Predicting Chronic Comorbid Conditions of Type 2 Diabetes in Newly-Diagnosed Diabetic Patients

Narges Razavian\textsuperscript{PhD,NYU}, Saul Blecker\textsuperscript{MD,NYU}, Ann Marie Schmidt\textsuperscript{MD,NYU}, Aaron Smith-McAllen\textsuperscript{PhD,IBC}, Somesh Ngiam\textsuperscript{PhD,IBC}, David Sontag\textsuperscript{MD,NYU}

Background\textsuperscript{1}

Comorbid Complications:\n- Myocardial infarction, Ventricular fibrillation, cardiac arrest, ischemic heart disease, Heart failure
- Cerebrovascular: Subarachnoid, Intracerebral, Nontraumatic extradural, Subdural hemorrhaage, Occlusion of cerebral arteries, Stroke, Transient cerebral ischemia, Occlusion and stenosis of precentral arteries, Other and ill-defined cerebrovascular disease
- Renal: Hemodialysis, Peritoneal dialysis, Acute or chronic kidney failure, kidney transplant
- Eye: Background retinopathy, Macular edema, Proliferative retinopathy, Retinal detachment or Vitreous hemorrhage, Operations of vitreous

Methods

Our prediction task:
- Given Patient Variables up to T2D, can we predict the risk of 4 comorbidities within 1 to 5 years after the onset?
- What variables are predictive?
- What variables differ between genders?

Our cohort:
- 4.1 million insurance beneficiaries of age 218 years between 2005 and 2013. We only included patients newly-diagnosed with T2D.
- Continuous enrollment for 6 months pre- and 5 years post-onset.

- Without the comorbid condition diagnosis before T2D.

Prediction Framework

Patients are only included if they are continuously enrolled throughout the patient enrollment window, and do not have comorbidity onset before the prediction condition window. We use 67% of the cohort to train our system, and evaluate the quality on 33% remaining patients data not used during training.

Top Lab Variables
- Alanine aminotransferase - high in the past 6 months
- Urea nitrogen - high in the entire past history
- Creatinine - high in the entire past history
- Hypertension

Top Lab factors
- Platelets - low in the past 2 years
- Thyrotopin - high in the entire past history
- Hemoglobin – low in the entire past history
- Urea nitrogen – high in the past 24 months
- Sinoatrial node dysfunction

Diabetes known risk factors
- All variables except ICD-9 diagnosis evaluated in 6 months, 2 years and entire history prior to T2D onset.

Software for patient variable extraction on public CMS data:
https://github.com/nyuvis/patient-viz
For software package including patient visualization visit:
https://github.com/nyuvis/patient-viz/tree/master/Feature_extraction
For additional information visit razavian@cs.nyu.edu or demirski@cs.nyu.edu

Results

Comorbidity Total included in training Total Positive cases during training
Cardiovascular 2,426 2,678 32,937
Cerebrovascular 1,948 2,660 1,043
Renal 32,498 3,560
Eye 3,438 1,780

Prediction Quality and Risk Factors

Comorbidity Cardiovascular Cerebrovascular Renal Eye

Comorbidity Events Predictive Variables (Total 317 variables)*

Eye Complication Predictive Variables (Total 236 variables)

Eye Complication Factors

Comorbidity Factors

Renal Events Predictive Variables (Total 272 variables)

Diabetes variables

Machine Learning Method

Prediction model: L1-regularized Logistic Regression

- Total Variable Size: 42,000 with diagnosis variables, and 27,000 without.

- All variables except ICD-9 diagnosis evaluated in 6 months, 2 years and entire history prior to T2D onset.

- Prediction model prediction can be developed elsewhere and plugged into administrative or EHR infrastructure.

For software package including patient visualization visit:
https://github.com/nyuvis/patient-viz/tree/master/Feature_extraction
For additional information visit razavian@cs.nyu.edu or demirski@cs.nyu.edu

Conclusions

- Type 2 Diabetes and its comorbidities increasingly strain healthcare sectors and societies worldwide, and population-level early intervention to avoid severe complications is necessary.
- Using longitudinal patient data, driven data approaches enable automatic risk factor identification and risk assessment, without the need for unaffordable, large scale screening tests.
- Using our method we can predict risk of each comorbidity for newly diagnosed T2D patients, and potentially novel risk factors, significantly differing in different subpopulations.
- Next step includes intervention allocation based on the risk prediction.

References

1. Razavian N, Davis C. Predicting Chronic Comorbid Conditions of Type 2 Diabetes in Newly-Diagnosed Diabetic Patients. 2016
4. Sontag S, Sontag D, Batalov S, Sandhu N, Davis C. The use of automatic data curation techniques to support the development of a comorbidities risk prediction model. PRIMING 2016
5. Sontag S, Sontag D, Batalov S, Sandhu N, Davis C. The use of automatic data curation techniques to support the development of a comorbidities risk prediction model. PRIMING 2016