515 removed (numbers in brackets represent primary exclusion reason):

17 (17) no abstract

19 (19) case reports

406 (383) cancer incidence not the outcome

386 (93) diabetes treatments not an exposure

53 (2) non-diabetic populations

29 (0) diabetic therapy use in cancer patients

3 (0) restricted to type 1 diabetes

8 (0) non-clinical cancer outcome

88 (1) reviews/comments/editorial articles

Relevant non-original data articles kept at this stage and reference lists checked.

107 no original data

36 cancer incidence not an outcome (solely mortality or effect of treatment after cancer diagnosis)

51 No age adjusted effect of metformin presented

14 comparison to non-diabetics

10 not in general population (e.g. post menopausal women, those undergoing colonoscopy)

4 no exact quantification of risk for comparison of interest. E.g. summary statistics only, comparison of incidence between trials.

1 same data source and analysis as another paper

4 identified from reference list search of systematic reviews and meta-analyses

1 identified from expert knowledge

822 studies exported

779 after removal of duplicates

264 remaining for full text scan

41 Met inclusion critreria

**46 Included**

**Figure 1: Flow chart of screening process detailing number of studies excluded at each stage and reason for exclusions.**

**Figure 2: Estimated relative risk (odds Ratio or hazard Ratio) with 95% CI for the 21 studies examining use of metformin and risk of all cancers, and corresponding assessment of bias according to pre-specified criteria. a**Represents the hazard ratio for cancer risk per one extra prescription of metformin

**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.



**Figure 4: 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.

****

**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.