

**Anti-Diabetic Drug Repurposing Using Electronic Health Records:
Design, Emulation and Analysis of a Synthetic In-Silico Clinical Trial for Alzheimer's Disease**

*Stan N. Finkelstein, MD¹, Shenbo Xu¹, Bowen Su², Bang Zheng², Marie-Laure Charpignon¹,
Ioanna Tzoulaki, PhD², Lefkos T. Middleton, MD² and Roy E. Welsch, PhD¹*

¹MIT, Cambridge, USA and ²Imperial College, London, UK

Background. Alzheimer's Disease affects more than 50 million persons worldwide, with a global health cost rising to a trillion US dollars in 2018. Yet, there is currently no known disease-modifying treatment for this condition. Hence the strong appeal in mining Electronic Health Records (EHR) to identify potential drug repurposing candidates in silico. Although the two conditions may be largely independent of each other, earlier studies have found that people with type 2 diabetes are more likely to develop Alzheimer's disease. Diabetes is also a risk factor for other types of dementia, because of their shared association with vascular and cardiovascular problems.

Data. Using the UK Clinical Practice Research Datalink (CPRD) primary-care database, we emulated a clinical trial of two anti-diabetic drugs and assessed their comparative effectiveness in slowing down dementia progression. Dementia diagnosis, death or change of general practitioner indicates the "end of trial" for a given patient. In an intention-to-treat analysis, we studied participants aged 50+ at their first drug exposure between 1987-2018 and with more than one prescription during an initial 12-month treatment assessment period, dated at least one year after patient registration date to ensure new drug taker status. Follow-up time ranges from 1 to 31 years (median: 8). Dementia incidence is defined by the presence of at least one of the following: medical code for dementia diagnosis, dementia-specific drug prescription, linked flag from the Hospital Episode Statistics (HES) or explicitly mapped label in the Office of National Statistics (ONS) database. After exclusion of cases with dementia diagnosis prior to their first anti-diabetic prescription, with polytherapy at baseline or insulin prescription within a year, the study cohort results in 130,369 and 57,244 patients on metformin and sulfonylureas monotherapy respectively. A total of 15,089 dementia cases is observed in the sample (6.5% of metformin- vs. 11.4 % of sulfonylureas-takers), with an onset age ranging from 52 to 105 years old (median: 81).

Methods. We designed a Cox-proportional hazard model with ATE Winsorized propensity-score weighting to rebalance treatment (metformin) and control (sulfonylureas), and Inverse Probability Weighting to account for missingness. With the exception of four covariates, most features used in the logistic regression model calibrating treatment assignment are complete: age, gender, geographical region, year of first prescription and comorbidities (hypertension, cancer, CVD, CKD, COPD and cancer history). The lack of completeness is reduced and only affects the Index of Multiple Deprivation which is a proxy metric for socio-economic status (6%), five-year latest smoking status (10%), and BMI (12%) as well as two-year latest HbA1c (36%) at baseline. Furthermore, treatment assignment is independent from missingness (p -value < 0.001). This justifies multiplying ATE weights resulting from the propensity-score rebalancing strategy with inverse completeness probability weights obtained via a second logistic regression, thus forming a single weight for each complete case.

Results. Our analysis provides weak evidence for weak protective effect of metformin on dementia 1.007 (0.889-1.141), compared with sulfonylureas. The results of this double-weighting procedure are robust to both cohort composition and covariate adjustment complexity. Full adjustment with complete cases leads to an HR of 1.057 (0.898-1.245) and minor adjustments (age, sex and region) to an HR of 1.145 (1.088-1.204).

Conclusion. Whether metformin influences fundamental aging factors that underlie Alzheimer's disease and other age-related conditions, including cancer and heart disease, is still to be determined. But what longitudinal observational studies based on EHR data unlock is the possibility to uncover (non-)relationships between drug treatments and disease progression, and the case of anti-diabetics and dementia is just one of many yet to be investigated.