

Impact of glycaemic traits, type 2 diabetes and metformin use on breast and prostate cancer risk: a Mendelian randomization study

Shiu Lun Au Yeung ,¹ Catherine Mary Schooling ^{1,2}

To cite: Au Yeung SL, Schooling CM. Impact of glycaemic traits, type 2 diabetes and metformin use on breast and prostate cancer risk: a Mendelian randomization study. *BMJ Open Diab Res Care* 2019;**7**:e000872. doi:10.1136/bmjdr-2019-000872

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2019-000872>).

Received 1 September 2019
Revised 12 November 2019
Accepted 3 December 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

²Graduate School of Public Health and Health Policy, City University of New York, New York City, New York, USA

Correspondence to
Dr Shiu Lun Au Yeung;
ayslyan@hku.hk

ABSTRACT

Objectives Observational studies suggest glycaemic traits and type 2 diabetes are positively associated, and metformin inversely associated with breast and prostate cancer risk. However, observational studies are susceptible to unmeasured confounding while studies of metformin use are also vulnerable to immortal time bias. The use of Mendelian randomization may reduce confounding due to random allocation of relevant genetic markers at birth, and may reduce immortal time bias (for metformin-related variants analysis) since the start of exposure is at birth.

Research design and methods We identified strong genetic predictors of fasting glucose, glycated hemoglobin, and type 2 diabetes from the Meta-Analyses of Glucose and Insulin-related traits Consortium and Diabetes Genetics Replication And Meta-analysis Consortium (n=140 595 for glucose; n=123 665 for HbA1c; n=898 130 for type 2 diabetes) and of *AMPK*-instrumented HbA1c reduction as a proxy of metformin and applied them to large genome-wide association studies of breast cancer (Breast Cancer Association Consortium; BCAC) and prostate cancer (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome; PRACTICAL). We used inverse variance weighting to obtain estimates. Sensitivity analyses included use of MR-Egger, weighted median, exclusion of pleiotropic instruments, and validation using UK Biobank (breast cancer only).

Results There was no association of fasting glucose (OR 1.03 per mmol/L, 95% CI 0.85 to 1.25), HbA1c (OR 1.02 per %, 95% CI 0.73 to 1.45), or type 2 diabetes (OR 0.98 per log odds, 95% CI 0.95 to 1.01) with breast cancer in BCAC, with similar findings from UK Biobank. There was no association of fasting glucose (OR 0.93 per mmol/L, 95% CI 0.73 to 1.17), HbA1c (OR 0.90 per %, 95% CI 0.58 to 1.40) or type 2 diabetes (OR 1.02 per log odds, 95% CI 0.97 to 1.07) with prostate cancer in PRACTICAL. No strong evidence was observed for *AMPK*-instrumented HbA1c reduction on cancer risk.

Conclusion Glycaemic traits and type 2 diabetes unlikely cause breast and prostate cancer. Whether metformin can be repurposed for cancer prevention remains unclear.

INTRODUCTION

Previous studies have suggested a potential relation of glycaemic traits with cancer, particularly hormone-related cancers, such as breast and prostate cancer.^{1–2} However,

Significance of this study

What is already known about this subject?

► Previous observational studies suggest a role of glycaemic traits in cancer risk. However, these studies could be susceptible to confounding and selection bias. Metformin use is also related to lower risk of cancer but these pharmacological studies can be susceptible to immortal time bias. The use of Mendelian randomization may help clarify the role of these related exposures in cancer risk.

What are the new findings?

► Our Mendelian randomization study shows a lack of association between glycaemic traits and breast and prostate cancer risk.
► When using genetic variants in *AMPK*, the target of metformin, no strong evidence was observed for *AMPK*-instrumented HbA1c reduction on breast and prostate cancer risk.

How might these results change the focus of research or clinical practice?

► Controlling glycaemic traits unlikely impacts cancer risk. However, whether the association between metformin use and cancer risk is causal remains to be explored in future studies, but the mechanism is likely to be independent of glycaemic traits.

the positive associations are not always consistently observed,^{1–3} and hence may be indicative of potential biases or residual confounding in previous studies. Studies on metformin, the first line medication for type 2 diabetes which reduces hepatic glucose production, have also shown a potential inverse relation with cancers,⁴ such as breast and prostate cancer,^{5–7} although these studies could be vulnerable to immortal time bias due to misallocation of exposure time.⁸ Some studies which corrected for immortal time bias showed no relation of metformin with cancer risk,⁹ but these studies remain open to confounding and selection bias. Others argued the approach to correct for immortal

time bias remains controversial and requires additional validations.¹⁰ Potentially stronger study designs, such as the marginal structural models applied to electronic health records with inverse probability weighting,¹¹ and Mendelian randomization using biomarkers of metformin¹² did not provide strong evidence for a causal role of metformin in breast cancer, with similar findings for prostate cancer in a recent meta-analysis.¹³ Given the relation between antidiabetes medications and breast cancer risk appears to vary across classes of medication with different modes of actions (eg, metformin vs sulfonylurea),¹⁴ this may imply glycaemic traits per se are not the main factors accounting for any reduction in cancer risk among metformin users. However, whether glycaemic traits are a potential target of intervention to reduce these cancers is unclear.

Mendelian randomization may provide more credible evidence concerning the role of glycaemic traits in breast and prostate cancer. Mendelian randomization uses the random allocation of genotype during conception to infer the causal relations, and hence reduces confounding.¹⁵ A previous Mendelian randomization study suggested a potentially positive relation of 2-hour glucose and fasting insulin with breast cancer although the study may be biased by the choice of instruments.^{16,17} No Mendelian randomization studies of glycaemic indicators have considered prostate cancer as the outcome. We conducted a Mendelian randomization study to address the causal role of glycaemic traits (ie, fasting glucose and glycated hemoglobin (HbA1c)) and type 2 diabetes in breast and prostate cancer using relevant genetic consortia and UK Biobank. Lastly, to further assess the potential for repurposing metformin to reduce risk of these cancers, we chose the genetic variants functionally relevant to metformin, that is, genes related to AMP-activated protein kinase (*AMPK*),¹⁸ in a separate analysis. The advantage of using genetics is it may circumvent immortal time bias given the start of higher/lower exposure is at birth for all participants.

METHODS

Study design

This is a two-sample Mendelian randomization design using summary statistics from genome-wide association studies (GWAS) and large biobanks.¹⁹ The assumptions of the Mendelian randomization design include: (1) instruments are related to exposures of interest; (2) instruments are not related to the confounders of the exposure–outcome relation; (3) instruments' links with the outcome are only via the exposure of interest.¹⁹ Online supplementary figure 1A shows the framework for a Mendelian randomization design.

Genetic predictors of glycaemic traits (fasting glucose, HbA1c) and type 2 diabetes

Fasting glucose and HbA1c

We extracted genetic instruments for fasting glucose and HbA1c from the Meta-Analyses of Glucose and

Insulin-related traits Consortium (MAGIC), which combined several GWAS to identify genetic loci related to glycaemic traits.

For fasting glucose, we used the most recent GWAS, which included up to 140 595 participants (52% women). In brief, participants were excluded from the GWAS if they had physician-diagnosed diabetes, were on diabetes treatment or had fasting plasma glucose >7 mmol/L. Imputation on up to 78 739 participants was based on the HapMap II CEU reference panel while direct genotyping was performed using iSELECT MetaboChip array (up to 61 422) and custom variant sets (up to 20 531). We only used the female-specific estimates (n=73 089) for the breast cancer, and the male-specific estimates (n=67 506) for the prostate cancer. The GWAS were adjusted for age, study site, and geographic covariates, and corrected for genomic control. Fasting glucose (mmol/L) was untransformed. Single-nucleotide polymorphisms (SNP) were not in high linkage disequilibrium ($r^2 < 0.001$).

For HbA1c, we used the most recent GWAS of HbA1c, which included up to 159 940 participants without diabetes, imputed using the International HapMap Project Phase 2 reference panel.²⁰ The GWAS was adjusted for age, sex, and study-specific covariates and corrected for genomic control. HbA1c (%) was untransformed. We only included loci that reached genome-wide significance among participants of European descent (up to 123 665), and were not in high linkage disequilibrium ($r^2 < 0.001$). We also excluded rs1800562 which violated Hardy Weinberg equilibrium ($p = 0.0007$) as previously.^{21,22}

Type 2 diabetes

We used the most recent GWAS of type 2 diabetes, which included up to 898 130 participants of European descent (74 124 cases and 824 006 controls) from the DIAbetes Genetics Replication And Meta-analysis consortium.²³ Imputation was performed using Haplotype Reference Consortium (HRC) reference panel except deCODE GWAS, which was imputed using a population-specific reference panel. Type 2 diabetes was defined in various ways, including diagnostic fasting glucose ≥ 7.0 mmol/L, previous type 2 diabetes diagnosis, the use of antidiabetic medication, or from electronic health records. The genetic associations, apart from adjusting for study-specific covariates, controlled for population stratification and relatedness via adjustment for principal components after exclusion of related individuals or the use of a mixed model with random effects for kinship from a genetic relation matrix, with genomic control applied to association summary statistics. To avoid selection bias from stratifying on a collider which may invalidate estimates, we used loci from the analyses without adjustment for body mass index. SNPs were not in high linkage disequilibrium ($r^2 < 0.001$).

Genetic associations with breast cancer from the Breast Cancer Association Consortium

We used summary genetic associations from a large GWAS of breast cancer among women of European ancestry from Breast Cancer Association Consortium (BCAC; 122977 cases and 105974 controls).²⁴ Imputation was based on the 1000 Genomes Project Phase 3 reference panel. The genetic analysis adjusted for study-specific covariates and up to 10 principal components to control for population stratification. To explore potential differences by breast cancer subtypes, we also examined the relation by estrogen receptor status (estrogen receptor positive and estrogen receptor negative).

Genetic associations with prostate cancer from the Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome consortium

We used summary genetic associations from a large GWAS of prostate cancer from Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome (PRACTICAL; 79 148 cases and 61 106 controls) of European descent,²⁵ where majority of the cases were aged 60+ years. Imputation was based on 1000 Genomes Project Phase 3 reference panel. The OncoArray and each GWAS were adjusted for principal components and study-relevant covariates, while the overall analysis only adjusted for principal components.

Genetic variants functionally relevant to metformin

Although the mechanisms by which metformin confers beneficial effects on glucose metabolism remains unclear, *AMPK* likely has a role.¹⁸ We chose genetic variants covering the subunits of *AMPK*, including *PRKAA1*, *PRKAA2*, *PRKAB1*, *PRKAB2*, *PRKAG1*, *PRKAG2*, and *PRKAG3*.²⁶ Specifically, we included all SNPs in the relevant gene regions, as identified by the corresponding positions (hg19 build) based on PubMed (<https://www.ncbi.nlm.nih.gov/gene/>). Following a previous study,²⁷ we checked the validity of these genetic predictors by (1) checking each SNP's relation with HbA1c ($p < 0.05$) using MAGIC²⁰; and (2) checking whether the SNP relates to relevant gene expression using the Genotype-Tissue Expression (GTEx) project data (<https://gtexportal.org/home/>, version 7), which gives expression quantitative trait loci analyses for 48 tissues among 620 donors ($p < 0.05$).²⁸ Specifically, we checked if the expression increasing allele (association with the lowest p value if expression was reported across tissues) corresponded to the HbA1c decreasing allele in MAGIC. Online supplementary figure 1B shows the framework of a Mendelian randomization design, with respect to exploring the impact of metformin using *AMPK* variants.

Exposure

The primary exposures were HbA1c (%), fasting glucose (mmol/L) and type 2 diabetes (log odds). The secondary exposure was HbA1c (%) reduction instrumented by *AMPK* to proxy metformin use.

Outcomes

The outcomes were breast and prostate cancer risk.

Statistical analyses for glycemic traits and type 2 diabetes

We approximated the F statistic of each instrument for fasting glucose, HbA1c and type 2 diabetes to indicate instrument strength, that is, the square of the gene-exposure association divided by the square of the corresponding SE of the association,²⁹ where higher F statistic indicates lower risk of weak instrument bias. We used inverse variance weighting (IVW) with multiplicative random effects to meta-analyze the Mendelian randomization estimates for each SNP, so as to derive an overall causal estimate for glycemic traits and type 2 diabetes on breast and prostate cancer. We used I^2 to assess the consistency of the Mendelian randomization estimates, where high I^2 may indicate the presence of invalid instruments. Given IVW assumes balanced pleiotropy and may be invalid in the presence of invalid instruments, we also conducted additional sensitivity analyses which rely on other assumptions.³⁰ Consistency across sensitivity analyses further strengthens the estimates,³¹ an approach used in previous studies.^{32 33}

MR-Egger regression

MR-Egger regression is similar to IVW, except the intercept is not fixed to zero, and hence allows for horizontal pleiotropy. The intercept also captures the degree of horizontal pleiotropy, and a $p < 0.05$ indicates the presence of such pleiotropy, which in turn suggests the IVW estimate may be invalid.³⁴ MR-Egger requires the Instrument Strength Independent of Direct Effect assumption and has lower statistical power.

Weighted median

The weighted median produces valid estimates as long as 50% of the information is derived from valid instruments.³⁵

Exclusion of pleiotropic instruments

We searched for pleiotropic effects of the instruments via PhenoScanner (but excluding UK Biobank),³⁶ which may also be a cause of breast cancer, such as sex hormone-binding globulin,³⁷ estrogen and pubertal stage,^{38 39} height,⁴⁰ lipids,⁴¹ obesity,¹⁷ and sleep,³³ and repeated the analyses without these instruments. For prostate cancer, we considered Microseminoprotein-beta,⁴² and testosterone and pubertal stage⁴³ as potential causes. We assume these are horizontal, not vertical, pleiotropic effects.¹⁶

Replication in the UK Biobank (breast cancer only)

We repeated the analyses using the UK Biobank, which had a different sampling strategy than BCAC,⁴⁴ and hence can be used to test for robustness of results due to potential biases introduced by study design. We used the summary genetic associations from UK Biobank generated by the Scalable and Accurate Implementation of Generalized (SAIGE) mixed model, restricted to British

White of European ancestry.⁴⁵ Imputation was based on the HRC reference panel. Breast cancer (12 671 female cases and 388 549 controls) was defined using the corresponding PheWAS code (174 for breast cancer), which corresponds to the relevant International Classification of Diseases codes. SAIGE controls for sample relatedness and was adjusted for sex, birth year, and the first four principal components.⁴⁵

Statistical analyses for *AMPK*-instrumented HbA1c to mimic metformin use

We used IVW with fixed effects to meta-analyze the Mendelian randomization estimates for each SNP, so as to derive an overall causal estimate for *AMPK*-instrumented HbA1c reduction, as proxy of metformin use, on breast and prostate cancer. Given these SNPs were correlated, we used a correlation matrix, obtained from 1000 Genomes via MR Base in the IVW analysis, as per our previous study.¹²

All analyses were performed using R V.3.5.2 (R Development Core Team, Vienna, Austria) using R packages ‘TwoSampleMR’ and ‘MendelianRandomization’.^{46 47}

RESULTS

The genetic instruments (up to 21 for fasting glucose, up to 34 for HbA1c, up to 159 for type 2 diabetes) used in this study are shown in online supplementary appendices 1–3. All SNP-specific F statistics for all exposures were >10, indicating that weak instrument bias was unlikely. The mean F statistics for fasting glucose, HbA1c, and type 2 diabetes were 98.7, 75.6, and 84.7, respectively, in the breast cancer analyses. The mean F statistics for fasting glucose, HbA1c, and type 2 diabetes were 79.8, 60.5, and 72.4, respectively, in the prostate cancer analyses. Based on PhenoScanner, the majority of the pleiotropic effects for breast cancer were adiposity-related traits such as fat and muscle mass. For prostate cancer, only rs1260326 and rs780093 were considered as pleiotropic because

these SNPs relate to testosterone. Among the 428 SNPs in the genes related to *AMPK*, 49 were associated with HbA1c ($p < 0.05$), with 43 SNPs within the *PRKAG2* region. Further screening based on the GTEx suggested 27 SNPs were related to the gene expression. We removed an additional five SNPs because the expression-increasing allele did not correspond to HbA1c-reducing allele. Nine SNPs were retained after excluding highly correlated SNPs ($r^2 \geq 0.8$) (online supplementary appendix 4).

Table 1 shows that there was no association of fasting glucose (OR 1.03 per mmol/L, 95% CI 0.85 to 1.25), HbA1c (OR 1.02 per %, 95% CI 0.73 to 1.45), or type 2 diabetes (OR 0.98 per log odds, 95% CI 0.95 to 1.01) with overall breast cancer in BCAC. Table 2 shows glycemic traits and type 2 diabetes were similarly not associated with estrogen receptor positive or negative breast cancer risk. Exclusion of potential pleiotropic SNPs did not change the conclusion. Table 3 shows that glycemic traits and type 2 diabetes were not associated with overall breast cancer risk in UK Biobank, with similar associations with or without pleiotropic SNPs. Exclusion of pleiotropic SNPs generally reduced I^2 (tables 1–3).

Table 4 shows that there was no association of fasting glucose (OR 0.93 per mmol/L, 95% CI 0.73 to 1.17), HbA1c (OR 0.90 per %, 95% CI 0.58 to 1.40), or type 2 diabetes (OR 1.02 per log odds, 95% CI 0.97 to 1.07) with prostate cancer in PRACTICAL, with similar observations in the sensitivity analyses. Exclusion of pleiotropic SNPs did not change the conclusion. Exclusion of pleiotropic SNPs generally reduced I^2 .

Table 5 shows that *AMPK*-instrumented HbA1c reduction was not associated with breast cancer in BCAC or UK Biobank, nor with prostate cancer in PRACTICAL although the estimates had wide CIs.

DISCUSSION

This Mendelian randomization study did not support a causal role of fasting glucose, HbA1c, or type 2

Table 1 The impact of glycemic traits on breast cancer risk in Breast Cancer Association Consortium using Mendelian randomization

All SNPs	Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
	OR	95% CI	I^2	OR	95% CI	Intercept	Intercept p value	OR	95% CI
Glucose (mmol/L) (n=21)	1.03	0.85 to 1.25	0.763	1.02	0.66 to 1.57	0.001	0.94	0.99	0.85 to 1.14
HbA1c (%) (n=32)	1.02	0.73 to 1.45	0.843	1.23	0.64 to 2.37	-0.004	0.53	1.07	0.86 to 1.33
Type 2 diabetes (n=147)	0.98	0.95 to 1.01	0.707	1.01	0.95 to 1.08	-0.003	0.24	1.02	0.99 to 1.06
Excluding pleiotropic SNPs									
Glucose (mmol/L) (n=11)	1.00	0.81 to 1.23	0.72	1.09	0.69 to 1.71	-0.004	0.68	0.94	0.81 to 1.10
HbA1c (%) (n=19)	1.06	0.91 to 1.23	0.00	1.12	0.85 to 1.49	-0.002	0.62	1.06	0.85 to 1.32
Type 2 diabetes (n=108)	0.97	0.94 to 1.01	0.585	0.90	0.80 to 1.01	0.005	0.18	0.97	0.93 to 1.01

SNP, single-nucleotide polymorphism.

Table 2 The impact of glycemc traits on breast cancer risk in Breast Cancer Association Consortium using Mendelian randomization, stratified by estrogen receptor subtypes

All SNPs	Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
	OR	95% CI	I ²	OR	95% CI	Intercept	Intercept p value	OR	95% CI
Estrogen receptor (+ve)									
Glucose (mmol/L) (n=21)	1.03	0.85 to 1.24	0.638	1.05	0.69 to 1.60	-0.001	0.90	1.02	0.87 to 1.18
HbA1c (%) (n=32)	1.06	0.75 to 1.50	0.776	1.15	0.60 to 2.23	-0.002	0.78	1.00	0.79 to 1.27
Type 2 diabetes (n=147)	0.98	0.95 to 1.02	0.651	1.00	0.93 to 1.07	-0.001	0.61	1.01	0.97 to 1.05
Excluding pleiotropic SNPs									
Glucose (mmol/L) (n=11)	1.00	0.81 to 1.24	0.603	1.06	0.67 to 1.67	-0.003	0.79	1.02	0.86 to 1.20
HbA1c (%) (n=19)	1.06	0.89 to 1.27	0.00	1.08	0.77 to 1.51	0.000	0.92	1.00	0.78 to 1.27
Type 2 diabetes (n=108)	0.98	0.93 to 1.02	0.555	0.87	0.76 to 0.99	0.007	0.08	0.99	0.94 to 1.05
Estrogen receptor (-ve)									
Glucose (mmol/L) (n=21)	1.06	0.82 to 1.38	0.577	1.06	0.59 to 1.90	0.000	0.98	1.03	0.80 to 1.31
HbA1c (%) (n=32)	1.12	0.68 to 1.84	0.751	1.76	0.69 to 4.52	-0.01	0.28	1.26	0.86 to 1.85
Type 2 diabetes (n=147)	0.97	0.93 to 1.01	0.467	1.00	0.92 to 1.09	-0.003	0.39	1.00	0.94 to 1.06
Excluding pleiotropic SNPs									
Glucose (mmol/L) (n=11)	1.01	0.80 to 1.26	0.200	1.28	0.81 to 2.02	-0.011	0.27	1.02	0.79 to 1.31
HbA1c (%) (n=19)	1.31	0.87 to 1.96	0.547	1.26	0.57 to 2.78	0.001	0.92	1.27	0.87 to 1.84
Type 2 diabetes (n=108)	0.95	0.89 to 1.00	0.342	0.87	0.74 to 1.03	0.005	0.31	0.93	0.87 to 1.00

SNP, single-nucleotide polymorphism.

diabetes in breast cancer, with similar observations by estrogen receptor subtype. Similarly, we did not find strong evidence for a role of glycemc traits in prostate cancer, which has not been investigated previously. As such, results from previous observational studies may be confounded by lifestyle factors, socioeconomic position or closely correlated traits.

Our study is consistent with a previous Mendelian randomization study, which showed no relation of fasting glucose with breast cancer.¹⁷ Our study adds by showing that the null findings were unlikely due to the use of

sex-adjusted instruments given we had used instruments for glucose in women only from the MAGIC. However, the previous study suggested a potential positive relation of 2-hour glucose and insulin (when pleiotropic SNPs were removed) with breast cancer.¹⁷ Such differences suggest that factors related to glycemc control (eg, insulin and glucagon) may be more relevant than the glycemc traits themselves to cancer development and would deserve more investigations in future studies.

Pharmacological studies have suggested a potential inverse relation of metformin with breast and prostate

Table 3 The impact of glycemc traits on breast cancer risk in UK Biobank using Mendelian randomization

All SNPs	Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
	OR	95% CI	I ²	OR	95% CI	Intercept	Intercept p value	OR	95% CI
Glucose (mmol/L) (n=21)	1.05	0.82 to 1.33	0.365	0.85	0.50 to 1.42	0.008	0.38	1.06	0.80 to 1.39
HbA1c (%) (n=34)	1.08	0.64 to 1.81	0.719	1.11	0.42 to 2.94	-0.001	0.94	1.28	0.81 to 2.02
Type 2 diabetes (n=159)	0.98	0.95 to 1.03	0.292	1.00	0.92 to 1.09	-0.001	0.63	1.04	0.98 to 1.10
Excluding pleiotropic SNPs									
Glucose (mmol/L) (n=11)	1.02	0.77 to 1.35	0.35	0.82	0.46 to 1.47	0.011	0.42	1.05	0.77 to 1.42
HbA1c (%) (n=21)	1.15	0.71 to 1.86	0.608	1.00	0.41 to 2.43	0.004	0.72	1.28	0.81 to 2.00
Type 2 diabetes (n=120)	1.00	0.95 to 1.06	0.238	1.02	0.88 to 1.17	-0.001	0.83	1.03	0.96 to 1.11

SNP, single-nucleotide polymorphism.

Table 4 The impact of glycaemic traits on prostate cancer risk in Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome consortium using Mendelian randomization

All SNPs	Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
	OR	95% CI	I ²	OR	95% CI	Intercept	Intercept p value	OR	95% CI
Glucose (mmol/L) (n=11)	0.93	0.73 to 1.17	0.464	0.73	0.41 to 1.29	0.008	0.38	0.88	0.70 to 1.12
HbA1c (%) (n=19)	0.90	0.58 to 1.40	0.575	1.26	0.32 to 4.92	-0.005	0.62	1.03	0.65 to 1.63
Type 2 diabetes (n=112)	1.02	0.97 to 1.07	0.690	0.98	0.86 to 1.10	0.003	0.50	0.97	0.93 to 1.02
Excluding pleiotropic SNPs*									
Glucose (mmol/L) (n=10)	1.01	0.81 to 1.24	0.263	0.75	0.48 to 1.19	0.01	0.20	0.90	0.72 to 1.13
Type 2 diabetes (n=111)	1.02	0.97 to 1.07	0.686	0.98	0.87 to 1.11	0.003	0.48	0.98	0.93 to 1.03

*No pleiotropic SNPs identified for HbA1c. SNP, single-nucleotide polymorphism.

cancer.⁴⁻⁶ Metformin reduces glycaemic traits suggesting a role of glycaemic traits in these cancers. However, not all antidiabetic medications show an inverse relation. For example, use of insulin and glucagon-like peptide-1 analog is not related to breast cancer.^{48,49} This discrepancy suggests that metformin's potential anticancer property could be due to other pathways rather than its glycaemic property. Baseline analysis of a randomized controlled trial suggested metformin may have pleiotropic effects, among which growth differentiation factor 15 (GDF-15) was the strongest biomarker⁵⁰ although we previously showed that GDF-15 did not appear to be a cause of breast cancer in BCAC.¹² Our findings on *AMPK*-instrumented HbA1c reduction also suggest no strong evidence for metformin use affecting cancer outcomes although they were wide CIs, and hence it would be premature to conclude there is no effect of metformin use on cancer risk. For example, *AMPK*-instrumented HbA1c reduction may increase breast cancer risk (OR 1.28, 95% CI 0.21 to 7.87) although the findings were not consistent in BCAC (OR 0.70, 95% CI 0.29 to 1.69). Such differences could be driven by differences in biases stemming from study design, such as differing levels of selection bias induced

by inevitably only selected survivors of the exposure and competing risk. As such, it remains inconclusive whether the observed inverse relation between metformin and various cancers is a reflection of immortal time bias, notwithstanding the observations concerning metformin and prostate cancer are unclear.¹³ Results from randomized controlled trials and Mendelian randomization studies using genetic risk scores for *AMPK* in large well-designed biobanks would provide a more definitive answer to this important clinical question.

Strengths of this study included the use of Mendelian randomization, which is more robust to confounding than observational studies, and the use of validation data sets. Nevertheless, there are some limitations. First, the Mendelian randomization design has its own assumptions. However, we used genetic loci identified from GWAS. Genetic variants should be unconfounded. Sensitivity analyses with different assumptions about pleiotropy gave consistent estimates. Second, we were only able to use sex-specific genetic estimates for fasting glucose due to data availability and hence we could not ascertain whether the null findings for HbA1c and type 2 diabetes were driven by the use of sex-adjusted genetic estimates.

Table 5 The impact of AMP-activated protein kinase (*AMPK*)-instrumented HbA1c reduction (%) on breast and prostate cancer using Mendelian randomization

Nine single-nucleotide polymorphisms	Data sets	Inverse variance weighting with fixed effects	
		OR	95% CI
Breast cancer (overall)	Breast Cancer Association Consortium	0.70	0.29 to 1.69
Breast cancer (estrogen receptor +ve)	Breast Cancer Association Consortium	0.77	0.27 to 2.20
Breast cancer (estrogen receptor -ve)	Breast Cancer Association Consortium	1.69	0.34 to 8.32
Breast cancer (overall)	UK Biobank	1.28	0.21 to 7.87
Prostate cancer	Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome consortium	0.60	0.19 to 1.85

Notwithstanding this possibility, the GWAS on glucose suggested a lack of heterogeneity by sex regarding the SNPs predicting glucose ($p \geq 0.001$). Third, we relied on the summary statistics within the genes concerning *AMPK* to estimate the impact of metformin use on cancer risk. Such approach may have less statistical power compared with using genetic risk scores. Furthermore, we could not rule out the existence of other targets by which metformin may confer its potential anticancer properties. Lastly, we were unable to assess potential non-linearity between these traits and breast cancer risk as we did not have access to individual-level data. However, the possibility of non-linear relation is unlikely. Furthermore, we were unable to explore other cancers given lack of access to relevant summary statistics.

This Mendelian randomization study suggested a lack of relation of fasting glucose, HbA1c and type 2 diabetes with breast and prostate cancer while it remains unclear whether metformin reduces cancer risk. The putative anticancer property of metformin, if causal, is likely via glycemia-independent pathways. Further research with better designs may help elucidate the underlying mechanisms, and hence identify new targets of intervention for cancer prevention.

Acknowledgements This research has been conducted using the UK Biobank Resource under application number 24460 (for breast cancer-related analyses). Data on sex-specific glucose have been contributed by MAGIC and have been downloaded from <https://www.magicinvestigators.org/> (Vasiliki Lagou, Reedik Mägi, Jouke-Jan J Hottenga, et al. Fasting glucose and insulin variability: sex-dimorphic genetic effects and novel loci (2019) In Preparation). Data on breast cancer have been contributed by BCAC investigators and have been downloaded from <http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/>. All studies and funders are listed in the study of Michailidou et al. Data on prostate cancer have been contributed by PRACTICAL investigators and have been downloaded from http://practical.icr.ac.uk/blog/?page_id=8164. We also thank the following for funding support: the Institute of Cancer Research and the Everyman Campaign, the Prostate Cancer Research Foundation, Prostate Research Campaign UK (now PCUK), the Orchid Cancer Appeal, Rosetrees Trust, the National Cancer Research Network UK, and the National Cancer Research Institute (NCRI) UK. We are grateful for support of the NIHR funding to the NIHR Biomedical Research Centre at the Institute of Cancer Research and the Royal Marsden NHS Foundation Trust. The Prostate Cancer Program of Cancer Council Victoria also acknowledges grant support from the National Health and Medical Research Council Australia, VicHealth, Cancer Council Victoria, the Prostate Cancer Foundation of Australia, the Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK from the original paper (Schumacher et al.) acknowledge the Intramural Program of the National Human Genome Research Institute for their support. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 15 August 2019.

Contributors SLAY designed the study and wrote the analysis plan. SLAY undertook analyses with feedback from CMS, and all authors interpreted the results in the study. SLAY wrote the first draft of the manuscript with critical feedback and revisions from CMS. All authors gave final approval of the version to be published. SLAY is the guarantor of this work.

Funding The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, the National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and the European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). The prostate cancer genome-wide association analyses are supported by the Canadian Institutes of Health Research, European Commission's Seventh Framework Programme grant

agreement number 223175 (HEALTH-F2-2009-223175), Cancer Research UK grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, and C16913/A6135, and the National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: 1U19 CA 148537-01 (the GAME-ON initiative). Grant support was granted by the National Health and Medical Research Council Australia (126402, 209057, 251533, 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296), VicHealth, Cancer Council Victoria, the Prostate Cancer Foundation of Australia, the Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) (U19 CA 148537 for ELucidating Loci Involved in Prostate cancer Susceptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008) and by Cancer Research UK grant A8197/A16565. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). Funding for the ICOGS infrastructure came from the European Community's Seventh Framework Programme under grant agreement number 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The BPC3 was supported by the US National Institutes of Health, National Cancer Institute (cooperative agreements U01-CA98233 to DJ Hunter, U01-CA98710 to SM Gapstur, U01-CA98216 to E Riboli, and U01-CA98758 to BE Henderson, and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics). CAPS GWAS study was supported by the Swedish Cancer Foundation (grant numbers 09-0677, 11-484, 12-823), the Cancer Risk Prediction Center (CRiSP; www.crispcenter.org), and a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council (grant numbers K2010-70X-20430-04-3, 2014-2269). PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, and National Institutes of Health. The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study only used publicly available data and hence no ethics approval was required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Shiu Lun Au Yeung <http://orcid.org/0000-0001-6136-1836>

Catherine Mary Schooling <http://orcid.org/0000-0001-9933-5887>

REFERENCES

- de Beer JC, Liebenberg L. Does cancer risk increase with HbA1c, independent of diabetes? *Br J Cancer* 2014;110:2361–8.
- Tseng C-H. Diabetes and risk of prostate cancer: a study using the National health insurance. *Diabetes Care* 2011;34:616–21.
- Hope C, Robertshaw A, Cheung KL, et al. Relationship between HbA1c and cancer in people with or without diabetes: a systematic review. *Diabet Med* 2016;33:1013–25.
- Yang X, So WY, Ma RCW, et al. Low HDL cholesterol, metformin use, and cancer risk in type 2 diabetes: the Hong Kong diabetes registry. *Diabetes Care* 2011;34:375–80.
- Bodmer M, Meier C, Krähenbühl S, et al. Long-Term metformin use is associated with decreased risk of breast cancer. *Diabetes Care* 2010;33:1304–8.
- Preston MA, Riis AH, Ehrenstein V, et al. Metformin use and prostate cancer risk. *Eur Urol* 2014;66:1012–20.

- 7 Murtola TJ, Tammela TLJ, Lahtela J, *et al.* Antidiabetic medication and prostate cancer risk: a population-based case-control study. *Am J Epidemiol* 2008;168:925–31.
- 8 Suijsa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care* 2012;35:2665–73.
- 9 Azoulay L, Dell'Aniello S, Gagnon B, *et al.* Metformin and the incidence of prostate cancer in patients with type 2 diabetes. *Cancer Epidemiol Biomarkers Prev* 2011;20:337–44.
- 10 Hanley JA, Foster BJ. Avoiding blunders involving 'immortal time'. *Int J Epidemiol* 2014;43:949–61.
- 11 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;48:527–37.
- 12 Au Yeung SL, Luo S, Schooling CM. The impact of GDF-15, a biomarker for metformin, on the risk of coronary artery disease, breast and colorectal cancer, and type 2 diabetes and metabolic traits: a Mendelian randomisation study. *Diabetologia* 2019;62:1638–46.
- 13 Feng Z, Zhou X, Liu N, *et al.* Metformin use and prostate cancer risk: a meta-analysis of cohort studies. *Medicine* 2019;98:e14955.
- 14 Onitilo AA, Stankowski RV, Berg RL, *et al.* Type 2 diabetes mellitus, glycemic control, and cancer risk. *Eur J Cancer Prev* 2014;23:134–40.
- 15 Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018;362.
- 16 Tan VY, Yarmolinsky J, Lawlor DA, *et al.* Letter regarding article, "Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis". *Int J Epidemiol* 2019;48:1014–5.
- 17 Shu X, Wu L, Khankari NK, *et al.* Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. *Int J Epidemiol* 2019;48:795–806.
- 18 Rena G, Hardie DG, Pearson ER. The mechanisms of action of metformin. *Diabetologia* 2017;60:1577–85.
- 19 Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016;45:908–15.
- 20 Wheeler E, Leong A, Liu C-T, *et al.* Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med* 2017;14:e1002383.
- 21 Au Yeung SL, Luo S, Schooling CM. The Impact of Glycated Hemoglobin (HbA_{1c}) on Cardiovascular Disease Risk: A Mendelian Randomization Study Using UK Biobank. *Diabetes Care* 2018;41:1991–7.
- 22 Au Yeung SL, Luo S, Schooling CM. The impact of glycated hemoglobin on risk of hypertension: a Mendelian randomization study using UK Biobank. *J Hypertens* 2020;38:38–44.
- 23 Mahajan A, Taliun D, Thurner M, *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50:1505–13.
- 24 Michailidou K, Lindström S, Dennis J, *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;551:92–4.
- 25 Schumacher FR, Al Olama AA, Berndt SI, *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018;50:928–36.
- 26 Jablonski KA, McAteer JB, de Bakker PIW, *et al.* Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. *Diabetes* 2010;59:2672–81.
- 27 Walker VM, Kehoe PG, Martin RM, *et al.* Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian randomization study. *Int J Epidemiol* 2019:dyy155.
- 28 Battle A, Brown CD, Engelhardt BE, *et al.* Genetic effects on gene expression across human tissues. *Nature* 2017;550:204–13.
- 29 Bowden J, Del Greco M F, Minelli C, *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol* 2016;45:dyw220–1974.
- 30 Bowden J, Del Greco M F, Minelli C, *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017;36:1783–802.
- 31 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;45:1866–86.
- 32 Au Yeung SL, Borges M-C, Lawlor DA. Association of genetic instrumental variables for lung function on coronary artery disease risk: a 2-Sample Mendelian randomization study. *Circ Genom Precis Med* 2018;11:e001952.
- 33 Richmond RC, Anderson EL, Dashti HS, *et al.* Investigating causal relations between sleep traits and risk of breast cancer in women: mendelian randomisation study. *BMJ* 2019;365:l2327.
- 34 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
- 35 Bowden J, Davey Smith G, Haycock PC, *et al.* Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;40:304–14.
- 36 Staley JR, Blackshaw J, Kamat MA, *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016;32:3207–9.
- 37 Dimou NL, Papadimitriou N, Gill D, *et al.* Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study. *Int J Epidemiol* 2019;48:807–16.
- 38 Rossouw JE, Anderson GL, Prentice RL, *et al.* Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health Initiative randomized controlled trial. *JAMA* 2002;288:321–33.
- 39 Burgess S, Thompson DJ, Rees JMB, *et al.* Dissecting causal pathways using Mendelian randomization with summarized genetic data: application to age at menarche and risk of breast cancer. *Genetics* 2017;207:481–7.
- 40 Lai FY, Nath M, Hamby SE, *et al.* Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis. *BMC Med* 2018;16:187.
- 41 Nowak C, Ärnlöv J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk. *Nat Commun* 2018;9:3957.
- 42 Smith Byrne K, Appleby PN, Key TJ, *et al.* The role of plasma microseminoprotein-beta in prostate cancer: an observational nested case-control and Mendelian randomization study in the European prospective investigation into cancer and nutrition. *Ann Oncol* 2019;30:1093/annonc/mdz121. [Epub ahead of print: 08 Apr 2019].
- 43 Bonilla C, Lewis SJ, Martin RM, *et al.* Pubertal development and prostate cancer risk: Mendelian randomization study in a population-based cohort. *BMC Med* 2016;14:66.
- 44 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017;186:1026–34.
- 45 Zhou W, Nielsen JB, Fritsche LG, *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* 2018;50:1335–41.
- 46 Hemani G, Zheng J, Elsworth B, *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;7. doi:10.7554/eLife.34408. [Epub ahead of print: 30 May 2018].
- 47 Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;46:1734–9.
- 48 Wu JW, Filion KB, Azoulay L, *et al.* Effect of long-acting insulin analogs on the risk of cancer: a systematic review of observational studies. *Diabetes Care* 2016;39:486–94.
- 49 Hicks BM, Yin H, Yu OHY, *et al.* Glucagon-Like peptide-1 analogues and risk of breast cancer in women with type 2 diabetes: population based cohort study using the UK clinical practice research Datalink. *BMJ* 2016;355:i5340.
- 50 Gerstein HC, Pare G, Hess S, *et al.* Growth differentiation factor 15 as a novel biomarker for metformin. *Diabetes Care* 2017;40:280–3.