Supplementary Online Content

Powell M, Clark C, Alyakin A, Vogelstein JT, Hart B. Exploration of residual confounding in analyses of associations of metformin use and outcomes in adults with type 2 diabetes. *JAMA Netw Open*. 2022;5(11):e2241505. doi:10.1001/jamanetworkopen.2022.41505

eAppendix. Supplemental Literature Review

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Supplemental Literature Review

A central claim of this study is that the various design and modeling choices used in this study are reflective of what is commonly seen in the literature, including in high-impact medical journals. A literature review was performed to support this claim and was conducted as follows:

1. We searched the *British Medical Journal* (BMJ) network, *Journal of the American Medical Association* (JAMA) network, *Lancet* family, and *New England Journal of Medicine* (NEJM) webpages for all studies including "metformin" in the article title.
2. We limited the results to observational studies of humans from 2015 to present; we excluded RCTs, meta-analyses, research letters, and animal/non-human studies. This returned 26 studies (N = 5 JAMA; N = 18 BMJ; N = 3 Lancet; N = 0 NEJM).
3. We further excluded studies of endpoints/indications/outcomes listed as on-label indication and/or contra-indication per the metformin package insert (N = 6).[1]
4. This left 20 studies for review (N=4 JAMA; N=14 BMJ; N = 2 Lancet; N = 0 NEJM). A high-level summary of these studies is below, followed by a more detailed listing of all 20 studies in table form.
    - Incident user design was specified in 25% (N=5).
    - The most common comparators were no metformin (with a background of any non-metformin diabetes drugs, N=3; without specification if a background of other diabetes drugs were permitted, N=13;), other non-metformin diabetes drug monotherapy (N=3 [sulfonylureas (N=2) and DPP-4i (N=1)] or no drugs (N=1).
    - The off-label outcomes evaluated by these studies were oncology-related outcomes, including incidence, recurrence, or treatment response (N=7), ocular-related diseases (N=4), cardiovascular-related events (N=3), post-operative complications/mortality/readmission (N=2), neurodegenerative disease (N = 1), benign prostatic hyperplasia (N = 1), infectious disease (N=1), and in-hospital COVID-19 mortality (N=1).

eTable 1. Summary of 20 observational studies of metformin for off-label outcomes (2015 to 2022) from the *BMJ*, *JAMA*, *Lancet*, and *NEJM* journal families.

Article	Journal Family	Design	Population	Intervention	Comparison	Outcome	Covariates
Chang (2015)	*Lancet*	Retrospective cohort study with uni- and multivariate Cox proportional hazards model	N = 2,003 veterans with type 2 diabetes and monoclonal gammopathy of undetermined significance (MGUS)	Metformin; prevalent use	No metformin	Progression to multiple myeloma	Age, sex, race, comorbidities, MGUS type, BMI, and level of HbA1c, serum M-protein, creatinine
Lin (2015)	*JAMA*	Retrospective Cohort w/ uni-and multivariate Cox proportional regression models	N = 150,016 people with diabetes mellitus age >= 40 years	metformin (by dosage quartiles); prevalent use	No metformin	incident open angle glaucoma	Vary by model. Age, sex, race, region, SES, type of diabetes, ocular comorbidities, other comorbidities, Charlson comorbidity index score,* cataract and retinal surgery, diabetes medication classes, HbA1c
Soffer (2015)	*BMJ*	Retrospective cohort study with multivariable Cox regression model	N = 66,778 women with type 2 diabetes	Metformin (monotherapy, in combination)	Non-metformin diabetes drugs, no drugs	Incidence of breast, endometrial, and ovarian cancer	Age, race/ethnicity, income, prior estrogen replacement therapy, statin use, Charlson comorbidity index, number of outpatient visits, maximum HbA1c
Hall (2016)	*BMJ*	Retrospective cohort study with descriptive analysis	N = 351 women with endometrial cancer and obesity	Metformin; prevalent use	No metformin	Type I endometrial cancer recurrence	None
Mor (2016)	*BMJ*	Retrospective cohort study with Cox regression	N = 131,949 people with type 2 diabetes initiation glucose-lowering pharmacotherapy >= 30 years	Metformin; incident use	Other glucose-lowering drugs; incident use	Community-based antibiotic use, hospital-treated infection	Charlson comorbidity index score, micro- and macrovasular complications, diabetes duration, alcohol-related disorders, obesity, medications (immuno-suppressives, oral corticosteroids, statins), marital status, calendar period.
Hanpra-sertpong (2017)	*BMJ*	Retrospective cohort study with uni- and multivariate Cox proportional hazards regression.	N = 248 women with cervical cancer and type 2 diabetes	Metformin	No metformin	Disease-free survival and overall survival	Age, BMI, hypertension, diabetes treatment, tumor features (size, histology, stage), cancer treatment, recurrence status, hemoglobin.
Takiuchi (2017)	*BMJ*	Retrospective cohort study with	N = 785 women with cervical cancer	Metformin, stratified by diabetes status	No metformin	Progression-free survival, cervical	Age (<60 vs ≥60 years), histologic subtype (squamous

		univariate analysis and multivariate Cox Proportional hazards model				cancer-specific overall survival	cell, adenocarcinoma, adenosquamous, and others), and stage (early, locally advanced, and distant metastasis).
Afzal (2018)	*BMJ*	Retrospective cohort study with Cox regression	N = 55 people with metastatic malignant melanoma and treated with an immune checkpoint inhibitor (ICI; ipilim-, nivol- and/or pembrolizumab)	Metformin, prevalent use, concurrent with ICI	No metformin	Objective response rate (partial + complete response)	Age at diagnosis, sex, any other malignancy, prior cancer therapy.
Roumie (2019)	*JAMA*	Retrospective cohort with adjusted Cox proportional hazards models	N = 49,478 veterans with new-onset type 2 diabetes and reduced kidney function (eGFR < 60 mL/min/1.73m2) age >= 18 years	Metformin; incident users, persisting with treatment after reduced renal function	sulfonylurea; incident users, persisting with treatment after reduced renal function	incident Major Adverse Cardiovascu-lar Events (MACE)	Age, sex, race, fiscal year, number of months from initial antidiabetic medication to kidney threshold (diabetes duration), body mass index, blood pressure, HbA1c, low-density lipoprotein, hemoglobin, proteinuria, creatinine, healthcare utilization, smoking status, comorbidities, select non-diabetes medications (statins, anti-hypertensives, other CVD, antipsychotics, oral glucocorticoids)
Shi (2019)	*BMJ*	Retrospective cohort study with multivariate Cox proportional hazards model	N = 5528 veterans aged>= 50 years with type 2 diabetes taking insulin	Metformin; prevalent use, stratified by duration of use	No metformin	Incident neurode-generative disease (dementia, Alzheimer's, Parkinson's, Huntington's, mild cognitive impairment)	Age, sex, race, medication history (diabetes drugs, antihypertensives, lipid-lowering), microvascular comorbidities, macrovascular comorbidities, hypertension, hyperglycemia, hyperlipidemia, kidney disease, mental health, obesity, tobacco use status.
Lin (2020)	*BMJ*	Retrospective cohort study with multiple logistic regression	N = 91,356 people with diabetes undergoing major surgery >= 20 years	Metformin; prevalent use	No metformin	Post-operative septicemia, renal failure, and 30-day in-hospital mortality	Age, sex, low income, volume of the hospital, types of surgery, types of anesthesia, hypertension, mental disorders, ischemic heart disease, chronic obstructive pulmonary disease, hyperlipidemia, liver cirrhosis, heart failure, alcohol-

							related illness, renal dialysis, Parkinson's disease, and Charlson comorbidity index
Matsuo (2020)	*BMJ*	Retrospective cohort study with Cox proportional hazard model	N = 245 women with obesity and complex atypical hyperplasia treated with oral or intrauterine progesterone	Metformin; prevalent use	No metformin	Time to treatment response determined by endometrial biopsies, stratified by oral vs intrauterine progesterone	Diabetes status, hyperlipidemia, polycystic ovarian syndrome, BMI, infertility.
Nørgaard (2020)	*BMJ*	Retrospective cohort study with cumulative incidence rate	N = men with type 2 diabetes age >= 30 initiating monotherapy with metformin or sulfonylurea	Metformin, incident use	Sulfonylurea, incident use	Incident benign prostatic hyperplasia	Diabetes duration, micro- and macrovascular comorbidities, HbA1c, Charlson comorbidity index, other comorbidities, medications (immunosup-pressants, oral corticosteroids, statins), marital status, calendar period.
Reitz (2020)	*JAMA*	Retrospective cohort with uni- and multivariable Cox proportional hazards regression model	N = 10088 people with type 2 diabetes age >= 18 years undergoing major surgical intervention	Metformin; prevalent use	No metformin	Postoperative mortality and readmission	Age, sex, race/ethnicity, insurance coverage, index surgical intervention, American Society of Anesthesiologists score and associated emergent status for the index operation, markers of operative complexity, length of stay, and discharge disposition, comorbidities, medications (insulin, statins, blood thinners, other CV), operative year, surgical specialty, HbA1c, eGFR, hemoglobin, prior year visit with primary care physician, endocrinology, cardiology, for colonoscopy*, surgeon specialty, operative year.
Sutton (2020)	*BMJ*	Retrospective cohort study with adjusted Cox proportional hazards model	N = 123,440 people, with and without type 2 diabetes, diagnosed with abdominal aortic aneurysm	Metformin	No metformin, stratified by diabetes status	Abdominal aortic aneurysm progression to surgery and/or death.	Age, race, Charlson comorbidity index score, comorbidities (hyper-cholesterolemia, lipidemia, triglyceridemia,

							hypertension), smoking status, body mass index, HbA1c, index year.
Blitzer (2021)	*JAMA*	Retrospective case-control study with uni- and multivariable logistic regression	312,404 people age >= 55 years with newly diagnosed age-related macular degeneration.	Metformin; prevalent use, and stratified by dose	No metformin	Incident age-related macular degeneration	Age, sex, geographic region, select comorbidities (hypertension, hyperlipidemia, obesity, diabetes), diabetic retinopathy, smoking status, Charlson comorbidity index score, diabetes medications, statins.
Bramante (2021)	*Lancet*	Retrospective cohort study with Cox proportional hazards model	N = 6256 people with type 2 diabetes or obesity with inpatient admission for COVID-19	Metformin; prevalent use	No metformin	In-hospital COVID-19 mortality	Comorbidities, medications.
Gokhale (2022)	*BMJ*	Retrospective cohort study with extended Cox proportional hazards regression	N = 173,689 people with newly diagnosed type 2 diabetes aged >= 40	Metformin, with or without other T2D drugs; incident use	All other diabetes medications	Incident age-related macular degeneration	Age, sex, ethnicity, SES, smoking status, body mass index, blood pressure, HbA1c, Charlson comorbidity index conditions, diabetes-related complications,
Jiang (2022)	*BMJ*	Retrospective cohort study with uni- and multivariate logistic regression models	N = 324 people with type 2 diabetes for at least 10 years, aged >= 50	Metformin, prevalent use	No metformin (+/- other T2D meds)	Incident age-related macular degeneration	Age, sex, diabetes duration, hypertension, hyperlipidemia, smoking status, diabetic retinopathy, body mass index, HbA1c, fasting blood glucose, cholesterol, triglycerides, HDL, LDL, uric acid, Creatinine.
Nishimura (2022)	*BMJ*	Retrospective cohort study used Cox proportional hazards	N = 8775 (database 1), 5141 (database 2) people diagnosed with type 2 diabetes at age >= 18 having no prior history of heart attack or cerebrovascular disease	Metformin (first-line), incident use	DPP-4 inhibitor (first-line), incident use	Hospitaliza-tion for myocardial infarction or stroke	Age, sex, Charlson comorbidity index, hypertension, dyslipidemia, antithrombotic drugs, HbA1c, eGFR.

eReferences

1. Glucophage (metformin) [package insert]. Princeton, NJ: Bristol-Myers Squibb Company. https://packageinserts.bms.com/pi/pi_glucophage.pdf Published May 2018.
2. Soffer D, Shi J, Chung J, Schottinger JE, Wallner LP, Chlebowski RT, Lentz SE, Haque R. Metformin and breast and gynecological cancer risk among women with diabetes. BMJ Open Diabetes Res Care. 2015 Jan 24;3(1):e000049. doi: 10.1136/bmjdrc-2014-000049. PMID: 25664181; PMCID: PMC4316195.
3. Hall C, Stone RL, Gehlot A, Zorn KK, Burnett AF. Use of Metformin in Obese Women With Type I Endometrial Cancer Is Associated With a Reduced Incidence of Cancer Recurrence. Int J Gynecol Cancer. 2016 Feb;26(2):313-7. doi: 10.1097/IGC.0000000000000603. PMID: 26588235.
4. Mor A, Petersen I, Sørensen HT, Thomsen RW. Metformin and other glucose-lowering drug initiation and rates of community-based antibiotic use and hospital-treated infections in patients with type 2 diabetes: a Danish nationwide population-based cohort study. BMJ Open. 2016 Aug 19;6(8):e011523. doi: 10.1136/bmjopen-2016-011523. PMID: 27543589; PMCID: PMC5013429.
5. Hanprasertpong J, Jiamset I, Geater A, Peerawong T, Hemman W, Kornsilp S. The Effect of Metformin on Oncological Outcomes in Patients With Cervical Cancer With Type 2 Diabetes Mellitus. Int J Gynecol Cancer. 2017 Jan;27(1):131-137. doi: 10.1097/IGC.0000000000000855. PMID: 27870711.
6. Takiuchi T, Machida H, Hom MS, Mostofizadeh S, Frimer M, Brunette LL, Matsuo K. Association of Metformin Use and Survival Outcome in Women With Cervical Cancer. Int J Gynecol Cancer. 2017 Sep;27(7):1455-1463. doi: 10.1097/IGC.0000000000001036. PMID: 29049093; PMCID: PMC7526033.
7. Afzal MZ, Mercado RR, Shirai K. Efficacy of metformin in combination with immune checkpoint inhibitors (anti-PD-1/anti-CTLA-4) in metastatic malignant melanoma. Journal for ImmunoTherapy of Cancer 2018;6:64. doi: 10.1186/s40425-018-0375-1
8. Shi Q, Liu S, Fonseca VA, Thethi TK, Shi L. Effect of metformin on neurodegenerative disease among elderly adult US veterans with type 2 diabetes mellitus. BMJ Open. 2019 Jul 30;9(7):e024954. doi: 10.1136/bmjopen-2018-024954. PMID: 31366635; PMCID: PMC6677947.
9. Matsuo K, Mandelbaum RS, Ciccone M, Khoshchehreh M, Pursuwani H, Morocco EB, Matsuzaki S, Dancz CE, Ozel B, Paulson RJ, Roman L. Route-specific association of progestin therapy and concurrent metformin use in obese women with complex atypical hyperplasia. Int J Gynecol Cancer. 2020 Sep;30(9):1331-1339. doi: 10.1136/ijgc-2020-001362. Epub 2020 May 5. Erratum in: Int J Gynecol Cancer. 2020 Dec;30(12):2022. PMID: 32376736; PMCID: PMC7521080.
10. Nørgaard M, Darvalics B, Thomsen RW. Metformin use and long-term risk of benign prostatic hyperplasia: a population-based cohort study. BMJ Open 2020;10:e041875. doi: 10.1136/bmjopen-2020-041875
11. Sutton SS, Magagnoli J, Cummings TH, Hardin JW. Association between metformin and abdominal aortic aneurysm in diabetic and non-diabetic US veterans. J Investig Med. 2020 Jun;68(5):1015-1018. doi: 10.1136/jim-2019-001177. Epub 2020 Apr 8. PMID: 32273298.
12. Lin CS, Chang CC, Yeh CC, Chang YC, Chen TL, Liao CC. Outcomes after surgery in patients with diabetes who used metformin: a retrospective cohort study based on a real-world database. BMJ Open Diabetes Res Care. 2020 Nov;8(2):e001351. doi: 10.1136/bmjdrc-2020-001351. PMID: 33257420; PMCID: PMC7705543.
13. Gokhale KM, Adderley NJ, Subramanian A, Lee WH, Han D, Coker J, Braithwaite T, Denniston AK, Keane PA, Nirantharakumar K. Metformin and risk of age-related macular degeneration in individuals with type 2 diabetes: a retrospective cohort study. Br J Ophthalmol. 2022 Feb 3:bjophthalmol-2021-319641. doi: 10.1136/bjophthalmol-2021-319641. Epub ahead of print. PMID: 35115301.

14. Jiang J, Chen Y, Zhang H, Yuan W, Zhao T, Wang N, Fan G, Zheng D, Wang Z. Association between metformin use and the risk of age-related macular degeneration in patients with type 2 diabetes: a retrospective study. BMJ Open. 2022 Apr 26;12(4):e054420. doi: 10.1136/bmjopen-2021-054420. PMID: 35473747; PMCID: PMC9045056.
15. Nishimura R, Takeshima T, Iwasaki K, Aoi S. Comparison of the effects on cardiovascular events between use of metformin and dipeptidyl peptidase-4 inhibitors as the first-line hypoglycaemic agents in Japanese patients with type 2 diabetes mellitus: a claims database analysis. BMJ Open. 2022 Mar 11;12(3):e045966. doi: 10.1136/bmjopen-2020-045966. PMID: 35277396; PMCID: PMC8919442.
16. Chang SH, Luo S, O'Brian KK, Thomas TS, Colditz GA, Carlsson NP, Carson KR. Association between metformin use and progression of monoclonal gammopathy of undetermined significance to multiple myeloma in US veterans with diabetes mellitus: a population-based retrospective cohort study. Lancet Haematol. 2015 Jan;2(1):e30-6. doi: 10.1016/S2352-3026(14)00037-4. Erratum in: Lancet Haematol. 2015 Feb;2(2):e54. PMID: 26034780; PMCID: PMC4448731.
17. Bramante CT, Ingraham NE, Murray TA, Marmor S, Hovertsen S, Gronski J, McNeil C, Feng R, Guzman G, Abdelwahab N, King S, Tamariz L, Meehan T, Pendleton KM, Benson B, Vojta D, Tignanelli CJ. Metformin and risk of mortality in patients hospitalised with COVID-19: a retrospective cohort analysis. Lancet Healthy Longev. 2021 Jan;2(1):e34-e41. doi: 10.1016/S2666-7568(20)30033-7. Epub 2020 Dec 3. PMID: 33521772; PMCID: PMC7832552.
18. Lin HC, Stein JD, Nan B, Childers D, Newman-Casey PA, Thompson DA, Richards JE. Association of Geroprotective Effects of Metformin and Risk of Open-Angle Glaucoma in Persons With Diabetes Mellitus. JAMA Ophthalmol. 2015 Aug;133(8):915-23. doi: 10.1001/jamaophthalmol.2015.1440.
19. Roumie CL, Chipman J, Min JY, Hackstadt AJ, Hung AM, Greevy RA Jr, Grijalva CG, Elasy T, Griffin MR. Association of Treatment With Metformin vs Sulfonylurea With Major Adverse Cardiovascular Events Among Patients With Diabetes and Reduced Kidney Function. JAMA. 2019 Sep 24;322(12):1167-1177. doi: 10.1001/jama.2019.13206.
20. Reitz KM, Marroquin OC, Zenati MS, Kennedy J, Korytkowski M, Tzeng E, Koscum S, Newhouse D, Garcia RM, Vates J, Billiar TR, Zuckerbraun BS, Simmons RL, Shapiro S, Seymour CW, Angus DC, Rosengart MR, Neal MD. Association Between Preoperative Metformin Exposure and Postoperative Outcomes in Adults With Type 2 Diabetes. JAMA Surg. 2020 Jun 1;155(6):e200416. doi: 10.1001/jamasurg.2020.0416.
21. Blitzer AL, Ham SA, Colby KA, Skondra D. Association of Metformin Use With Age-Related Macular Degeneration: A Case-Control Study. JAMA Ophthalmol. 2021 Mar 1;139(3):302-309. doi: 10.1001/jamaophthalmol.2020.6331.

Data Sources

Standardization of Data Entry and Data Structure

Medical and pharmacy claims data are captured, predominantly electronically, from sites of care seeking third-party reimbursement for both Medicare and commercial plans using the industry standard data collection forms HCFA/CMS-1500 for facility claims, UB04/CMS-1450 for professional services and outpatient claims, and NCPDP for pharmacy claims or their electronic equivalents. Structured data from these standardized forms are coded using the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), National Drug Codes (NDC), Current Procedural Terminology (CPT) codes, and Logical Observation Identifiers Names and Codes (LOINC) codes, and Diagnosis Related Groups (DRG). This nomenclature ensures consistency of data collection across geographic regions, health systems, and payers throughout the United States.

Other Drugs for Type 2 Diabetes Treatment

Beyond metformin, there are many other drug classes used to treat type 2 diabetes. The places in therapy for these pharmacologic treatment options are well characterized in randomized controlled trials, with treatment recommendations largely standardized among national and international diabetes care organizations.[15–17] Notably, diabetes guidelines typically recommend using pharmacologic therapies additively, rather than substitutively, and differential recommendations for second-line diabetes drug classes exist for certain subpopulations of people with type 2 diabetes, including for people with or at high risk for atherosclerotic cardiovascular disease (ASCVD), those with heart failure, and those with chronic kidney disease.[15–17] While the variety of available diabetes drug class options may give an appearance of a suite of active comparators for consideration in observational studies, this conclusion ignores the connections to disease severity that certain drug classes may have, as well as the consistency of metformin as a guideline-recommended, first-line therapy in the background of most treatment regimens as other drugs are added over time. A full list of diabetes medications appears in Supplemental eTable 7.

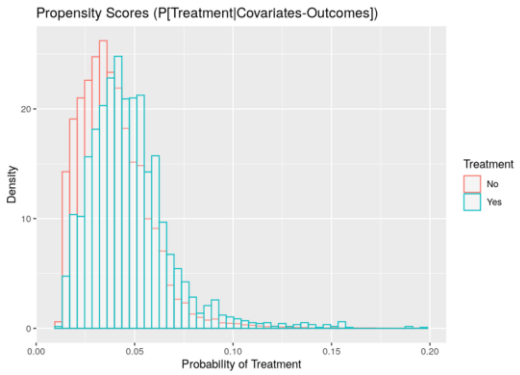
Validation Tools

Covariate Balance Diagnostics

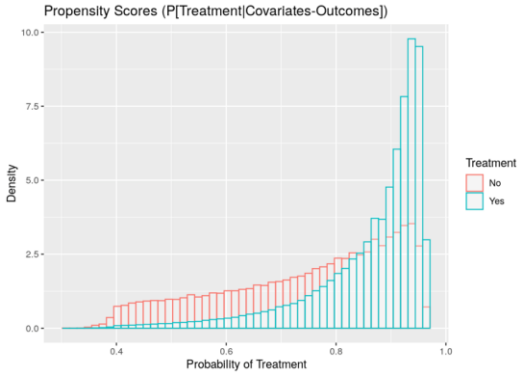
Observational studies often attempt to demonstrate that acceptable covariate balance has been achieved between the exposure groups. Whether it comes through matching, inverse propensity weighting, or some other method seeking covariate balance, a demonstration that balance has been achieved is necessary to convince the reader that two groups that are clearly different (as expected by their different prescribed treatments) have been manipulated in such a way that a weighted or reduced sample shows similar covariate distributions on a set of covariates deemed important for minimizing confounding. Figure 1 presents a diagnostic plot showing the pre- and post-adjustment covariate balance achieved in models for the two example outcomes; acceptable covariate balance generally requires the standardized mean differences for all covariates to have absolute values <0.1. An additional step that can help illustrate whether or not the various comparison groups are appropriate for comparison is to look at the propensity distribution plots for each group. Below we show examples of the propensity distribution plots for the primary analysis. The approach described in the main text ensures overlap in the propensity score distributions by trimming observations with extremely low probabilities of metformin treatment.

Metformin Propensity Distributions in Prediabetes and Type 2 Diabetes Cohorts

Prediabetes (Metformin vs. Nonusers)



Type 2 Diabetes (Metformin vs. Insulin)



Negative Control Experiments

A negative control experiment is one where there exists no causal relationship between the treatment and the outcome.[24] In particular, negative control outcomes, also known as falsification endpoints, preserve the same treatment/control group designations as the primary outcome for each individual in the study, but they are outcomes that cannot reasonably be impacted by the exposure. In our metformin study, outcomes like dry eye syndrome and low back pain are candidate negative control outcomes because there is no known mechanism by which metformin could directly impact these events. Importantly, negative control outcomes should be subject to the same residual confounding as the primary outcomes, which our study assumes to be exclusively related to overall health.

By looking at a wide range of outcomes with no direct, mechanistic connection to the treatment, we seek to expose differences in overall health not accounted for by the treatment or the other observed covariates. For a comparison of metformin and insulin users in a type 2 diabetes cohort, a pattern of nonzero treatment effects for metformin on the negative control outcomes is evidence that we have not adequately controlled for the underlying differences in the overall health status of these two groups.

Negative Control Outcome Criteria

The objective in a negative control outcome experiment is to detect significant relationships between the treatment of interest (e.g., metformin) and a mechanistically unrelated outcome (e.g., low back pain); detecting such a relationship raises serious concerns about residual confounding related to the study design. Importantly, these negative control outcomes should be evaluated against these five criteria:

1. There is no mechanistic connection to the treatment under investigation (i.e., no established mechanism of action for this treatment to affect the negative control outcome).
2. Negative control outcomes must be reasonably prevalent; the statistical power associated with the negative control experiment increases as the prevalence of the outcome increases in the population under study. As a rule of thumb, consider giving preference to negative control outcomes at least as prevalent as the primary outcome. A rare negative control outcome will likely be less informative, typically returning a null result even in the presence of significant residual confounding. The noisy results of a large collection of rare negative control outcomes can still be informative, however, even if an individual outcome occurs too rarely to confidently estimate a treatment effect.
3. Negative control outcomes are suspected of being subject to the same residual confounding as the primary outcome under investigation (e.g., does not require a different level of health care access, health insurance benefit design, etc.). This is unverifiable due to the nature of residual confounding, but the point is that whatever may introduce bias in the outcome of interest should be a potential source of similar bias (expected to be the same direction and magnitude) for the negative control outcome.
4. (Optional) There is no causal relationship to disease severity (e.g., not a known indicator of disease severity for the disease indicating this medication).
5. (Optional) The negative control outcome is not an indicator of health-seeking behavior (e.g., some of the most common recorded “diagnoses” in claims data are screenings or exams that could be sex-specific or age-specific). Exceptions are appropriate when the primary outcome is a health-seeking behavior.

Of these five criteria, (1), (2), and (3) are required, and (4) and (5) are desirable in order to further distance the negative outcomes from obvious group differences in disease severity or health care utilization. The entire collection of negative control outcomes will be used to identify meaningful ways in which the

comparison groups differ even after adjustments are made for observed covariates. Negative control experiments -- well-accepted and commonly recommended -- remain infrequently conducted components of observational study designs.[23] One explanation for the lack of widespread adoption is that identifying the perfect negative control experiment is often quite challenging. In practice, however, a wide range of negative control experiments do not have to individually be perfect to collectively reveal the residual confounding we seek to expose. For this reason, one might empirically determine a host of acceptable negative control outcomes by reviewing the most frequently observed diagnoses among individuals in the cohort. The focus of this approach is using the collective body of evidence from many possibly imperfect negative experiments rather than relying on any single negative experiment's ability to survive heavy scrutiny.

Negative Control Outcome Selection Algorithm

In a randomized trial, thoughtful selection of negative controls is necessary because that data must be intentionally collected, possibly at significant cost. In an observational study, we must draw candidate negative controls from data that has already been collected as acquiring new data on the selected individuals is highly unlikely. To find outcomes present in our data that meet the criteria for negative controls, we proceed as follows (example results depicted in eTable 2):

1. Identify the 500 most common diagnosis codes in 2018 in terms of affected individuals in the primary cohort, ignoring multiple diagnoses for the same condition for the same individual. Repeat for the complementary cohort.
2. Filter the observed diagnoses to only retain diagnoses observed in both cohorts.
3. Rank the diagnosis codes in each cohort and then sum the ranks for a composite rank sum (e.g., low back pain is #13 in prediabetes and #15 in type 2 diabetes for a rank sum of 28).
4. Order the diagnosis codes by rank sums from smallest to largest.
5. Take the top 10/20/50/etc. outcomes that meet the criteria for negative control outcomes. This requires domain expertise to individually consider each outcome for potential mechanistic connections to the treatment, disease severity, and health-seeking behavior. Additionally, a power analysis can help establish a prevalence minimum.

It is likely that these negative control outcomes will span a variety of body systems and will not all be highly correlated, reducing the impact of a single negative control outcome in the group that may have an unrecognized connection to the drug or disease under investigation. Other approaches to identifying negative control outcomes have been developed, including the ATLAS tool created by the Observational Health Data Sciences and Informatics (OHDSI) organization, which also uses a combination of automated discovery and expert review to identify 50-100 negative controls.[23]

Observed Condition		Empirical Prevalence					Expert Review			
ICD Code	ICD Code Description	T2D Rank	T2D Total Cases	Prediabetes Rank	Prediabetes Total Cases	Prediabetes and T2D Ranks Summed	well-established mechanistic connection to metformin	impacted by diabetes severity	general health-seeking behavior	negative control candidate
i10	essential (primary) hypertension	2	1419426	2	146023	4	N	Y	N	N
e785	hyperlipidemia unspecified	3	755895	4	90418	7	N	Y	N	N
z0000	encounter for general adult medical examination without abnormal findings	4	679139	3	124587	7	N	N	Y	N
z23	encounter for immunization	5	575125	5	83550	10	N	N	Y	N
e782	mixed hyperlipidemia	7	479440	7	55719	14	N	Y	N	N
z1231	encounter for screening mammogram for malignant neoplasm of breast	8	369026	6	68582	14	N	N	Y	N
e039	hypothyroidism unspecified	10	301838	9	36737	19	N	N	N	Y
r05	cough	12	271000	14	30625	26	N	N	N	Y
e7800	pure hypercholesterolemia unspecified	17	248728	11	34292	28	N	Y	N	N
m545	low back pain	15	262643	13	31102	28	N	N	N	Y

eTable 2. Negative control outcomes were selected through an automated generation of candidate outcomes followed by an expert review. Prevalence of outcomes in both the primary and complementary cohorts is emphasized in this approach, resulting in an ordering of candidates from which experts can identify the first 10/20/50/etc. candidates that satisfy multiple negative control outcome criteria. Here we find three acceptable negative control outcomes in the 10 most common diagnoses. We found 50 suitable negative control outcomes in the ~100 most prevalent diagnoses in our data set.

Complementary Cohorts

Complementary cohorts provide a second tool to stress test the primary results by nullifying or reversing any overall health advantage the treatment group has in the primary cohort. If we suspect the treatment group may be healthier in some unmeasurable way than the comparison group (aside from the possible effect of the treatment), we construct another cohort in which the treatment group is expected to be less healthy than the comparison group (aside from the possible effect of the treatment). The construction of this cohort requires relevant domain expertise in order to satisfy the following criteria:

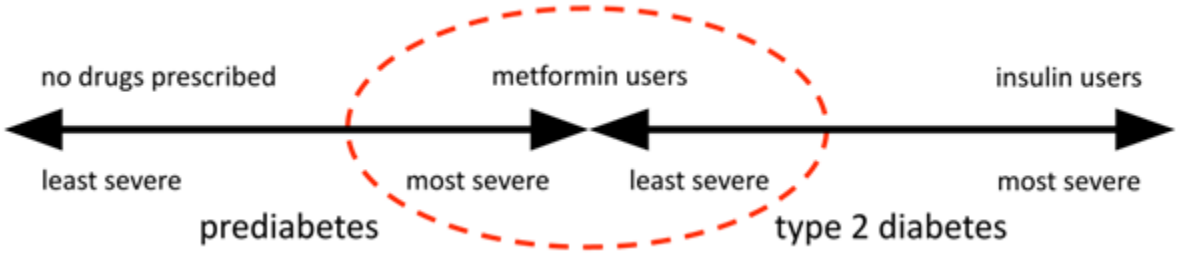
1. The primary treatment of interest must be reasonably prevalent in the complementary cohort. If there are too few users of the primary treatment in the complementary cohort, there will be limited power to detect an effect.
2. If treatment is concentrated among the healthiest members of the primary cohort (in both measurable and unmeasurable ways), use of the treatment in the complementary cohort should be

concentrated among individuals with worse overall health relative to the rest of the cohort, and vice versa. This creates a mirror image of the primary cohort and is critical if the goal is to nullify or reverse the overall health advantage (or disadvantage) suspected in the treated group in the primary cohort.

3. There should be no unnecessary differences in the cohort-identifying disease. This is easiest to satisfy in diseases with a commonly diagnosed "predisease" stage (e.g., prediabetes/type 2 diabetes, albuminuria/chronic kidney disease, osteopenia/ osteoporosis), a framework that best ensures that any residual confounding present in the complementary cohort will be of the same nature as that of the primary cohort (i.e., related to overall health). Introducing a complementary cohort from a completely different disease is still possible, but it may introduce complicated disease differences that must be addressed.

In a study of individuals diagnosed with type 2 diabetes, there is a natural complementary cohort in the population of individuals diagnosed with prediabetes. Critically, the drug metformin is prescribed to individuals in both cohorts -- extensively in type 2 diabetes and more modestly in prediabetes. What makes this primary/complementary cohort specification ideal is where the concentration of metformin users exists in each cohort. In the American Diabetes Association's (ADA) published "Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes—2020," metformin monotherapy is recommended as the first-line treatment for type 2 diabetes along with comprehensive lifestyle modifications.[15] If comorbidities like atherosclerotic cardiovascular disease, heart failure, or chronic kidney disease are present, other drugs may augment or replace metformin. If metformin and/or other drugs cannot effectively control blood glucose, an individual may ultimately be prescribed insulin. Thus, individuals with a metformin claims history, but no history of insulin use, are earlier in the spectrum of type 2 diabetes severity than those who have progressed to using insulin.

In prediabetes, the ADA Standards of Medical Care recommend considering treatment with metformin for individuals at risk for developing type 2 diabetes, particularly for high-risk individuals, including those with a history of gestational diabetes, BMI >= 35 kg/m^2, or age less than 60 years old.[14] Whereas the metformin users were the least severe cases in the type 2 diabetes cohort, they hold the opposite position in the prediabetes cohort. As depicted in eFigure 1, prediabetes thus presents an ideal complementary cohort by eliminating or potentially reversing any metformin exposure group advantage that could be attributable to overall patient health in the primary cohort.



eFigure 1. Based on guideline-driven treatment recommendations, metformin users diagnosed with prediabetes are assumed to be on the opposite end of their respective diabetes severity spectrum compared to metformin users diagnosed with type 2 diabetes. The prediabetes cohort thus reverses the overall health advantage enjoyed by the metformin users in the type 2 diabetes cohort, which makes it an ideal candidate for a complementary cohort aiming to expose residual confounding related to overall health.

The role of this complementary cohort is to validate whether the covariate selections and method choices in the primary cohort analysis are indeed effective at addressing confounding related to disease severity and overall health. If they are, we should expect to reproduce the primary cohort finding in the complementary cohort. Conflicting findings suggest that the result in the primary cohort may be a result of residual confounding and not a true treatment effect. Other explanations exist for conflicting findings, specifically a heterogeneous treatment effect, a possibility that goes unexplored when the entire focus is on the primary cohort.

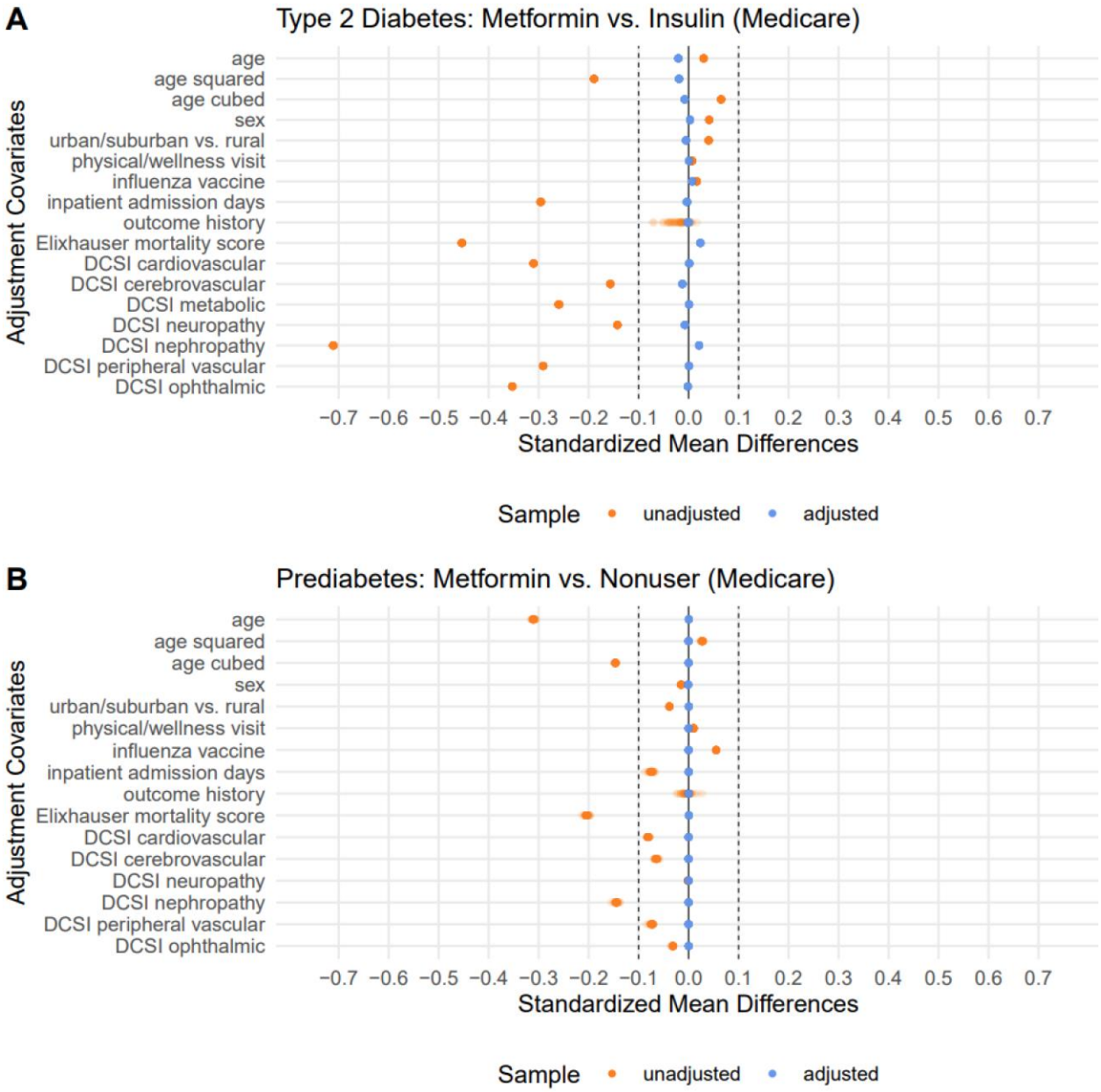
Negative Controls in Complementary Cohorts

Negative control outcome results may reveal residual confounding in the primary cohort (e.g., through a pattern of protective associations between the treatment and the negative control outcomes), which further motivates the use of a complementary cohort to test the primary result. These same negative control outcomes must also test the selection of the complementary cohort, ensuring it exhibits the most desired quality of a complementary cohort: no residual confounding-induced advantage in negative control outcome experiments. The strength of a complementary cohort is defined by how much the negative control outcome associations are nullified or reversed in comparison with the primary cohort results. Larger reversals indicate that the complementary cohort provides a more robust validation of the initial results. If the pattern of bias (i.e., nonzero negative control outcome effect sizes) is similar in the two groups, then the second analysis has little to add aside from validating the result in another population.

When adequately powered, null results across all negative control outcome experiments in both the primary and complementary cohorts are a good indication that residual confounding may be fairly minimal in the identified cohorts and the results are likely trustworthy. When the primary and complementary cohorts yield conflicting results across a host of negative control outcomes, we attribute that difference to a difference in residual confounding in the two cohorts -- this is exactly what we hope to uncover if it exists. In Figure 2, we capture these residual confounding differences in an easily digestible diagnostic plot; it is this diagnostic plot that provides the necessary backdrop to interpret a study's primary result from a more informed position.

Covariate Balance in Negative Control Outcome Experiments

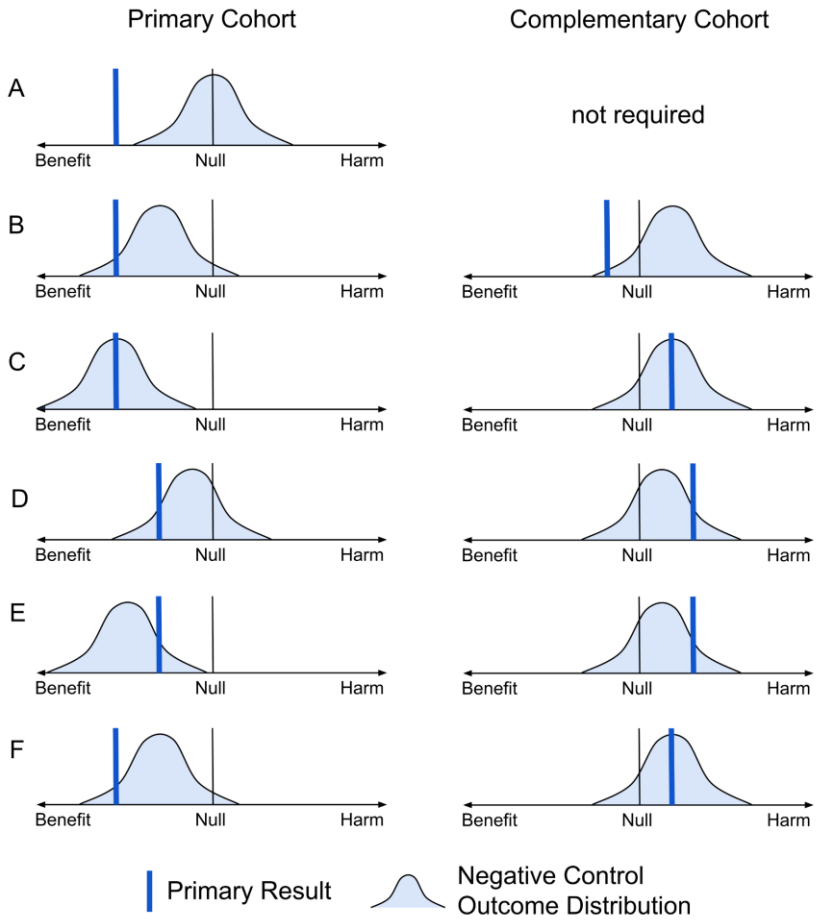
Covariate balance was demonstrated for the primary outcomes in the primary and complementary cohorts in the main text. In eFigure 2 we show that acceptable covariate balance was achieved using IPW for 50 negative control outcome experiments conducted in both cohorts.



eFigure 2. Here we show 50 overlaid balance plots for the Medicare Advantage type 2 diabetes cohort (panel A) and the prediabetes cohort (panel B) where the only covariate changing from one negative control outcome experiment to the next is the history of each respective outcome. The fact that adjusted balance changes negligibly with the exception of outcome history suggests that the sample under analysis (after propensity trimming) is largely the same from one experiment to the next. The noticeable leftward shift in unadjusted balance in panel A reflects the healthier nature of the metformin group in the type 2 diabetes cohort (fewer inpatient days, lower Elixhauser in-hospital mortality score, lower prevalence of diabetes complications in every DCSI dimension despite being slightly older as a group). In panel B, we see better unadjusted balance that appears to meaningfully reduce the metformin advantage. Across both cohorts, the IPW approach achieved satisfactory balance after adjustment for every observed covariate of interest.

Final Results Plot Interpretation

Figure 4 in the main text (as well as eFigure 7, eFigure 11, and eFigure 15 in the supplement) depicts the primary and complementary cohorts results for a primary outcome against a backdrop of negative control outcome experiment results. The discussion surrounding these figures presents an interpretation for each figure, but not all possible scenarios were observed in the real data example. In eFigure 3 we explore a more comprehensive set of possible findings that may appear in a main results figure, and we provide a recommended interpretation for each set of results. The interpretation of each scenario centers on examining each cohort's primary result in the context of its negative control outcome distribution; we then check for agreement between the cohorts.



eFigure 3. Possible Primary Outcome Results: In each of these scenarios, a primary cohort is depicted with a blue bar identifying the result for the primary outcome, and a light blue distribution of negative control outcome results appears behind it. The complementary cohort results are depicted in the same manner. The discussion in the supplement text examines each row A-F for the level of evidence it provides to support a claim of a beneficial treatment effect.

In eFigure 3 row A, we see a strong primary result and no pattern of bias in the negative control outcome experiments. With no evidence of bias, there is no requirement to have a complementary cohort, and we have reason to trust the primary result. We would similarly trust a result indicating harm the farther right it is from the null. Our confidence in either result would decrease as it moves toward the null.

In eFigure 3 row B, we observe a strong primary result, but there is a clear pattern of bias in the negative control outcome experiments, which weakens the evidence supplied by the primary result. In the complementary cohort (validated by observed bias reversal), we see a more modest primary result, but it is quite strong compared to the negative control outcome experiments. Taken together, all the evidence from both cohorts points to a beneficial treatment effect. In this case, the complementary cohort analysis strengthened our confidence in the primary cohort result.

In eFigure 3 row C, a strong primary result is nullified by the negative control outcome experiments. An unfavorable result in the complementary cohort is also nullified by negative control outcome experiments. Taken together, there is no strong evidence of any effect.

In eFigure 3 row D, we see a modest primary result in the primary cohort, and this result exceeds a large majority of the negative control outcome experiments. In the complementary cohort, everything is reversed such that the primary result is now harmful and exceeds a similarly large majority of the negative control outcome experiments. Taken together, these conflicting findings present no strong evidence of any effect. In this case, the complementary cohort analysis erased our confidence in the primary cohort result.

In eFigure 3 row E, a modest primary result in the primary cohort actually appears harmful compared to the distribution of negative control outcome experiments. In the complementary cohort, the result indicating harm is worse than a large majority of the negative control outcome distribution. Taken together, these results suggest there may be a harmful effect. In this case, the complementary cohort analysis strengthened our confidence in the primary cohort result.

In eFigure 3 row F, a strong primary result in the primary cohort exceeds a large majority of the negative control outcome results, indicating a potentially beneficial treatment effect. In the complementary cohort, a result indicating harm is squarely in the middle of the negative control outcome results, effectively indicating a null effect. Taken together, these results are inconclusive. There's some evidence supporting benefit and other evidence suggesting no effect. In this case, the complementary cohort analysis lessened our confidence in the primary cohort result.

Supplementary Analyses

The main text focused on a comparison group of insulin users, but nonusers are also a frequent comparison group in studies of metformin and other drugs. Nonusers are a difficult group to conceptualize when virtually every stage of treatment for a condition involves prescription medication (as seen in the type 2 diabetes treatment recommendations). The “nonuser” population can also be hard to describe when the drug under investigation is available inexpensively without using insurance (e.g., metformin). Since metformin is so widely prescribed in the type 2 diabetes population, there is a reasonable chance that some “nonusers” are taking metformin; they are simply purchasing it outside the visibility of their insurance plans, making them only appear as nonusers in our study despite obtaining the medication through alternate means such as cash pay. This has the effect of biasing any effect estimate toward the null and was the primary reason we selected insulin users as the comparison group for the main analysis.

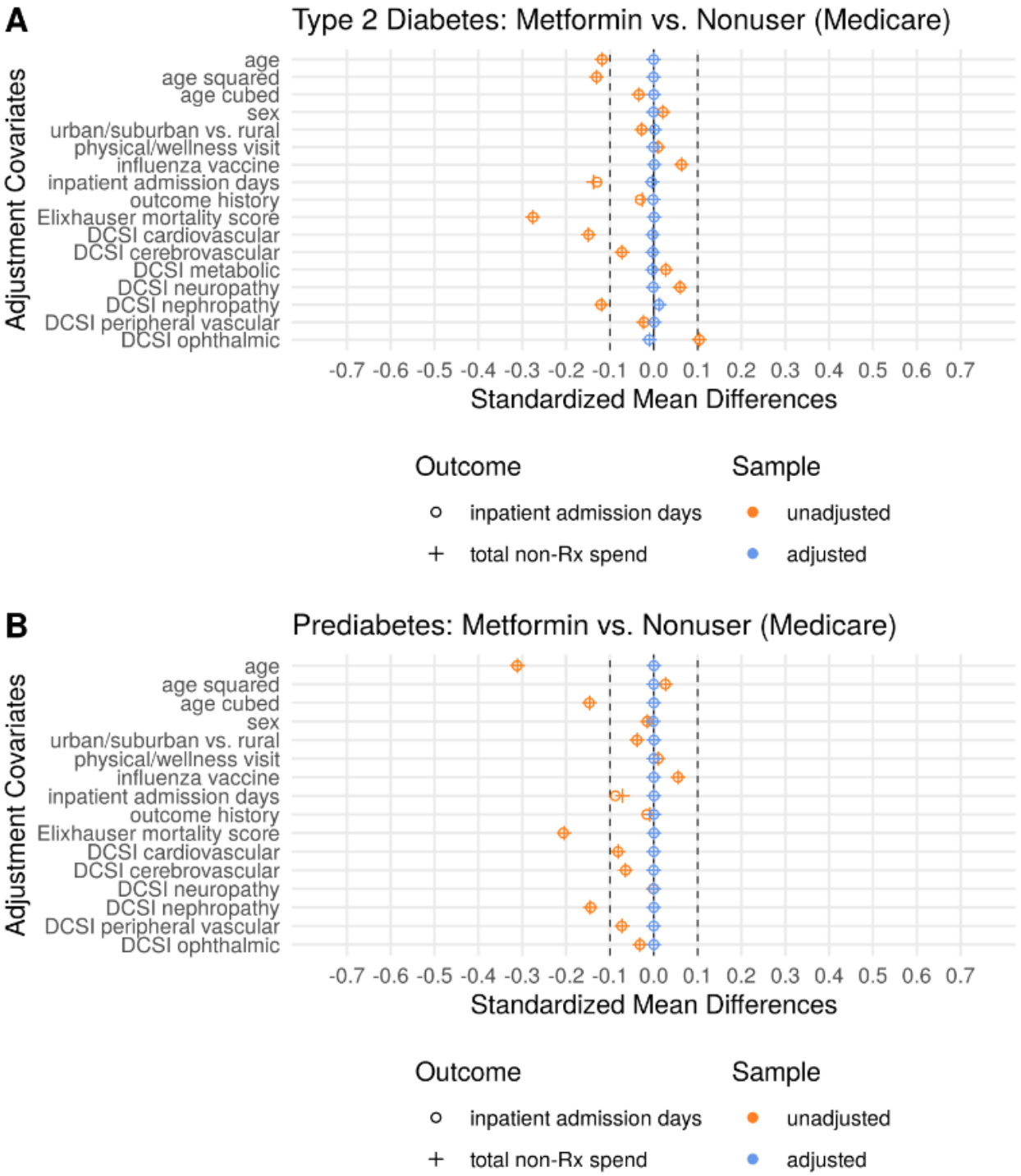
In eTables 3-5 and eFigures 4-15 we show three supplemental analyses not presented in the main text. Each of the three analyses is presented in one results table and four figures. eTable 3, eFigure 4, eFigure 5, eFigure 6, and eFigure 7 present a metformin analysis in the Medicare Advantage population with a comparison group of nonusers (different comparison group from the main text). eTable 4, eFigure 8, eFigure 9, eFigure 10, and eFigure 11 present a metformin analysis in the commercially insured population with a comparison group of insulin users (different population from the main text). eTable 5, eFigure 12, eFigure 13, eFigure 14, and eFigure 15 present a metformin analysis in the commercially insured population with a comparison group of nonusers (different population and comparison group from the main text). Each four-figure group shows two Love plot figures (one for the example outcomes and one for the 50 negative control outcome experiments) depicting acceptable covariate balance, a residual confounding plot, and finally the primary outcome results plotted on a distribution of negative control outcome effect estimates in both the primary and complementary cohorts. We see bias in the same direction emerge in the distributions of negative control outcome effect estimates in every combination of population and comparison group definition. In summary, none of the populations and comparison groups we explored appear immune to a concerning amount of residual confounding.

Medicare Advantage Beneficiaries (metformin users vs. nonusers)

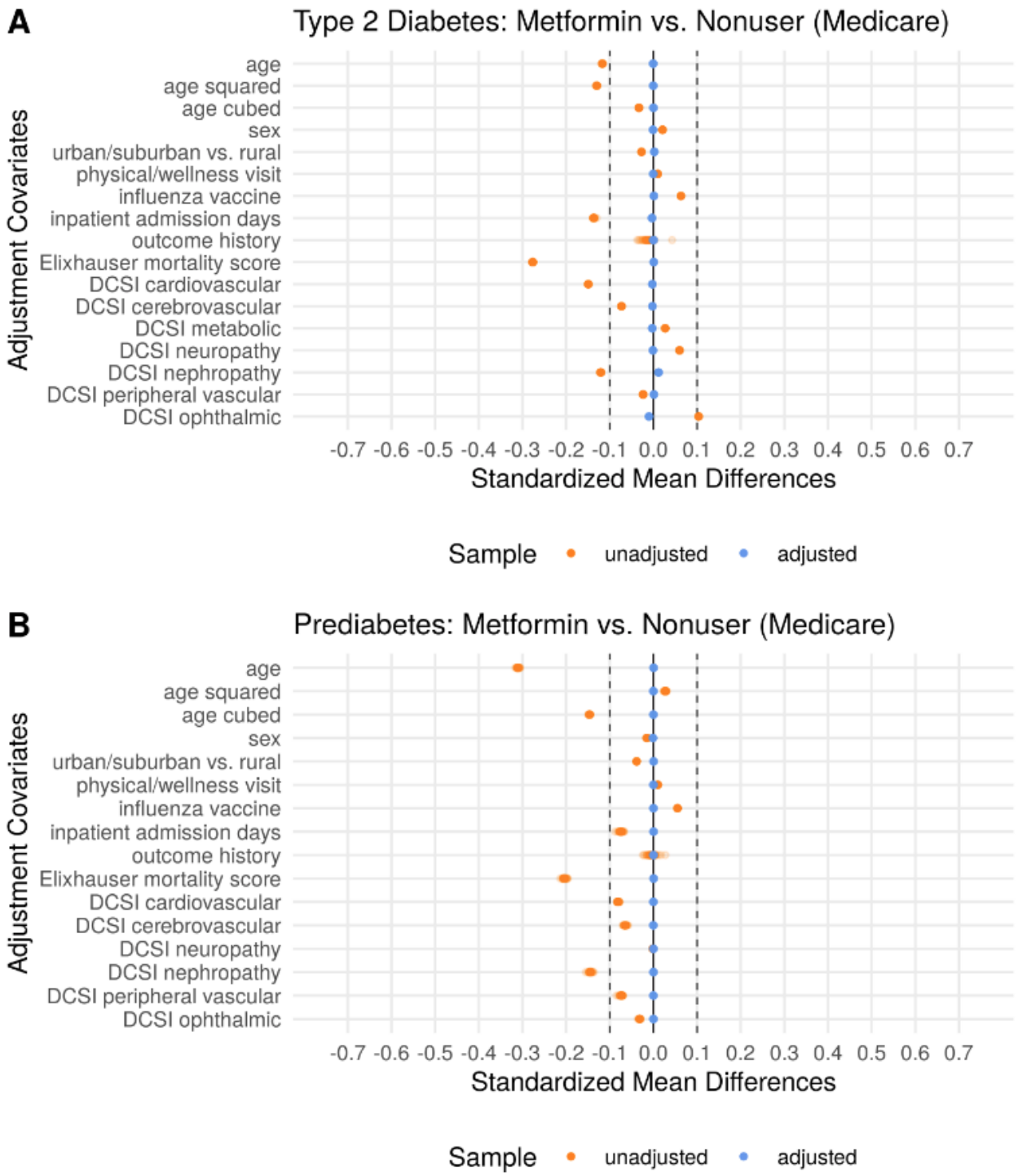
Examining the nonuser comparison group in the Medicare Advantage population produced favorable results supporting a metformin benefit (see eTable 3). The large Medicare Advantage population makes these conservative results highly confident. The pre-adjustment covariate balance in eFigure 4 and eFigure 5 indicates a slightly healthier metformin user population, and eFigure 6 confirms once again through the negative control outcome experiments that a strong bias exists favoring metformin in type 2 diabetes while also showing a weaker bias against metformin in the prediabetes cohort. Interestingly, eFigure 7 shows that the example outcome effect estimates in type 2 diabetes are at best as strong as those seen for an average negative control outcome experiment, but they are far worse than an average negative control outcome experiment in the prediabetes cohort. Thus, while eTable 3 may indicate a favorable treatment effect estimate, no such conclusion can be supported by the total evidence supplied by the complementary cohort design.

outcome	model	log OR (base 2)	95% CI	p	E-value
inpatient days	unadjusted	-0.39	(-0.41, -0.36)	$<10^{-216}$	1.94
inpatient days	IPW logistic	-0.09	(-0.12, -0.06)	$<10^{-10}$	1.33
medical spend	unadjusted	-0.55	(-0.58, -0.53)	$<10^{-323}$	2.30
medical spend	IPW logistic	-0.17	(-0.20, -0.14)	$<10^{-23}$	1.50

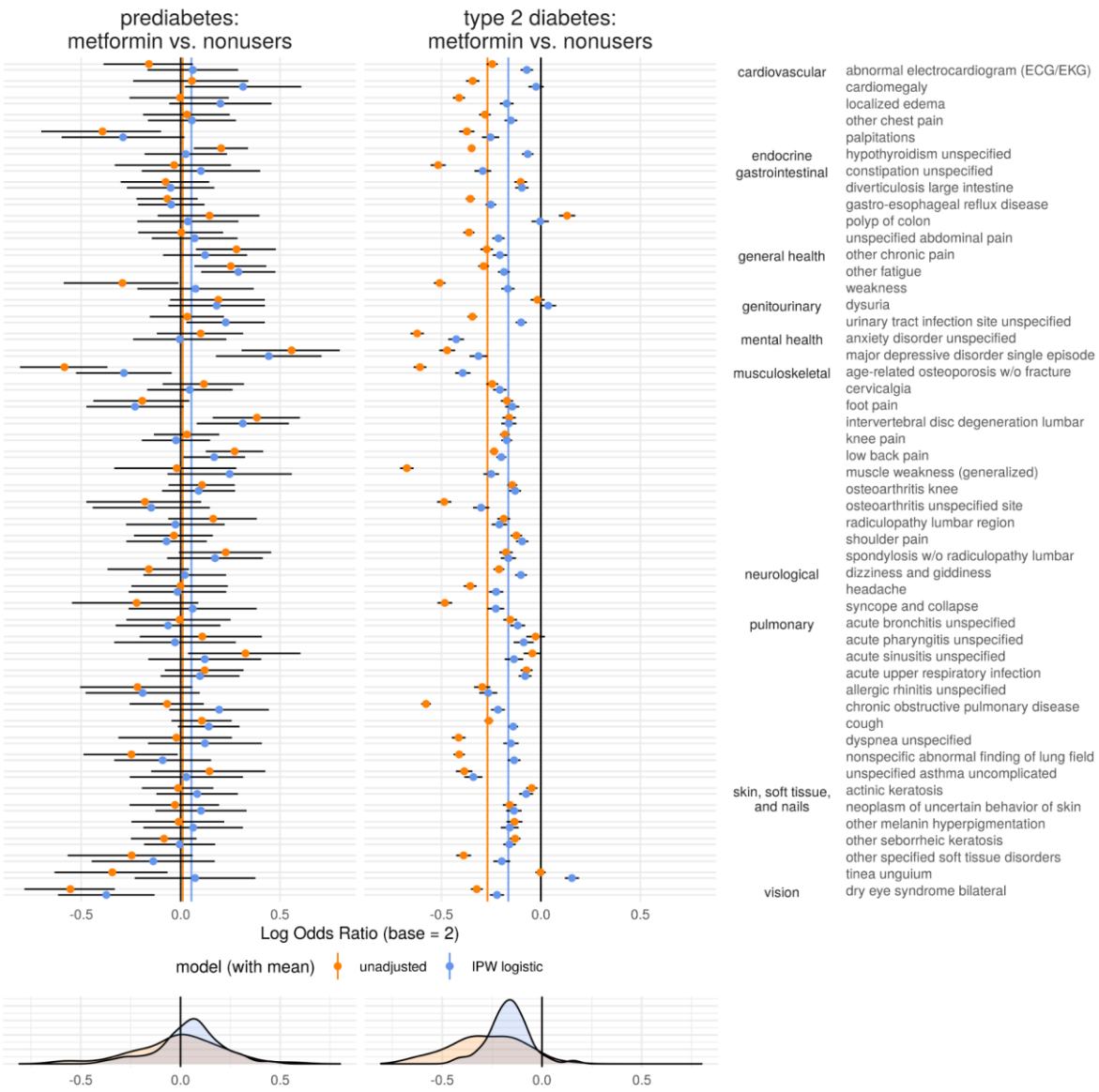
eTable 3. Treatment effect estimates for metformin: this study was conducted in a Medicare Advantage type 2 diabetes population comparing metformin users to a control group of nonusers. The outcomes represent >0 inpatient admission days in 2019 and a total medical spend (insurance payouts to health care providers) exceeding the 90th percentile of all type 2 diabetes patient expenditures (>$25,793). Metformin appears associated with fewer inpatient admission days and lower health care costs, even after adjustment for a range of relevant covariates.



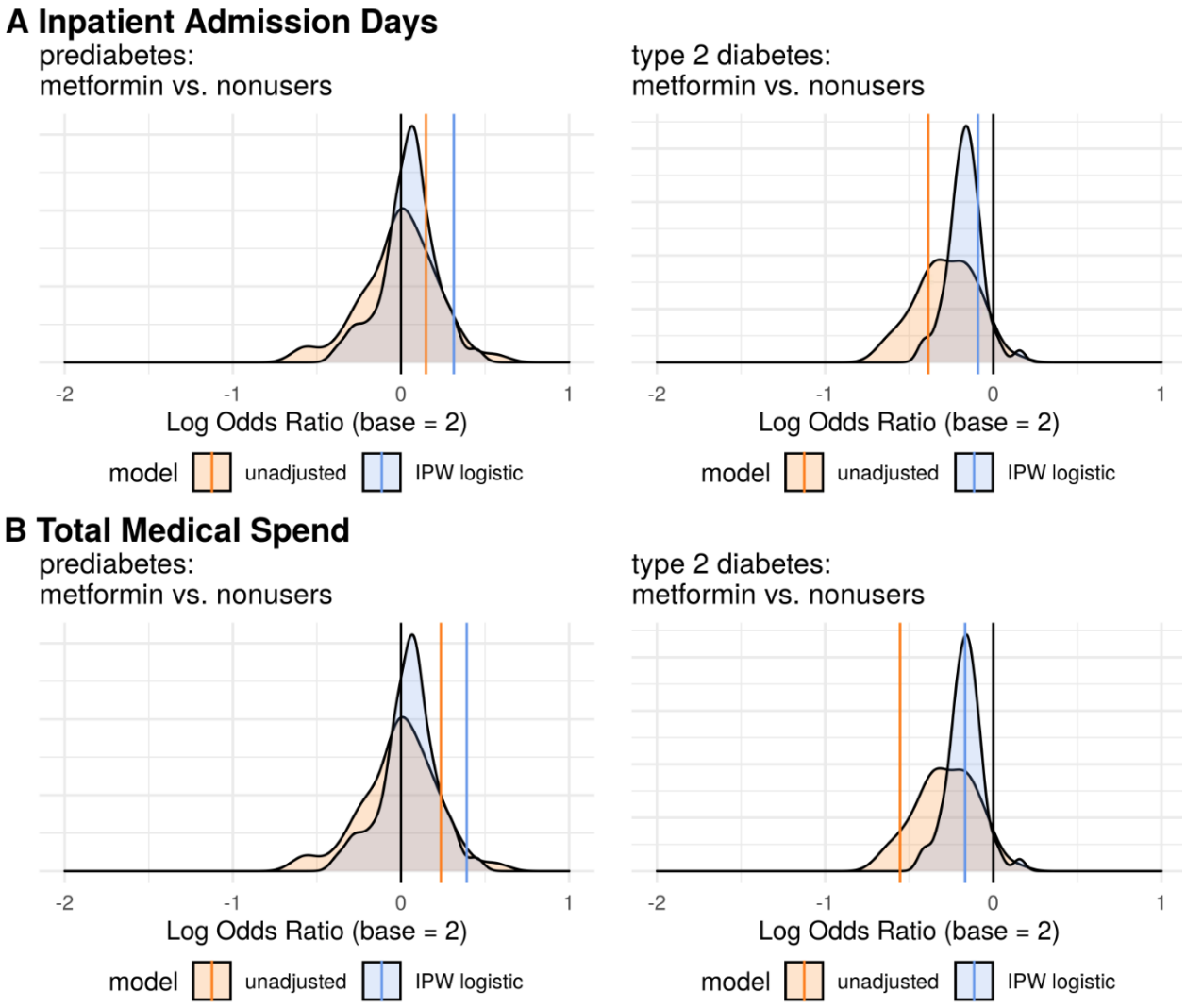
eFigure 4. Covariate Balance in Example Outcomes: Medicare Advantage Beneficiaries (metformin users vs. nonusers). This is a different comparison group compared to Figure 1, and in both the prediabetes and type 2 diabetes cohorts, there appears to be a slight health advantage among the metformin users (pre-adjustment standardized mean differences <0). In all cases, the post-adjustment balance is excellent.



eFigure 5. Covariate Balance in Negative Control Outcome Experiments: Medicare Advantage Beneficiaries (metformin users vs. nonusers). This is a different comparison group compared to eFigure 2, and in both the prediabetes and type 2 diabetes cohorts, there appears to be a slight health advantage among the metformin users (pre-adjustment standardized mean differences <0). In all cases, the post-adjustment balance is excellent.



eFigure 6. Residual Confounding Plot: Medicare Advantage Beneficiaries (metformin users vs. nonusers of any diabetes drug). This is a different comparison group compared to Figure 2 in the main text. The residual confounding again appears consistent with the primary analysis, strongly favoring metformin users in the type 2 diabetes cohort while maintaining a smaller bias against metformin in the prediabetes cohort.



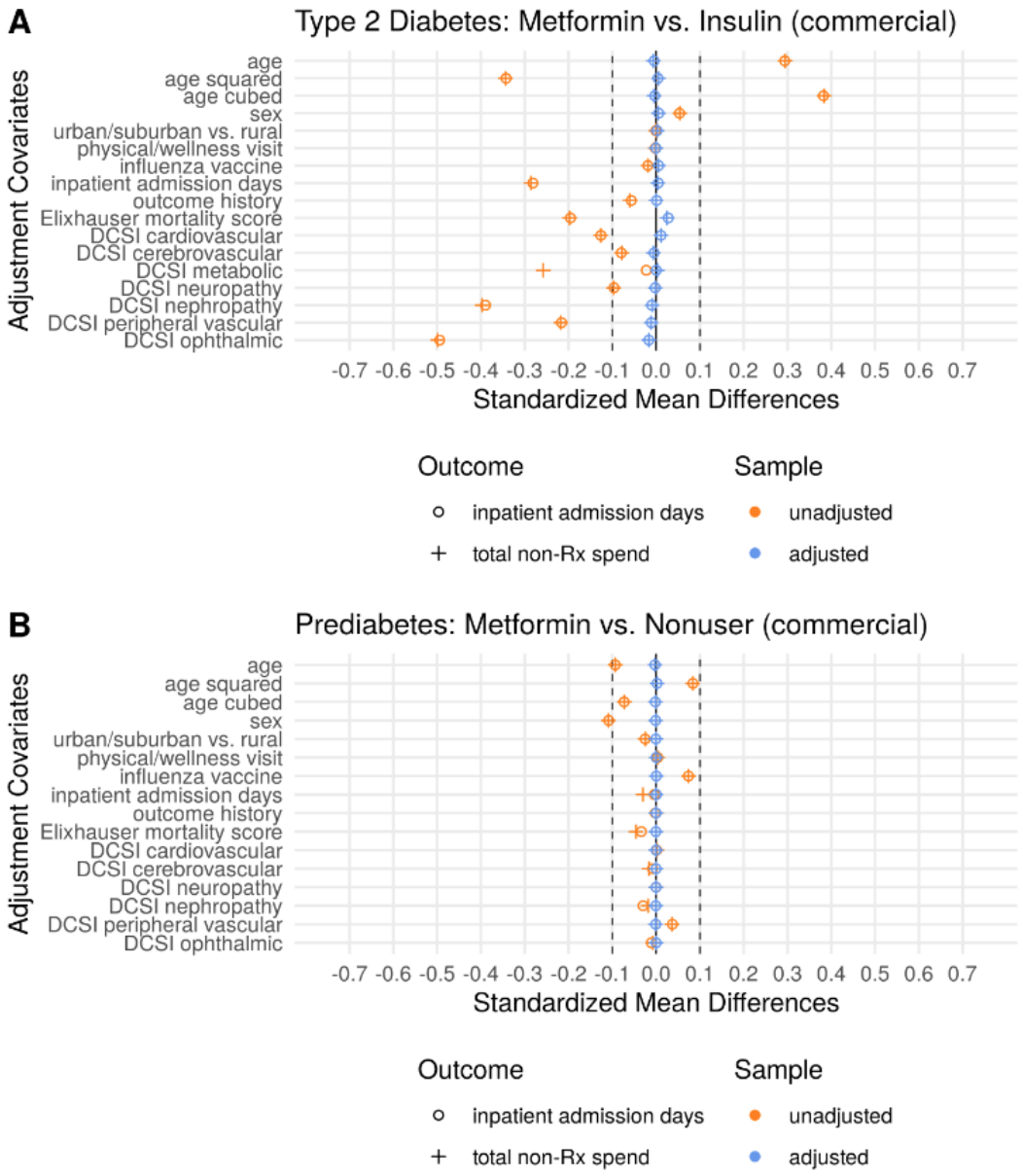
eFigure 7. Primary Outcome Results: Medicare Advantage Beneficiaries (metformin users vs. nonusers of any diabetes drug). This is a different comparison group compared to Figure 4 in the main text. We see adjusted treatment effect estimates that are at best on par with an average effect size for a negative control outcome in type 2 diabetes and considerably worse in prediabetes; together, these observations should elicit doubt about any claims of a real effect in the primary analysis.

Commercial Insurance Beneficiaries (metformin users vs. insulin users)

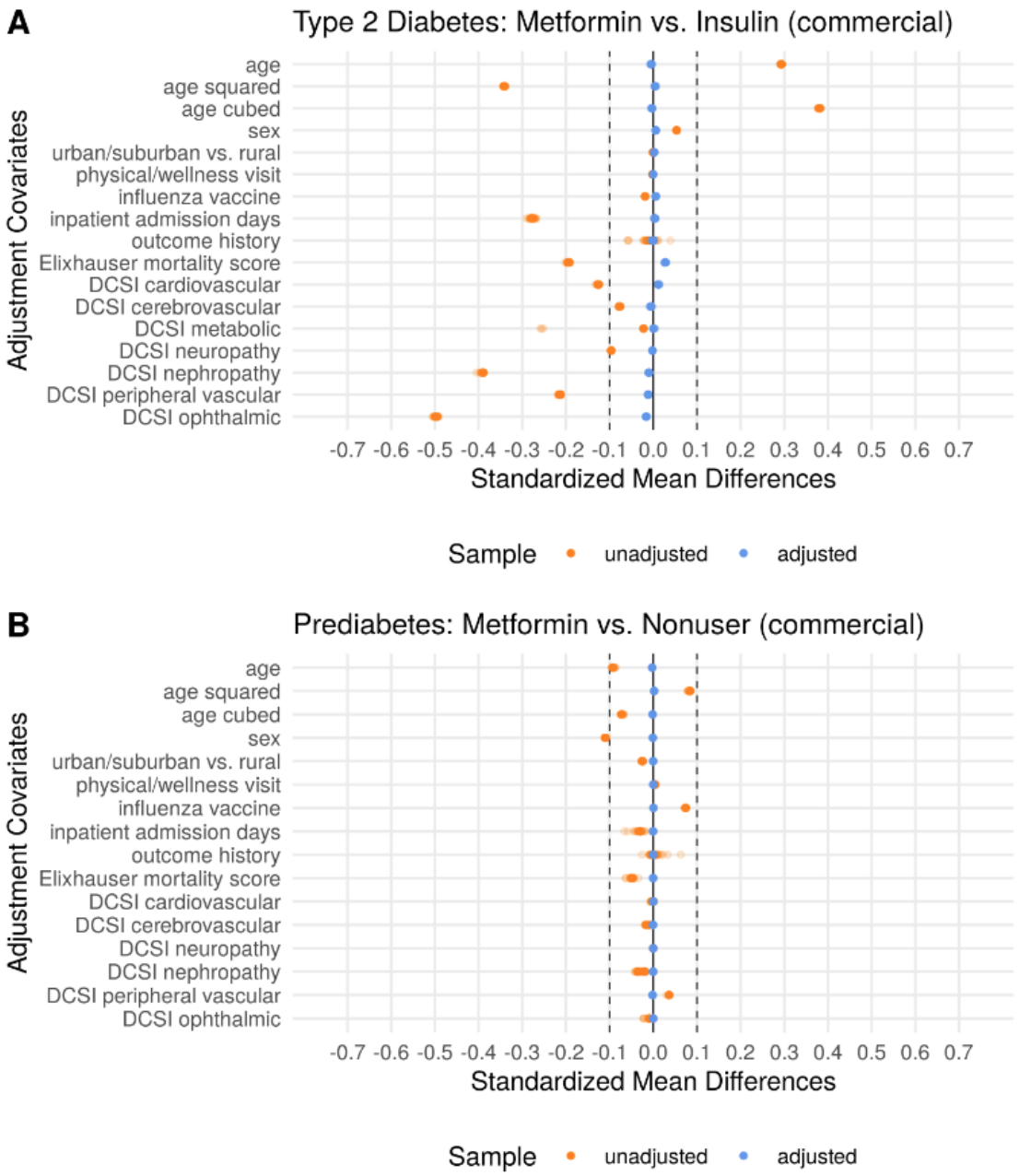
In an alternate population of commercially insured beneficiaries, we see strong results in eTable 4 favoring metformin usage that are quite confident despite the much smaller population under study. eFigure 8 and eFigure 9 show excellent covariate balance, but we continue to see in eFigure 10 negative control outcome effect estimates biased in favor of metformin in the type 2 diabetes population and biased against metformin in the prediabetes population. Interestingly, while the adjusted effect estimates for the example outcomes indicate a potential treatment effect that appears relatively strong compared to the negative control outcome effect estimates in eFigure 11, one effect goes to 0 while the other substantially reverses in the prediabetes population. Together, these results suggest that the findings in eTable 4 are likely products of significant residual confounding.

outcome	model	log OR (base 2)	95% CI	p	E-value
inpatient days	unadjusted	-1.49	(-1.62, -1.36)	$<10^{-106}$	5.07
inpatient days	IPW logistic	-0.71	(-0.89, -0.52)	$<10^{-13}$	2.65
medical spend	unadjusted	-1.51	(-1.63, -1.40)	$<10^{-134}$	5.15
medical spend	IPW logistic	-0.65	(-0.82, -0.48)	$<10^{-13}$	2.51

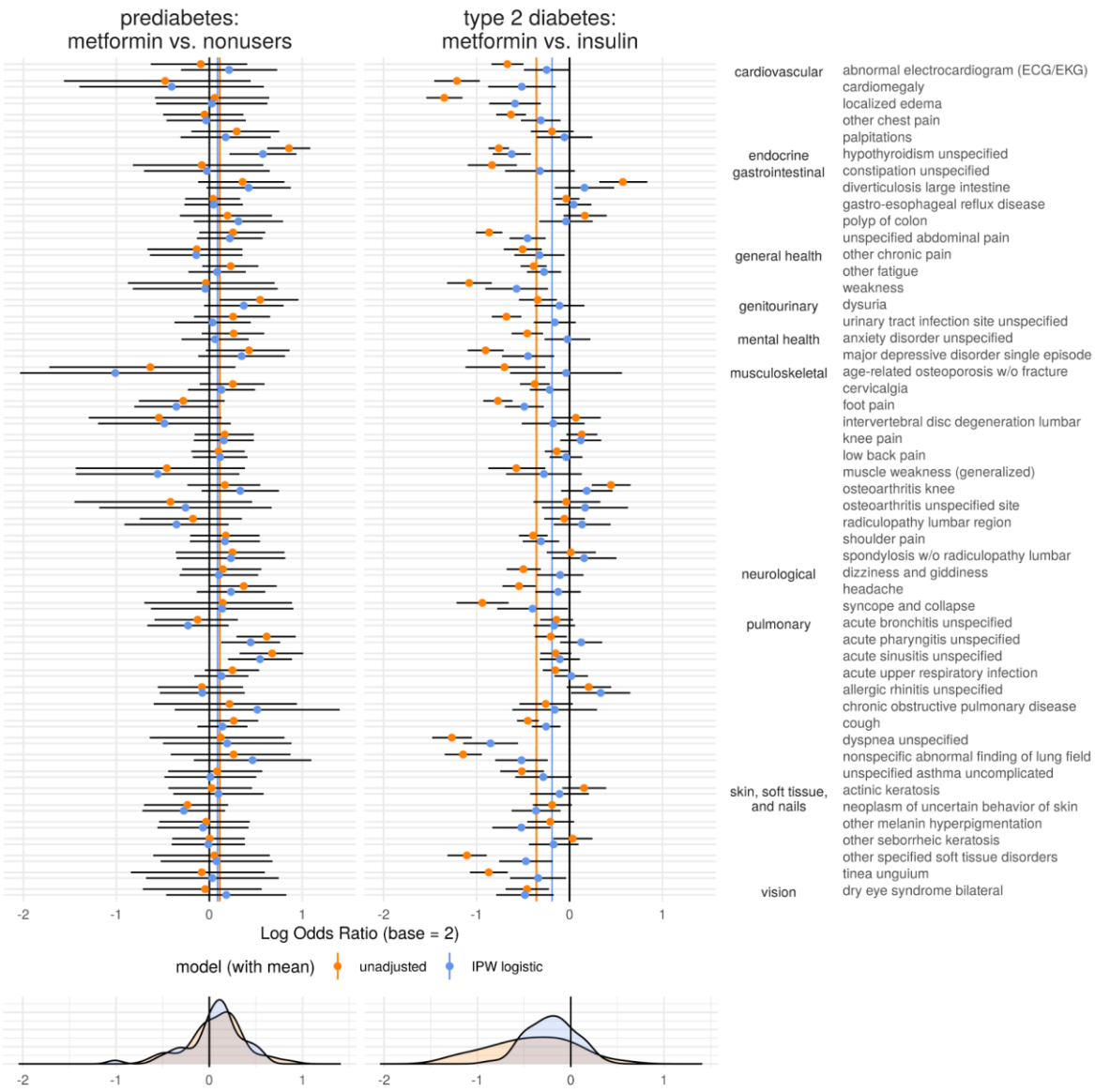
eTable 4. Treatment effect estimates for metformin: this study was conducted in a commercially insured type 2 diabetes population comparing metformin users to a control group of insulin users. The outcomes represent >0 inpatient admission days in 2019 and a total medical spend (insurance payouts to health care providers) exceeding the 90th percentile of all type 2 diabetes patient expenditures (>$21,433). Metformin appears strongly associated with fewer inpatient admission days and lower health care costs, even after adjustment for a range of relevant covariates.



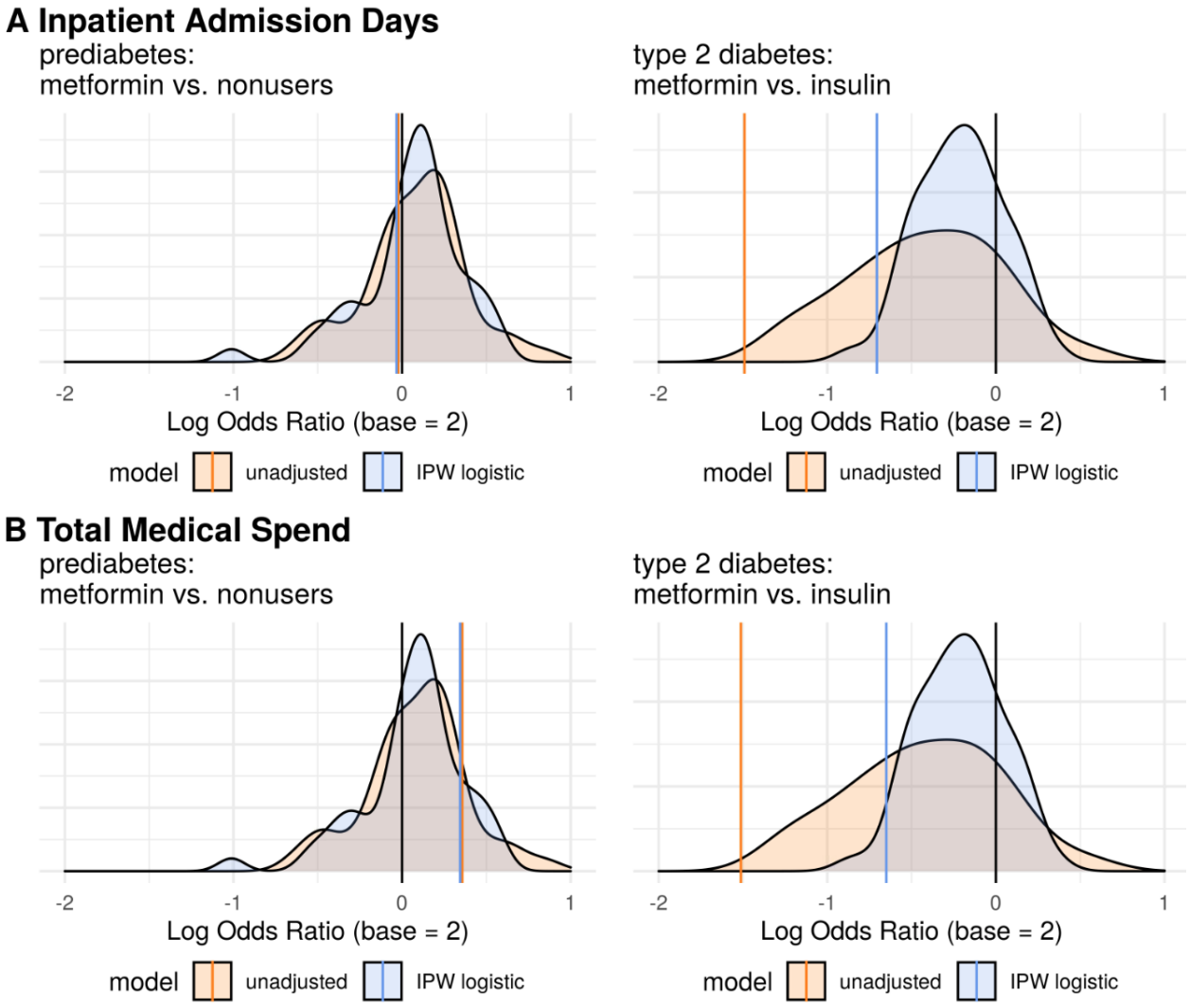
eFigure 8. Covariate Balance in Example Outcomes: Commercially Insured Beneficiaries (metformin users vs. insulin). This is a different population compared to Figure 1. In the type 2 diabetes cohort, the metformin users have a noticeable health advantage that essentially disappears in the prediabetes population. In both the prediabetes and type 2 diabetes cohorts, the post-adjustment balance is excellent.



eFigure 9. Covariate Balance in Negative Control Outcome Experiments: Commercial Insurance Beneficiaries (metformin users vs. insulin users). This is a different population compared to eFigure 2. Here we show 50 overlaid balance plots for the type 2 diabetes cohort (panel A) and the prediabetes cohort (panel B) where the only covariate changing from one negative control outcome experiment to the next is the history of each respective outcome. Excellent post-adjustment covariate balance is achieved for all negative control outcome experiments.



eFigure 10. Residual Confounding Plot: Commercially Insured Beneficiaries (metformin users vs. insulin users). This is a different population compared to Figure 3 in the main text. The residual confounding again appears consistent with the primary analysis, strongly favoring metformin users in the type 2 diabetes cohort while appearing to work against metformin users in the prediabetes cohort.



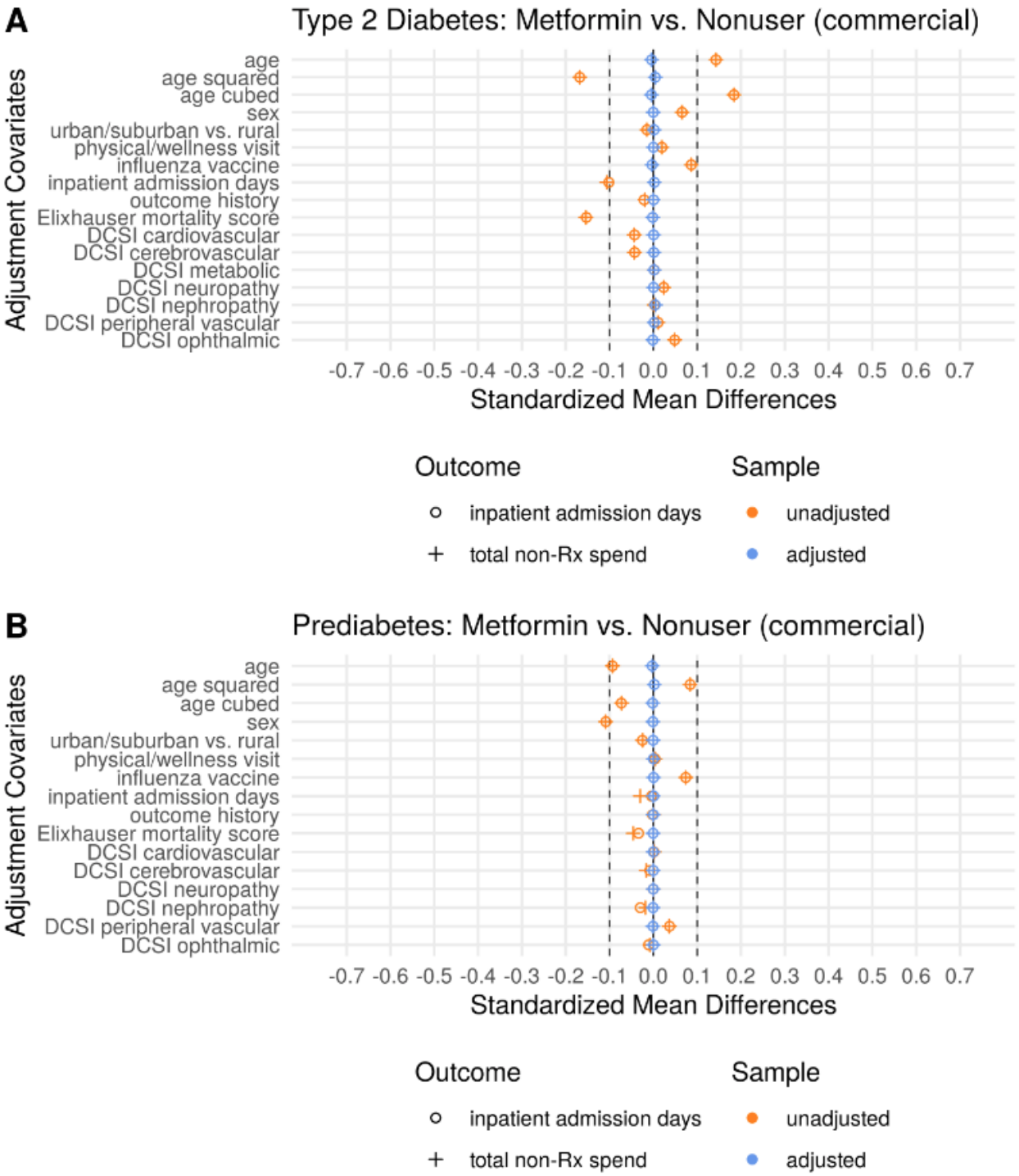
eFigure 11. Primary Outcome Results: Commercially Insured Beneficiaries (metformin users vs. insulin users). This is a different population compared to Figure 4 in the main text. We see adjusted treatment effect estimates that exceed a large majority of the estimated effect sizes from the negative control outcome experiments in type 2 diabetes. In prediabetes, however, these effect estimates become null in one case and substantially reverse in the other, which together should elicit some doubt about any claims in the primary analysis.

Commercial Insurance Beneficiaries (metformin users vs. nonusers)

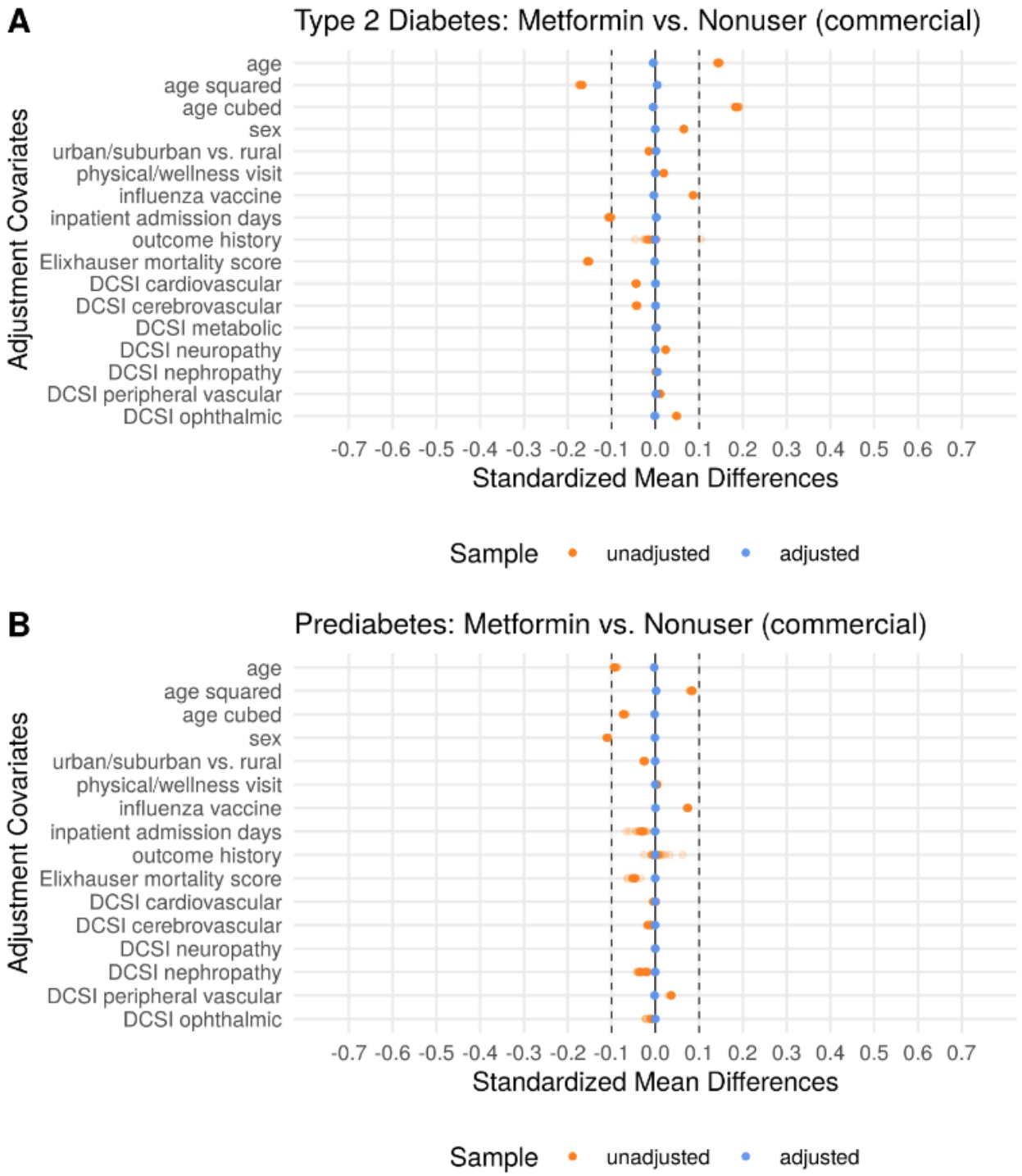
In an alternate comparison group of nonusers of any diabetes drugs among the commercially insured, we see more conservative results favoring metformin usage in eTable 5. While the nonusers may be hard to completely explain, we can be relatively confident they are not insulin users (and thus more advanced type 2 diabetes cases) due to the generally high list price of insulins. eFigure 12 and eFigure 13 show excellent covariate balance, but we continue to see in eFigure 14 negative control outcome effect estimates biased in favor of metformin in the type 2 diabetes population and biased against metformin in the prediabetes population. Interestingly, while the adjusted effect estimates for the example outcomes indicate a potential treatment effect with varying levels of confidence, neither type 2 diabetes effect estimate is stronger than even half of the negative control outcome effect estimates in eFigure 15. This observation strongly challenges the results in eTable 5 as nothing more than products of significant residual confounding.

outcome	model	log OR (base 2)	95% CI	p	E-value
inpatient days	unadjusted	-0.39	(-0.49, -0.29)	$<10^{-13}$	1.95
inpatient days	IPW logistic	-0.15	(-0.26, -0.04)	<0.01	1.45
medical spend	unadjusted	-0.26	(-0.35, -0.17)	$<10^{-7}$	1.69
medical spend	IPW logistic	-0.09	(-0.19, 0.01)	0.08	1.33

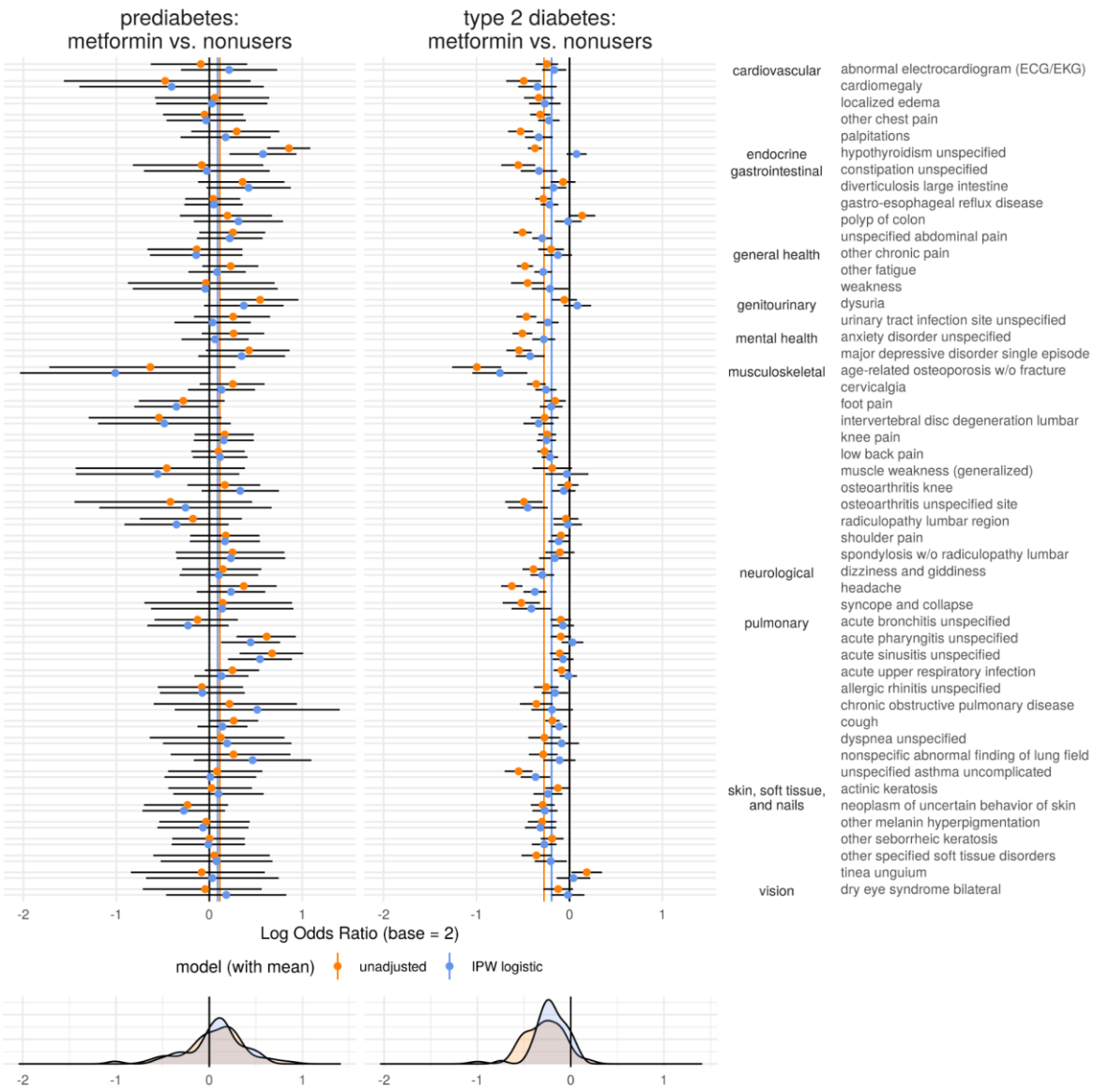
eTable 5. Treatment effect estimates for metformin: this study was conducted in a commercially insured type 2 diabetes population comparing metformin users to a control group of nonusers. The outcomes represent >0 inpatient admission days in 2019 and a total medical spend (insurance payouts to health care providers) exceeding the 90th percentile of all type 2 diabetes patient expenditures (>$21,433). Metformin appears strongly associated with fewer inpatient admission days and lower health care costs, even after adjustment for a range of relevant covariates.



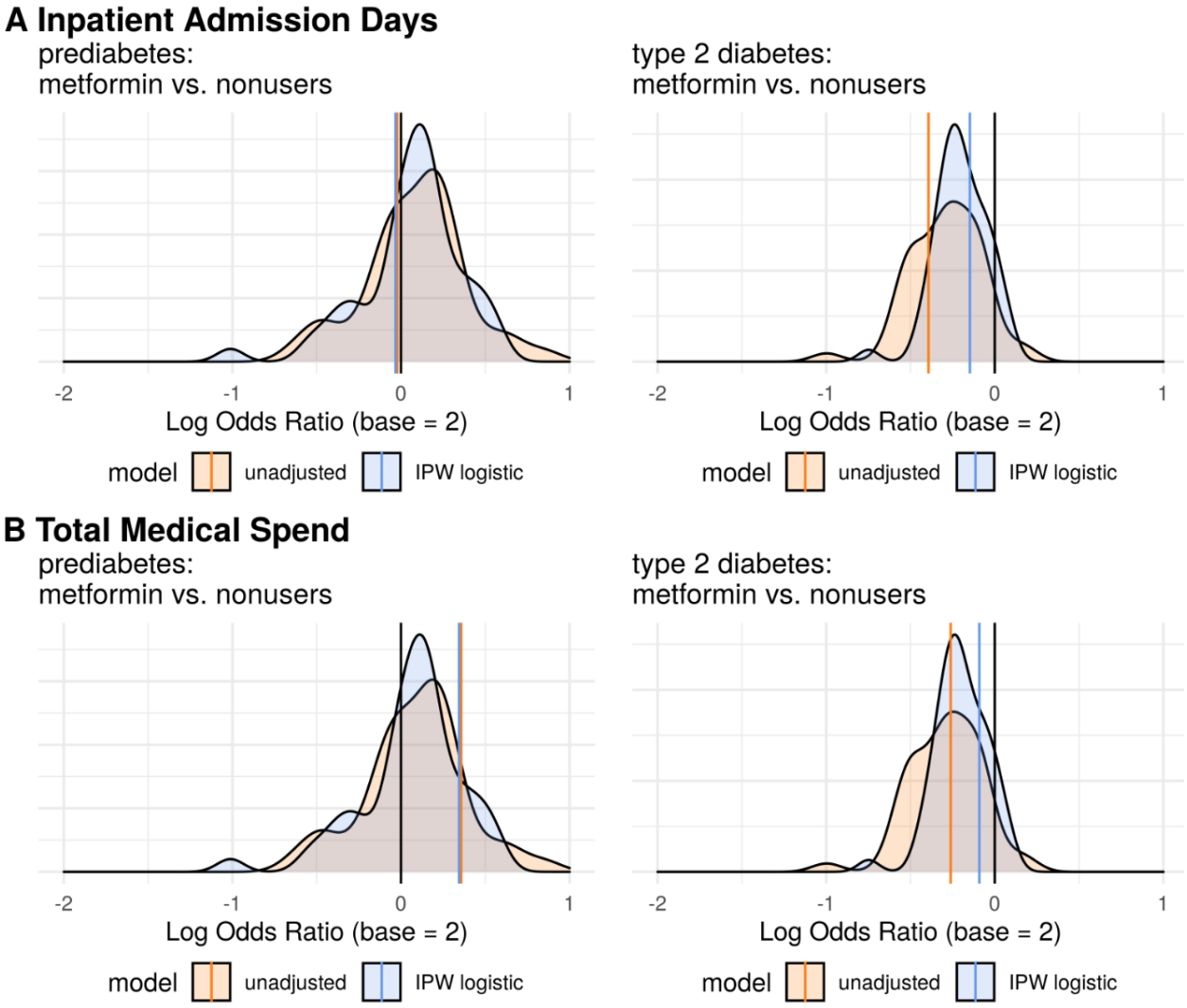
eFigure 12. Covariate Balance in Example Outcomes: Commercial Insurance Beneficiaries (metformin users vs. nonusers). This is the same analysis as Figure 1 in the main text, but it considers a different population and comparison group. Compared to Figure 1, the unadjusted balance in the type 2 diabetes population indicates metformin users are much more comparable to nonusers than insulin users.



eFigure 13. Covariate Balance in Negative Control Outcome Experiments: Commercial Insurance Beneficiaries (metformin users vs. nonusers). This is a different population and comparison group compared to eFigure 2. The type 2 diabetes metformin users and nonusers are considerably better balanced pre-adjustment here compared to eFigure 2, but that is not enough to eliminate the bias we see in eFigure 14.



eFigure 14. Residual Confounding Plot: Commercial Insurance Beneficiaries (metformin users vs. nonusers of any diabetes drug). This is the different population and comparison group compared to Figure 3 in the main text. The effect sizes are more conservative with this comparison group (though still biased to favor metformin in type 2 diabetes and oppose metformin in prediabetes), possibly due to the number of metformin users mixed into the nonuser population.



eFigure 15. Primary Outcome Results: Commercial Insurance Beneficiaries (metformin users vs. nonusers of any diabetes drug). This is a different population (commercially insured) compared to Figure 4 in the main text, and it considers a different type 2 diabetes comparison group: nonusers. In this population and comparison group setting, the adjusted effect estimates in the type 2 diabetes setting are not even as favorable as what we observe for an average negative control outcome, which is an immediate indicator that the observed association may be spurious.

Deviations from Preregistration

The preregistered analysis plan for this study can be found at https://osf.io/qf49p.

Deviations from this plan are listed and explained below:

1. Covariates - The preregistration states that “health-seeking behavior will be indicated by the presence of at least one immunization (typically a flu shot).” We instead used *only* flu shots because of the widespread eligibility and anticipated annual frequency of flu shots not common to all vaccinations.
2. Outcomes - We did not specify any primary outcomes for the example analysis in the preregistration. We realized after conducting all the negative control experiments that we were missing the opportunity to illustrate interpreting a real result. We only ever tried two example outcomes, and both are reported in the four analyses spanning the main text and supplement. As the stated criteria in the supplement state, the negative control outcomes should not be known indicators of type 2 diabetes severity. We thus removed “essential (primary) hypertension” and “hyperlipidemia/hypercholesterolemia” from our list of negative control outcomes due to their known association with cardiovascular comorbidities, an index component of the Diabetes Complications and Severity Index. We replaced those two negative control outcomes with the two next-most prevalent outcomes satisfying our criteria: “syncope and collapse” and “unspecified asthma uncomplicated.”

Supplementary Tables to Support Replication

eTable 6. Cohort Criteria

Cohort	ICD-10 Codes
Prediabetes	R73%
Type 2 Diabetes	E11%

eTable 7. Drug Class Members

Medication Class	Generic Name
Biguanides (metformin)	alogliptin-metformin hcl, canagliflozin-metformin hcl, dapagliflozin-metformin hcl, empagliflozin-metformin hcl, ertugliflozin-metformin hcl, glipizide-metformin hcl, glyburide-metformin, linagliptin-metformin hcl, metformin hcl, pioglitazone hcl-metformin hcl, repaglinide-metformin hcl, rosiglitazone maleate-metformin hcl, saxagliptin-metformin hcl, sitagliptin-metformin hcl
Insulins	insulin aspart, insulin aspart (with niacinamide), insulin aspart protamine & aspart (human), insulin glulisine, insulin lispro, insulin lispro protamine & lispro, insulin lispro-aabc, insulin nph isophane & reg (human), insulin reg (human) buffered,insulin regular, insulin regular (human), insulin regular (human) in sodium chloride, insulin regular (pork), insulin degludec, insulin degludec-liraglutide, insulin detemir, insulin glargine, insulin glargine-lixisenatide, insulin isophane, insulin isophane (pork), insulin nph (human) (isophane), insulin zinc, insulin zinc (human), insulin zinc (pork), insulin zinc extended (human)
Other Diabetes Drugs	pioglitazone hcl/glimepiride, glyburide, chlorpropamide, glipizide, glimepiride, tolbutamide, tolazamide, pioglitazone hcl, rosiglitazone maleate, miglitol, acarbose, pramlintide acetate, bromocriptine mesylate, sitagliptin phosphate, linagliptin, alogliptin benzoate, saxagliptin hcl, ertugliflozin/sitagliptin, dapagliflozin/saxagliptin hcl, empagliflozin/linagliptin, alogliptin benz/pioglitazone, dulaglutide, exenatide microspheres, lixisenatide, exenatide, albiglutide, liraglutide, semaglutide, nateglinide, repaglinide, empagliflozin, canagliflozin, dapagliflozin propanediol, ertugliflozin pidolate

eTable 8. Covariate Logic

Covariate	**Logic**
Influenza Vaccination	American Hospital Formulary Service (AHFS) therapeutic class code 80120000 and generic name containing “flu” (pharmacy claims) -OR- any of the following procedure codes: 90630, 90653, 90656, 90662, 90673, 90674, 90682, 90685, 90686, 90687, 90688, 90756, Q2039, Q2035, Q2037 (medical claims)
Physician Visit / Wellness Visit	either of the following health care encounter service type descriptions: “physician visits” or “wellness visits” (medical claims)
Elixhauser In-hospital Mortality Score	The Elixhauser In-Hospital Mortality Score follows the guidelines presented by Moore et al.[18]
Diabetes Complications Severity Index (DCSI)	The DCSI component scores follow logic presented by Glasheen et al. (appendices A1-A7, B).[19] In addition to standard insurance claims data, the database used for the study also had the necessary lab results to compute the nephrology component score per the indicated reference.

eTable 9. Negative Control Outcomes (50 total)

Number	ICD Code(s)	Condition
1	b351	tinea unguium
2	d485	neoplasm of uncertain behavior of skin
3	e039	hypothyroidism unspecified
4	f329	major depressive disorder single episode unspecified
5	f419	anxiety disorder unspecified
6	g8929	other chronic pain
7	h04123	dry eye syndrome of bilateral lacrimal glands
8	i517	cardiomegaly
9	j0190	acute sinusitis unspecified
10	j029	acute pharyngitis unspecified
11	j069	acute upper respiratory infection unspecified
12	j209	acute bronchitis unspecified
13	j309	allergic rhinitis unspecified
14	j449	chronic obstructive pulmonary disease unspecified
15	j45909	unspecified asthma uncomplicated
16	k219	gastro-esophageal reflux disease without esophagitis
17	k5730	diverticulosis of large intestine without perforation or abscess without bleeding
18	k5900	constipation unspecified
19	k635	polyp of colon
20	l570	actinic keratosis
21	l814	other melanin hyperpigmentation
22	l821	other seborrheic keratosis
23	m170	bilateral primary osteoarthritis of knee
--	m1711	unilateral primary osteoarthritis right knee
--	m1712	unilateral primary osteoarthritis left knee
24	m1990	unspecified osteoarthritis unspecified site
25	m25511	pain in right shoulder
--	m25512	pain in left shoulder
26	m25561	pain in right knee
--	m25562	pain in left knee

--	m25569	pain in unspecified knee
27	m47816	spondylosis without myelopathy or radiculopathy lumbar region
28	m5136	other intervertebral disc degeneration lumbar region
29	m5416	radiculopathy lumbar region
30	m542	cervicalgia
31	m545	low back pain
32	m6281	muscle weakness (generalized)
33	m79671	pain in right foot
--	m79672	pain in left foot
34	m7989	other specified soft tissue disorders
35	m810	age-related osteoporosis without current pathological fracture
36	n390	urinary tract infection site not specified
37	r002	palpitations
38	r05	cough
39	r0600	dyspnea unspecified
40	r0789	other chest pain
41	r109	unspecified abdominal pain
42	r300	dysuria
43	r42	dizziness and giddiness
44	r51	headache
45	r531	weakness
46	r5383	other fatigue
47	r55	syncope and collapse
48	r600	localized edema
49	r918	other nonspecific abnormal finding of lung field
50	r9431	abnormal electrocardiogram [ecg] [ekg]