
Downloaded from: http://researchonline.lshtm.ac.uk/3316201/

DOI: 10.1093/ije/dyw275
Metformin and cancer in type 2 diabetes: a systematic review and comprehensive bias evaluation

Ruth E. Farmer,1,2* Deborah Ford,2 Harriet J. Forbes,1 Nishi Chaturvedi,3 Richard Kaplan,2 Liam Smeeth1 and Krishnan Bhaskaran1

1Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK, 2MRC Clinical Trials Unit at UCL, London, UK and 3Institute of Cardiovascular Science, University College London, London, UK

*Corresponding author. Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: Ruth.farmer@lshtm.ac.uk

Accepted 2 September 2016

Abstract

Background: Existing observational studies provide conflicting evidence for the causal effect of metformin use on cancer risk in patients with type-2 diabetes, and there are concerns about bias affecting a number of studies.

Methods: MEDLINE was used to identify observational studies investigating the association between metformin and overall or site-specific cancer in people with type-2 diabetes. A systematic data extraction and bias assessment was conducted, in which risk of eight bias domains (outcome, exposure, control selection, baseline confounding, time-dependent confounding, immortal time, missing data, censoring methods) were assessed against pre-defined criteria, and rated as unlikely, low, medium or high.

Results: Of 46 studies identified, 21 assessed the effect of metformin on all cancer. Reported relative risks ranged from 0.23 to 1.36, with 10/21 reporting a statistically significant protective effect and two a harmful effect. The range of estimates was similar for site-specific cancers; 3/46 studies were rated as low or unlikely risk of bias in all domains. Two of these had results consistent with no effect of metformin; one observed a moderate protective effect overall, but presented further analyses that the authors concluded were inconsistent with causality. However, 28/46 studies were at risk from bias through exposure definition, 22 through insufficient baseline adjustment and 35 from possible time-dependent confounding.

Conclusions: Observational studies on metformin and cancer varied in design, and the majority were at risk of a range of biases. The studies least likely to be affected by bias did not support a causal effect of metformin on cancer risk.

Key words: Pharmacoepidemiology, diabetes, cancer, confounding, bias, causality
**Key Messages**

- Many existing observational studies investigating the effect of metformin use on cancer incidence in patients with type 2 diabetes have risk of bias.
- No studies to date have used appropriate statistical models to estimate the effect of time-varying treatment correctly controlling for time-dependent confounders which may be affected by previous treatment.
- Studies at lowest risk of bias do not support the hypothesis that metformin is protective against cancer.
- Previously reported large protective associations are unlikely to be causal.

**Introduction**

Research exists to suggest type 2 diabetes mellitus (T2DM) may be a risk factor for cancer,¹,² and observational studies have suggested that diabetic therapies could also affect this risk.³–⁵ Multiple observational studies have reported an apparent protective effect of metformin, a common first-line therapy for T2DM, against incidence of any cancer.⁵–⁹ However, a number of potential biases have been identified within some of these studies.¹⁰ There are limited data from clinical trials comparing metformin with other treatments, though one meta-analysis of adverse events from trials did not find any association between metformin use and cancer occurrence.¹¹

Particular difficulties arise for observational studies in this context because treatment with metformin for T2DM changes through time (is ‘time varying’), and is influenced by disease severity. This means that disease severity may be a confounder between metformin use and cancer, but will also be on the causal pathway since metformin is prescribed in order to control disease severity. For example, glycated haemoglobin (HbA1c), a measure of long-term blood glucose control, and body mass index (BMI) are predictive of metformin use according to well-defined treatment guidelines for T2DM,¹² but use of metformin will likely influence future HbA1c and BMI. There is also evidence that both BMI¹³ and HbA1c¹⁴ affect cancer risk. In this situation, standard statistical models cannot provide unbiased estimates of the total causal effect of time-varying treatment.¹⁵ Throughout this paper, such time-updated variables that may be both confounders of, and on the causal pathway for the association between exposure and outcome are referred to as ‘time dependent confounders affected by prior treatment’ (Box 1).

Reviews to date have examined existing evidence for the link between metformin use and cancer; however, some were not comprehensive¹⁰ and others have not systematically evaluated or presented a detailed evaluation of bias.¹⁶–²¹

The aim of this study was to summarize existing observational studies investigating possible associations between metformin use and cancer risk in patients with T2DM, and to systematically examine the research design and analysis methods with regard to risk of bias. A secondary aim was to use meta-regression to estimate the extent to which these biases may account for the differences between study estimates.

**Methods**

**Search strategy**

MEDLINE was searched by R.F. using OvidSP on 30 May 2014 for all English-language published articles on cancer.
risk and type 2 diabetes treatments from 1946 onwards. The search involved using MeSH headings as well as keyword searches in the title and abstract. The full search terms are presented in Supplementary data (available at IJE online). Conference abstracts and unpublished studies were excluded.

Screening strategy

Articles were included in the review if they were of a standard epidemiological design and presented original observational research. Reviews and meta-analysis were not included. Studies were required to present a measure of effect of metformin on risk of cancer incidence (either all cancer or site-specific) in patients with T2DM, with age adjustment as a minimum. Studies restricted to populations with additional comorbidities or diseases were excluded.

During an initial title and abstract screen, reviews, meta analyses and editorial pieces that looked at metformin and cancer were retained so that reference lists could be checked. Additionally, papers that appeared not to meet inclusion criteria, for example those that had primarily compared cancer incidence between diabetics and non-diabetics, were kept for full-text screening in case the required measure of effect was reported as a secondary analysis. A full-text screening was then applied to the remaining papers, and the reference lists of relevant reviews and meta analyses searched. A 10% random sample of the extracted studies were screened by an additional researcher (H.F.) to test the reliability of the inclusion criteria. A Cohens kappa score was calculated to give a quantitative measure of rater reliability, with a value of 0.75 used as the threshold for ‘excellent agreement’.

Data extraction and bias assessment

The data extraction table was piloted on five studies (by R.F., K.B. and D.F.) and subsequently refined to ensure systematic documentation of the relevant information. An example extraction table is supplied in Supplementary Table 1 (available as Supplementary data at IJE online).

Although none of the investigators were blinded to the aims of the review, detailed criteria for assessment of bias were produced in order to consider risks of bias for each study. The eight domains assessed for bias were: (i) outcome definition; (ii) exposure definition (including choice of comparator); (iii) control selection (case control studies only); (iv) consideration of HbA1c, BMI and other antidiabetic drugs as time-dependent confounders affected by previous treatment (Box 1); (v) adjustment for baseline (study entry) confounders (smoking, diabetes severity, age, gender); (vi) immortal time (cohort studies only); (vii) missing data; and (viii) censoring methods (cohort studies only). For each bias domain, pre-defined criteria allowed categorization into high, medium, low or unlikely risk of bias. Bias in study estimate occurs when aspects of the design or data analysis either induce or fail to eliminate non-causal imbalances in risk of cancer between those who are exposed or unexposed. How this may occur is dependent on the bias domain in question, and detailed criteria for each domain are presented in Supplementary Table 2 (available as Supplementary data at IJE online). Broadly, studies were considered at unlikely risk of bias in a particular domain if the design and analysis methods were unlikely to induce a systematic difference between risk of cancer between metformin users and non-users. Low risk meant that there was small possibility of bias but the potential magnitude of the bias was unlikely to materially affect the overall study conclusions. Medium and high risk of bias meant that there was potential for some or substantial effect of bias on the study estimate, respectively. Although the specific criteria for each bias domain may have left some room for subjectivity, they were developed and agreed in advance by R.F., K.B. and D.F. to make them as objective as possible.

Time-dependent confounders affected by previous treatment were considered as a separate domain in addition to baseline confounding, to highlight the difference between baseline confounding that could be easily adjusted for in a standard analysis, and the more subtle bias that may arise if time-dependent confounders affected by previous treatment are not correctly adjusted for. If studies omitted a particular confounder because they found it did not alter the estimate of metformin on cancer risk in a multivariable model, then they were not deemed to be at risk of bias due to its exclusion. However, the timing and accuracy of the confounder were still considered as sources of bias, since these aspects could have resulted in its incorrect omission.

Bias from outcome and exposure definition encompassed both misclassification bias, biases induced by timing of measurement, and whether the definitions may introduce selection bias. Potential bias induced by using time-varying exposure without consideration for the time needed for exposure to plausibly cause cancer, could be considered as inappropriate censoring or as inappropriate exposure definition; to avoid double counting, this was considered a censoring bias.

Some studies provided multiple estimates based on dose-response categories (13 studies), or differing comparators (five studies). In this situation, the main estimate used for our analyses was that deemed to be most comparable to other studies. For multiple estimates from a dose-response model, if an overall exposed vs non-exposed...
estimate was not presented (five studies), a middle category best representing a moderate level of exposure was taken.

Meta-regression
As an exploratory analysis designed to investigate whether between-study heterogeneity in the observed effect of metformin could be explained by bias and other study level factors, a random effects meta-regression was performed. Separate regressions were performed for the five most common outcomes: all cancer, colorectal/bowel cancer, lung cancer, breast cancer and pancreatic cancer. Studies that reported only stratum-specific results (three studies) were each entered into a meta-analysis to generate a pooled estimate for that study, which was subsequently used in the meta-regression.

Study characteristics evaluated in the meta-regression were a subset of all available, based on a priori assumptions about which might have the largest impact on study estimates. Characteristics included were comparator exposure [diet only, other oral antidiabetic drugs (OADs), less metformin and no metformin (diet and other OADs combined)], bias in exposure definition, bias in outcome definition, bias from baseline adjustments, bias from time-dependent confounders, immortal time bias and whether the cohort were incident users of diabetic drugs. Zero was assigned to studies rated as unlikely or low in the bias assessment, and one to those rated medium or high. A binary variable was used to reduce sparsity. Backwards stepwise selection was used to identify which characteristics best explained study heterogeneity. A P-value cut-off of 0.4 was used due to small sample size and the large number of parameters in the full model. Analysis was conducted using STATA v14.

Results
Search and screening
The numbers of studies included/excluded at each stage of the process are presented in Figure 1. From an initial 822 references (779 after removal of duplicates), 46 studies were included in the final review. Full texts were available for all studies. The random sample of 76 studies independently screened by two researchers against the inclusion criteria resulted in a Cohen’s kappa of 0.79, and only a single initial disagreement over inclusion of a study; it was agreed on discussion that this study did meet the inclusion criteria. One article examined adverse event reports from two randomized controlled trials and so was technically not observational; however, it was included as it could be considered a retrospective cohort study with a trial-based data source. It did not adjust for age, but this exclusion criterion was waived since treatment was randomized.

Table 1 summarizes the data sources, outcomes, exposure definitions, timing of exposure measurements and comparator exposures used. More detailed study-level information is presented in Supplementary Table 3 (available as Supplementary data at IJE online).

Study characteristics
Of the 46 studies, 22 were case-control design and 24 were cohort studies. Data from electronic health records were used by 37 (80%) of the studies: most commonly, the UK’s Clinical Practice Research Datalink (CPRD) (13 studies) and the Taiwan National Health Insurance Claims Database (eight studies). As previously mentioned, one paper used data from two randomized controlled trials. The remaining eight (all case-control) collected data from a specific cancer or diabetes clinic.

A total of 22 studies (46%) defined exposure to metformin as any exposure, without considering overall duration. Three further studies refined this definition by requiring a minimum time period or number of prescriptions before an individual was considered exposed. Nine studies (20%) looked at monotherapy with metformin and 10 studies (22%) used total exposure to enable dose-response analyses. The remaining two studies looked at metformin in combination with specific OADs, with a comparator group that allowed the estimation of the effect of just metformin. The most frequently used comparator group was no metformin, used in 24 studies (52%). Use of sulphonylurea [another popular first-line oral agent; 11 studies (24%)] was also a common comparator.

There were 116 estimates presented for the effect of metformin on risk of cancer when considering separate estimates for different cancer sites. A total of 21 studies examined the outcome of all cancers excluding non-melanoma skin cancer (NMSC). Colorectal and/or bowel (14 studies) were the most common sites studied, followed by pancreas (13 studies), breast (13 studies), lung (12 studies) and prostate (11 studies). Other sites had less than 10 estimates each.

Effect of metformin on cancer risk
Figure 2 displays the study estimates and 95% confidence intervals (CIs) for relative risk [odds ratio (OR) or hazard ratio (HR)] of metformin use on incidence of all cancer. Estimates and 95% CIs for the four most commonly studied site-specific cancers are presented in Figure 3.

For all cancer, 16/21 studies estimated a protective effect of metformin, with 10/16 having upper confidence limits below 1. The magnitude of the effect estimates
ranged from just a 0.04% reduction in risk\textsuperscript{45} to a 77% reduction in risk.\textsuperscript{49} Two studies estimated a statistically significant harmful effect of metformin, with estimates of 1.36 (1.1 –1.54)\textsuperscript{3} and 1.10 (1.00–1.20).\textsuperscript{47} For site-specific cancers, estimates were also highly variable across studies (Figure 3).

Bias evaluation

Only three studies\textsuperscript{51,61,62} scored low or unlikely for risk of bias in all categories. One further study, which looked at lung cancer only, scored unlikely or low in all categories, except missing data\textsuperscript{43} where it was rated unknown. Three of these studies saw no evidence of an effect of
metformin. One study estimated a modest protective effect of long-term use (> 60 months) in comparison with short-term use (0-6 months) with a hazard ratio (HR) of 0.82 (0.75-0.90), but ultimately concluded that there was no evidence for a causal effect due to patterns of risk that were inconsistent with causality.62
Study-specific results of bias assessment for studies assessing all cancer as an outcome are displayed alongside risk estimates in Figure 2. Of the 10 studies that estimated a statistically significant protective effect of metformin, nine had at least medium risk of bias in at least two domains; seven had medium or high risk of bias from exposure definition; and seven had medium or high risk of bias for treatment of HbA1c, BMI and other OADs. Bias assessments for all other studies are presented in Supplementary Table 4 (available as Supplementary data at IJE online).

Time-dependent confounders affected by previous treatment (HbA1c, BMI other antidiabetics)

Only four studies were considered as unlikely to be affected by bias due to how HbA1c, BMI and other antidiabetic treatments were accounted for in the analysis. These studies considered exposure to metformin as fixed from baseline ['intention to treat' (ITT) principle], and had confounders measured immediately before baseline.

Only 16/47 studies included HbA1c as a confounder in the final model. Six further studies reported considering it as a potential confounder, but did not include it in their final model due to lack of statistical significance or because it did not alter the results of the multivariable model.25,27,59,61 All but one of these studies were still considered at risk since it was questionable whether the HbA1c used was representative of HbA1c at the time of starting treatment; 26 studies accounted for BMI in their final model. In most case-control studies, the measurement of HbA1c and BMI preceded the date of cancer diagnosis (or matched date for the control) but it was rarely clear where this occurred in relation to the measurement of exposure, and therefore the potential for these studies to have adjusted for factors on the causal pathway between metformin and cancer was high. For the cohort studies, most used BMI and HBA1c measurements at or close to the time of cohort entry, which therefore either preceded or coincided with exposure classification. None of the studies reviewed used time-updated values of either HbA1c or BMI, though some used averages across follow-up.

The appropriate adjustment for other antidiabetic drugs is dependent upon the exposure and comparator group definitions. In six of the cohort studies examined, adjustment for use of other diabetic drugs was not necessary.8,9,51,59,61,64 In the remaining studies, 22 accounted for OADs. Tables 2 (case-control) and 3 (cohort) detail which adjustments were made, and the timing of the measurement within the follow-up period for each study separately.

Other sources of bias

Exposure definition (n = 28) and baseline adjustments (n = 22) were the other most common reasons for medium or high risk of bias. The exposure definition was most likely to have introduced bias in case-control studies by having different time windows to measure exposure, meaning the
Figure 3. Estimated relative risk (odds ratio or hazard ratio) with 95% CI for 4 most commonly studied site specific cancers. Case control studies are represented by hollow triangle, Cohort studies by filled circles.
overall chance of seeing individuals exposed to metformin is systematically different between the cases and controls. Bias was most often introduced into cohort studies because future information was used to inform exposure definition; 7/24 cohort studies were considered to have high risk of immortal time bias. In all, 22 studies were considered at risk of bias from confounding due to incomplete or inappropriate baseline adjustment because either the comparator used may have resulted in comparing patients at differing disease stages without adjustment for baseline disease severity, or measures of severity used in the adjustment could be on the causal pathway between exposure and outcome, therefore not correctly accounting for differences in disease severity that may have influenced choice of treatment at baseline.

In addition, 36 studies were considered at risk of bias due to not considering a latency period for cancer (outcome definition). Since the effect of this bias is probably small in magnitude, this was considered to be low risk. This was supported by the five studies that considered different latency periods in sensitivity analyses, concluding that estimates did not differ substantially.7,24,43,60,61

Many studies were considered as at unknown risk of bias for censoring (12/24 cohort studies) and missing data (16 studies) due to a lack of information. Particularly for censoring, few cohort studies reported the numbers lost to follow-up or for what reason. Four studies were rated medium or high for risk of bias from missing data, three of these because the missing indicator method was used, which will increase the risk of residual confounding.65 With these three studies having > 20% missing data, the effect of residual confounding could be large.

Meta-regression

Table 4 presents estimates and model diagnostics for the final meta-regression models obtained. For the outcome of all cancer, after backwards stepwise selection, the only

<table>
<thead>
<tr>
<th>Study name</th>
<th>HbA1c</th>
<th>BMI</th>
<th>Other diabetic medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azoulay et al. (2011)24</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Becker et al. (2013)25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bodmer et al. (2011)29</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bodmer et al. (2010)30</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Bodmer et al. (2012) (Lung)26</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Bodmer et al. (2012) (Pancreatic)28</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Bodmer et al. (2012) (Colorectal)27</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Bosco et al. (2011)31</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chaiteerakij et al. (2013)32</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Dabrowski et al. (2013)33</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Donadon et al. (2010)35</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Li et al. (2009)37</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Evans et al. (2005)7</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hassan et al. (2012)38</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margel et al. (2013)39</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Mazzone et al. (2012)40</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Monami et al. (2009)42</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Monami et al. (2011)41</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Smiechowski et al. (2013)43</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2013)34</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Chen et al. (2013)31</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Donadon et al. (2010)36</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

*Index date, time of cancer diagnosis/matched date for control.

*Sensitivity analysis assessed whether there was a difference between adjusting for covariates measured before exposure or any time between 1 year before exposure and index date.
study level predictor that remained in the model was the comparator group. The model estimated that using a comparator group of diet, as opposed to no metformin, made metformin appear more protective, whereas using other OADs or less metformin as a reference group made metformin appear less protective. However, this model was only predicted to explain 7% of the between-study variance. The comparator group was also retained in the models for lung, breast and pancreatic cancer; however in these models, using other OADs as the comparator was estimated to make metformin appear more protective.

The strongest predictor of heterogeneity for studies of lung cancer was risk of bias from exposure definition which, if present, was estimated to reduce the log risk ratio by 0.44, 95% CI (0.17, 0.72) \( P = 0.007 \), making metformin appear more protective. For breast cancer, the strongest predictor was use of an incident user cohort, which made metformin look less protective. This predictor was also identified for studies of lung and pancreatic cancer, but the estimates had much less precision. Presence of both time-dependent and baseline confounding was also estimated to influence study heterogeneity for breast cancer, with presence of these biases estimated to have equal and opposite effects on the log risk ratio. For colorectal cancer, the final model included only immortal time bias. As expected, risk of this bias was estimated to make metformin appear more protective.

### Table 3. Adjustment method for key time-dependent confounders affected by previous treatment: cohort studies

<table>
<thead>
<tr>
<th>Study name</th>
<th>HbA1c</th>
<th>BMI</th>
<th>Other diabetic medication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted for value at cohort entry (at time of or prior to first exposure)</td>
<td>Adjusted as a time-updated variable</td>
<td>Measured as an average of values at any point after exposure</td>
</tr>
<tr>
<td>Currie et al. (2009)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Currie et al. (2013)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Geraldine et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Home et al. (2010)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hsieh et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lai et al. (2012) (HCC)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lai et al. (2012) (LUNG)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lee et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Libby et al. (2009)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Qi et al. (2013)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Redaniel et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ruit et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tsilidis et al. (2014)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Yang et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Buch and Silverman (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Oliveira et al. (2008)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hense et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Chiu et al. (2013)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ferrara et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lehman et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Morden et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Neumann et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Van Staa et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Morgan et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

*Weight used instead of BMI.
*Measured within 3 months/1 year of cohort entry (either side of first exposure).
*Diabetes complications used as proxy measures for severity.
*Treatment randomised so no adjustment necessary.
*Adjustment for use of other OADs not necessary as study looked at incident users of diabetes medications and censored at change in medication.
<table>
<thead>
<tr>
<th>Comparator group</th>
<th>All cancer</th>
<th>Colorectal/Bowel</th>
<th>Lung</th>
<th>Breast</th>
<th>Pancreatic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>P-value</td>
<td>Estimate</td>
<td>P-value</td>
<td>Estimate</td>
</tr>
<tr>
<td>No metformin</td>
<td>0 (ref)</td>
<td></td>
<td>0 (ref)</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Diet only</td>
<td>-1.19 (-2.43, 0.05)</td>
<td>0.075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less metformin</td>
<td>0.18 (-0.27, 0.63)</td>
<td></td>
<td>0.05 (-0.34, 0.44)</td>
<td>4.0.107</td>
<td>0.22 (-0.41, -0.03)</td>
</tr>
<tr>
<td>Other OAD</td>
<td>0.24 (-0.06, 0.54)</td>
<td></td>
<td>-0.15 (-0.31, 0.02)</td>
<td></td>
<td>-0.22 (-0.41, -0.03)</td>
</tr>
<tr>
<td>Low risk</td>
<td>0 (ref)</td>
<td></td>
<td></td>
<td></td>
<td>0 (ref)</td>
</tr>
<tr>
<td>High risk</td>
<td>-0.44 (-0.72, -0.17)</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>0 (ref)</td>
<td></td>
<td>-0.17 (-0.48, 0.14)</td>
<td>0.234</td>
<td></td>
</tr>
<tr>
<td>High risk</td>
<td>-0.39 (-0.85, 0.08)</td>
<td>0.097</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immortal time bias</td>
<td>Low risk</td>
<td>0 (ref)</td>
<td></td>
<td></td>
<td>0 (ref)</td>
</tr>
<tr>
<td>High risk</td>
<td>0.17 (-0.48, 0.14)</td>
<td>0.234</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias from time-</td>
<td>Low risk</td>
<td>0 (ref)</td>
<td></td>
<td></td>
<td>0 (ref)</td>
</tr>
<tr>
<td>dependent confounding</td>
<td>High risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias from baseline</td>
<td>Low risk</td>
<td>0 (ref)</td>
<td></td>
<td></td>
<td>0 (ref)</td>
</tr>
<tr>
<td>confounding</td>
<td>High risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incident users</td>
<td>Yes</td>
<td></td>
<td>0 (ref)</td>
<td>-0.22 (-0.45, 0.02)</td>
<td>0.063</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td>0.018 (-0.14, 0.5)</td>
<td>0.218</td>
<td>0.017 (-0.08, 0.41)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.28 (-0.51, -0.05)</td>
<td>0.019</td>
<td>0.06 (-0.22, 0.11)</td>
<td>0.483</td>
<td>0.018 (-0.14, 0.5)</td>
</tr>
<tr>
<td>I squared</td>
<td>88.60%</td>
<td></td>
<td>75.00%</td>
<td>0%</td>
<td>75.4%</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>6.90%</td>
<td></td>
<td>30.40%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Tau^2</td>
<td>0.059</td>
<td></td>
<td>0.036</td>
<td>0</td>
<td>0.017 (0.08, 0.41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimate represents the expected change in the log risk ratio (either HR or OR, depending on analysis method) for the effect of metformin on cancer, for each study level predictor. For example, a study of metformin and lung cancer, in which there is high risk of bias from exposure definition, is estimated to have a log risk ratio 0.44 lower than a study not at risk of bias from exposure definition.

*I squared is the estimate of residual variation due to study heterogeneity.

**Adjusted R^2 is the estimated proportion of between study variance explained by the covariates in the meta regression.

*Tau^2 is the estimate of the remaining between-study variance.

*Joint p value for test of all levels comparator group.
Discussion

The 46 studies examined in this review did not provide consistent evidence to support a protective effect of metformin on risk of cancer. Two of three studies with low or unlikely risk of bias for all categories had estimates consistent with no effect of metformin. The third study had an estimate consistent with a moderate protective effect; however this study included many analyses, and also reported that when comparing metformin exposure with other classes of oral antidiabetics, the risk of cancer did not differ between drugs. The authors also found that the incidence rates of cancer were higher in the first 3 months after therapy initiation, which they suggested might be due to detection bias, which would also explain why longer exposure appears protective when compared with the first 6 months of therapy.

The estimates of effect reported across the 46 studies were highly variable for all outcomes studied. Many studies were at high risk of bias from exposure definition which, for reasons already outlined by Suissa and Azoulay,\(^\text{10}\) can have a large effect on estimates of risk. Within studies considered to be at low or unlikely risk of such bias, effect estimates were closer to the null but there was still variation in point estimates, albeit with some wide confidence intervals. Figure 4 displays the study estimates from Figure 2 ordered by risk of bias from exposure definition (left) to demonstrate this. It is possible that confounding by disease severity, and in particular confounding from time-dependent variables affected by previous treatment, could partly explain the remaining heterogeneity in observed estimates. By assigning values of 0, 1, 2 and 3 to unlikely, low, medium and high risk, respectively, and summing over all domains, an overall bias score was calculated. When ordered by this score [Figure 4 (right)] it is clear that heterogeneity increases as risk of bias increases, and the strongest protective effects are from those studies with the highest risk of bias overall.

The bias evaluation performed was detailed and thorough, and every effort was made to agree in advance the criteria for risk of bias in each of the eight domains examined. However, as in all studies of this kind, it was not possible to eliminate all subjectivity from this process.

Figure 5A represents the total causal effect of metformin use on cancer risk that we wish to estimate in a simple example where we assume HbA1c is the only time-dependent confounder affected by previous treatment (as previously defined in Box 1). Figure 5 B, C and D illustrates the causal pathways that are actually being estimated under the three approaches most commonly used in the studies examined in this review. In Figure 5B, studies adjust for HbA1c but the measurement is taken any time during follow-up, which may result in ‘adjusting out’ any effect of metformin that is mediated through HbA1c. In Figure 5C, because treatment may change after baseline, the single adjustment at time 0 may lead to residual confounding by post-baseline HbA1c.

![Figure 4.](http://ije.oxfordjournals.org) Estimates of relative risk of cancer from metformin use, ordered by risk of bias from exposure assessment only (left) and by overall risk of bias (right). Overall bias score is sum of bias risk over all domains, with unlikely = 0, low = 1, medium = 2, high = 3. Case control studies are represented by hollow triangle, Cohort studies by filled circle.
In Figure 5D, the fixing of exposure from baseline removes the issue of time-dependent confounding and therefore allows the total effect of exposure on cancer to be estimated, but typically estimates an ITT effect only, which may not be appropriate given that patients are unlikely to adhere to a single treatment throughout follow-up. One study adjusted for non-adherence using a method that produces an unbiased estimate if there are no unmeasured confounders of the association between non-adherence and outcome, but the validity of this assumption is questionable. This approach is also limited by considering comparisons between active drugs only. When applied and analysed carefully, it will give an unbiased estimate of the effect of initiating metformin compared with initiating (as an example) sulphonylurea on development of cancer. However, this is not necessarily equivalent to estimating causal pharmacological effect of metformin use on cancer incidence and may be inappropriate if the comparator in question may itself affect risk of cancer. Most studies with low risk of other biases used the approach outlines in 5D. The lack of variation in how time-dependent confounders were adjusted for in these studies mean that it is not possible with the current literature alone to assess whether there is a meaningful impact of time-dependent confounders affected by previous treatment on the estimated effect of metformin on cancer risk.

In order to estimate the causal pharmacological effect of metformin on risk of cancer, the ideal would be to emulate a randomized controlled trial where patients are randomized to either metformin or diet only. This would involve comparison of patients initiating metformin with those controlling their disease by diet only, and correctly adjusting for disease severity at time of initiation while maintaining the effect of previous treatment on future disease severity measures. Causal methodology has been successfully used in other areas to overcome issues with

Figure 5. Directed Acyclic Graphs (DAGs) to represent estimated causal pathways for A) the desired total causal effect of treatment on cancer risk, and B)-D) the estimated effect under different methods of adjustment for time dependent confounders affected by prior treatment. Box indicates adjustment. Dotted line represents causal associations that are present but not included in the desired/estimated effect.

A Solid lines represent the pathways needed to estimate the total causal effect of time varying treatment on cancer.
B HbA1c measured at a single time point during the measurement window (usually the most recent value). Exposure may be time updated or assumed fixed from cohort entry. Solid line represents the pathways included in the estimate of effect under this approach.
C HbA1c measured once at/before cohort entry, exposure modelled as time varying. Solid line represents the pathways included in the estimate of effect under this approach.
D Exposure is assigned at cohort entry and assumed fixed (intention to treat (ITT) principle), HbA1c measured once at/before cohort entry. Solid line represents the pathways included in the estimate of effect under this approach.
time-dependent confounders affected by previous treatment, and could be applied to this question as a valuable addition to the current literature.

These causal inference methods (marginal structural models with inverse probability of treatment weighting or the g-computation formula) may be required to fully guard against some of the potential biases we identified, notably time-dependent confounders affected by previous treatment. However, even with standard analytical approaches, careful study design and analysis can minimize the risk of bias being introduced. For example, it is desirable to clearly identify incident users of oral diabetes medications, ideally in patients with newly diagnosed diabetes, and to ensure that important confounders such as HbA1c, BMI and other disease severity measures are recorded and adjusted for at study entry—either before or at the time of medication initiation. This will ensure that disease severity is broadly balanced at study entry, and that the effect of medication on future values of important covariates is not eliminated. In addition, if medication use is not assumed to be fixed from baseline, then it is important to classify time before first exposure as unexposed in order to avoid introducing immortal time bias. Secondary analyses to look at effects of cumulative exposure, and sensitivity analysis with exclusion of periods in which un-diagnosed cancer may be affecting probability of treatment, would also be advisable to establish whether observed associations are likely to be causal.

This review has systematically identified and assessed the existing literature on the pharmacoepidemiological question of metformin use and cancer risk. The search identified a large number of studies from varying countries and journals, and the inclusion criteria were shown to have good reliability between raters. Only one database was used in the search, and therefore some relevant literature may have been omitted from the review. However, by searching reference lists of other meta-analyses and systematic reviews, the majority of studies will still have been identified. Since performing the original search, it is likely that new studies will also have been published on this topic; however, a brief updated search did not identify any new studies that used methods substantially different from those covered in this review, though one study used slightly more sophisticated methods to deal with baseline confounding by indication. The meta-regression aimed to establish whether any of the potential sources of bias could explain the heterogeneity in risk estimates. A comparator group was selected for the final model in four of the five analyses as a predictor of heterogeneity, but the direction of effect was inconsistent between models. Use of a non-incident user cohort was also identified in three models as a predictor of heterogeneity, but estimates of how this would affect study results were imprecise.

The overall reliability of the meta-regression results is questionable. For all cancer there were 21 studies contributing to this analysis, and even after selecting only key study level predictors, there were nine parameters in the initial model. The analysis was likely underpowered, and backward selection may not have produced reliable results. Additionally, many of these estimates lacked precision. For the site-specific cancers, since the sample sizes for the metaregressions were smaller, these issues may be enhanced further and individual studies with extreme estimates are likely to have had a large influence. Furthermore for some biases, two high-risk studies could be rated as such for different reasons, which would bias the estimate in opposite directions, resulting in the bias appearing to have no effect overall. In addition, the ability to examine only published studies may itself introduce a publication bias which cannot be accounted for in a meta-regression. Also as previously mentioned, the bias evaluation could not be perfectly objective, which adds further uncertainty to any results of this analysis. Therefore, overall the results of this exploratory analysis should be interpreted cautiously.

Overall, the existing literature provides inconsistent answers to the question of metformin use and cancer risk in type 2 diabetes. Variation in design of studies and the potential for many kinds of bias make it difficult to explain the differences in risk estimates, particularly in terms of the potential impact of less easily detectable bias such as that from time-dependent confounders affected by previous treatment. It is likely that the largest protective effects that have been observed are a result of immortal time bias and other issues relating to how metformin use is defined. Studies without such biases tend to have estimates closer to the null, and whereas an effect of metformin use on risk of subsequent cancer in patients with type 2 diabetes cannot be excluded, the previously reported large protective associations are unlikely to be causal.

**Supplementary Data**

Supplementary data are available at IJE online.

**Funding**

This work was supported by the Medical Research Council London Hub for Trials Methodology Research [PhD studentship to R.F.], Wellcome Trust/Royal Society Sir Henry Dale Fellowship [107731/Z/15/Z to K.B.] and Wellcome Trust Senior Research Fellowship in Clinical Science [098504/Z/12/Z to L.S.].

**Conflict of interest:** There is no conflict of interest to report.
References


51. Home PD, Kahn SE, Jones NP et al. Experience of malignancies with oral glucose-lowering drugs in the randomised controlled ADOPT (A Diabetes Outcome Progression Trial) and RECORD (Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycaemia in Diabetes) clinical trials. Diabetologia 2010;53:1838-45.
67. HIV-Causal Collaboration; Ray M, Logan R, Sterne JA et al. The effect of combined antiretroviral therapy on the

