Report Supplementary Material 3: Statistical model specification

Objective 1

In the Cox regressions modelling the time to type 2 diabetes (T2DM) onset, patients became at risk on the diagnosis date of SMI. For those patients whose follow-up started after SMI diagnosis, they were treated as 'delayed entry' and entered the analysis at the start of follow-up. Patients exited the analysis at the time when T2DM was diagnosed. Otherwise they were right-censored at the end of the follow-up period. Time of analysis was measured in years and financial years were adjusted in these regressions. We tested the proportionality assumption for covariates using the Schoenfeld residuals[118] and, for covariates violating this assumption, we included interaction terms between the covariate in question and the analysis time. Only interaction terms with statistically significant coefficients were kept in our final models.

Objective 2

In addition to the candidate explanatory variables, we adjusted for the length of follow-up and financial years in the regressions of macrovascular complications and mental health outcomes. We also specified cluster robust standard errors to account for the correlation by practices in these regressions. In the Cox proportional hazards regressions for microvascular complications and all-cause mortality, patients became 'at risk' and entered the analysis on the recorded diagnosis date of SMI or T2DM, whichever was later. They were treated as 'delayed entry' if their follow-up started later than the second diagnosis date. Patients exited the analysis on the date when the first clinical event was recorded for retinopathy, nephropathy or neuropathy in the analysis of microvascular complications (these being irreversible), and on the date of death in the analysis of mortality. In the absence of these events, patients were right-censored at the end of follow-up. Financial years were adjusted in these regressions. We tested the proportional hazard assumption using the Schoenfeld residuals and, for covariates with violated assumption, we included an interaction term between the analysis time and the covariates in question. Only interactions with statistically significant coefficients were kept in the final models.

Objective 3

In the regressions of macrovascular complications, SMI relapses and physical health checks, we further adjusted for the length of follow-up and financial years to account for the impact of time, in addition to the candidate explanatory variables. We specified robust standard errors to account for correlations by practice in these regressions. For all-cause mortality, we applied Cox proportional hazards regressions to estimate the association between diabetes status and the time from SMI diagnosis to death. Patients became 'at risk' upon the diagnosis of SMI and entered the analysis at the start of their follow-up periods (as defined above). Those people with follow-up started after SMI diagnosis were treated as 'delayed entry'. The exit point was the date of death for patients who died before the end of follow-up. Otherwise they were right-censored at the end of follow-up period. Our previous findings (from objectives 1 and 2) suggested that most people with both SMI and T2DM were diagnosed with SMI first, and that people with different diagnosis orders differed in many aspects.

We therefore accounted for the impact of diagnosis timing and treated the status of T2DM, our exposure variable, as time-dependent. We estimated robust standard errors to account for the clustering by practices and tested the proportional hazard assumption using the Schoenfeld residuals. For risk factors rejecting this assumption, we interacted these variables with analysis time and kept interaction terms with significant coefficients in the final models.

The impact of T2DM on the time to onset of microvascular complications was analysed using Cox proportional hazards regressions, following the same estimation strategy as all-cause mortality. Patients exited the analysis upon the earliest diagnosis of microvascular complications, otherwise they became right-censored at the end of follow-up period. The timing of T2DM diagnosis was accounted for by treating the status of T2DM as time-dependent.

Objective 4

We adjusted for the effect of time by including length of follow-up and financial years in the regressions of macro-vascular complications, depression and physical health checks. In the regressions of mortality, we treated the SMI status as time-dependent to account for the diagnosis timing and adjusted for financial years. Standard errors robust for correlations by practices were specified in all these regressions.