Assessing Variation in First-Line Type 2 Diabetes Treatment across eGFR Levels and Providers

Christina X Ji, PhD[a]*, Saul Blecker, MD[b], Michael Oberst, PhD[a], Ming-Chieh Shih, MD PhD[a], Leora I Horwitz, MD[b], David Sontag, PhD[a]

[a]MIT CSAIL and IMES, 32 Vassar Street, Cambridge, MA 02142

[b]Department of Population Health, NYU Grossman School of Medicine, 550 First Avenue, New York, NY 10016 and Department of Medicine, NYU Grossman School of Medicine, 550 First Avenue, New York, NY 10016

*Corresponding author: cji@alum.mit.edu

Keywords: Epidemiology, Guideline Adherence, Metformin, Observational Study, Physician Preferences, Prescribing Trends, Treatment Preference, Type 2 Diabetes

Abstract

Objective: This study examines the decision between metformin and a DPP-4 inhibitor or sulfonylurea for first-line type 2 diabetes treatment. We performed a retrospective cohort study on an insurance claims dataset. Since metformin is contraindicated by kidney damage, we first examined variation across estimated glomerular filtration rate (eGFR) levels. Then, we assessed variation due to the prescribing provider.

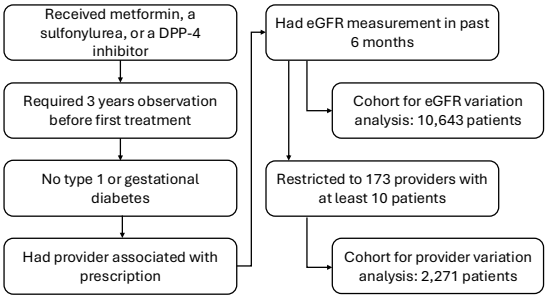
Research Design and Methods: Cohort criteria included initiation of first-line treatment, three years of prior observation, eGFR results, and no type 1 or gestational diabetes. With these 10,643 patients, we performed a chi-squared test for association between eGFR and treatment. Then, on a restricted cohort of 2,271 patients seen by 173 providers with at least 10 patients, we fitted restricted cubic spline models to predict treatment from patient characteristics (age, eGFR, sex, history of heart failure, and treatment date) and performed a generalized likelihood ratio test (GLRT) to assess whether including provider-specific random effects improved fit.

Results: Treatment choice was significantly associated with eGFR level ($p < 0.0001$). The GLRT found significant variation existed across providers even after accounting for patient characteristics ($p < 0.0001$).

Conclusions: As first-line type 2 diabetes treatments vary across eGFR levels and prescribing providers, further studies can elucidate whether such variation across providers is appropriate. Our data-driven approaches can also be applied to other treatment decisions to identify potential areas for improvement in guidelines and provider-specific treatment practices for diabetes management.

Assessing Variation in First-Line Type 2 Diabetes Treatment across eGFR Levels and Providers

1. Defined cohort with first-line type 2 diabetes treatment



2. Tested for significant treatment variation across eGFR levels

Number of patients in each eGFR category who receive each treatment class as first-line treatment

eGFR level	Metformin	DPP-4i / Sulfonylurea	Total
eGFR < 30	9	55	64
30 ≤ eGFR < 45	92	204	296
45 ≤ eGFR < 60	709	356	1,065
60 ≤ eGFR < 90	3,851	662	4,513
90 ≤ eGFR	4,178	527	4,705
Total	8,839	1,804	10,643

Chi-squared test: Is there significant association between metformin prescription and eGFR level?

Yes, there is significant variation in first-line treatment across eGFR levels ($p < 0.0001$).

3. Tested for significant treatment variation across providers

Fit generalized linear models predicting treatment Y from patient characteristics X with and without provider-specific random effects

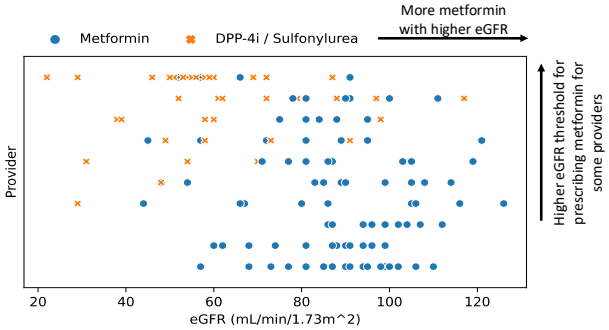
Patient characteristics X:
- Estimated glomerular filtration rate (eGFR)
- Age
- Treatment date
- Heart failure
- Sex

First-line type 2 diabetes treatment Y:
- Metformin vs Sulfonylurea / DPP-4 inhibitor

Generalized likelihood ratio test: Does data have higher likelihood under model with provider-specific random effects compared to model without random effects?

Yes, there is significant variation in first-line treatment across providers ($p < 0.0001$).

4. Visualized provider decisions with respect to eGFR



Article Highlights

Why did we undertake this study? Metformin is contraindicated when estimated glomerular filtration rate (eGFR) is low. However, there is no clear consensus around the eGFR level at which metformin should no longer be prescribed.

What is the specific question(s) we wanted to answer? Is first-line type 2 diabetes treatment significantly associated with eGFR and preferences from the prescribing provider?

What did we find? First-line treatment is significantly associated with eGFR levels. Significant variation exists across providers even after accounting for patient age, eGFR, sex, history of heart failure, and treatment date.

What are the implications of our findings? Since some variation cannot be explained by guidelines, our results suggest future studies about whether variation in first-line treatment decisions across providers is appropriate may help improvement diabetes management.

1. Introduction

Clinical practice guidelines have been developed to promote evidence-based practice in medicine and improve quality of care and clinical outcomes.[1] For instance, the American Diabetes Association has published guidelines for diabetes management.[2] Actual practice often deviates from guidelines though. This variation may be due to incomplete or unclear guidelines, barriers to adoption, different practice styles of individual providers and health systems, or other factors that influence patient-provider interactions.[3-4] For example, guidelines may not be applicable for complex patients with multiple co-morbidities. In this case, providers may deviate from the guidelines to provide more appropriate care.[5-6] In other cases though, deviation may be due to poor dissemination of guidelines or lack of resources to support adoption.[3-4] Such variability in guideline compliance may result in suboptimal quality of care or clinical outcomes for patients.

Large clinical datasets offer opportunities to empirically study clinical decision-making and practice variation across providers. In particular, models fit to real-world evidence may be better at capturing decision-making processes that are not covered by or differ from guidelines. With observational data, we can understand when variation is captured by guidelines, seems appropriate based on patient characteristics, or may be part of a consistently flawed decision-making process that needs to be remedied.[1,7] In the latter case, we must be careful when flagging a provider or a health system as an outlier that needs attention. A provider who takes on more challenging cases by seeing patients who have more co-morbidities or are more likely to have contraindications for evidence-based care should not be unduly punished. Even if their patients have worse outcomes, much of that variation may still be clinically appropriate.[8] To account for these nuances, we propose first examining variation that can be explained by patient characteristics and then accounting for these characteristics when studying variation across providers. Past works with similar approaches have examined whether the choice of physician has an impact on cancer treatment and hospice enrollment.[9-11]

Metformin is the guideline-recommended initial treatment for type 2 diabetes. It is effective at improving glycemic control and promoting weight loss without increasing risk of hypoglycemia.[12] The medication is also cost-effective. American Diabetes Association (ADA) guidelines recommend starting metformin unless "there are contraindications."[12] However, these guidelines are only suggestions for providers to consider. The provider ultimately decides which treatment to prescribe. For example, ADA guidelines currently suggest that metformin is safe in chronic kidney disease (CKD) so long as eGFR is at least 30 mL/min/1.73m^2.[2] However, this has not always been the case. For patients with mild renal impairment, defined as eGFR levels between 30 and 60, metformin only started being recommended in 2016 after several large-scale cohort studies found that metformin did not increase the risk of lactic acidosis in these patients.[13-15] Because there is uncertainty around the specific eGFR threshold at which the risk of metformin in CKD outweighs the benefits in diabetes, the decision to prescribe another medication because metformin is contraindicated falls on the individual provider, leading to practice variation.

Using a large insurance database, we examined how providers decided to initiate metformin versus another diabetes medication for patients with type 2 diabetes. First, we assessed whether there is significant dependence between the choice to initiate metformin and eGFR levels. This allowed us to verify that the patient characteristics used to define treatment guidelines indeed play a role in real-world decision-making. Second, we modelled how eGFR and other factors influenced the likelihood of prescribing metformin in general and for each individual provider. We then used these models to assess whether significant variation exists across providers. This part of our analysis established the existence of variation that cannot be explained by treatment guidelines and relevant patient characteristics. We hope this work will encourage future studies that examine how this variation affects patient outcomes and, if needed, how to issue better guidance to reduce harmful variation.

2. Research Design and Methods

2.1 Outcome Definition

The outcome of interest was a binary variable for whether the patient was prescribed metformin versus a sulfonylurea or a DPP-4 inhibitor. Diabetes drugs were defined as drugs that are descendants of the "drugs used in diabetes" concept in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) and contain at least one of the following ingredients: metformin, sitagliptin, vildagliptin, saxagliptin, linagliptin, gemigliptin, anagliptin, teneligliptin, acetohexamide, carbutamide, chlorpropamide, glycyclamide, tolcyclamide, metahexamide, tolazamide, tolbutamide, glibenclamide, glyburide, glibornuride, gliclazide, glipizide, gliquidone, glisoxepide, glyclopyramide, or glimepiride.

2.2 Patient Characteristics

Our primary exposure was eGFR, which was based on the last measurement obtained in the 6 months prior to the initial medication fill. Estimated glomerular filtration rate (eGFR) measurements were defined by all lab concepts that started with "Glomerular filtration rate/1.73 sq M". This included both CKD-EPI that was commonly used through 2017 and MDRD that became more predominant starting in 2018. Because some eGFR measurements are race-adjusted, for Black or African American patients, we prioritized measurement concepts that specified "among Blacks" over concepts that did not specify race and only used concepts that specified "among non-Blacks" if no other measurements were available. For patients of other races or without race specified, the priorities of concepts that specified "among Blacks" and "among non-Blacks" were flipped. If multiple measurement values at the same priority level exist on the most recent measurement date, the maximum value was taken.

Other variables included age, sex, and the presence of a diagnosis for heart failure. Heart failure was defined by concepts that are descendants of the SNOMED concept "Heart failure". These variables were chosen as they may contribute to variation in decision making related to use of metformin.[16-17] We also included treatment date as a covariate since metformin usage increased over time, particularly when metformin was no longer contraindicated for patients with moderate renal failure in 2016.[13-15]

2.3 Cohort Criteria

We performed a retrospective cohort study using insurance claims data from a large insurance provider in the northeast United States, spanning from 2012 to 2021. The database includes laboratory test results when the insurance company has a contract with the laboratory center filling the claim.

Patients who received an initial medication fill for metformin, a sulfonylurea, or a dipeptidyl peptidase-4 (DPP-4) inhibitor were included in the cohort. Patients who were prescribed multiple treatment classes on the initial treatment date, did not have exactly one provider associated with the prescription, or did not have an eGFR measurement in the past 6 months were excluded. To focus on type 2 diabetes, we excluded patients who had a type 1 diabetes, gestational diabetes, neonatal diabetes, or pregnancy-related code at any point in their history. To ensure that the prescription is the first diabetes drug a patient receives, the patient must be observed for at least 95% of the 3 years preceding the treatment date. Although this requirement significantly reduced the cohort size, it is important for guaranteeing we are indeed studying first-line treatment decisions.

For the analysis on provider-based variability, we additionally limited our cohort to patients who were seen by providers who had at least 10 patients who met the inclusion and exclusion criteria. Treatment decisions were attributed to the provider listed on the insurance claim for the medication fill. Figure 1 is a flowchart of the cohort criteria and the number of patients after each criterion is applied.

2.4. Statistical Approach to Test for Variation across eGFR Levels

Because first-line diabetes treatment guidelines are heavily dependent on eGFR, we first verified that the general treatment policy observed in real-world data aligned with guidelines. We created the following 5 eGFR categories based on the definitions for the different stages of CKD: below 30, 30-44, 45-59, 60-89, and at least 90.[7] We calculated the number of patients in each eGFR category who were prescribed each medication class. These counts were used in a chi-squared test to determine if there was a difference in rates of metformin use by eGFR category.

2.5. Statistical Approach to Test for Variation across Prescribing Providers

To assess whether providers have different treatment policies with respect to eGFR, we needed to account for how eGFR and other factors affected the likelihood of prescribing metformin. This is important for two reasons: 1) Some providers see more challenging patients.[8] For example, patients with severe CKD are more likely to seek out specialists than general practitioners for treatment. Modeling patient features accounts for these differences in patient populations across providers. 2) We are interested not just in how providers prescribe the treatments at different rates but in how providers differ in how they take these factors into account when making treatment decisions. Thus, these features need to be included when predicting the likelihood of prescribing metformin.

To account for these features when modeling the treatment decisions, we first found the best model in each of the following two model families: 1) The first model family included

generalized linear models with restricted cubic spline features for eGFR, age, and treatment date and binary indicators for heart failure and sex. The cubic spline features allow us to model non-linear relationships and avoid extrapolation issues.[18] For eGFR, age, and treatment date, we allowed up to 4 knots set at the quantiles suggested by Harrell et al.[18] 2) The second model family included generalized linear models with the previously mentioned features, random intercepts, and random slopes for the eGFR features. The random intercepts and random slopes account for differences in how providers account for eGFR and weigh the risks and benefits of each treatment when making decisions.

We performed a generalized likelihood ratio test (GLRT) to assess whether the data has a significantly higher likelihood under the model with provider-specific random effects.[19-20] The model families with and without provider-specific random effects are defined by 18 and 12 parameters, respectively, so the G-statistic follows a χ^2 distribution with 6 degrees of freedom under the null hypothesis. By evaluating the likelihood only at the observed samples, this test focused on decisions for the types of patients who were actually seen by each provider.

The method we propose joins a large body of work on identifying outlying providers, including mixture models of provider effects, hierarchical and Bayesian approaches to modeling random effects, and Markov chain Monte Carlo simulations of normal behavior.[21-25] Unlike these prior works, our method tests whether variation exists across all providers rather than whether a single provider is an outlier. This is particularly useful when there are few samples per provider. While there may be insufficient power to identify individual providers as outliers, the total number of samples across all providers may provide sufficient power to identify whether variation exists across all providers.

As an additional analysis, we also examined whether any individual provider deviated significantly from the average treatment policy. For this part, we conducted separate GLRTs using only the samples from patients seen by a particular provider. We used the Benjamini-Hochberg procedure to keep the expected false discovery rate at 5%.[26] In contrast to the simulation approach for identifying outlying providers proposed by Ohlssen et al,[21] our method does not exaggerate the significance of a small number of decisions from a single provider.

2.6. Visualizations

To visualize how the treatment decisions made by each provider vary with eGFR, we first plotted the treatments observed for patients in the dataset against each eGFR value, with the decisions for each provider in a separate row. Then, we plotted the treatment policies learned by the models fit for the GLRT. The model predicts the likelihood of prescribing metformin given eGFR level, age, sex, treatment date, and history of heart failure. We held all features besides eGFR constant, so the policy shown is for a 50-year-old female given treatment on 2014-05-25 with no history of heart failure. These features were chosen arbitrarily to fall within the observed population. To create this second plot, we varied the eGFR value from the minimum to the maximum observed (3 to 155) and plotted the predicted likelihood of prescribing metformin in general and for each provider. Unlike the observed decisions in the first plot, the second figure shows the policy a provider may follow at any eGFR value.

2.7 Data and Resource Availability

The data that support the findings of this study were obtained from Independence Blue Cross but restrictions apply to the availability of these data, which were used under license for the current study and therefore are not publicly available. This research was ruled exempt by MIT's IRB (protocol E-4025).

Data was extracted using SQL from a postgres database. Models were fit using the glm and glmer packages in R. All other statistical procedures were carried out using Python. All code is publicly available at https://github.com/clinicalml/t2dm_provider_variation_analysis.

3. Results

3.1. Descriptive Statistics and Variation across eGFR levels

A total of 10,643 patients met the inclusion and exclusion criteria. Metformin was prescribed to 83.0% of these patients, and the other 17.0% were prescribed a sulfonylurea or a DPP-4 inhibitor. The mean age was 60.6 (standard deviation 12.9). 52.8% of patients were male. 6.7% of patients had heart failure in the past 730 days. Metformin was prescribed for only 63.9% of patients who had heart failure, compared to 84.4% of patients who did not have heart failure. The mean eGFR level was 84.0, with a standard deviation of 20.5.

As shown in Table 1, both categories of medications were prescribed across the range of eGFR values. However, there was a noticeable and significant increase in the prevalence of metformin prescriptions as eGFR values increased ($p<0.0001$ from chi-squared test). Only 14.1% of patients with eGFR less than 30 were prescribed metformin. This rate increased to 31.1% among patients with eGFR 30-44 and 66.6% among patients with eGFR 45-59. For patients whose eGFR measurements indicated no sign of kidney damage, metformin prescription rates were 85.3% among patients with eGFR 60-89 and 88.8% among patients with eGFR ≥ 90.

3.2. Variation across Providers

There were 173 providers who saw at least 10 patients in the dataset. Among these 173 providers, the mean number of patients per provider was 13.1 with a range of 10 to 41. The total number of patients in this restricted cohort was 2,271. The mean per-provider metformin prescription rate was 82.2% with a range from 22.7% to 100.0%. The maximum likelihood model with and without random effects both had 4 knots each for eGFR, age, and treatment date. The GLRT comparing the data likelihood under the model with random effects and the model without random effects indicated significant variation in metformin usage across providers even when accounting for eGFR, age, sex, prior heart failure, and treatment date (G-statistic 387.3, $p<0.0001$).

While there was variation across all providers, we were not able to identify any single provider as an outlier because each provider only had a few samples. The smallest p-value for a single provider was 0.059.

Figure 2 illustrates that metformin use increased as eGFR increases. However, there was heterogeneity among providers. Most of this variation among providers occurred at low eGFR values. Some providers still prescribed metformin when eGFR was low, while other providers did not use metformin even when eGFR was high. Providers at the top of the figure prescribed no metformin. Those at the bottom prescribed entirely metformin.

The black line in Figure 3 that depicts the model learned without provider-specific random effects illustrates how the metformin prescription probability started around 20% for small eGFR values and increased to around 90% as eGFR levels started to indicate normal kidney function. The blue lines in Figure 3 depict the policies learned for each provider. Again, when eGFR was low, some providers still prescribed metformin with high probability, while other providers were less inclined to give metformin. Other values for patient age, sex, and treatment date yielded similar plots.

4. Conclusions

Using a large insurance claims dataset, we studied how the choice of first-line treatment for type 2 diabetes varies with different levels of kidney damage and with prescribing provider. The 83.0% metformin prescription rate in our cohort is fairly consistent with the 58-77% metformin use rate found in prior studies.[27-29] We also observed that metformin was prescribed significantly more frequently among patients with higher eGFR levels, as would be expected based on treatment guidelines. However, eGFR level, age, sex, history of heart failure and treatment date did not fully explain variation in metformin prescription. We also found significant variation across providers. This suggests that if a patient sees different providers, they may receive different care, be prescribed different treatments, and ultimately have different outcomes.

The degree of provider level variation is particularly interesting because of its novelty and the potential for interventions to improve patient care. Most prior research in this area relied on self-reported interviews and surveys, which may not always reflect clinical practice.[30] One study demonstrated that providers are not always aware of the appropriate timing for and contraindications to metformin.[31] Another survey found that only a third of providers are aware that the eGFR cutoff typically used for prescribing metformin is 30.[16] Past studies related to variation in clinical practice for type 2 diabetes focused on process-of-care indicators. These procedures included annual assessments of labs such as estimated glomerular filtration rate (eGFR) and hemoglobin A1c (HbA1c), annual examinations for diabetic retinopathy or neuropathy, and self-management training from nurses.[32-34] These studies found that poor adherence to these process-of-care indicators was associated with worse patient-reported outcomes.

To our knowledge, this is the first study that uses a large observational dataset to empirically assess how first-line type 2 diabetes treatment varies across providers while accounting for patient characteristics. Using the observational dataset allowed us to evaluate variation with the decisions that were actually made.

Because this study demonstrated that variation exists across providers who prescribe first-line type 2 diabetes treatments, it raises the question of whether this variation needs to be addressed. If the variation is inappropriate, how can the negative effects be mitigated? A non-specific intervention would be to provide brief education about first-line diabetes treatments to providers whose treatment decisions do not align with the guidelines related to eGFR and metformin. More data-driven and individualized interventions such as reviewing provider practice patterns to identify areas for improvement or imposing financial penalties for unwarranted variation would require accurate assessment of the performance of individual clinicians, identification of outlying providers, and good intervention design.[35-37] For such interventions, there must be sufficient evidence based on decisions for a large number of patients that the provider is indeed making poor treatment decisions to avoid incorrectly penalizing providers.[38-39]

Our study should be interpreted in the context of its limitations. First, this study was limited to a single insurance provider whose beneficiaries are primarily located in the northeast United States. As a result, the findings may not be generalizable. Second, while our models included several patient characteristics, they may not have accounted for all patient-level factors that contribute to treatment decisions. Third, while we chose to evaluate the performance of individual providers, variation may also be studied at the level of provider groups or health care facilities.[40]

The statistical approach in this study is more broadly applicable to other treatment decisions. The models in our method can be built with relevant patient characteristics for those decisions. Identifying variation in treatment practices that cannot be explained by guidelines opens the door to improvements in guidelines and practices for diabetes management.

Acknowledgments

Personal Thanks: We would like to thank Rebecca Boiarsky and Justin Lim for setting up the databases used in this study. We would also like to thank James Denyer, Aaron Smith-McLallen, Stephanie Gervasi, and the rest of the data science group at Independence Blue Cross for providing the dataset.

Funding and Assistance: This work was supported in part by Independence Blue Cross, Office of Naval Research Award No. N00014-21-1-2807, and the LEAP program from the Ministry of Science and Technology in Taiwan.

Conflict of Interest: There are no competing interests.

Author Contributions and Guarantor Statement: Christina X Ji: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; Saul Blecker: Conceptualization, Validation, Writing – original draft, Writing – review & editing; Michael Oberst: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization; Ming-Chieh Shih: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization; Leora I Horwitz: Conceptualization, Validation, Writing – review & editing; David Sontag: Conceptualization, Funding acquisition, Supervision

References

1. Mercuri M, Gafni A. Examining the role of the physician as a source of variation: Are physician-related variations necessarily unwarranted? *J Eval Clin Pract*. 2018;24(1):145-151. doi:10.1111/jep.12770
2. American Diabetes Association. Standards of Medical Care in Diabetes—2021. 2021;44(Supplement 1).
3. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282(15):1458-65. doi:10.1001/jama.282.15.1458
4. Grol R, Dalhuijsen J, Thomas S, et al. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ*. 1998;317(7162):858-861.
5. Giugliano D, Esposito K. Clinical inertia as a clinical safeguard. *JAMA*. 2011;305(15):1591-2. doi:10.1001/jama.2011.490
6. Safford MM, Shewchuk R, Qu H, et al. Reasons for not intensifying medications: differentiating "clinical inertia" from appropriate care. *J Gen Intern Med*. 2007;22(12):1648-55. doi:10.1007/s11606-007-0433-8
7. Shashar S, Codish S, Ellen M, et al. Determinants of medical practice variation among primary care physicians: protocol for a three phase study. *JMIR Res Protoc*. 2020;9(10):e18673. doi:10.2196/18673
8. Landon BE, O'malley AJ, Keegan T. Can choice of the sample population affect perceived performance: implications for performance assessment. *J Gen Intern Med*. 2010;25(2):104-109. doi: 10.1007/s11606-009-1153-z.
9. Lipitz-Snyderman A, Sima CS, Atoria CL, et al. Physician-driven variation in nonrecommended services among older adults diagnosed with cancer. *JAMA Intern Med.* 2016;176(10):1541-1548. doi: 10.1001/jamainternmed.2016.4426.
10. Hawlet ST, Hofer TP, Janz NK, et al. Correlates of between-surgeon variation in breast cancer treatments. *Med Care.* 2006 Jul;44(7):609-16. doi: 10.1097/01.mlr.0000215893.01968.f1.
11. Obermeyer Z, Powers BW, Makar M, et al. Physician characteristics strongly predict patient enrollment in hospice. *Health Affairs.* 2015;34(6):993-1000. doi: 10.1377/hlthaff.2014.1055.
12. Choi JG, Winn AN, Skandari MR, et al. First-line therapy for type 2 diabetes with sodium-glucose cotransporter-2 inhibitors and glucagon-like peptide-1 receptor agonists: a cost-effectiveness study. *Ann Intern Med*. 2022;175(10):1392-1400. doi:10.7326/m21-2941
13. Engler C, Leo M, Pfeifer B, et al. Long-term trends in the prescription of antidiabetic drugs: real-world evidence from the Diabetes Registry Tyrol 2012-2018. *BMJ Open Diabetes Research & Care.* 2020;8:e001279. doi:10.1136/ bmjdrc-2020-001279.
14. Tanner C, Wang G, Liu N, et al. Metformin: time to review its role and safety in chronic kidney disease. *Medical Journal of Australia.* 2019;211(1):37-42.
15. Imam TH. Changes in metformin use in chronic kidney disease. *Clinical Kidney Journal.* 2017;10(3):301-304.
16. Goldberg T, Kroehl ME, Suddarth KH, et al. Variations in metformin prescribing for type 2 diabetes. *The Journal of the American Board of Family Medicine*. 2015;28(6):777-784. doi:10.3122/jabfm.2015.06.150064
17. Pilla SJ, Segal JB, Alexander GC, et al. Differences in national diabetes treatment patterns and trends between older and younger adults. *J Am Geriatr Soc*. 2019;67(5):1066-1073. doi:10.1111/jgs.15790

18. Harrell FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. *Springer*. 2015.
19. Neyman J, Pearson ES. On the problem of most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Seria A, Containing Papers of a Mathematical or Physical Character.* 1933;231(694-706):289-337.
20. Vonesh EF, Chinchilli VM, Pu K. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*. 1996;572-587.
21. Ohlssen DI, Sharples LD, Spiegelhalter DJ. A hierarchical modelling framework for identifying unusual performance in health care providers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2007;170(4):865-890. doi:https://doi.org/10.1111/j.1467-985X.2007.00487.x
22. Marshall EC, Spiegelhalter DJ. Comparing institutional performance using Markov chain Monte Carlo methods. *Statistical analysis of medical data: new developments.* 1998;229-249.
23. Normand SLT, Glickman ME, Gatsonis, CA. Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association.* 1997;92(439):803-814.
24. Austin PC. A comparison of Bayesian methods for profiling hospital performance. *Medical Decision Making.* 2002;22(2):163-172.
25. Guthrie B, Donnan PT, Murphy DJ, et al. Bad apples or spoiled barrels? Multilevel modelling analysis of variation in high-risk prescribing in Scotland between general practitioners and between the practices they work in. *BMJ Open*. 2015;5(11):e008270. doi:10.1136/bmjopen-2015-008270
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (methodological)*. 1995;57(1):289-300.
27. Berkowitz SA, Krumme AA, Avorn J, et al. Initial choice of oral glucose-lowering medication for diabetes mellitus: a patient-centered comparative effectiveness study. *JAMA Intern Med*. 2014;174(12):1955-62. doi:10.1001/jamainternmed.2014.5294
28. Desai NR, Shrank WH, Fischer MA, et al. Patterns of medication initiation in newly diagnosed diabetes mellitus: quality and cost implications. *Am J Med*. 2012;125(3):302.e1-7. doi:10.1016/j.amjmed.2011.07.033
29. Montvida O, Shaw J, Atherton JJ, et al. Long-term trends in antidiabetes drug usage in the U.S.: Real-world evidence in patients newly diagnosed with type 2 diabetes. *Diabetes Care*. 2018;41(1):69-78. doi:10.2337/dc17-1414
30. Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & quantity*. 2013;47(4):2025-2047.
31. Trinkley KE, Malone DC, Nelson JA, et al. Prescribing attitudes, behaviors and opinions regarding metformin for patients with diabetes: a focus group study. *Ther Adv Chronic Dis*. 2016;7(5):220-8. doi:10.1177/2040622316657328
32. Oude Wesselink SF, Lingsma HF, Robben PB, et al. Guideline adherence and health outcomes in diabetes mellitus type 2 patients: a cross-sectional study. *BMC health services research*. 2015;15: 1-8.
33. Ackermann RT, Thompson TJ, Selby JV, et al. Is the number of documented diabetes process-of-care indicators associated with cardiometabolic risk factor levels, patient satisfaction, or self-rated quality of diabetes care? The Translating Research in Action for Diabetes (TRIAD) study. *Diabetes Care.* 2006;29(9):2108-2113.

34. Kaplan SH, Griffith JL, Price LL, et al. Improving the reliability of physician performance assessment: identifying the “physician effect” on quality and creating composite measures. *Med Care*. 2009;47:378-387.
35. Loeb JM. The current state of performance measurement in health care. *Int J Qual Health Care*. 2004;16(suppl 1):i5-i9.
36. Eijkenaar F. Key issues in the design of pay for performance programs. *Eur J Health Econ*. 2013;14:117-131.
37. de Bruin SR, Baan CA, Struijs JN. Pay-for-performance in disease management: a systematic review of the literature. *BMC Health Serv Res*. 2011;11:272.
38. Hofer TP, Hayward RA, Greenfield S, et al. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281(22):2098-105. doi:10.1001/jama.281.22.2098
39. Fung V, Schmittdiel JA, Fireman B, et al. Meaningful variation in performance: a systematic literature review. *Med Care*. 2010;48:140-148.
40. Krein SL, Hofer TP, Kerr EA, et al. Whom should we profile? Examining diabetes care practice variation among primary care providers, provider groups, and health care facilities. *Health Serv Res*. 2002;37:1159-1180.

Table 1. Number of patients in each eGFR category who were prescribed metformin or DPP-4i / sulfonylurea as their initial diabetes medication.

eGFR level	Metformin	DPP-4i / Sulfonylurea	Total
eGFR < 30	9	55	64
30 ≤ eGFR < 45	92	204	296
45 ≤ eGFR < 60	709	356	1,065
60 ≤ eGFR < 90	3,851	662	4,513
90 ≤ eGFR	4,178	527	4,705
Total	8,839	1,804	10,643

Figure 1. Inclusion-exclusion criteria for cohort.

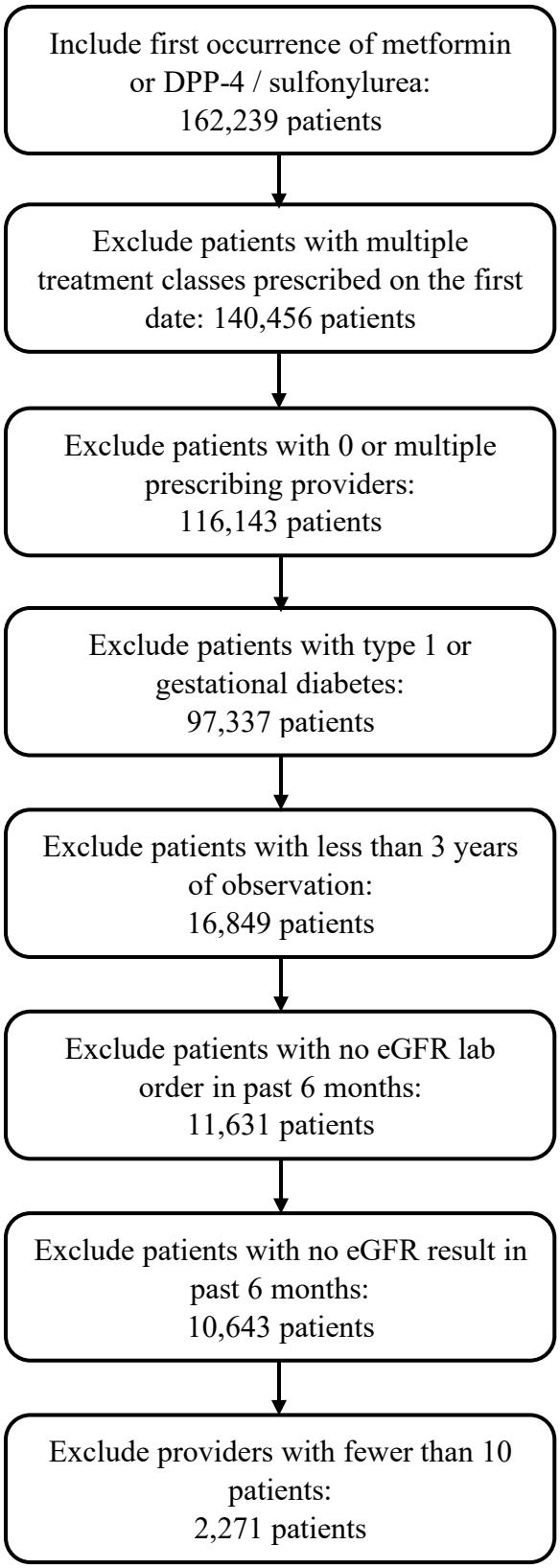
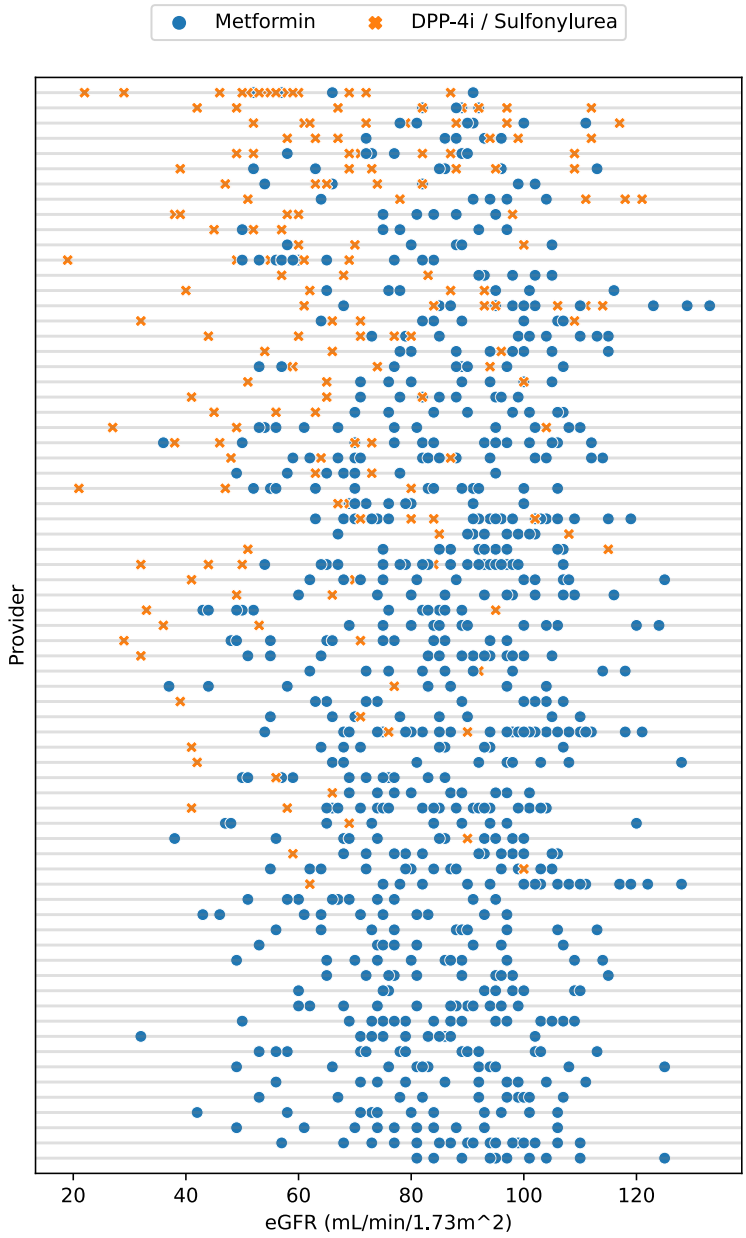
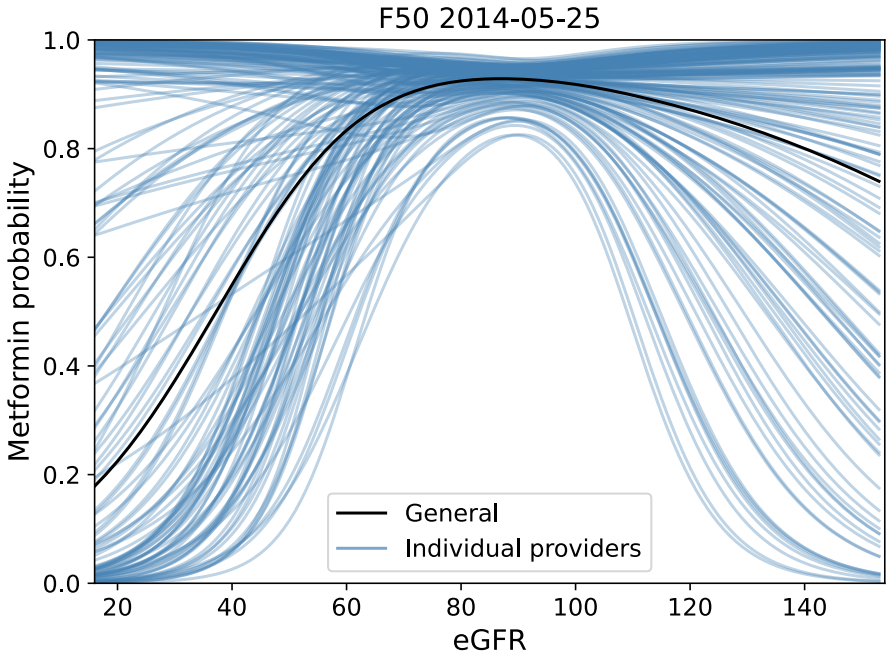


Figure 2. Individual provider choice of initial treatment for diabetes, by most recent eGFR on x-axis.



Each row is one provider. Providers ordered by metformin prescription rates. A subset of the 173 providers with at least 10 patients are shown: 10 with smallest rates, 10 with largest rates, and every third in between. A blue dot indicates the provider prescribed metformin to a patient with that eGFR value. An orange X is the corresponding decision to prescribe DPP-4i / sulfonylurea.

Figure 3. Predicted probability of prescribing metformin versus eGFR for a 50-year-old female without heart failure whose first-line treatment is initiated on 2014-05-25.



Metformin probabilities estimated from models without random effects (black) and with provider-specific random effects (blue).