**Supplementary file 2. Example of a primary care electronic health record data extraction template: for type 2 diabetes.**This template includes relevant background demographic variables, diagnoses, clinical measurements and laboratory investigations

| Variable name | Field type | READ code (if relevant) | Excluded READ codes | Time window of extraction |
|---|---|---|---|---|
| Date of data extract | Date | | | |
| Patient ID | Numeric | NA | | |
| Gender | Text | NA | | |
| Age | Numeric | NA | | |
| Year of birth | Year | NA | | |
| Ethnicity | code | 9S%, 9i% | | latest ever |
| Country of birth | code | 13e% , 13Zq , 13k% , 13g% , 13d% , 13f% , 13j% , 13h% | | latest ever |
| IMD2010 | Numeric | NA | | latest ever |
| LSOA2011 | code | NA | | latest ever |
| Registration Start Date | Date | NA | | |
| Registration End Date | Date | NA | | |
| Family History of Diabetes | code/date | 1253 | | latest ever |
| Type 1 Diabetes | code/date | C10E% , C10ER | 21263, 212H  Diabetes resolved | earliest ever |
| Type 2 Diabetes | code/date | C10F%, C109J | C10F8 , 21263, 212H Diabetes resolved | earliest ever |
| Secondary Diabetes | code/date | C10H%, C10B, C10N%, C10G% | | earliest ever |

| | | | | |
|---|---|---|---|---|
| Other types of diabetes | code/date | C10A, C10C, C10D, C1A%, C10M%, C326%, A3A2, C135, PKyP, Q441, C10FS, C150%, C3500, C10N1, C370%, PKyF, C1zy4, PH3yA, PKy93, PKy1 | | earliest ever |
| Pancreatic disease/surgery | code/date | J670%, J671%, J6710, J6711, J67y6, 9b8G | | earliest ever |
| Diabetes in pregnancy | code/date | L1808, L1809 | | earliest ever |
| Diabetes emergencies | code/date | C10EM, C109K | | latest ever |
| Bariatric surgery | code/date | 76110, 76111, 76112, 76113, 76114, 76115, 76116, 76130, 76131, 76132, 76133, 76134, 76135, 76136, 76140, 76141, 76160, 76166 | | earliest ever |
| Diabetes risk: Prediabetes code OR glucose intolerance codes | code/date (x2) | C11y2, C11y3, C11y5, R10E, R10D0, R102 | | earliest AND latest ever |
| Current smoker | code/date | 1371, 1372,1373,1374,1375,1376,1377,1378,1379,137a,137A,137b,137B,137c, 137C, 137d, 137e, 137f, 137F, 137g, 137G, 137h, 137H, 137j, 137J, 137K, 137l, 137L, 137m,137M,137n,137N,137O,137P, 137Q, 137R, 137S, 137T, 137V, 137X,137Y,137Z | 137D, 137E, 137i, 137I, 137k, 137o,137U,137W, EMISNQEX4, EMISNQSM14,EMISNQSN1,EMISNQWA11 | latest ever |
| On examination weight | code/date/value/unit/age@event (x3) | 22A | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |
| On examination height | code/date/value/unit/age@event (x3) | 229 | | earliest ever AND latest up to + 6 months |

| | | | | |
|---|---|---|---|---|
| | | | | after T1 or T2 diagnosis |
| BMI | code/date/value/age@event (x3) | 22K | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |
| HbA1c | code/date/value (x3) | 42W5 , 42W4 | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |
| Total cholesterol | code/date/value (x3) | 44P | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |
| HDL cholesterol | code/date/value (x3) | 44P5 | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |

| | | | | |
|---|---|---|---|---|
| Serum LDL cholesterol (LDL-C) | code/date/value (x3) | 44P6 | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |
| Serum triglycerides | code/date/value (x3) | 44Q | | earliest ever AND in +/- 6 month window around T1 or T2 diagnosis AND latest ever AND highest ever |
| Urine protein/creatinine ratio | code/date/value (x2) | 44ID | | earliest ever AND latest ever |
| Urine albumin/creatinine ratio | code/date/value (x2) | 46TC | | earliest ever AND latest ever |
| Urine albumin | code/date/value (x2) | 46N4 | | earliest ever AND latest ever |
| Systolic Blood Pressure | code/date/value (x2) | 2469 | | earliest ever AND latest ever AND highest ever |
| Diastolic Blood Pressure | code/date/value (x2) | 246A | | earliest ever AND latest ever AND highest ever |
| Serum creatinine | code/date/value (x2) | 44J3 | | earliest ever AND latest ever |

| eGFR | code/date/value (x2) | 451E | | earliest ever AND latest ever |
|------|----------------------|------|--|-------------------------------|