UKB CHD-Sex-Metabolite Project Update

Margaret Janiczek

2023-05-11

Summary

We gained access to download the UKBiobank NMR Metabolomics data on April 27, in addition to having access to the full cohort of UKBiobank data. Our first objective was to summarize some baseline characteristics in both the full UKB sample (N = 502,411) as well as the subset with available metabolomics (N = 121,716). We also summarized the occurrences of various cardiac-related outcomes after initial assessment in both the full sample and subset with metabolites. We stratified the summaries by sex (field 31) as well as included overall summaries.

Methods

I summarized number and percent of total for categorical variables such as smoking status, hysterectomy, menopause, and hormone replacement therapy (HRT) status. I calculated median and interquartile range (IQR) of several continuous measurements taken at first assessment visit (height, weight, body mass index (BMI), age at first assessment, body and trunk fat percentage, whole body impedance (estimated muscle mass), HbA1c, several blood pressure (BP) and pulse measurements (both automated and manual), and various bone mineral density (BMD) measures). Field code is included in the table in case definitions are needed for clarity.

By matching ICD10 codes (as described in the Said 2018 paper) in field 41270, I identified incident diagnosis of Hypertension, Type 2 Diabetes, Atrial Fibrilation, Stroke, and Coronary Artery Disease (CAD). I used the corresponding date of diagnoses (field 41280) and compared it to the date of initial assessment (field 53) to identify diagnoses that occurred after first assessment. Some subjects had more than one condition (e.g. both hypertension and diabetes), all distinct diagnoses are included. In cases where a subject had multiple dates for the same diagnosis (e.g. multiple dates recording hypertension), I used the first diagnosis date to compare to initial assessment at UKB assessment center.

I made an indicator variable "Has Metabolites" based on if subjects had any values entered in fields 23400-23948 to get the subset of 121,716 subjects with NMR metabolite measurements.

Next steps

The incident table should be thought of as a "lower bound" since it only includes diagnoses from the field of ICD10 codes, which were captured during in-patient hospital visits. We also need to look at the mortality, outpatient surgery, and self-reported diagnoses fields to capture a more precise estimate for the incidence of outcomes.

Next we will also need to do processing of the metabolite data itself before summarizing it and moving forward with analysis.

Questions

- What additional demographic/baseline assessment variables should we consider?
- What definitions of demographic/baseline assessment variables should we use (e.g. in cases where there are self-reported, automatic measurements, and manual entry, we should determine which measurement(s) to use)?
- Are there additional outcomes we should examine?

Tables

ICD10 Codes

The following ICD10 codes were used to identify outcomes of interest.

disease	ICD10
Coronary Artery Disease	I21-25, Z951, Z955
Atrial Fibrillation	I48
Stroke	I60, I61, I629, I63, I64, I678, I690, I693, G951, H341, H342, S066
Hypertension	I10-I13, I15, O10
Diabetes Mellitus Type 2	E10-E14

Summary of incidence of outcomes

The below table describes the number and percent of incident diagnoses (defined as ICD10 diagnosis that occurred after date of first assessment). "Samples with Metabolites" is defined as the subset of subjects with any entry in the NMR metabolite fields (23400-23948).

Characteristic	**Overall**, N = 502,411	**Female**, N = 273,325	**Male**, N = 229,086	**Overall**, N = 121,716	**Female**, N = 65,774	**Male**, N = 55,942
Hypertension	111,578 (22%)	53,379 (20%)	58,199 (25%)	27,032 (22%)	12,796 (19%)	14,236 (25%)
Type 2 Diabetes	35,049 (7.0%)	14,593 (5.3%)	20,456 (8.9%)	8,408 (6.9%)	3,436 (5.2%)	4,972 (8.9%)
Atrial Fibrillation	29,934 (6.0%)	11,442 (4.2%)	18,492 (8.1%)	7,284 (6.0%)	2,791 (4.2%)	4,493 (8.0%)
Stroke	11,827 (2.4%)	5,013 (1.8%)	6,814 (3.0%)	2,877 (2.4%)	1,200 (1.8%)	1,677 (3.0%)
CAD	41,565 (8.3%)	15,193 (5.6%)	26,372 (12%)	10,023 (8.2%)	3,646 (5.5%)	6,377 (11%)

Summary of Baseline characteristics

The below table describes the distribution of various baseline characteristics and measurements, defined as measured at the first date of assessment.

Note that certain characteristics only have entry for subjects with Sex = Female (e.g. hysterectomy, menopause, and HRT status).

		All Samples	
Characteristic	**N**	**Overall**, N = 502,411[1]	**Female**, N = 273,325[1]

Current Smoker	501,518	NA	NA
Prefer not to answer	NA	428 (<0.1%)	220 (<0.1%)
no	NA	448,128 (89%)	248,293 (91%)
yes	NA	39,231 (7.8%)	18,681 (6.8%)
occasionally	NA	13,731 (2.7%)	5,680 (2.1%)
Hysterectomy (ever)	241,682	NA	NA
Prefer not to answer	NA	272 (0.1%)	272 (0.1%)
Not Sure	NA	257 (0.1%)	257 (0.1%)
no	NA	221,237 (92%)	221,237 (92%)
yes	NA	19,916 (8.2%)	19,916 (8.2%)
HRT Use (ever)	272,846	NA	NA
Not Sure	NA	799 (0.3%)	799 (0.3%)
Prefer not to answer	NA	297 (0.1%)	297 (0.1%)
no	NA	167,845 (62%)	167,845 (62%)
yes	NA	103,905 (38%)	103,905 (38%)
Had menopause	272,849	NA	NA
Prefer not to answer	NA	534 (0.2%)	534 (0.2%)
no	NA	64,043 (23%)	64,043 (23%)
yes	NA	165,380 (61%)	165,380 (61%)
Not Sure (Hysterectomy)	NA	31,165 (11%)	31,165 (11%)
Not Sure (Other)	NA	11,727 (4.3%)	11,727 (4.3%)
Pulse (automated) (102)	467,984	68 (62, 76)	69 (63, 77)
BMI (21001)	499,304	26.7 (24.1, 29.9)	26.1 (23.5, 29.7)
Baseline age (21003)	502,411	58 (50, 63)	57 (50, 63)
Pulse wave arterial stiffness (21021)	169,759	8.99 (6.87, 11.16)	8.27 (6.31, 10.47)
Bodyfat % (23099)	492,003	31 (25, 38)	37 (32, 41)
Whole body impedance (23106)	492,224	595 (533, 663)	652 (605, 701)
Trunk fat % (23127)	491,981	31 (26, 37)	35 (29, 40)
had_menopause_f2724_0_0	272,849	NA	NA
-3	NA	534 (0.2%)	534 (0.2%)
0	NA	64,043 (23%)	64,043 (23%)
1	NA	165,380 (61%)	165,380 (61%)
2	NA	31,165 (11%)	31,165 (11%)
3	NA	11,727 (4.3%)	11,727 (4.3%)
ever_used_hormonereplacement_therapy_hrt_f2814_0_0	272,846	NA	NA
-1	NA	799 (0.3%)	799 (0.3%)
-3	NA	297 (0.1%)	297 (0.1%)
0	NA	167,845 (62%)	167,845 (62%)
1	NA	103,905 (38%)	103,905 (38%)
Age at Hysterectomy (2824)	51,031	43 (38, 49)	43 (38, 49)
PEF (3064)	453,610	379 (304, 474)	332 (275, 384)
HbA1c (30750)	466,414	35.2 (32.8, 37.9)	35.2 (32.7, 37.7)
Age started HRT (3536)	103,905	48 (42, 50)	48 (42, 50)
Age last used HRT (3546)	103,905	52 (-1, 57)	52 (-1, 57)
Diastolic BP (auto) (4079)	467,984	82 (75, 89)	80 (73, 88)
Systolic BP (auto) (4080)	467,971	138 (126, 152)	135 (122, 150)
Heel BMD (auto, Tscore) (4125)	164,169	-0.39 (-1.09, 0.40)	-0.61 (-1.27, 0.13)
Heel BMD (manual, Tscore) (4138)	1,228	0.60 (0.20, 1.20)	0.60 (0.10, 1.30)
Waist circumference (cm) (48)	500,248	90 (80, 99)	83 (75, 92)
Hip circumference (cm) (49)	500,189	102 (97, 108)	102 (96, 109)
Height (cm) (50)	499,870	168 (162, 175)	162 (158, 167)
Heel bone ultrasound (manual, Tscore) (77)	42,723	-0.30 (-1.10, 0.50)	-0.50 (-1.20, 0.30)
Heel BMD (automated, Tscore) (78)	279,042	-0.46 (-1.15, 0.32)	-0.65 (-1.31, 0.07)
Systolic BP (manual) (93)	33,102	139 (126, 153)	136 (123, 150)

Diastolic BP (manual) (94)	33,102	82 (75, 90)	81 (74, 88)
Pulse during BP (manual) (95)	33,102	70 (63, 78)	71 (64, 78)

[1]n (%); Median (IQR)