Epidemiology Publish Ahead of Print

DOI: 10.1097/EDE.0000000000001626

Evaluating metformin strategies for cancer prevention: a target trial emulation using electronic health records

Barbra A. Dickerman,[1,2] Xabier García-Albéniz,[1,2,3] Roger W. Logan,[1,2] Spiros Denaxas,[4,5,6] Miguel A. Hernán[1,2,7]

[1] CAUSALab, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US

[2] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US

[3] RTI Health Solutions, Barcelona, Spain

[4] Institute of Health Informatics Research, University College London, London, UK

[5] Health Data Research UK (HDR UK) London, University College London, London, UK

[6] The Alan Turing Institute, London, UK

[7] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US

Correspondence: Dr. Barbra A. Dickerman, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, 8th Floor, Department of Epidemiology, Boston, MA, 02115. Email: bad788@mail.harvard.edu.

Type of manuscript: Original Research Article

Running head: Emulating target trials of metformin and cancer

Conflicts of interest: None declared.

Source of funding: This work was supported by grant R00 CA248335 to B.A.D. from the National Institutes of Health.

Acknowledgments: The authors thank Dr. Sonia Hernández-Diaz, Harvard T.H. Chan School of Public Health, Boston, MA, for helpful discussions related to this work.

Keywords: Cancer prevention, causal inference, comparative effectiveness, electronic health records, intention-to-treat analysis, inverse-probability weighting, per-protocol analysis, pharmacoepidemiology, target trial

Data availability: This study is based in part on data from the CPRD obtained under license from the UK Medicines and Healthcare Products Regulatory Agency. The data are provided by patients and collected by the UK National Health Service as part of their care and support. The interpretation and conclusions contained in this study are those of the authors alone. Because electronic health records are classified as sensitive data by the UK Data Protection Act, information governance restrictions (to protect patient confidentiality) prevent data sharing via public deposition. Data are available with approval through the individual constituent entities controlling access to the data. Specifically, the primary care data can be requested via application to the CPRD (https://www.cprd.com).

Word counts: Abstract, 247; Text, 3,751.

ABSTRACT

Background: Metformin users appear to have a substantially lower risk of cancer than nonusers in many observational studies. These inverse associations may be explained by common flaws in observational analyses that can be avoided by explicitly emulating a target trial.

Methods: We emulated target trials of metformin therapy and cancer risk using population-based linked electronic health records from the UK (2009-2016). We included individuals with diabetes, no history of cancer, no recent prescription for metformin or other glucose-lowering medication, and hemoglobin A1C (HbA1c) <64 mmol/mol (<8.0%). Outcomes included total cancer and 4 site-specific cancers (breast, colorectal, lung, prostate). We estimated risks using pooled logistic regression with adjustment for risk factors via inverse-probability weighting. We emulated a second target trial among individuals regardless of diabetes status. We compared our estimates with those obtained using previously applied analytic approaches.

Results: Among individuals with diabetes, the estimated 6-year risk differences (metformin – no metformin) were -0.2% (95% CI: -1.6%, 1.3%) in the intention-to-treat analysis and 0.0% (95% CI: -2.1%, 2.3%) in the per-protocol analysis. The corresponding estimates for all site-specific cancers were close to zero. Among individuals regardless of diabetes status, these estimates were also close to zero and more precise. By contrast, previous analytic approaches yielded estimates that appeared strongly protective.

Conclusions: Our findings are consistent with the hypothesis that metformin therapy does not meaningfully influence cancer incidence. The findings highlight the importance of explicitly emulating a target trial to reduce bias in the effect estimates derived from observational analyses.

INTRODUCTION

Observational studies suggest that users of metformin, a first-line treatment for diabetes, have a substantially lower risk of cancer compared with nonusers.[1-10] The prospect of reducing cancer risk with a safe and affordable medication such as metformin is very appealing. However, secondary analyses of randomized trials in diabetes prevention suggest that metformin does not have a cancer-protective effect. The effect estimates from randomized trials are imprecise, and thus difficult to interpret conclusively, because they are based on a relatively small number of cases of total cancer.[11]

Evaluating metformin for the prevention of site-specific cancers using randomized trials may not be feasible given the large sample size and long follow-up that would be required. Observational datasets, such as the ones available in electronic health records, can be used to explicitly emulate (hypothetical) target trials that address these limitations. However, the use of observational databases requires adequate emulation procedures, including the comparison of clinically realistic treatment strategies, the avoidance of biases related to mishandling of time zero of follow-up (selection bias and immortal time bias), and sufficient adjustment for confounding for treatment initiation.[12,13]

Selection bias, due to the inclusion of prevalent users, and immortal time bias, due to the use of postbaseline treatment information to assign treatment groups, can be eliminated by a sound emulation of the target trial, as described in detail previously.[12,14] Confounding can be reduced, for example, by restricting the analysis to individuals with indications for treatment initiation if these indications are strong risk factors for the outcome of interest. Specifically, if diabetes (an indication for metformin initiation) were a risk factor for cancer, the observational analysis would restrict eligibility to individuals with diabetes to adjust for confounding by diabetes; otherwise, no restriction to individuals with diabetes would be necessary even though the prevalence of diabetes is expected to be much higher among individuals who receive metformin than among those who do not receive it.

In this study, we used a large database of linked electronic health records from primary care, hospitalizations, and mortality registrations to emulate target trials of clinically relevant strategies of metformin therapy for the prevention of total and site-specific cancer. We conducted separate analyses among individuals with type 2 diabetes and among individuals regardless of diabetes status.

METHODS

Specification of the target trials

We designed this observational analysis to emulate target trials (i.e., hypothetical pragmatic trials that would have answered the causal questions of interest) of metformin as compared with no metformin for the prevention of cancer. The key protocol components of these target trials are summarized in **Table 1**.

Eligibility criteria for the target trial among individuals with diabetes include age ≥30 years between April 1, 2009 and February 29, 2016, diagnosis of type 2 diabetes, no history of cancer (except nonmelanoma skin cancer), no metformin contraindication (hepatic or renal impairment or lactic acidosis), HbA1c <64 mmol/mol (<8.0%), no prescription for metformin or other glucose-lowering medication within the past year, at least 1 year of up-to-standard data in a Clinical Practice Research Database (CPRD) general practice (defined as high-quality data deemed suitable for use in research[15]), and at least 1 year of potential follow-up, as well as known HbA1c measured within the past year and known smoking and body-mass index measured within the past 4 years. Baseline is defined as the first month in which all eligibility criteria are met. The target trial among individuals regardless of diabetes status has the same eligibility criteria except for type 2 diabetes and otherwise shares the same protocol.

The dynamic strategies to be compared are (1) initiation of metformin therapy at baseline and continuation over follow-up until the development of a contraindication (hepatic or renal impairment or lactic acidosis) or cancer diagnosis and (2) no initiation of metformin therapy over follow-up until the development of an indication (HbA1c ≥64 mmol/mol [≥8.0%]). When clinically warranted during the follow-up (i.e., upon the development of these conditions), individuals and their clinicians would decide whether to start, stop, or switch therapy. These are clinically relevant strategies, in contrast to the static strategies evaluated in previous observational studies under which individuals were not allowed to deviate from their assigned treatment strategy when clinically appropriate.[16,17]

The outcomes of interest are incident total cancer and the 4 most common site-specific invasive cancers in this population: female breast, colorectal, lung (non-small cell), and prostate. Previous validation studies have confirmed 95% of cancers recorded in this database.[18]

For each eligible individual, follow-up starts at treatment assignment (baseline) and ends upon the outcome of interest, death, loss to follow-up (transfer out of the practice, or incomplete follow-up [2 years after the last recorded lab prognostic factors or 4 years after the last recorded lifestyle prognostic factors]), 6 years after baseline, or the administrative end of follow-up (end of practice data collection or February 29, 2016), whichever happens first.

The causal estimands of interest are the intention-to-treat effect of being assigned to the treatment strategies and the per-protocol effect of adhering to them.

In the intention-to-treat analysis, risks (cumulative incidences) can be estimated nonparametrically via the Kaplan–Meier estimator or parametrically via a pooled logistic regression model for the monthly probability of the outcome that includes an indicator of assigned strategy, a flexible function of months since randomization (linear and quadratic terms), and a product term between the treatment indicator and time. The predicted values from this model are used to estimate 6-year cancer risks under each strategy. If the model also needs to include baseline covariates, the risks will be standardized to the distribution of the baseline covariates (see **eMethods 1 http://links.lww.com/EDE/C36** for details). The same model without the product term can be used to approximate the hazard ratio (for comparison with estimates from previous studies) because the monthly risk of the outcome is low.[19]

In the per-protocol analysis, this pooled logistic regression model is fit to the data after censoring individuals if and when they deviate from their assigned treatment strategy. Specifically, individuals in the initiator group are censored when they stop metformin (unless they develop a contraindication or cancer) and individuals in the non-initiator group are censored when they start metformin (unless they develop an indication). To adjust for factors associated with adherence, time-varying inverse-probability weights are estimated via a pooled logistic regression model for the monthly probability of treatment that includes baseline and time-varying factors. After the development of one of the above conditions, the weights for adherence remain constant until the end of follow-up. Estimated weights are truncated at their 99th percentile to prevent outliers from having an undue influence on the analyses.

Nonparametric bootstrapping with 500 samples can be used to calculate percentile-based 95% confidence intervals for risk estimates, and robust variances can be used to calculate conservative 95% confidence intervals for hazard ratio estimates.

To identify potential subgroups of individuals for whom the treatment strategies may be most beneficial, analyses are conducted separately in subsets of the eligible population defined at baseline according to age (<70 vs. ≥70 years), sex (male vs. female), and, in the target trial among individuals with diabetes, time since diabetes diagnosis (<1 vs. ≥1 year).

Emulation of the target trials

We explicitly emulated the target trials described above using observational data from the CPRD, Hospital Episode Statistics, and Office of National Statistics. These population-based datasets are comprised of longitudinal UK electronic health records from primary care consultations, admitted hospitalization episodes, and death registrations for approximately 15 million individuals, accessed through the CALIBER resource.[15,20] Longitudinal primary care data on demographics, lifestyle factors, symptoms, diagnoses, clinical examination findings, laboratory test results, referrals, and prescriptions were recorded by general practitioners in the CPRD. Hospitalization data were obtained through linkage with Hospital Episode Statistics. Mortality data were obtained through linkage with the Office of National Statistics. Disease phenotypes were derived using algorithms that combine information on diagnoses, symptoms, laboratory values, physiologic measures, prescriptions, and procedures, which were created and validated using an established methodology.[21,22]

We used the observational data to emulate each protocol component of the target trials as closely as possible (**Table 1**). We classified individuals into 1 of 2 treatment groups according to their prescription records at baseline and assumed these groups were exchangeable at baseline conditional on the covariates in **Table 2**. The analysis to estimate the observational analogues of the intention-to-treat and per-protocol effects proceeded as for the target trials, with adjustment for these baseline covariates to emulate randomization (and incorporation of their time-varying values into the inverse-probability weights for the per-protocol analysis, as in the target trial) and with sequential emulation for statistical efficiency (see **eMethods 1** **http://links.lww.com/EDE/C36**).

Specifically, we emulated the target trial as a sequence of trials[23-25] starting at each of the 71 months between April 2009 and February 2015. This accommodates the fact that individuals may meet the eligibility criteria at several times over follow-up and is more statistically efficient than choosing just one of those times as time zero.[26] Separately for each of the 71 months, eligible individuals were classified into a treatment group and followed until the outcome of interest, death, loss to follow-up, 6 years after baseline, or the administrative end of follow-up, whichever happened first. We then conducted a pooled analysis over all 71 emulated trials and estimated the observational analogues of intention-to-treat and per-protocol effects.

Sensitivity analyses

We performed several sensitivity analyses to address potential misclassification, residual confounding, and selection bias. Specifically, we (1) increased the maximum gap between successive prescriptions from 30 to 60 days, (2) additionally adjusted for practice region (at the Strategic Health Authority level), family history of cancer, cancer screening in the past year, and influenza vaccination in the past year (a marker of health care seeking behavior) as potential confounders, (3) truncated weights at their 99.5th percentile, and (4) additionally applied weights for censoring due to loss to follow-up. We also allowed individuals with diabetes to discontinue metformin upon the initiation of insulin therapy. To explore the potential influence of reverse causation, we lagged treatment values by 6 months.

It might be argued that the protective effect of metformin reported by previous observational analyses is not metformin-specific but the result of glycemic control more generally. We therefore emulated a third target trial of intensification to metformin–sulfonylurea dual therapy (a second-line diabetes treatment that adds an oral hypoglycemic medication to metformin) vs. continuation of metformin monotherapy and cancer incidence, among individuals receiving metformin monotherapy for diabetes (see **eMethods 2 http://links.lww.com/EDE/C36** for details). The eligibility criteria are the same as those described above, except they require current use of metformin therapy and additionally require HbA1c ≥48 mmol/mol (≥6.5%), an indication that treatment intensification from metformin monotherapy to dual therapy may be needed. Given these eligibility criteria, all individuals in this comparison of first- and second-line treatments are expected to have a similar stage and severity of diabetes.

Conventional analyses

Some of the estimates from previous observational studies may be partly explained by 3 types of deviations from target trial emulation: (1) mishandling of time zero by comparing ever-users vs. never-users of metformin therapy over the follow-up,[27,28] (2) comparison of unrealistic (static) strategies of metformin vs. no metformin therapy, regardless of clinical indications for stopping or starting treatment,[17] and (3) failure to apply the same eligibility criteria to all treatment groups under study.[17,29] The latter occurred when comparing initiators of metformin monotherapy who had received no prescription for any glucose-lowering medication in the past 6 months (predominantly individuals diagnosed with diabetes recently) vs. initiators of metformin–sulfonylurea dual therapy who had received metformin monotherapy but no prescription for other glucose-lowering medications (predominantly individuals who were diagnosed with diabetes some time ago and for whom metformin was insufficient).[17] Previous studies subsequently censored individuals in each treatment group when they switched from their baseline treatment.[17] We replicated each of these analytic decisions in our own data among individuals with diabetes.

Ethical approval

The CPRD has been granted generic ethical approval for observational studies that make use of only anonymized data and linked anonymized National Health Service healthcare data (Multiple Research Ethics Committee ref. 05/MRE04/87). This study was approved by the Medicines and Healthcare Products Regulatory Agency Independent Scientific Advisory Committee (protocol 16_221) and the Harvard T.H. Chan School of Public Health Institutional Review Board.

RESULTS

Figure 1 shows a flowchart of patient selection, and **Table 2** shows the baseline characteristics of the 44,237 individuals eligible for the emulated trial among individuals with type 2 diabetes, and the 216,785 individuals eligible for emulated trial among individuals regardless of diabetes status. Compared with metformin non-initiators at baseline, metformin initiators were, on average, younger, had higher HbA1c and body-mass index, and included a higher proportion of individuals with any recent specialist referral. Among individuals with diabetes, metformin initiators also had a shorter time since diabetes diagnosis. Among individuals regardless of diabetes status (who satisfied the requirement for a recent HbA1c measurement, among the other eligibility criteria), 89% of metformin initiators and 34% of metformin non-initiators had type 2 diabetes (data not tabulated).

In the emulated trial among individuals with diabetes, 2,777 individuals developed cancer, including 272 female breast, 365 colorectal, 368 lung, and 416 prostate cancers, over the 6-year follow-up (median 3.3 years, interquartile range 2.0-4.9 years). In the emulated trial among individuals regardless of diabetes status, 7,507 individuals developed cancer, including

821 female breast, 934 colorectal, 1,001 lung, and 1,214 prostate cancers, over the 6-year follow-up (median 2.1 years, interquartile range 1.3-2.6 years).

Table 3 shows the estimated 6-year risks for cancer comparing metformin with no metformin. In the emulated trial among individuals with diabetes, the estimated observational analogue of the intention-to-treat 6-year risk difference was -0.2% (95% CI: -1.6%, 1.3%) for total cancer, and ranged from -0.4% to 0.7% across cancer sites. The estimated observational analogue of the per-protocol 6-year risk difference was 0.0% (95% CI: -2.1%, 2.3%) for total cancer, and ranged from -0.2% to 1.6% across cancer sites. Estimates were similar in the emulated trial among individuals regardless of diabetes status, but confidence intervals were narrower. Risk curves under each strategy were almost overlapping (**Figure 2**). Estimates for total cancer were similar (1) in subgroups defined at baseline according to age, sex, and time since diabetes diagnosis (**eTable 2 http://links.lww.com/EDE/C36**), (2) under several sensitivity analyses for potential misclassification, residual confounding, and selection bias due to loss to follow-up (**eTables 3-6 http://links.lww.com/EDE/C36**), (3) when allowing individuals with diabetes to discontinue metformin upon the initiation of insulin therapy (**eTable 7 http://links.lww.com/EDE/C36**), (4) when lagging treatment values by 6 months (intention-to-treat hazard ratio 1.08 among individuals with diabetes and among individuals regardless of diabetes status), and (5) when only adjusting for age (among individuals with diabetes: intention-to-treat hazard ratio 1.01, per-protocol hazard ratio 1.01; among individuals regardless of diabetes status: intention-to-treat hazard ratio 1.05, per-protocol hazard ratio 1.04); and identical when additionally adjusting for use of other glucose-lowering medications.

In the emulated trial among individuals receiving metformin monotherapy for diabetes, the estimated effect of treatment intensification to metformin–sulfonylurea dual therapy vs.

continuation of metformin monotherapy was also near null (intention-to-treat hazard ratio for total cancer 1.03, 95% CI: 0.71, 1.50; per-protocol hazard ratio for total cancer 0.89, 95% CI: 0.45, 1.75; see **eMethods 2 http://links.lww.com/EDE/C36**).

Conventional analyses

In analyses that replicated the analytic approaches of some previous observational studies in our data among individuals with diabetes, estimates were near null when we compared initiators of metformin monotherapy vs. initiators of metformin–sulfonylurea dual therapy, identified by applying different eligibility criteria to each treatment group (hazard ratio for total cancer 1.17, 95% CI: 0.92, 1.48). When we compared static treatment strategies of metformin vs. no metformin therapy, estimates for total cancer were near null (hazard ratio 0.97, 95% CI: 0.84, 1.13), but estimates for lung cancer were further from the null (hazard ratio 0.64, 95% CI: 0.39, 1.05) than in the primary analysis (**eTable 8 http://links.lww.com/EDE/C36**). Analyses that compared ever-users vs. never-users of metformin therapy over the follow-up also resulted in strong inverse associations (hazard ratio for total cancer 0.54, 95% CI: 0.49, 0.60) (data not tabulated). We found a similar pattern when comparing ever-users vs. never-users of sulfonylureas (hazard ratio for total cancer 0.67, 95% CI: 0.57, 0.79).

DISCUSSION

After emulating target trials using the electronic health records of 216,785 individuals, we found little indication that metformin therapy influences cancer incidence over the study period, regardless of whether eligibility was restricted to having diabetes. These findings are consistent with secondary analyses of randomized trials in diabetes prevention,[11] but inconsistent with previous observational studies that have reported a substantially lower risk of cancer among users of metformin compared with nonusers.[5,6,8,9]

The approach of explicitly specifying the protocol of the target trial and its observational emulation prevents common biases in observational analyses. Specifically, the extreme, apparently beneficial, effect estimates from previous observational studies may be partly explained by mishandling of time zero and by the comparison of unrealistic static strategies.

Unhitching eligibility assessment and treatment assignment from time zero can lead to substantial selection bias and immortal time bias, as previously discussed.[12,30-33] When we replicated this flaw by classifying individuals according to their observed treatment use over follow-up via a comparison of ever-users vs. never-users of metformin, we obtained an apparently protective estimate for cancer of an implausible magnitude (hazard ratio 0.54).

In the real world, treatment strategies are dynamic. Static strategies are unrealistic because they require that individuals continue taking treatment even after the onset of contraindications or toxicity. As a result, using real world data to compare static strategies can lead to positivity violations and bias. When we replicated this flaw by comparing static strategies of metformin vs. no metformin over follow-up, we found a more "protective" estimate for lung cancer than in our analyses comparing more realistic dynamic strategies (hazard ratio 0.64 vs. 0.75). Further exploration of this issue was limited by the imprecision of our estimates for lung cancer.

Following the basic principles of study design can avoid time-related biases in observational studies. Indeed, a systematic review showed that previous observational studies identified as least likely to be affected by these biases also suggested no effect of metformin on cancer risk, as in the present study.[7] The proposed target trial approach can be viewed as a guide to implement sound principles of causal inference and study design,[34] as well as a way to estimate appropriately adjusted measures of absolute risk and to evaluate clinically realistic dynamic treatment strategies. Our study had additional strengths. The electronic health records capture rich longitudinal data on demographic and clinical features that allowed us to characterize individuals with high resolution and adjust for many potential confounders. Unlike some previous studies on this topic, we were able to distinguish type 1 from type 2 diabetes, adjust for time since diabetes diagnosis and HbA1c to minimize potential confounding by disease duration and severity, and incorporate HbA1c into our eligibility criteria and treatment strategies to reduce potential confounding as well as positivity violations. Also, our approach to emulate a sequence of target trials is more statistically efficient than emulating a single target trial.[26]

Our study also had some potential limitations. First, as in any observational analysis, assignment to a treatment strategy was not randomized. If the 2 treatment groups had different distributions of risk factors, then the effect estimates would be confounded. However, much less confounding by indication is expected when evaluating unintended effects (e.g., cancer outcomes) vs. intended effects (e.g., coronary heart disease, death)[35]; the treatment groups were similar at baseline with respect to their demographic characteristics and medical history; and we adjusted for many potential baseline and time-varying confounders. Some unmeasured variables that may be imbalanced between the treatment groups are diet and physical activity (e.g., metformin non-initiators may have achieved diabetes control via improved diet and increased exercise) in the diabetes-only analysis and non-diabetes indications for metformin (e.g., polycystic ovary syndrome) in the general analysis. Smoking is a strong risk factor for lung cancer that was coarsely measured (as never, former, or current smoker); however, estimates for lung cancer were similar when only adjusting for age, suggesting a potentially limited role for confounding in this setting (intention-to-treat hazard ratio among individuals with diabetes 1.00

vs. 1.00 in the primary analysis, among individuals regardless of diabetes status 0.99 vs. 0.93 in the primary analysis). Second, we were limited by our reliance on prescription records and diagnosis codes, which may contribute to measurement error and residual confounding. However, previous validation studies have confirmed a high proportion of recorded cancers (95%) and other diagnoses in our study data.[18,36] Third, the length of follow-up may have been insufficient to capture slowly progressing cancers. However, previous observational studies with comparable or shorter follow-up have reported a substantially lower risk of total and site-specific cancers,[37] including prostate cancer,[38] among metformin users.

Most previous studies of metformin and cancer were restricted to individuals with diabetes, an indication for metformin initiation.[8] This restriction protects against bias that would arise if diabetes were a risk factor for cancer, conditional on the other measured clinical features. In the present study, estimates were similar regardless of whether eligibility was restricted to this indication though, as expected, estimates were more precise when not making this restriction. The choice of eligibility criteria when emulating any target trial using observational data will be guided by these considerations about the comparability of the treatment groups. When evaluating intended effects of treatment (e.g., statins and risk of death), confounding by indication may be a larger concern, and it may therefore be important to restrict eligibility to individuals with an indication for treatment (e.g., coronary heart disease). When evaluating unintended effects of treatment, as in the present study, confounding by indication may be a smaller concern[35] and lower variance may be achieved by omitting the indication from the eligibility criteria. Note that, to adjust for potential confounding by glycemic status, we required a recent measure of HbA1c as an eligibility criterion, even for individuals without diabetes, which increased the proportion of individuals in our study who had a reason to have their HbA1c assessed (e.g., a cardiometabolic disorder).

In summary, our findings suggest that metformin therapy does not meaningfully influence cancer incidence over 6 years. Our explicit emulation of a target trial helped to reduce bias that may contribute to discrepancies between the effect estimates derived from observational analyses and randomized trials. Our analysis also highlights how more precise effect estimates may be obtained by omitting the indication for treatment from the eligibility criteria in cases where this restriction is not necessary to achieve comparability between the 2 treatment groups, as will often be the case in observational analyses that evaluate unintended effects of medications.

AUTHOR CONTRIBUTIONS

B.A.D., X.G.-A., S.D. and M.A.H. conceived the overall study. All authors contributed to the design and analysis. B.A.D. analyzed the data. R.W.L. provided key input in processing data from the database. All authors contributed to the interpretation of the results. B.A.D. wrote the first draft of the manuscript, and all authors reviewed, revised, and approved the final version of the manuscript.

References

1. Chikermane SG, Sharma M, Abughosh SM, Aparasu RR, Trivedi MV, Johnson ML. Dose-dependent relation between metformin and the risk of hormone receptor-positive, her2-negative breast cancer among postmenopausal women with type-2 diabetes. *Breast Cancer Res Treat.* 2022;195(3):421-430.
2. Lee JW, Choi EA, Kim YS, et al. Metformin usage and the risk of colorectal cancer: a national cohort study. *Int J Colorectal Dis.* 2021;36(2):303-310.
3. Freedman LS, Agay N, Farmer R, Murad H, Olmer L, Dankner R. Metformin Treatment Among Men With Diabetes and the Risk of Prostate Cancer: A Population-Based Historical Cohort Study. *Am J Epidemiol.* 2022;191(4):626-635. PMC8971081
4. Zhang ZJ, Bi Y, Li S, et al. Reduced risk of lung cancer with metformin therapy in diabetic patients: a systematic review and meta-analysis. *Am J Epidemiol.* 2014;180(1):11-14.
5. Col NF, Ochs L, Springmann V, Aragaki AK, Chlebowski RT. Metformin and breast cancer risk: a meta-analysis and critical literature review. *Breast Cancer Res Treat.* 2012;135(3):639-646.
6. Zhang ZJ, Zheng ZJ, Kan H, et al. Reduced risk of colorectal cancer with metformin therapy in patients with type 2 diabetes: a meta-analysis. *Diabetes Care.* 2011;34(10):2323-2328. PMC3177711
7. Farmer RE, Ford D, Forbes HJ, et al. Metformin and cancer in type 2 diabetes: a systematic review and comprehensive bias evaluation. *Int J Epidemiol.* 2017;46(2):728-744. PMC5837266

8. Decensi A, Puntoni M, Goodwin P, et al. Metformin and cancer risk in diabetic patients: a systematic review and meta-analysis. *Cancer Prev Res (Phila).* 2010;3(11):1451-1461.

9. Noto H, Goto A, Tsujimoto T, Noda M. Cancer risk in diabetic patients treated with metformin: a systematic review and meta-analysis. *PLoS One.* 2012;7(3):e33411. PMC3308971

10. Franciosi M, Lucisano G, Lapice E, Strippoli GF, Pellegrini F, Nicolucci A. Metformin therapy and risk of cancer in patients with type 2 diabetes: systematic review. *PLoS One.* 2013;8(8):e71583. PMC3732236

11. Stevens RJ, Ali R, Bankhead CR, et al. Cancer outcomes and all-cause mortality in adults allocated to metformin: systematic review and collaborative meta-analysis of randomised clinical trials. *Diabetologia.* 2012;55(10):2593-2603.

12. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nature Medicine.* 2019;25:1601-1606.

13. Hernán MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70-75. PMC5124536

14. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol.* 2016;183(8):758-764. PMC4832051

15. Denaxas S, Gonzalez-Izquierdo A, Direk K, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc.* 2019;26(12):1545-1559.

16. Farmer RE, Ford D, Mathur R, et al. Metformin use and risk of cancer in patients with type 2 diabetes: a cohort study of primary care records using inverse probability weighting of marginal structural models. *Int J Epidemiol.* 2019;48(2):527-537. PMC6469299

17. Currie CJ, Poole CD, Gale EA. The influence of glucose-lowering therapies on cancer risk in type 2 diabetes. *Diabetologia.* 2009;52(9):1766-1777.

18. Margulis AV, Fortuny J, Kaye JA, et al. Validation of cancer cases using primary care, cancer registry, and hospitalization data in the United Kingdom. *Epidemiology.* 2018;29(2):308-313. PMC5794229

19. Thompson WA, Jr. On the treatment of grouped observations in life studies. *Biometrics.* 1977;33(3):463-470.

20. Denaxas SC, George J, Herrett E, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol.* 2012;41(6):1625-1638. PMC3535749

21. Morley KI, Wallace J, Denaxas SC, et al. Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. *PLoS One.* 2014;9(11):e110900. PMC4219705

22. Kuan V, Denaxas S, Gonzalez-Izquierdo A, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *The Lancet Digital Health.* 2019;1(2):e63-e77.

23. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med.* 2019;25(10):1601-1606. PMC7076561

24. Danaei G, Garcia Rodriguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol.* 2018;96:12-22. PMC5847447

25. Danaei G, Rodriguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res.* 2013;22(1):70-96. PMC3613145

26. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol.* 2017;32(6):495-500. PMC5759953

27. Geraldine N, Marc A, Carla T, et al. Relation between diabetes, metformin treatment and the occurrence of malignancies in a Belgian primary care setting. *Diabetes Res Clin Pract.* 2012;97(2):331-336.

28. Lai SW, Liao KF, Chen PC, Tsai PY, Hsieh DP, Chen CC. Antidiabetes drugs correlate with decreased risk of lung cancer: a population-based observation in Taiwan. *Clin Lung Cancer.* 2012;13(2):143-148.

29. Bowker SL, Majumdar SR, Veugelers P, Johnson JA. Increased cancer-related mortality for patients with type 2 diabetes who use sulfonylureas or insulin. *Diabetes Care.* 2006;29(2):254-258.

30. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care.* 2012;35(12):2665-2673. PMC3507580

31. Hernández-Diaz S, Adami HO. Diabetes therapy and cancer risk: causal effects and other plausible explanations. *Diabetologia.* 2010;53(5):802-808.

32. Hernández-Diaz S. Name of the bias and sex of the angels. *Epidemiology*. 2011;22(2):232-233.

33. Golozar A, Liu S, Lin JA, Peairs K, Yeh HC. Does Metformin Reduce Cancer Risks? Methodologic Considerations. *Curr Diab Rep.* 2016;16(1):4.

34. Hernán MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. *N Engl J Med.* 2021;385(15):1345-1348.

35. Miettinen OS. The need for randomization in the study of intended effects. *Stat Med.* 1983;2(2):267-271.

36. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol.* 2010;69(1):4-14. PMC2805870

37. Lee MS, Hsu CC, Wahlqvist ML, Tsai HN, Chang YH, Huang YC. Type 2 diabetes increases and metformin reduces total, colorectal, liver and pancreatic cancer incidences in Taiwanese: a representative population prospective cohort study of 800,000 individuals. *BMC Cancer.* 2011;11:20. PMC3031263

38. Tseng CH. Diabetes and risk of prostate cancer: a study using the National Health Insurance. *Diabetes Care.* 2011;34(3):616-621. PMC3041193

FIGURE LEGENDS

Figure 1. Selection and flow of eligible individuals when emulating a target trial of metformin therapy and cancer risk (a) among individuals regardless of diabetes status and (b) among individuals with diabetes, 2009-2016. Panel B shows the flow of individuals after applying the additional eligibility criterion of type 2 diabetes. Numbers in parentheses represent unique individuals in each group. Counts of initiator and non-initiator individuals do not sum to the total number of eligible individuals because some eligible individuals contributed to both groups in different nested trials.

Figure 2. Estimated risk of cancer by metformin therapy among individuals with diabetes (observational analogue to an intention-to-treat [a] and per-protocol [b] analysis), and among individuals regardless of diabetes status (observational analogue to an intention-to-treat [c] and per-protocol [d] analysis), using linked electronic health records from Clinical Practice Research Datalink, Hospital Episode Statistics, and Office of National Statistics, 2009-2016. Shaded areas represent pointwise 95% confidence intervals.

Table 1. Specification and emulation of pragmatic target trials of metformin therapy and cancer risk using linked electronic health records from Clinical Practice Research Datalink, Hospital Episode Statistics, and Office of National Statistics.

Protocol	Target trial specification	Target trial emulation
Eligibility criteria	Target trial among individuals with diabetes • Aged ≥30 years between April 1, 2009, and February 29, 2016 • Type 2 diabetes mellitus (ascertained using diagnosis codes) • No history of cancer (except nonmelanoma skin cancer) • No metformin contraindication (hepatic or renal impairment or lactic acidosis). Hepatic impairment is ascertained using a diagnosis code for hepatic failure or ALT ≥120 IU/L; renal impairment using a diagnosis code for renal failure, end-stage renal disease, or eGFR <30 mL/min/1.73m^2 using the MDRD equation[a]; and lactic acidosis using a diagnosis code for lactic acidosis. • HbA1c <64 mmol/mol (<8.0%) • No prescription for metformin or other glucose-lowering medication within the past year • At least 1 year of up-to-standard data in a Clinical Practice Research Datalink general practice • At least 1 year of potential follow-up, based on the planned end of follow-up on February 29, 2016 • Information on lab values (HbA1c) measured during the past year and lifestyle factors (body-mass index, smoking status) during the past 4 years Target trial among individuals regardless of diabetes status • All criteria from target trial 1, except for type 2 diabetes mellitus	Same as for the target trial.
Treatment strategies	(1) Initiation of metformin at baseline and continuation over follow-up until the development of a contraindication (hepatic or renal impairment or lactic acidosis) or diagnosis of cancer (2) No initiation of metformin over follow-up until the development of an indication (HbA1c ≥64 mmol/mol [≥8.0%]) Treatment is considered continuous if there is a gap of <30 days between successive prescriptions. When clinically warranted during the follow-up (i.e., upon the development of these indications and contraindications), patients and their physicians will decide whether to start, stop, or switch therapy. Participants must have a primary care consultation at least once every 2 years to assess lab prognostic factors and at least once every 4 years to assess lifestyle prognostic factors associated with adherence and loss to follow-up.	Same as for the target trial. We defined the date of medication initiation to be the first date of a prescription. We calculated discontinuation dates using the daily dose and quantity of pills in the prescription.
Treatment assignment	Individuals are randomly assigned to a strategy at baseline. Individuals and their treating physicians will be aware of the assigned treatment strategy.	We classified individuals into 1 of 2 groups according to the strategy that their data were compatible with at baseline and assumed randomization conditional on baseline covariates.
Outcomes	Total cancer and the 4 most common site-specific invasive cancers in this population: female breast, colorectal, lung (non-small cell), prostate. Cancer diagnoses are ascertained via medical records using Read codes (version 2) and ICD-10 codes.	Same as for the target trial.
Follow-up	For each eligible individual, follow-up starts at treatment assignment and ends on the month of the cancer outcome of interest, death, loss to follow-up (transfer out of the practice or incomplete follow-up [2 years after the last recorded lab prognostic	Same as for the target trial.

Table 1. Specification and emulation of pragmatic target trials of metformin therapy and cancer risk using linked electronic health records from Clinical Practice Research Datalink, Hospital Episode Statistics, and Office of National Statistics.

	factors or 4 years after the last recorded lifestyle prognostic factors]), 6 years after baseline, or administrative end of follow-up (end of practice data collection or February 29, 2016), whichever happens first.	
Causal contrasts	Intention-to-treat effect and per-protocol effect.	Observational analogues of the intention-to-treat and per-protocol effects.
Statistical analysis	Pooled logistic regression to estimate hazard ratios and standardized risk curves. Intention-to-treat analysis: apply inverse-probability weights to adjust for pre- and post-baseline prognostic factors associated with loss to follow-up. Per-protocol analysis: censor individuals if and when they deviate from their assigned treatment strategy and apply inverse-probability weights to adjust for pre- and post-baseline prognostic factors associated with adherence and loss to follow-up. Subgroup analyses by age (<70 vs. ≥70 years), sex, and, for the target trial among individuals with diabetes, time since diabetes diagnosis (<1 vs. ≥1 year).	Same as for the target trial with sequential emulation and adjustment for baseline covariates. See **eTable 1** and **eMethods 1** for details on covariates and models. Weights to adjust for potential selection bias due to loss to follow-up had a negligible influence on the point estimates and were omitted from the primary analysis for simplicity.

Abbreviations: eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; MDRD; Modification of Diet in Renal Disease.

[a] eGFR (mL/min/1.73m^2) = 175 * (serum creatinine [μmol/L]/88.4)$^{-1.154}$ * age$^{-0.203}$ * 0.742 (if female) * 1.210 (if Black). We note that alternative prediction algorithms (e.g., the CKD-EPI 2021 eGFR creatinine equation) have been more recently defined.

Table 2. Baseline characteristics of eligible individuals with type 2 diabetes mellitus and regardless of diabetes status when emulating target trials of metformin therapy and cancer risk using linked electronic health records from Clinical Practice Research Datalink, Hospital Episode Statistics, and Office of National Statistics, 2009-2015[a].

	Among individuals with diabetes		**Among individuals regardless of diabetes status**	
Characteristic[b]	**Metformin initiators** (N=9,835)	**Non-initiators** (N=1,021,112)	**Metformin initiators** (N=11,919)	**Non-initiators** (N=3,100,055)
Age (years) – mean (SD)	63.4 (12.1)	68.6 (12.2)	63.1 (12.6)	63.9 (14.0)
Sex – no. (%)				
Female	4,501 (46)	475,710 (47)	5,582 (47)	1,582,518 (51)
Male	5,334 (54)	545,402 (53)	6,337 (53)	1,517,537 (49)
Body-mass index (kg/m^2) – mean (SD)	32.3 (6.7)	30.2 (6.0)	32.3 (6.7)	29.3 (6.1)
Hemoglobin A1c (mmol/L) – mean (SD)	53.7 (6.2)	47.1 (6.5)	53.2 (6.5)	42.1 (6.8)
Time since type 2 diabetes diagnosis (months) – mean (SD)	31.5 (35.4)	50.3 (37.9)	--	--
Smoking status – no. (%)				
Never	4,692 (48)	494,622 (48)	5,747 (48)	1,540,828 (50)
Former	3,552 (36)	387,066 (38)	4,246 (36)	1,051,163 (34)
Current	1,591 (16)	139,424 (14)	1,926 (16)	508,064 (16)
Comorbidities – no. (%)				
Coronary heart disease	709 (7)	73,391 (7)	823 (7)	173,094 (6)
Hypertension	3,085 (31)	338,909 (33)	3,531 (30)	784,611 (25)
Cerebrovascular disease	163 (2)	19,911 (2)	184 (2)	50,306 (2)
Other cardiovascular disease[c]	2,377 (24)	285,398 (28)	2,856 (24)	799,490 (26)
Medications – no. (%)				
Antihypertensive use[d]	6,541 (67)	717,506 (70)	7,813 (66)	1,637,751 (53)
Aspirin use	2,453 (25)	276,140 (27)	2,981 (25)	569,094 (18)
Nonsteroidal anti-inflammatory drug use	906 (9)	78,921 (8)	1,126 (9)	237,721 (8)
Hormonal replacement therapy – no. (% of women)	80 (2)	5,195 (1)	95 (2)	31,558 (2)
Oral contraceptive use – no. (% of women)	83 (2)	5,469 (1)	108 (2)	44,635 (3)
Any specialist referral in the past 3 months – no. (%)	2,205 (22)	106,862 (10)	2,631 (22)	314,076 (10)

Table 2. Baseline characteristics of eligible individuals with type 2 diabetes mellitus and regardless of diabetes status when emulating target trials of metformin therapy and cancer risk using linked electronic health records from Clinical Practice Research Datalink, Hospital Episode Statistics, and Office of National Statistics, 2009-2015[a].

[a] Baseline ranges from April 2009 to February 2015. Phenotype definitions are available at https://www.caliberresearch.org/

[b] Each individual may contribute to more than 1 emulated trial.

[c] Other cardiovascular disease includes acute rheumatic fever, chronic rheumatic heart disease, pulmonary heart disease, and other circulatory disease.

[d] Antihypertensive use includes all primary care prescriptions from British National Formulary chapters 2.2.1 thiazides and related diuretics, 2.2.3 potassium-sparing diuretics and aldosterone antagonists, 2.2.4 potassium-sparing diuretics with other diuretics, 2.4 beta-adrenoceptor blocking drugs, 2.5 hypertension and heart failure, 2.6.2 calcium-channel blockers.

Table 3. Estimated 6-year standardized risks and hazard ratios[a] for cancer comparing metformin therapy with no metformin therapy, using linked electronic health records from Clinical Practice Research Datalink, Hospital Episode Statistics, and Office of National Statistics, 2009-2016.

	No. of incident cancers[b]		6-year risk (%) (95% CI)		Risk difference (%) (95% CI)	Hazard ratio (95% CI)
	Initiators	Non-initiators	Initiators	Non-initiators		
Among individuals with diabetes						
Intention-to-treat[c]						
Total cancer	467	2,694	12.5 (10.9, 14.2)	12.6 (11.5, 13.7)	-0.2 (-1.6, 1.3)	1.00 (0.90, 1.10)
Breast, female	48	265	3.3 (2.0, 5.0)	2.7 (1.9, 3.6)	0.7 (-0.5, 2.0)	1.00 (0.74, 1.36)
Colorectal	60	355	1.4 (1.0, 2.1)	1.4 (1.1, 1.7)	0.1 (-0.4, 0.7)	1.02 (0.77, 1.33)
Lung	58	360	2.3 (1.5, 3.3)	1.8 (1.4, 2.3)	0.5 (-0.3, 1.4)	1.00 (0.76, 1.33)
Prostate	79	399	3.2 (2.3, 4.3)	3.5 (2.8, 4.5)	-0.4 (-1.4, 0.7)	1.07 (0.85, 1.35)
Per-protocol[d]						
Total cancer	309	2,350	12.7 (10.8, 15.0)	12.7 (11.5, 13.9)	0.0 (-2.1, 2.3)	0.99 (0.86, 1.13)
Breast, female	32	227	4.1 (2.1, 6.7)	2.4 (1.7, 3.3)	1.6 (-0.2, 4.3)	1.15 (0.74, 1.78)
Colorectal	45	309	1.7 (1.0, 2.7)	1.4 (1.1, 1.8)	0.4 (-0.3, 1.4)	1.20 (0.83, 1.73)
Lung	28	320	1.6 (0.8, 2.8)	1.8 (1.3, 2.4)	-0.2 (-1.2, 0.9)	0.75 (0.49, 1.17)
Prostate	59	342	3.5 (2.4, 5.0)	3.6 (2.7, 4.6)	-0.1 (-1.7, 1.4)	1.20 (0.86, 1.66)
Among individuals regardless of diabetes status						
Intention-to-treat[c]						
Total cancer	558	7,430	10.1 (9.0, 11.3)	10.5 (9.8, 11.2)	-0.4 (-1.5, 0.8)	0.97 (0.89, 1.06)
Breast, female	56	815	2.5 (1.7, 3.8)	2.3 (1.8, 2.9)	0.2 (-0.6, 1.5)	0.94 (0.72, 1.23)
Colorectal	73	922	1.1 (0.8, 1.5)	1.1 (1.0, 1.3)	0.0 (-0.4, 0.3)	0.92 (0.73, 1.17)
Lung	67	991	1.7 (1.2, 2.4)	1.5 (1.2, 1.8)	0.2 (-0.3, 0.9)	0.93 (0.72, 1.20)
Prostate	95	1,200	2.8 (2.2, 3.7)	3.1 (2.6, 3.7)	-0.3 (-1.0, 0.6)	1.04 (0.85, 1.29)
Per-protocol[d]						
Total cancer	361	6,985	10.0 (8.5, 11.6)	10.5 (9.8, 11.3)	-0.5 (-2.2, 1.2)	0.95 (0.84, 1.07)
Breast, female	37	767	3.1 (1.9, 5.1)	2.2 (1.7, 2.8)	1.0 (-0.4, 2.8)	1.02 (0.70, 1.49)
Colorectal	52	860	1.1 (0.8, 1.7)	1.1 (0.9, 1.3)	0.1 (-0.4, 0.6)	1.02 (0.74, 1.41)
Lung	33	942	1.3 (0.7, 2.1)	1.5 (1.2, 1.9)	-0.2 (-0.9, 0.7)	0.72 (0.49, 1.05)
Prostate	69	1,124	3.1 (2.1, 4.2)	3.2 (2.6, 3.9)	-0.1 (-1.2, 1.1)	1.13 (0.85, 1.51)

Abbreviation: CI, confidence interval.

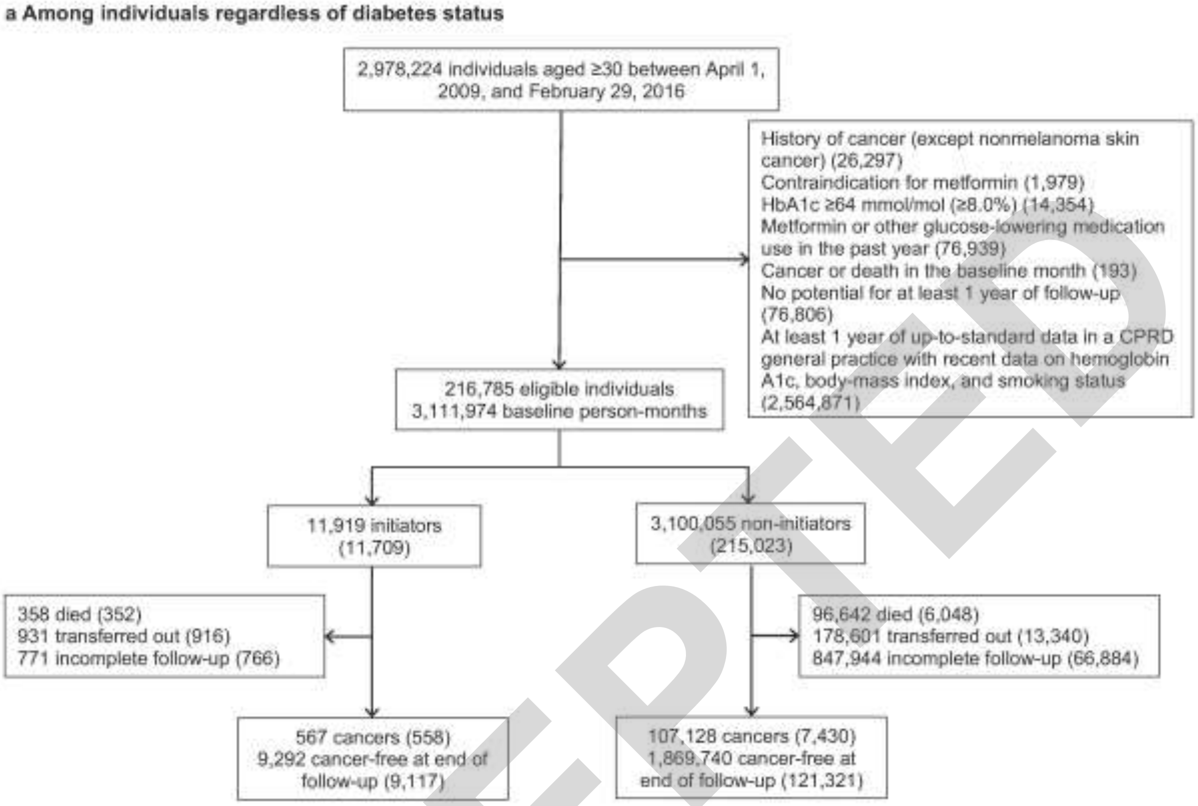
[a] Adjusted for age, sex, body-mass index, smoking status, hemoglobin A1c, months since last measure of hemoglobin A1c, coronary heart disease, hypertension, cerebrovascular disease, other cardiovascular disease, antihypertensive use, aspirin use, nonsteroidal anti-inflammatory drug use, any specialist referral in the past 3 months. Estimates for breast and colorectal cancer additionally adjusted for hormone replacement therapy and oral contraceptive use. The emulated trial among individuals with diabetes additionally adjusted for time since diabetes diagnosis. Estimated risk differences were standardized to the joint distribution of the baseline covariates.

[b] The number of events in the initiator and non-initiator groups do not sum to the total number of events because some individuals contributed as events to both groups in different nested emulated trials. The number of events is lower in the per-protocol analysis because of the censoring under this approach (see Methods).

[c] Comparing metformin initiation at baseline with no metformin initiation at baseline.

[d] Comparing metformin initiation at baseline and continuation over follow-up until the development of a contraindication or cancer with no metformin initiation over follow-up until the development of an indication.

Figure 1a


a Among individuals regardless of diabetes status

2,978,224 individuals aged ≥30 between April 1, 2009, and February 29, 2016
History of cancer (except nonmelanoma skin cancer) (26,297)
Contraindication for metformin (1,979)
HbA1c ≥64 mmol/mol (≥8.0%) (14,354)
Metformin or other glucose-lowering medication use in the past year (76,939)
Cancer or death in the baseline month (193)
No potential for at least 1 year of follow-up (76,806)
At least 1 year of up-to-standard data in a CPRD general practice with recent data on hemoglobin A1c, body-mass index, and smoking status (2,564,871)
216,785 eligible individuals
3,111,974 baseline person-months
11,919 initiators (11,709)
3,100,055 non-initiators (215,023)
358 died (352)
931 transferred out (916)
771 incomplete follow-up (766)
96,642 died (6,048)
178,601 transferred out (13,340)
847,944 incomplete follow-up (66,884)
567 cancers (558)
9,292 cancer-free at end of follow-up (9,117)
107,128 cancers (7,430)
1,869,740 cancer-free at end of follow-up (121,321)

Figure 1b

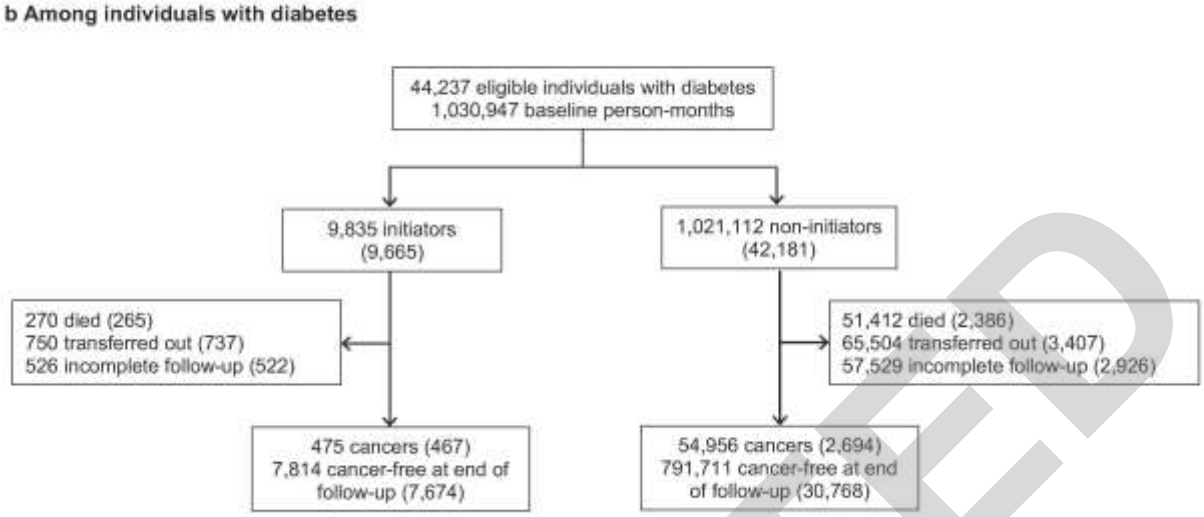
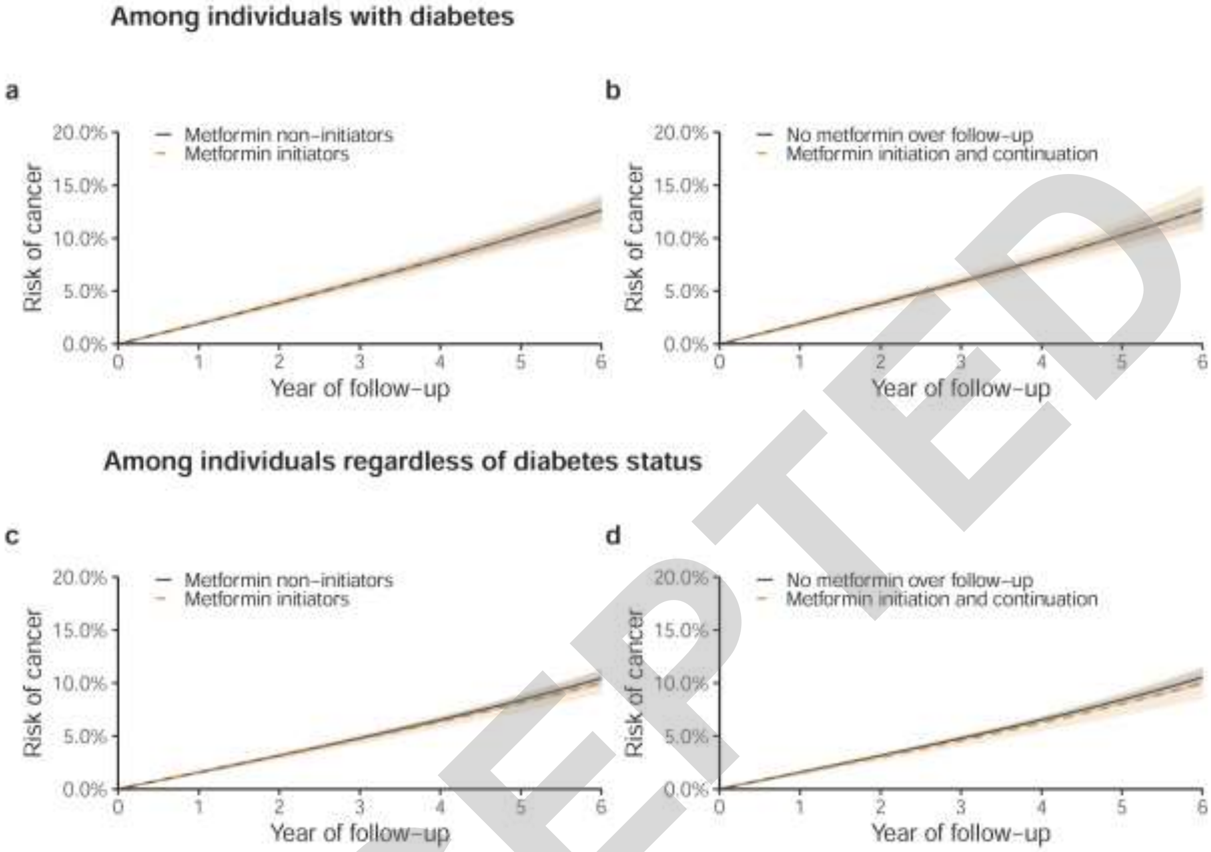

b Among individuals with diabetes
44,237 eligible individuals with diabetes
1,030,947 baseline person-months
9,835 initiators
(9,665)
1,021,112 non-initiators
(42,181)
270 died (265)
750 transferred out (737)
526 incomplete follow-up (522)
51,412 died (2,386)
65,504 transferred out (3,407)
57,529 incomplete follow-up (2,926)
475 cancers (467)
7,814 cancer-free at end of
follow-up (7,674)
54,956 cancers (2,694)
791,711 cancer-free at end
of follow-up (30,768)

Figure 2


Among individuals with diabetes
a
20.0%
15.0%
10.0%
5.0%
0.0%
Risk of cancer
— Metformin non-initiators
— Metformin initiators
0
1
2
3
4
5
6
Year of follow-up
b
20.0%
15.0%
10.0%
5.0%
0.0%
Risk of cancer
— No metformin over follow-up
— Metformin initiation and continuation
0
1
2
3
4
5
6
Year of follow-up
Among individuals regardless of diabetes status
c
20.0%
15.0%
10.0%
5.0%
0.0%
Risk of cancer
— Metformin non-initiators
— Metformin initiators
0
1
2
3
4
5
6
Year of follow-up
d
20.0%
15.0%
10.0%
5.0%
0.0%
Risk of cancer
— No metformin over follow-up
— Metformin initiation and continuation
0
1
2
3
4
5
6
Year of follow-up