	1	2885	41
5481	5481	1	9941870	2	1	29	1974	1	2695	37
5482	5482	1	9943257	1	1	35	1972	0		
5483	5483	1	9945338	2	1	38	1972	1	3014	46
5484	5484	1	9947765	2	1	26	1974	1	2815	34
5485	5485	1	9947948	2	1	46	1972	1	2948	54
5486	5486	1	9948725	2	1	41	1973	1	2803	49
5487	5487	1	9951579	1	1	43	1974	1	2802	51
5488	5488	1	9957447	1	1	45	1972	0		
5489	5489	1	9958150	2	1	47	1973	1	2590	54
5490	5490	1	9965528	1	1	25	1971	1	3110	34
5491	5491	1	9967264	2	1	35	1974	0		
5492	5492	1	9973742	1	1	22	1973	0		
5493	5493	1	9975906	2	1	28	1973	1	2800	35
5494	5494	1	9977699	1	1	27	1972	1	2962	35
5495	5495	1	9980291	2	1	27	1971	1	2962	35
5496	5496	1	9982799	2	1	28	1973	1	2775	36
5497	5497	1	9982997	2	1	28	1972	1	3452	38
5498	5498	1	9985882	1	1	35	1973	1	2897	43
5499	5499	1	9986747	1	1	48	1973	1	2773	56
5500	5500	1	9987740	1	1	17	1972	1	2963	25
5501	5501	1	9988354	2	1	32	1972	1	2992	41
5502	5502	1	9988615	2	1	38	1975	0		
5503	5503	1	9989304	1	1	47	1972	1	2993	55
5504	5504	1	9990078	1	1	49	1974	0		
5505	5505	1	9995187	2	1	34	1972	1	2957	42
5506	5506	1	9995880	1	1	53	1973	1	2889	61

	...	Smoke_2000	Steroid_2000	Beer_2000	Whitewine_2000	RedWine_2000	\
0	...	0.0	NaN	0.0	0.0	0.0	
1	...	NaN	NaN	NaN	NaN	NaN	
2	...	NaN	NaN	NaN	NaN	NaN	
3	...	0.0	NaN	28.0	0.0	0.0	
4	...	NaN	NaN	NaN	NaN	NaN	
5	...	0.0	NaN	0.0	1.0	0.0	
6	...	NaN	NaN	NaN	NaN	NaN	
7	...	0.0	NaN	0.0	0.0	0.0	
8	...	1.0	NaN	0.0	1.0	0.0	
9	...	NaN	NaN	NaN	NaN	NaN	
10	...	0.0	NaN	0.0	3.0	0.0	
11	...	NaN	NaN	NaN	NaN	NaN	
12	...	NaN	NaN	NaN	NaN	NaN	
13	...	NaN	NaN	NaN	NaN	NaN	

14	...	NaN	NaN	NaN	NaN	NaN
15	...	0.0	1.0	0.0	0.0	0.0
16	...	NaN	NaN	NaN	NaN	NaN
17	...	0.0	NaN	0.0	0.0	0.0
18	...	NaN	NaN	NaN	NaN	NaN
19	...	NaN	NaN	NaN	NaN	NaN
20	...	NaN	NaN	NaN	NaN	NaN
21	...	NaN	NaN	NaN	NaN	NaN
22	...	NaN	NaN	NaN	NaN	NaN
23	...	NaN	NaN	NaN	NaN	NaN
24	...	0.0	NaN	0.0	6.0	0.0
25	...	0.0	NaN	2.0	0.0	0.0
26	...	NaN	NaN	NaN	NaN	NaN
27	...	NaN	NaN	NaN	NaN	NaN
28	...	NaN	NaN	NaN	NaN	NaN
29	...	NaN	NaN	NaN	NaN	NaN
...
5477	...	0.0	NaN	1.0	1.0	3.0
5478	...	NaN	NaN	NaN	NaN	NaN
5479	...	NaN	NaN	NaN	NaN	NaN
5480	...	0.0	NaN	2.0	4.0	0.0
5481	...	1.0	NaN	0.0	14.0	0.0
5482	...	0.0	NaN	7.0	1.0	4.0
5483	...	0.0	NaN	0.0	0.0	0.0
5484	...	0.0	NaN	0.0	1.0	0.0
5485	...	1.0	NaN	0.0	0.0	0.0
5486	...	0.0	NaN	0.0	2.0	1.0
5487	...	NaN	NaN	NaN	NaN	NaN
5488	...	0.0	NaN	0.0	0.0	0.0
5489	...	0.0	NaN	0.0	0.0	0.0
5490	...	0.0	NaN	1.0	0.0	2.0
5491	...	0.0	NaN	0.0	0.0	0.0
5492	...	NaN	NaN	NaN	NaN	NaN
5493	...	1.0	NaN	0.0	0.0	0.0
5494	...	0.0	NaN	10.0	0.0	14.0
5495	...	0.0	NaN	0.0	0.0	0.0
5496	...	0.0	NaN	0.0	0.0	0.0
5497	...	0.0	NaN	0.0	0.0	0.0
5498	...	0.0	NaN	0.0	0.0	0.0
5499	...	0.0	NaN	0.0	0.0	7.0
5500	...	0.0	NaN	5.0	0.0	1.0
5501	...	0.0	NaN	0.0	0.0	0.0
5502	...	0.0	NaN	0.0	4.0	0.0
5503	...	NaN	NaN	NaN	NaN	NaN
5504	...	NaN	NaN	NaN	NaN	NaN
5505	...	NaN	NaN	NaN	NaN	NaN
5506	...	NaN	NaN	NaN	NaN	NaN

	Cocktail_2000	Estrogen_2000	Bisphosphonates_2000	Beta_2000	BMI_2000
0	0.0	0.0	0.0	0.0	25.32
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	0.0	8.0	0.0	1.0	33.33
4	NaN	NaN	NaN	NaN	NaN
5	1.0	2.0	0.0	1.0	NaN
6	NaN	NaN	NaN	NaN	NaN
7	0.0	0.0	0.0	0.0	32.03
8	0.0	0.0	0.0	0.0	NaN
9	NaN	NaN	NaN	NaN	NaN
10	0.0	1.0	0.0	0.0	20.08
11	NaN	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN	NaN
15	0.0	1.0	0.0	0.0	31.12
16	NaN	NaN	NaN	NaN	NaN
17	0.0	0.0	0.0	0.0	22.96
18	NaN	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN	NaN
24	0.0	1.0	1.0	0.0	26.25
25	0.0	8.0	0.0	0.0	30.15
26	NaN	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN
29	NaN	NaN	NaN	NaN	NaN
...
5477	14.0	8.0	0.0	0.0	30.99
5478	NaN	NaN	NaN	NaN	NaN
5479	NaN	NaN	NaN	NaN	NaN
5480	0.0	8.0	0.0	0.0	28.85
5481	0.0	1.0	0.0	0.0	23.36
5482	1.0	8.0	0.0	0.0	26.63
5483	0.0	1.0	0.0	1.0	27.57
5484	0.0	1.0	0.0	0.0	19.67
5485	0.0	0.0	0.0	1.0	20.18
5486	0.0	0.0	0.0	0.0	33.96
5487	NaN	NaN	NaN	NaN	NaN
5488	0.0	8.0	0.0	0.0	NaN
5489	0.0	1.0	0.0	0.0	24.91
5490	2.0	8.0	0.0	0.0	24.15
5491	0.0	0.0	0.0	1.0	35.95
5492	NaN	NaN	NaN	NaN	NaN

5493	0.0	0.0	0.0	0.0	19.91
5494	0.0	8.0	0.0	1.0	23.35
5495	0.0	1.0	0.0	0.0	23.39
5496	0.0	1.0	0.0	0.0	26.69
5497	0.0	0.0	0.0	0.0	24.37
5498	0.0	8.0	0.0	1.0	37.82
5499	7.0	8.0	0.0	0.0	28.80
5500	0.0	8.0	0.0	0.0	25.31
5501	0.0	1.0	0.0	0.0	26.95
5502	8.0	1.0	0.0	0.0	25.74
5503	NaN	NaN	NaN	NaN	NaN
5504	NaN	NaN	NaN	NaN	NaN
5505	NaN	NaN	NaN	NaN	NaN
5506	NaN	NaN	NaN	NaN	NaN

[5507 rows x 58 columns]

At this point I have the full offspring dataset in csv with columns renamed.

Need go back and rename the columns of the original dataset. Will do in excel. This was a bit dumb, because I didn't fork the excel sheet, and when I reran the python notebook, it erased my column replacements.'

```
In [90]: os.chdir('P:\Framingham hip\Original datasets of interest')
original_full=pd.read_csv('Original_lex17_bBlock_bmi_meo_diab_1990_2000.csv')
list(original_full.columns.values)
```

```
Out[90]: ['Unnamed: 0',
'Unnamed: 0.1',
'PID',
'idtype',
'sex',
'att1',
'age1',
'examyr1',
'att2',
'date2',
'examyr2',
'age2',
'att16',
'date16',
'age16',
'examyr16',
'att21',
'date21',
'age21',
'examyr21',
'att26',
'date26',
```



```

'age26',
'examyr26',
'BMI16',
'BMI21',
'bmi26',
'Diab_16',
'Diab_21',
'Diab_26',
'Estrogen_1980',
'Steroid_1980',
'Smoke_1980',
'Beer1980',
'Wine_1980',
'Cocktail_1980',
'FI43',
'AM2',
'AM3',
'AM4',
'AM5',
'IDTYPE',
'FN185',
'FN125',
'FN175',
'FN178',
'FN181',
'FN124',
'FN99',
'FS347',
'FS291',
'FS345',
'FS339',
'FS336',
'FS333',
'FS329',
'FS314',
'FS307',
'FS245']

```

```

In [103]: original_full.rename(columns={
'FI43': 'beta_1980',
'AM2': 'age_periods_stop',
'AM3': 'cause_periods_stop',
'AM4': 'ovaries_removed',
'AM5': 'exam_periods_stop',
'FN185': 'smoke_1990',
'FN125': 'steroid_1990',
'FN175': 'beer_1990',
'FN178': 'wine_1990',

```

```

'FN181': 'cocktail_1990',
'FN124': 'estrogen_1990',
'FN99': 'beta_1990',
'FS347': 'smoke_2000',
'FS291': 'steroid_2000',
'FS345': 'cocktail_2000',
'FS339': 'other_wine_2000',
'FS336': 'red_wine_2000',
'FS333': 'white_wine_2000',
'FS329': 'beer_2000',
'FS314': 'estrogen_2000',
'FS307': 'bisphosphonate_2000',
'FS245': 'beta_2000'

}

, inplace=True)

```

```
In [123]: list(original_full.columns.values)
```

```

Out[123]: ['Unnamed: 0',
'Unnamed: 0.1',
'PID',
'idtype',
'sex',
'att1',
'age1',
'examy1',
'att2',
'date2',
'examy2',
'age2',
'att16',
'date16',
'age16',
'examy16',
'att21',
'date21',
'age21',
'examy21',
'att26',
'date26',
'age26',
'examy26',
'BMI_1980',
'BMI_1990',
'BMI_2000',
'Diab_1980',
'Diab_1990',

```

```

'Diab_2000',
'Estrogen_1980',
'Steroid_1980',
'Smoke_1980',
'Beer1980',
'Wine_1980',
'Cocktail_1980',
'beta_1980',
'age_periods_stop',
'cause_periods_stop',
'ovaries_removed',
'exam_periods_stop',
'IDTYPE',
'smoke_1990',
'steroid_1990',
'beer_1990',
'wine_1990',
'cocktail_1990',
'estrogen_1990',
'beta_1990',
'smoke_2000',
'steroid_2000',
'cocktail_2000',
'other_wine_2000',
'red_wine_2000',
'white_wine_2000',
'beer_2000',
'estrogen_2000',
'biphosphonate_2000',
'beta_2000',
'total_drinks_1980']

```

```

In [106]: original_full.rename(columns={
    'BMI16': 'BMI_1980',
    'BMI21': 'BMI_1990',
    'bmi26': 'BMI_2000',
    'Diab_16': 'Diab_1980',
    'Diab_21': 'Diab_1990',
    'Diab_26': 'Diab_2000',

    },inplace=True)

```

```

In [ ]: list(original_full.columns.values)

```

```

In [110]: original_full.to_csv('original_full_columns_renamed.csv')

```

So here is my dataset for ORIGINAL. Now I'm going to calculate the actual risk factors from some of the columns

alcohol is considered a risk factor for osteoporosis if > 3 drinks per week. so we sum beer,wine and cocktail. If>3 ETOH_risk factor=1 else zero

```
In [122]: type(original_full['Beer1980'])
```

```
Out[122]: pandas.core.series.Series
```

got stuck here. Adding two columns together resulted in concatenation not sum.

```
In [124]: original_full['Smoke_1980'].describe()
```

```
Out[124]: count      2029
          unique       27
          top         88
          freq       957
          Name: Smoke_1980, dtype: object
```

```
In [134]: df['smoke_2000']
```

```
Out[134]: 0      NaN
          1      NaN
          2      NaN
          3      NaN
          4      NaN
          5      NaN
          6      NaN
          7      NaN
          8      NaN
          9      NaN
         10      NaN
         11      NaN
         12      NaN
         13      NaN
         14      NaN
         15      0.0
         16      NaN
         17      NaN
         18      NaN
         19      NaN
         20      NaN
         21      NaN
         22      0.0
         23      NaN
         24      NaN
         25      NaN
         26      NaN
         27      NaN
         28      NaN
         29      0.0
```

```

...
5049    NaN
5050    NaN
5051    NaN
5052    NaN
5053    0.0
5054    NaN
5055    NaN
5056    NaN
5057    NaN
5058    NaN
5059    NaN
5060    NaN
5061    NaN
5062    NaN
5063    NaN
5064    NaN
5065    NaN
5066    NaN
5067    NaN
5068    NaN
5069    NaN
5070    NaN
5071    0.0
5072    NaN
5073    NaN
5074    NaN
5075    NaN
5076    NaN
5077    NaN
5078    0.0
Name: smoke_2000, Length: 5079, dtype: float64

```

```
In [127]: df=original_full.apply(pd.to_numeric, errors='coerce')
```

```
In [128]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5079 entries, 0 to 5078
Data columns (total 60 columns):
Unnamed: 0          5079 non-null int64
Unnamed: 0.1        5079 non-null int64
PID                 5079 non-null int64
idtype              5079 non-null int64
sex                 5079 non-null int64
att1                5079 non-null int64
age1                5079 non-null int64
examyrl             5079 non-null int64

```

att2	5079 non-null int64
date2	4671 non-null float64
examyr2	4671 non-null float64
age2	4671 non-null float64
att16	5079 non-null int64
date16	2236 non-null float64
age16	2236 non-null float64
examyr16	2236 non-null float64
att21	5079 non-null int64
date21	1210 non-null float64
age21	1210 non-null float64
examyr21	1210 non-null float64
att26	5079 non-null int64
date26	542 non-null float64
age26	542 non-null float64
examyr26	542 non-null float64
BMI_1980	2190 non-null float64
BMI_1990	1019 non-null float64
BMI_2000	325 non-null float64
Diab_1980	2105 non-null float64
Diab_1990	997 non-null float64
Diab_2000	351 non-null float64
Estrogen_1980	2026 non-null float64
Steroid_1980	2028 non-null float64
Smoke_1980	1991 non-null float64
Beer1980	2016 non-null float64
Wine_1980	2016 non-null float64
Cocktail_1980	2015 non-null float64
beta_1980	2234 non-null float64
age_periods_stop	2658 non-null float64
cause_periods_stop	2705 non-null float64
ovaries_removed	2570 non-null float64
exam_periods_stop	2659 non-null float64
IDTYPE	2785 non-null float64
smoke_1990	1210 non-null float64
steroid_1990	1210 non-null float64
beer_1990	1206 non-null float64
wine_1990	1204 non-null float64
cocktail_1990	1205 non-null float64
estrogen_1990	1209 non-null float64
beta_1990	1210 non-null float64
smoke_2000	540 non-null float64
steroid_2000	540 non-null float64
cocktail_2000	454 non-null float64
other_wine_2000	458 non-null float64
red_wine_2000	458 non-null float64
white_wine_2000	458 non-null float64
beer_2000	455 non-null float64

```

estrogen_2000      542 non-null float64
bisphosphonate_2000 541 non-null float64
beta_2000          541 non-null float64
total_drinks_1980  2016 non-null float64
dtypes: float64(48), int64(12)
memory usage: 2.3 MB

```

In [129]: df

```

Out[129]:      Unnamed: 0  Unnamed: 0.1      PID  idtype  sex  att1  age1  examyr1  \
0              0              0  10000142      0      1      1      45      1949
1              1              1  10008641      0      1      1      36      1952
2              2              2  10009718      0      1      1      34      1952
3              3              3  10013675      0      2      1      47      1949
4              4              4  10014385      0      2      1      57      1949
5              5              5  10015218      0      2      1      45      1949
6              6              6  10018470      0      2      1      33      1951
7              7              7  10018519      0      2      1      53      1949
8              8              8  10021384      0      1      1      33      1950
9              9              9  10024359      0      2      1      43      1951
10             10             10  10028367      0      2      1      33      1951
11             11             11  10028966      0      2      1      33      1950
12             12             12  10033400      0      2      1      56      1951
13             13             13  10033434      0      1      1      49      1952
14             14             14  10034205      0      1      1      32      1949
15             15             15  10035343      0      2      1      36      1951
16             16             16  10037287      0      1      1      46      1949
17             17             17  10037330      0      1      1      51      1952
18             18             18  10037336      0      1      1      36      1951
19             19             19  10038395      0      2      1      43      1951
20             20             20  10040425      0      1      1      47      1951
21             21             21  10042702      0      2      1      44      1951
22             22             22  10046366      0      2      1      34      1952
23             23             23  10046544      0      2      1      40      1950
24             24             24  10048014      0      1      1      62      1949
25             25             25  10052831      0      1      1      55      1951
26             26             26  10055229      0      2      1      50      1949
27             27             27  10055254      0      2      1      49      1951
28             28             28  10056052      0      2      1      39      1951
29             29             29  10058420      0      1      1      31      1948
...             ...             ...      ...      ...      ...      ...      ...
5049           5049           5049  19955175      0      2      1      54      1949
5050           5050           5050  19956122      0      2      1      43      1952
5051           5051           5051  19962079      0      1      1      38      1951
5052           5052           5052  19962312      0      1      1      40      1950
5053           5053           5053  19964296      0      2      1      35      1952
5054           5054           5054  19966273      0      2      1      48      1948

```

5055	5055	5055	19966326	0	2	1	45	1951
5056	5056	5056	19967802	0	1	1	51	1951
5057	5057	5057	19968397	0	2	1	37	1949
5058	5058	5058	19970781	0	2	1	46	1949
5059	5059	5059	19970948	0	1	1	48	1951
5060	5060	5060	19971112	0	1	1	50	1951
5061	5061	5061	19974096	0	1	1	60	1950
5062	5062	5062	19975520	0	2	1	41	1948
5063	5063	5063	19976090	0	1	1	47	1951
5064	5064	5064	19976145	0	2	1	40	1948
5065	5065	5065	19976158	0	1	1	43	1950
5066	5066	5066	19976680	0	2	1	42	1950
5067	5067	5067	19979016	0	1	1	56	1951
5068	5068	5068	19981203	0	1	1	40	1949
5069	5069	5069	19981956	0	2	1	42	1949
5070	5070	5070	19982857	0	2	1	35	1949
5071	5071	5071	19983897	0	2	1	36	1951
5072	5072	5072	19986805	0	2	1	46	1949
5073	5073	5073	19989455	0	2	1	38	1952
5074	5074	5074	19995267	0	1	1	49	1949
5075	5075	5075	19995599	0	1	1	57	1952
5076	5076	5076	19996887	0	2	1	40	1951
5077	5077	5077	19997237	0	1	1	49	1951
5078	5078	5078	19997933	0	1	1	34	1951

	att2	date2	...	steroid_2000	cocktail_2000	\
0	1	924.0	...	NaN	NaN	
1	1	707.0	...	NaN	NaN	
2	1	763.0	...	NaN	NaN	
3	1	800.0	...	NaN	NaN	
4	1	906.0	...	NaN	NaN	
5	0	NaN	...	NaN	NaN	
6	1	729.0	...	NaN	NaN	
7	1	766.0	...	NaN	NaN	
8	1	734.0	...	NaN	NaN	
9	1	686.0	...	NaN	NaN	
10	0	NaN	...	NaN	NaN	
11	1	706.0	...	NaN	NaN	
12	0	NaN	...	NaN	NaN	
13	1	699.0	...	NaN	NaN	
14	1	787.0	...	NaN	NaN	
15	1	744.0	...	0.0	0.0	
16	1	637.0	...	NaN	NaN	
17	1	664.0	...	NaN	NaN	
18	1	645.0	...	NaN	NaN	
19	1	671.0	...	NaN	NaN	
20	1	666.0	...	NaN	NaN	
21	1	676.0	...	NaN	NaN	

22	1	717.0	...	0.0	0.0
23	1	722.0	...	NaN	NaN
24	1	782.0	...	NaN	NaN
25	1	670.0	...	NaN	NaN
26	1	913.0	...	NaN	NaN
27	1	746.0	...	NaN	NaN
28	1	700.0	...	NaN	NaN
29	1	971.0	...	0.0	0.0
...
5049	1	780.0	...	NaN	NaN
5050	0	NaN	...	NaN	NaN
5051	1	833.0	...	NaN	NaN
5052	0	NaN	...	NaN	NaN
5053	1	636.0	...	0.0	0.0
5054	1	953.0	...	NaN	NaN
5055	1	682.0	...	NaN	NaN
5056	1	663.0	...	NaN	NaN
5057	1	938.0	...	NaN	NaN
5058	1	778.0	...	NaN	NaN
5059	1	658.0	...	NaN	NaN
5060	1	893.0	...	NaN	NaN
5061	0	NaN	...	NaN	NaN
5062	1	907.0	...	NaN	NaN
5063	1	660.0	...	NaN	NaN
5064	1	1192.0	...	NaN	NaN
5065	1	841.0	...	NaN	NaN
5066	1	812.0	...	NaN	NaN
5067	1	666.0	...	NaN	NaN
5068	0	NaN	...	NaN	NaN
5069	1	933.0	...	NaN	NaN
5070	1	826.0	...	NaN	NaN
5071	1	746.0	...	0.0	0.0
5072	1	856.0	...	NaN	NaN
5073	0	NaN	...	NaN	NaN
5074	0	NaN	...	NaN	NaN
5075	1	697.0	...	NaN	NaN
5076	1	749.0	...	NaN	NaN
5077	1	667.0	...	NaN	NaN
5078	1	696.0	...	0.0	3.0

	other_wine_2000	red_wine_2000	white_wine_2000	beer_2000	\
0	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	
6	NaN	NaN	NaN	NaN	

7	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN
15	0.0	0.0	0.0	0.0
16	NaN	NaN	NaN	NaN
17	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN
22	0.0	1.0	0.0	0.0
23	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN
29	0.0	0.0	0.0	0.0
...
5049	NaN	NaN	NaN	NaN
5050	NaN	NaN	NaN	NaN
5051	NaN	NaN	NaN	NaN
5052	NaN	NaN	NaN	NaN
5053	0.0	0.0	0.0	0.0
5054	NaN	NaN	NaN	NaN
5055	NaN	NaN	NaN	NaN
5056	NaN	NaN	NaN	NaN
5057	NaN	NaN	NaN	NaN
5058	NaN	NaN	NaN	NaN
5059	NaN	NaN	NaN	NaN
5060	NaN	NaN	NaN	NaN
5061	NaN	NaN	NaN	NaN
5062	NaN	NaN	NaN	NaN
5063	NaN	NaN	NaN	NaN
5064	NaN	NaN	NaN	NaN
5065	NaN	NaN	NaN	NaN
5066	NaN	NaN	NaN	NaN
5067	NaN	NaN	NaN	NaN
5068	NaN	NaN	NaN	NaN
5069	NaN	NaN	NaN	NaN
5070	NaN	NaN	NaN	NaN
5071	0.0	0.0	0.0	0.0
5072	NaN	NaN	NaN	NaN

5073	NaN	NaN	NaN	NaN
5074	NaN	NaN	NaN	NaN
5075	NaN	NaN	NaN	NaN
5076	NaN	NaN	NaN	NaN
5077	NaN	NaN	NaN	NaN
5078	0.0	2.0	2.0	0.0

	estrogen_2000	bisphosphonate_2000	beta_2000	total_drinks_1980
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	77.0
2	NaN	NaN	NaN	0.0
3	NaN	NaN	NaN	0.0
4	NaN	NaN	NaN	0.0
5	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	0.0
7	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	44.0
12	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	1.0
15	0.0	1.0	1.0	0.0
16	NaN	NaN	NaN	0.0
17	NaN	NaN	NaN	70.0
18	NaN	NaN	NaN	110.0
19	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN
22	0.0	0.0	0.0	31.0
23	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN
29	8.0	0.0	0.0	1.0
...
5049	NaN	NaN	NaN	0.0
5050	NaN	NaN	NaN	NaN
5051	NaN	NaN	NaN	NaN
5052	NaN	NaN	NaN	201.0
5053	0.0	0.0	1.0	10.0
5054	NaN	NaN	NaN	NaN
5055	NaN	NaN	NaN	NaN
5056	NaN	NaN	NaN	NaN
5057	NaN	NaN	NaN	5214.0

5058	NaN	NaN	NaN	NaN
5059	NaN	NaN	NaN	NaN
5060	NaN	NaN	NaN	NaN
5061	NaN	NaN	NaN	NaN
5062	NaN	NaN	NaN	NaN
5063	NaN	NaN	NaN	NaN
5064	NaN	NaN	NaN	NaN
5065	NaN	NaN	NaN	NaN
5066	NaN	NaN	NaN	8.0
5067	NaN	NaN	NaN	NaN
5068	NaN	NaN	NaN	0.0
5069	NaN	NaN	NaN	NaN
5070	NaN	NaN	NaN	172.0
5071	0.0	0.0	0.0	4.0
5072	NaN	NaN	NaN	NaN
5073	NaN	NaN	NaN	NaN
5074	NaN	NaN	NaN	NaN
5075	NaN	NaN	NaN	NaN
5076	NaN	NaN	NaN	20.0
5077	NaN	NaN	NaN	NaN
5078	8.0	0.0	0.0	1211.0

[5079 rows x 60 columns]

In [131]: df['Smoke_1980']

Out[131]:

0	NaN
1	0.0
2	0.0
3	0.0
4	NaN
5	NaN
6	30.0
7	NaN
8	NaN
9	NaN
10	NaN
11	88.0
12	NaN
13	NaN
14	88.0
15	88.0
16	88.0
17	88.0
18	20.0
19	NaN
20	NaN
21	NaN

```

22      88.0
23      NaN
24      NaN
25      NaN
26      NaN
27      NaN
28      NaN
29      0.0
...
5049    88.0
5050    NaN
5051    NaN
5052     0.0
5053    88.0
5054    NaN
5055    NaN
5056    NaN
5057    88.0
5058    NaN
5059    NaN
5060    NaN
5061    NaN
5062    NaN
5063    NaN
5064    NaN
5065    NaN
5066     0.0
5067    NaN
5068    88.0
5069    NaN
5070     0.0
5071    88.0
5072    NaN
5073    NaN
5074    NaN
5075    NaN
5076     0.0
5077    NaN
5078    88.0

```

```
Name: Smoke_1980, Length: 5079, dtype: float64
```

ok we learned tha smoke 1980 is corrupt. lets go get it again.

Oh I forgot, smoke_1980 is the number of ciggartes do you now smoke. Need to change this to
>1 is 1

In []: