

Drug network classification and electronic health record analysis reveal insights into diabetes drug effects and patterns of use

Anjali Sivanandan, Jennifer L. Wilson
University of California, Los Angeles

Background

The treatment landscape changes for patients with chronic conditions, like diabetes, as new medicines are approved. This creates treatment trajectories that are distinct from the clinical trials where new drugs are originally tested, and may influence disease outcomes. Fortunately, real world data enables the study of diverse and complex treatment trajectories that complement assessment of safety and efficacy in clinical trials. Separately, protein-protein interaction (PPI) network methods predict patterns among drugs across traditional drug classes, and we have previously shown these pathway-based drug classes to be associated with diverse patient outcomes, like drug-drug interactions and overall survival (1,2). We hypothesize that studying diabetic patient outcomes in the EHR using PPI-derived drug groups may uncover novel outcomes in the context of complex patient treatment landscapes. To pilot our approach we studied diabetes drug pathways, assessed their impact on diabetes outcomes, and characterized diabetes treatment trajectories in health records data.

Methods

We used PathFX to generate PPI-based drug groups (3). The algorithm predicted connections to 181 diabetes-related pathway phenotypes for 36 diabetes drugs. Briefly, PathFX uses the available evidence supporting PPIs around drug targets to prioritize downstream proteins and uses statistical enrichment to predict pathway phenotypes, “effects”, for a drug’s target(s). We used hierarchical clustering to generate 4 pathway-based drug classes (PPI-classes) based on their diabetes network pathways. We assessed Gene Ontology (GO) enrichment of genes in each PPI-class using GOrilla with all unique PathFX proteins as the background set (4).

First, to mimic more traditional clinical development and to understand any clinical differences in drug PPI-classes, we conducted pairwise comparisons of all drug PPI-classes by restricting our analysis to patients on a single class and analyzing outcomes in a trial-relevant time window. We used the University of California Health Data Warehouse (UCHDW), a de-identified health record in the standard OMOP Common Data Model format. Without access to OHDSI tools, we developed pipelines using OHDSI best practices. We used a standard treatment-comparator study design with 1:1 propensity score matching after removing duplicates. Exposure was defined as first exposure to a PPI-class drug, lasting at least 90 days, with a prior diabetes diagnosis, and without prior exposure to the drugs in the comparator class. We collected non-diabetes diagnoses and non-study drugs that appeared in at least 10% of study patients in the year before the first drug exposure, and patient demographics (age, race, ethnicity, and gender) as covariates. We used logistic regression to ensure covariate balance and used propensity scores for patient matching. We measured the difference between an A1c measurement taken ≤ 2 weeks before the drug start date and one after ≥ 90 days on the drug. We calculated the average treatment effect on the treated (ATT) for A1c difference for each pairwise PPI-class comparison in both directions for a total of 12 comparisons.

In the ~15 years of available data, we observed that patients changed treatment throughout their disease. To better understand these patterns in diabetes, we investigated drug PPI-class trajectories for all diabetes patients in the UCHDW. We collected all diabetes drug exposures after a patient's diabetes diagnosis with ≥ 30 days exposure and binned them into 1 year intervals over 15 years. We defined 15 drug "states" to describe exposure to a single drug PPI-class or any combination of the 4 PPI-classes in 1 year intervals. We used K-Means clustering with cosine similarity to cluster patients based on their drug state over 15 years and used silhouette analysis to determine the optimal number of clusters. To characterize each patient cluster, we counted the number of patients experiencing a certain drug state per year and plotted heatmaps for each cluster.

Results

PathFX generated 254 pathway predictions connecting 36 diabetes drugs to diabetes phenotypes. Hierarchical clustering of these diabetes drugs based on their diabetes network pathways resulted in 4 PPI-classes. We then measured clinical outcomes in diabetes between drug PPI-classes in the EHR by performing pairwise comparisons of A1c difference before and after drug exposure (Table 1). While many of the comparisons had a p-value < 0.05 , they did not have a clinically meaningful effect size ($\geq 0.5\%$ change in A1c). The only comparisons that were both statistically significant with effect sizes slightly $> 0.5\%$ was between PPI-class 2 (mainly consisted of insulins) and PPI-class 3 (mainly consisted of GLP-1 analogues). However, the effect size was rather small ($\sim 0.6\%$). Overall, our results suggested that despite drug PPI-classes having unique network pathways, there was little to no difference in A1c change between drug classes when studying a time interval relevant to clinical trials.

Example Drugs	PPI Class	ATT for A1c % Difference			
		ATT-1	ATT-2	ATT-3	ATT-4
empagliflozin, glipizide	1		-0.4963 [-0.5010, -0.4917] (1.01E-09)	0.1183 [0.1166, 0.1200] (0.0049)	-0.0503 [-0.0518, -0.0488] (0.2202)
insulin glargine, insulin lispro	2	0.5765 [0.5721, 0.5808] (1.68E-13)		0.6566 [0.6516, 0.6615] (7.56E-15)	0.2672 [0.2635, 0.2709] (2.95E-04)
semaglutide, tirzepatide	3	-0.0571 [-0.0589, -0.0553] (0.1872)	-0.6735 [-0.6788, -0.6682] (5.52E-14)		-0.1735 [-0.1750, -0.1720] (1.72E-05)
metformin, tolbutamide	4	0.0755 [0.0740, 0.0770] (0.0678)	-0.3131 [-0.3170, -0.3091] (4.40E-05)	0.2074 [0.2059, 0.2090] (7.24E-07)	

Table 1. Pairwise comparisons reveal little to no significant difference in A1c difference between drug PPI-classes. Examples of drugs in each PPI-class are given in the far left column. ATT is provided for each class comparison in the format: Effect Size [95% confidence interval] (Welsh's p-value). ATTs with significant p-values (< 0.05) and clinical relevant effect size (A1c difference $\geq |0.5\%|$) are highlighted in gray.

Through the process of developing our cohorts for the A1c comparison, we found that diabetes treatment was often complex and included drug switching and polypharmacy. We characterized the "drug state" (single or combination PPI-class exposure) in 1 year intervals over 15 years for all patients in the UCHDW and generated 9 patient clusters based on their drug states over time. Analysis of the patient clusters revealed large variation in diabetes treatment trajectory. Figure 1 shows patient clusters 2 and 5 as examples to illustrate the following patterns we discovered through clustering: (1) the majority of patients were exposed to one drug PPI-class at a time as opposed to drug PPI-class combinations, (2) there were large gaps in treatment and clusters differed based on when those gaps

occurred, and (3) drug class combinations appeared more often in later years. Additionally, several patient clusters demonstrated a pattern of starting on either drug PPI-classes 1, 2, or 4, then later switching to drug PPI-class 3. Drug PPI-class 3 consisted mostly of the newer GLP-1 analogues which may indicate a changing treatment landscape due to new medicines.

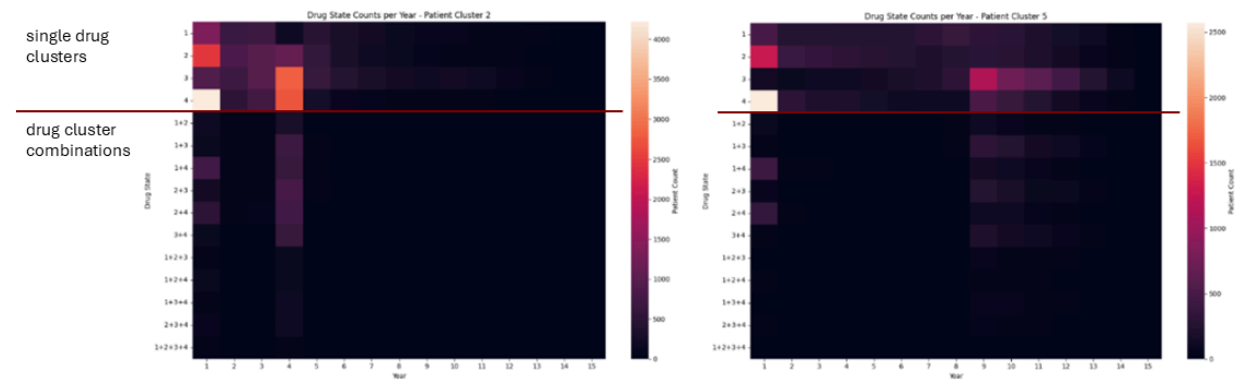


Figure 1. Patients clustered by diabetes drug trajectories reveal distinct patterns in treatment over time. Heatmaps show the number of patients in a certain drug state (y-axis) per year (x-axis) for patient clusters 2 (left) and 5 (right).

Finally, we conducted GO enrichment on network genes for each drug PPI-class and found that the drug classes were enriched for lipid, immune, vascular, and neural-related GO terms (Fig. 2). These terms are related to known longer-term comorbidities of diabetes such as angiopathies or cholesterol related conditions. While A1c difference was largely similar between PPI-classes, because drug groups were enriched for terms related to distinct comorbidities, we hypothesize that these drug classes could be associated with distinct long-term outcomes.

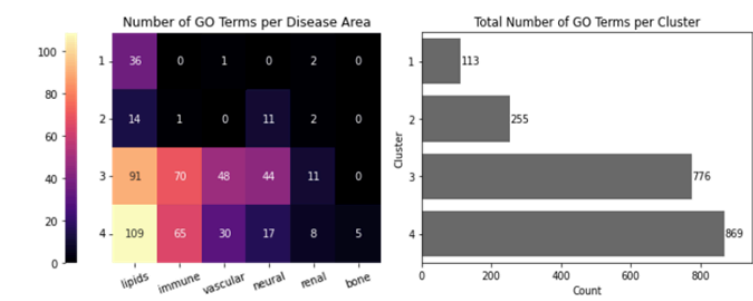


Figure 2. Diabetes drug PPI-classes are enriched for GO terms related to disease areas with associations to diabetes. Heatmap (left) shows the number of GO terms related to a certain disease area (x-axis) for each drug PPI-class (y-axis). Bar chart (right) shows the number of total GO terms per cluster.

Conclusion

In this study, we combined pathway analysis of PPI networks with EHR studies to learn more about diabetes drug effects on short-term clinical outcomes and their patterns of use. Our pathway analysis revealed that diabetes drugs grouped into distinct pathway-based classes and EHR studies suggested the equivalence of drug PPI-classes on diabetes outcomes, measured through A1c difference. We discovered patterns in diabetes drug trajectories which included drug “holidays”, use of single agents at a time, and switching between drug PPI-classes. In future studies, we plan to compare clinical outcomes among patients with similar trajectories. Studying long term and multi-drug exposures may provide important insight into long term treatment of diabetes, a chronic and progressive disease.

References

1. Wilson JL, Steinberg E, Racz R, Altman RB, Shah N, Grimes K. A network paradigm predicts drug synergistic effects using downstream protein–protein interactions. *CPT Pharmacomet Syst Pharmacol*. 2022 Nov;11(11):1527–38. doi:10.1002/psp4.12861
2. Alidoost M, Huang JY, Dermentzaki G, Blazier AS, Gaglia G, Hammond TR, et al. Uncovering New Therapeutic Targets for Amyotrophic Lateral Sclerosis and Neurological Diseases Using Real-World Data. *Clin Pharmacol Ther*. 2025 Jul;118(1):242–51. doi:10.1002/cpt.3682
3. Wilson JL, Racz R, Liu T, Adeniyi O, Sun J, Ramamoorthy A, et al. PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development. *PLoS Comput Biol*. 2018 Dec;14(12):e1006614. doi:10.1371/journal.pcbi.1006614 PubMed PMID: 30532240; PubMed Central PMCID: PMC6285459.
4. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009 Feb 3;10:48. doi:10.1186/1471-2105-10-48 PubMed PMID: 19192299; PubMed Central PMCID: PMC2644678.