Evaluating Clinical Guidelines without Randomization

Tony Liu [1] Patrick Lawlor [2] Lyle Ungar [1] Konrad Kording [1]

[1]University of Pennsylvania [2]Children's Hospital of Philadelphia

Introduction

Regression discontinuity design (RDD) is a method for estimating causal effects when randomized trials may not be feasible. RDD uses the fact that individuals close to a threshold for treatment have about the same probability of being just above or just below the cut-off, resulting in treatment assignment that is effectively random. Thus, it can be used to evaluate the causal effect of clinical guidelines when only observational data is available.

Objective: Evaluate Diabetes Diagnostic Criteria

- We use RDD to evaluate diabetes diagnostic criteria using claims data
- Our continuous variable: a patient's first recorded A1C measurement
- American Diabetic Association guidelines first set in 2010 [1] determined patients with A1C $\geq 6.5\%$ should be diagnosed with diabetes
- This change in guidelines resulted in a jump in diabetes diagnoses at the 6.5% level in records after 2010 (Figure 1)

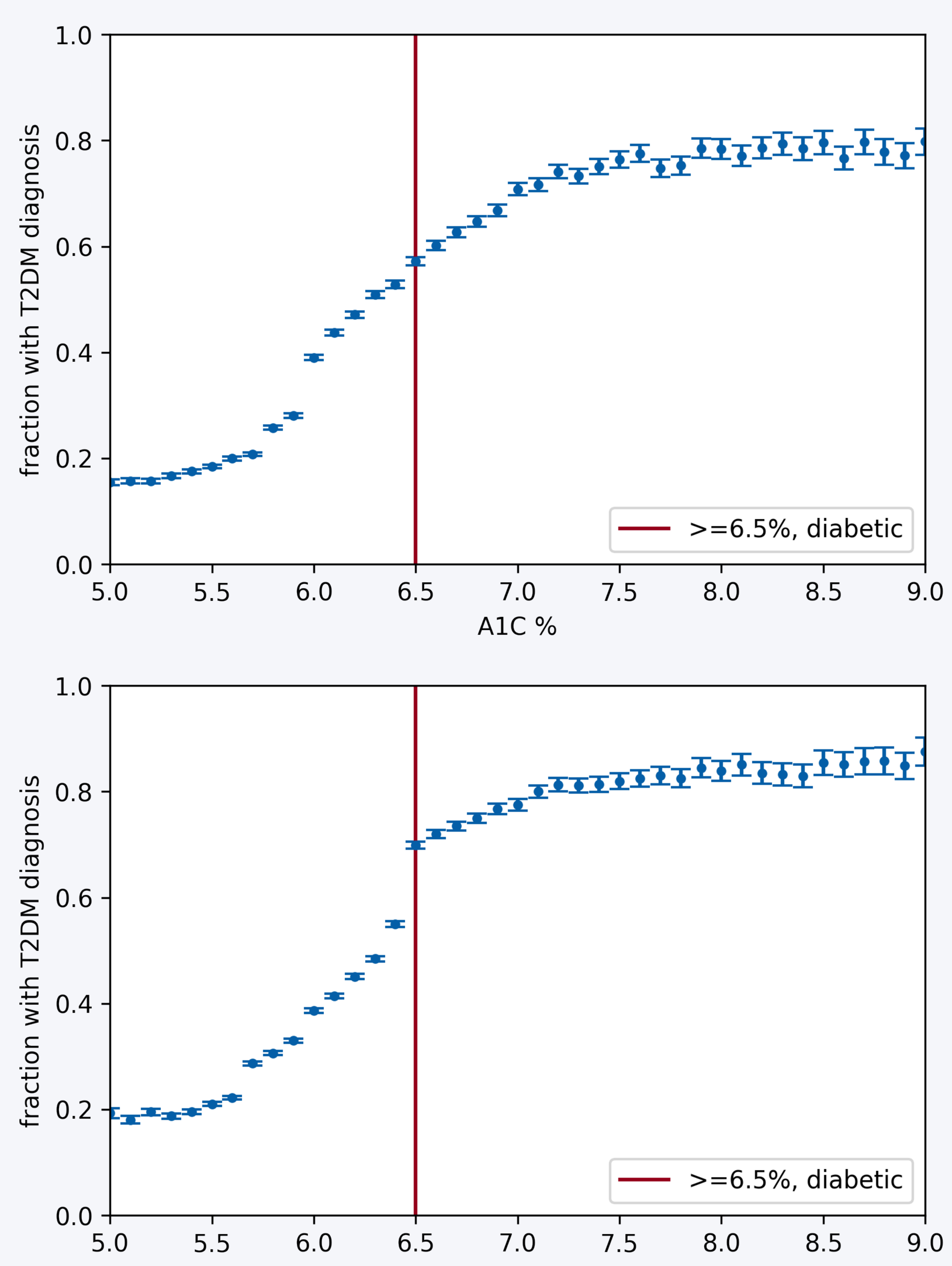

Figure 1. Diabetes diagnoses across A1C levels for patient records before 2010 (top) and after (bottom).

Claims Data for Diabetes

- Sourced from Optum's de-identified Clinformatics® Data Mart Database, 360k unique patients
- **Inclusion criteria:** First A1C measurement for patients after 2010
- **Exclusion criteria:** No prior prescription of diabetic drugs or prior diagnosis of type II diabetes
- Target outcome is heart attack incidence one year after A1C measurement

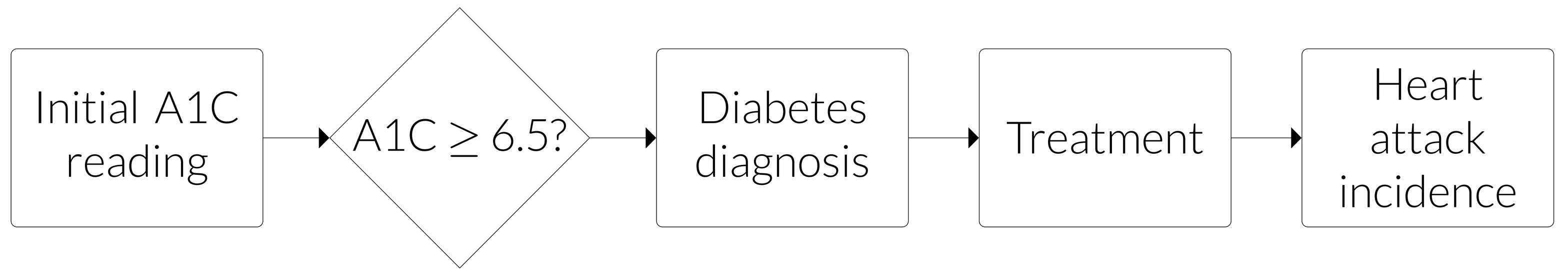

Figure 2. The causal chain of our RDD analysis.

Methods: Fuzzy Regression Discontinuity Design

We perform feasibility and covariate balance tests to validate the RDD:

- Is the running variable continuous around the threshold (Figure 3)?
- Is age, gender, and socioeconomic status balanced around the threshold?

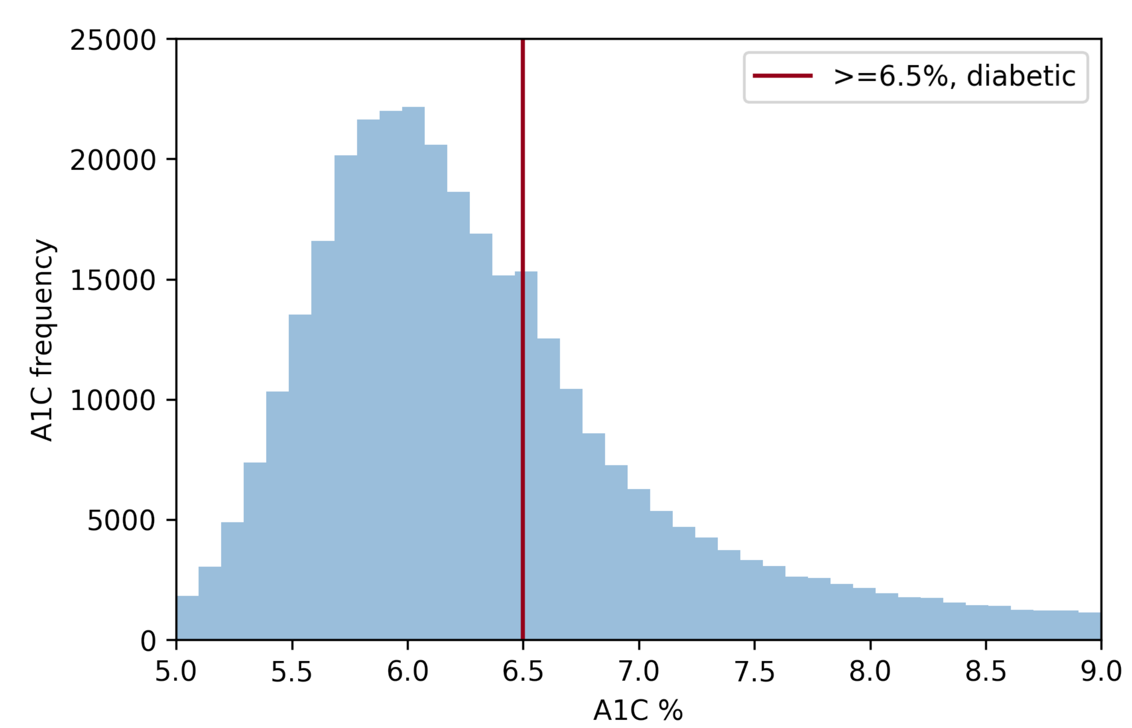

Figure 3. Density of A1C measures for our target population.

We use the "fuzzy" two stage regression with T as our treatment indicator [2]:

$$D_{\text{diagnosis}} = \alpha_0 + \alpha_1 T + \alpha_2(1-T)(X_{\text{A1C}} - 6.5) + \alpha_3(T)(X_{\text{A1C}} - 6.5) + \epsilon$$
$$Y_{\text{heart attack}} = \beta_0 + \beta_1 D + \beta_2(1-T)(X_{\text{A1C}} - 6.5) + \beta_3(T)(X_{\text{A1C}} - 6.5) + \nu$$

We also evaluate patient's *residualized risk* for heart attack:

1. $P(Y_{\text{heart attack}} = 1) = \hat{f}(X_{\text{age}}, X_{\text{race}}, X_{\text{gender}}, X_{\text{is smoker}}, X_{\text{HDL}})$
2. $Y_{\text{residual}} = Y_{\text{heart attack}} - \hat{f}$

1. model patient's probability of a heart attack through an ML model $\hat{f}$
2. use RDD to evaluate the effect of diabetes diagnosis on the residual

Results

We find a 12% increase in diabetes diagnoses at the A1C threshold as well as a 0.16% decrease in heart attack rates, both significant at $p < 0.001$ (Figure 4).

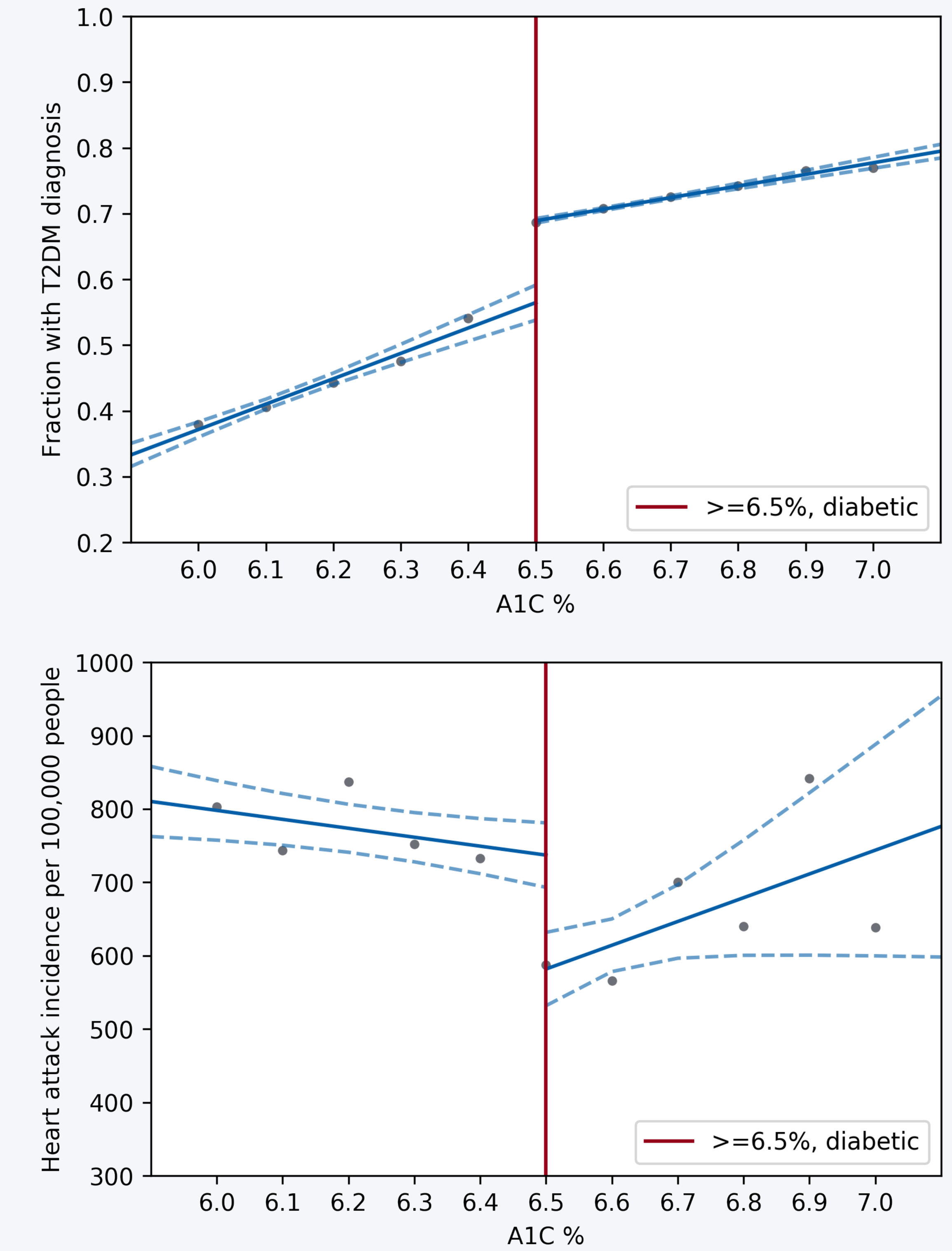

Figure 4. Our two stage regression for diabetes diagnosis (top) and heart attacks (bottom).

Conclusions and Future Work

- **Our RDD analysis shows that the 6.5% A1C guideline reduced heart attack incidence by 155 people per 100k at the threshold**
- An RDD approach works well on large, observational claims data; estimating the residualized risk increases statistical power
- Future RDD targets include lipid screening and ER management of infant fever
- Future research will also involve optimization of the threshold itself [3]

TL;DR Regression discontinuity design can be used to estimate causal effects without interventions. Here, we apply RDD to diabetes diagnostic criteria and find that the 2010 A1C guidelines reduce heart attack incidence at the threshold. RDD analyses show promise for causal inference particularly in large observational datasets when controlling on observables or conducting an experiment is not feasible.

References

[1] American Diabetes Association et al. Standards of medical care in diabetes—2010. *Diabetes care*, 33(Supplement 1):S11–S61, 2010.

[2] David S Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355, 2010.

[3] Ioana Elena Marinescu, Sofia Triantafillou, and Konrad Kording. Regression discontinuity threshold optimization. *Available at SSRN 3333334*, 2019.