Using Mendelian Randomization to model the causal effect of cancer on health economic outcomes and to simulate the costeffectiveness of anti-cancer interventions

Padraig Dixon^{1,2‡*}, Richard M Martin^{2,3,4}, Sean Harrison^{2,5‡}
1: Nuffield Department of Primary Care Health Sciences, University of Oxford
2: MRC Integrative Epidemiology Unit, University of Bristol e
(lu 2: MRC Integrative Epidemiology Unit, University of Bristol
3: Population Health Sciences, Bristol Medical School, University of Bristol
IR Bristol Biomedical Research Centre, University Hospitals Bristol and Westor
Founda 3: Population Health Sciences, Bristol Medical School, University of Bristol
2: Bristol Biomedical Research Centre, University Hospitals Bristol and Westr
6: Foundation Trust and the University of Bristol, Bristol
5: UK He 3: Pristol Biomedical Research Centre, University Hospitals Bristol and Wester
Foundation Trust and the University of Bristol, Bristol
5: UK Health Security Agency Foundation Trust and the University of Bristol, Bristol
5: UK Health Security Agency
January 2023 5: UK Health Security Agency
January 2023

5: UK Health Security Agency
1951
ABSTRACT January 2023

 \overline{a} $\frac{1}{t}$ lit termine ABSTRACT
BACKGROUND Cancer is associated with significant economic impacts. Quantifying the scale of
these impacts is challenged by confounding variables that jointly influence both cancer status BACKGROUND Cancer is associated with significant economic impacts. Quantifying the scale of
these impacts is challenged by confounding variables that jointly influence both cancer status
and economic outcomes such as healt and economic outcomes such as healthcare costs and quality of life. Moreover, the increasing
costs attributed to cancer drug development complicate the cost-effective provision of cancer
care.

and economic outcomes such as healthcare cost and quality of life. The interesting costs attributed to cancer drug development complicate the cost-effective provision of cancer care.
 METHODS We address both challenges i care.

METHODS We address both challenges in this paper by using germline genetic variation in the

risk of incident cancer as instrumental variables in Mendelian Randomization analyses of eight

cancers. We developed caus METHODS We address both challenges in this paper by using germline genetic variation in the
risk of incident cancer as instrumental variables in Mendelian Randomization analyses of eight
cancers. We developed causal estima METHODS We address both challenges in this paper by using germine genetic variation in the
risk of incident cancer as instrumental variables in Mendelian Randomization analyses of eight
cancers. We developed causal estimat cancers. We developed causal estimates of the genetically predicted effect of bladder, breast,
colorectal, lung, multiple myeloma, ovarian, prostate and thyroid cancers on healthcare costs
and quality adjusted life years (colorectal, lung, multiple myeloma, ovarian, prostate and thyroid cancers on healthcare costs
and quality adjusted life years (QALYs) using outcome data drawn from the UK Biobank cohort
We then used Mendelian Randomization and quality adjusted life years (QALYs) using outcome data drawn from the UK Biobank cohort
We then used Mendelian Randomization to model a hypothetical population-wide preventative
intervention based on a repurposed class We then used Mendelian Randomization to model a hypothetical population-wide preventative
intervention based on a repurposed class of anti-diabetic drugs known as sodium-glucose co-
transporter-2 (SGLT2) inhibitors very re intervention based on a repurposed class of anti-diabetic drugs known as sodium-glucose co-
transporter-2 (SGLT2) inhibitors very recently shown to reduce the odds of incident prostate
cancer.
RESULTS Genetic liability to

were associated with considerable uncertainty. SGLT2 inhibition was unlikely to be a cost-Theor (Scancer) increases an example and the same in teach cancer of the same in tests of the sets.
2 (SCLT2) in the setsember of the outselver and the origin of the original causal impacts of the odds of the o
19 (SCLT2) **RESULT**

on healt

were as

effective

the pric RESULTS GENETIC HADING TO PROSTATE CARTER AND TO DREAST CARTER IN THE ON INDICITS ON healthcare costs and QALYS. Mendelian Randomization results for the less common cancer were associated with considerable uncertainty. SGL on the associated with considerable uncertainty. SGLT2 inhibition was unlikely to be a cost-
effective preventative intervention for prostate cancer, although this conclusion depended on
the price at which these drugs woul

the price at which these drugs would be offered for a novel anti-cancer indication.
 IMPLICATIONS Our new causal estimates of cancer exposures on health economic outcomes

may be used as inputs into decision analytic mod THE PERTIONS Our new causal estimates of cancer exposures on health economic
The price at as inputs into decision analytic models of cancer interventions such as
1
NOTE: This preprint reports new research that has not been INTERTIONS OUT HEW CAUSAT ESTIMATES OF CANCET EXPOSUTES ON HEAlth ECONOMIC OUTCOMES
That way be used as inputs into decision analytic models of cancer interventions such as screening
1
NOTE: This preprint reports new resea

programmes or simulations or simulations or simulations and the ReCTs with short follow-ups. Our new method allows us to rapidly and efficiently estimate the
cost-effectiveness of a hypothetical population-scale anti-cance Cost-effectiveness of a hypothetical population-scale anti-cancer intervention to inform and
complement other means of assessing long-term intervention cost-effectiveness.
Keywords: Genetics, cancer, instrumental variables

complement other means of assessing long-term intervention cost-effectiveness.
Keywords: Genetics, cancer, instrumental variables, healthcare costs, cost-effectiveness,
quality of life, Mendelian Randomization complement other means of assessing long-term intervention cost-effectiveness. $\frac{1}{2}$

Reywords: Genetics, cancer, instrumental variables, healthcare costs, cost-effectiveness,
quality of life, Mendelian Randomization
*Corresponding author. Padraig Dixon, Nuffield Department of Primary Care Health Scien
Univ quality of life, Mendelian Randomization

*Corresponding author. Padraig Dixon, N

University of Oxford, Radcliffe Primary Ca
Rd, Oxford OX2 6GG, padraig dixon@phc _________________________ University of Oxford, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock
Rd, Oxford OX2 6GG, padraig.dixon@phc.ox.ac.uk
‡ These authors contributed equally University of Oxford, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock
Rd, Oxford, OX2 6GG, padraig.dixon@phc.ox.ac.uk
These authors contributed equally

These authors contributed equally

These authors contributed equally ‡ These authors contributed equally

It is made available under a CC-BY 4.0 International license. **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2023.02.06.23285521;](https://doi.org/10.1101/2023.02.06.23285521) this version posted February 8, 2023. The copyright holder for this preprint

1 Introduction

 $\frac{1}{1}$ economic impacts on patients, carers and health systems (5-8). Increases in the average
populations, and improvements in the detection and treatment of cancer, are creating ϵ
cohorts of individuals who have received can populations, and improvements in the detection and treatment of cancer, are creating growing
cohorts of individuals who have received cancer treatment, who are receiving treatment, or
who may receive some form of cancer in populations, and improvements in the detection and in elements is charger, are creating growing
cohorts of individuals who have received cancer treatment, who are receiving treatment, or
who may receive some form of cancer who may receive some form of cancer intervention for the rest of their lives (6, 9, 10). In turn
extended treatment modalities, new adjuvant regimes and increases in the costs of therapie
challenge the cost-effective deli extended treatment modalities, new adjuvant regimes and increases in the costs of therapies,
challenge the cost-effective delivery of care for cancer patients (8, 11-16).
Knowledge of how cancer status affects health econo

extended treatment modalities, new adjustment games and increases in the cost of mattepties,
challenge the cost-effective delivery of care for cancer patients (8, 11-16).
Knowledge of how cancer status affects health econo challenge of how cancer status affects health economic outcomes such a
patient quality of life is central to providing cost-effective care to patients.
designs that examine associations between cancer status and these outc patient quality of life is central to providing cost-effective care to patients. Conventional study
designs that examine associations between cancer status and these outcomes can be affected
by measurement error and by omi patient quality of the to contain to providing our care tractor patients. Conventional care,
designs that examine associations between cancer status and these outcomes can be affected
by measurement error and by omitted va by measurement error and by omitted variable bias. Any traits, behaviours, or disease
processes (including prodromal processes) that influence or are influenced by cancer status are
likely in most cases to independently in by measurement error and by control variable bias. Any trains, behaviours, or alleader
processes (including prodromal processes) that influence or are influenced by cancer s
likely in most cases to independently influence

processes (independently influence healthcare costs, quality of life or both.
For example, risk factors such as cigarette smoking or adverse weight profiles may co-occur, are
known to independently and jointly influence ri likely in most cases to intepermently inhering instances to supply prime of a single
For example, risk factors such as cigarette smoking or adverse weight profiles may co-o
known to independently and jointly influence risk For example, risk factors such as eigen extending or adverse weight profiles may be such, and
known to independently and jointly influences on health economic outcomes (22-25). Cohort
studies (for example (26)) that attrib quantitatively important causal influences on health economic outcomes (22-25). Cohort
studies (for example (26)) that attribute costs to cancer patients may account for direct costs of the studies (for example (26)) that attribute costs to cancer patients may account for direct conditionist
following a diagnosis of cancer, but cannot reflect all causal impacts associated with liabil
cancer or cancer d following a diagnosis of cancer, but cannot reflect all causal impacts associated with liability to
cancer or cancer diagnoses because in general it is not possible to identify all confounding
variables, to assess their co cancer or cancer diagnoses because in general it is not possible to identify all confounding
variables, to assess their contribution without measurement error, to condition away their
influence in quantitative analysis, or variables, to assess their contribution without measurement error, to condition away their
influence in quantitative analysis, or to include all downstream causal impacts.
A further challenge to providing affordable care r

variates, to assess their contribution with a measurement error, to contribute and, then
influence in quantitative analysis, or to include all downstream causal impacts.
A further challenge to providing affordable care rel influence in the relates to the high and increasing a further challenge to providing affordable care relates to the high and increasing
inssociated with anti-cancer pharmacotherapies (27). The inflation-adjusted laur
inewl associated with anti-cancer pharmacotherapies (27). The inflation-adjusted launch price
newly approved cancer drugs in the United States increased at 10% every year between
and 2013 (28), implying a doubling of prices roug newly approved cancer drugs in the United States increased at 10% every year between 1995
and 2013 (28), implying a doubling of prices roughly every 7 years, and amounting to average
3 and 2013 (28), implying a doubling of prices roughly every 7 years, and amounting to average and 2013 $\left(28\right)$, implying a doubling of prices roughly every 7 years, and amounting to average

annual increases of \$8,000. Annual average costs for novel anti-cancer drugs in the US are now
over \$100,000 and considerably more where such drugs are taken in combination (13, 29, 30).
This is despite weak or no evidence This is despite weak or no evidence of a connection between high prices, development costs
and therapeutic efficacy for drugs as a whole (31-34). These challenges are not confined to any
one country (8, 14, 15, 35-38) and and therapeutic efficacy for drugs as a whole (31-34). These challenges are not confined to are country (8, 14, 15, 35-38) and serve to emphasize the formidable costs associated with developing and remunerating anti-cance and therepeads challenge as a matter (31-34). These challenges are not country (8, 14, 15, 35-38) and serve to emphasize the formidable costs associated with
developing and remunerating anti-cancer pharmacotherapies.
Ident

developing and remunerating anti-cancer pharmacotherapies.
Identifying cost-effective cancer therapies, and improving the efficiency of drug development
Therefore offers considerable promise for improving the value of canc dentifying cost-effective cancer therapies, and improving the education of the refore offers considerable promise for improving the value and dresses both challenges. We report the first use of Mendelia Idertify therefore offers considerable promise for improving the value of cancer care. This paper
addresses both challenges. We report the first use of Mendelian Randomization (39, 40) to
estimate the causal effect of gene addresses both challenges. We report the first use of Mendelian Randomization (39, 40)
estimate the causal effect of genetic liability to different site-specific cancers on, respecti
healthcare cost and quality of life. and estimate the causal effect of genetic liability to different site-specific cancers on, respective
healthcare cost and quality of life.
Mendelian Randomization (39-43) is a type of instrumental variable analysis that re healthcare cost and quality of life.
Mendelian Randomization (39-43) is a type of instrumental variable analysis that relies on the

estimate the cost and quality of life.
Mendelian Randomization (39-43) is a type of instrumental variable analysis that relies on the
natural experiment of quasi-random allocation of genetic variation from parents to child matural experiment of quasi-random allocation of genetic variation from parents to children at
conception. Some of this genetic variation is known to influence susceptibility to cancer and is
in principle, and with assumpt conception. Some of this genetic variation is known to influence susceptibility to cancer and is –
in principle, and with assumptions that we describe below – independent of post-conception
influences that would otherwise in principle, and with assumptions that we describe below – independent of post-conception
influences that would otherwise confound the relationship between cancer status and health
economic outcomes in conventional analyt influences that would otherwise confound the relationship between cancer status and health
economic outcomes in conventional analytic modelling. We then use these Mendelian
Randomization estimates to efficiently assess the influences that would onticially and the relationship is considered that the matricially
economic outcomes in conventional analytic modelling. We then use these Mendelian
Randomization estimates to efficiently assess the c Randomization estimates to efficiently assess the cost-effectiveness of a novel and pote
affordable population-wide anti-cancer prophylactic drug therapy repurposed from the
treatment of diabetes. affordable population-wide anti-cancer prophylactic drug therapy repurposed from the
treatment of diabetes.
Together, these two work strands demonstrate the feasibility of using Mendelian

treatment of diabetes.
Together, these two work strands demonstrate the feasibility of using Mendelian
Randomization to study the impact of cancer exposures on health economic outcomes, Together, these two wo
Randomization to stud₎
as the practicality of th Together, the impact of cancer exposures on health economic outconstandomization to study the impact of cancer exposures on health economic outconstant as the practicality of these methods to rapidly and efficiently estima as the practicality of these methods to rapidly and efficiently estimate the cost-effectiveness of
anti-cancer interventions to complement, prioritise or inform the design of randomized
controlled trials or other evaluatio anti-cancer interventions to complement, prioritise or inform the design of randomized
controlled trials or other evaluations of cancer therapies.
2 Methods France intervention to complement, private or inform the design of randomized
controlled trials or other evaluations of cancer therapies.
2 Methods

controlled transformations of cancer therapies.
2 Methods **Methods**

2.1 Introduction

We used Mendelian Randomization to estimate the causal effect of cancer on healthcare costs
and on quality of life, and in turn, quality adjusted life years. Many introductions to Mendelian
Randomization are available (39, Randomization are available (39, 41-43). Briefly, Mendelian Randomization relies on Mendel's
first (random segregation of alleles at conception) and second (independent assortment of
alleles) laws of inheritance, which des Randomization are available (39, 12-43). Briefly, Mendelian Randomization relieve the internet of
first (random segregation of alleles at conception) and second (independent assortment of
alleles) laws of inheritance, whic alleles) laws of inheritance, which describe how genetic variants are acquired by children fre
their parents. An allele refers to the specific form of genetic variation found a particular
location in the genome. alleles) laws of inheritance, which describe the regions of an anti-dependency children in the senome.
In their parents. An allele refers to the specific form of genetic variation found a particular
Some of these genetic v

the individual in the genome.
Some of these genetic variants transmitted in this conditionally random manner are kno
influence the risk of expressing specific phenotypes, including disease phenotypes such a Some of these genetic vanishing
influence the risk of expression is
cancer. At a population l influence the risk of expressing specific phenotypes, including disease phenotypes such as
cancer. At a population level, the association of genetic variants that influence the risk of, for
example, a site-specific cancer, influence the risk of expressing specific phenotypes, including disease phenotypes such as
cancer. At a population level, the association of genetic variants that influence the risk of, for
example, a site-specific cancer, cancer. At a site-specific cancer, permits the statistical identification of groups that differ in the
levels of their exposure to the risk for that cancer.
Genetic variants inherited according to Mendelian principles are

Genetic variants inherited according to Mendelian principles are therefore strong candidates as
instrumental variables since their association with an exposure of interest can – in principle – levels of their exposure to the risk for the cancer.
Genetic variants inherited according to Mendelian
instrumental variables since their association with
be demonstrated (ideally from genome wide assoc instrumental variables since their association with an exposure of interest can – in principle –
be demonstrated (ideally from genome wide association studies), thus satisfying the first
assumption ("relevance") of valid instrumental variables since their assessment can also inputed since their in principle
be demonstrated (ideally from genome wide association studies), thus satisfying the first
assumption ("relevance") of valid instrument

assumption ("relevance") of valid instrumental variable analysis.
The second assumption necessary for valid instrumental variable analysis is that the instrument
be independent of all confounding variables; put differently assumption ("relevance") of valid instrumental variable analysis. The independent of all confounding variables; put differently, that the instrument should be as
"good as randomly assigned" across different groups defined by confounding variables. This
assumption cannot be tested, as thi be independent of assigned.

"Bood as randomly assigned" across different groups defined by confounding variables. This

assumption cannot be tested, as this requirement encompasses all manner of confounding

variables, in assumption cannot be tested, as this requirement encompasses all manner of confounding
variables, including those are unmeasured, unobserved, and unknown. However, the
assignment of genetic variants at conception means tha ariables, including those are unmeasured, unobserved, and unknown. However, the
assignment of genetic variants at conception means that postnatal influences are unlikely to
confound associations of interest, meaning that d variables, including these are although an extended, and until the terms of the assignment of genetic variants at conception means that postnatal influences are unli
confound associations of interest, meaning that differen confound associations of interest, meaning that differences in outcomes between groups
defined by genetic liability to cancer (in this case) may be more causally attributed to the can
exposure itself. defined by genetic liability to cancer (in this case) may be more causally attributed to the cancer
exposure itself. deprosure itself.
By generation of this case of this case of this case of this case of the case of the cancer cancer of the cancer
S exposure itself.
5

The third assumption for valid instrumental variable analysis, the exclusion restriction, requires
that the instrument influence the outcome only via the exposure of interest. This is generally
untestable since it implies that the instrument influence the calculate the instrument from all outcomes (including
potential outcomes) conditional on the exposure. Moreover, genetic instrument variables of
the type used in Mendelian Randomization ar potential outcomes) conditional on the exposure. Moreover, genetic instrument variable
the type used in Mendelian Randomization are very likely to violate this assumption in tv
ways. the type used in Mendelian Randomization are very likely to violate this assumption in two
ways.

the tirst potential violation is linkage disequilibrium, which refers to correlation between
The first potential violation is linkage disequilibrium, which refers to correlation between
genetic variants which may arise whe The first potential violation is linkage disequilibrium, which refers to correlation between
genetic variants which may arise when they are located near to each other on the genome
Linkage disequilibrium will violate the e genetic variants which may arise when they are located near to each other on the genome
Linkage disequilibrium will violate the exclusion restriction if correlated variants affect the
outcome other than via the exposure of general channel may also may are located the to each other on the general change.
Linkage disequilibrium will violate the exclusion restriction if correlated variants affect the
outcome other than via the exposure of inter outcome other than via the exposure of interest. The second potential violation of the excl
restriction may arise due to pleiotropy, which refers to the effect of a variant on more thar
phenotype. Pleiotropy will violate t restriction may arise due to pleiotropy, which refers to the effect of a variant on more than one
phenotype. Pleiotropy will violate the exclusion restriction if these other phenotypes also affect
the outcome through chann phenotype. Pleiotropy will violate the exclusion restriction if these other phenotypes also affect

phenotic reference through channels that operate independently of the exposure.
Below, we describe how we identified genetic variants to serve as instrumental variables for as
wide a variety of site-specific cancers as pos the channel indeptit and operate independently of the exposure.
Below, we describe how we identified genetic variants to serve as instrumen
wide a variety of site-specific cancers as possible in our Mendelian Randomiz
also wide a variety of site-specific cancers as possible in our Mendelian Randomization analysis. We
also describe the steps taken to ensure our results were robust to potential violations of the
instrumental variable assumptio also describe the steps taken to ensure our results were robust to potential violations of the
instrumental variable assumptions.
2.2 Polvgenic risk score instrumental variable models also describe the steps taken to ensure the steps to potential violations in the
instrumental variable assumptions.
We developed polygenic risk scores (PRSs) for each cancer, the specific steps in the

Polygenic risk score instrumental variable models

We developed polygenic risk scores (PRSs) for each cancer, the specific steps in the
construction of which are described in detail below. The PRSs, which are functions of the Construction of which are described in detail below. The PRSs, which are functions construction of which are described in detail below. The PRSs, which are functions of weighted log-odds of developing each specific cancer, weighted log-odds of developing each specific cancer, were used as instruments in "just
identified" two-stage-least square (2SLS) instrumental variable models. "Just identified" n
have as many instruments as there are expo identified" two-stage-least square (2SLS) instrumental variable models. "Just identified"
have as many instruments as there are exposures whereas "over identified" models have
instruments than exposures. We used the *ivreg* have as many instruments as there are exposures whereas "over identified" models have more
instruments than exposures. We used the *ivreg2* package in Stata (version 17.0) with robust
standard errors to estimate these mode

In the first stage of the 2SLS model for each cancer, a cancer exposure variable was regressed instruments than exposures. We used the *ivregz* package in Stata (version 17.0) with robust
standard errors to estimate these models.
In the first stage of the 2SLS model for each cancer, a cancer exposure variable was re In the first stage of the 2SLS model for each
on the PRSs with age at baseline assessmer
6 In the PRSs with age at baseline assessment, sex, 40 genetic principal components (to control
6 on the PRSs with all angles with all all all all all all all all principal components (to control to control to
6

for the effects of the coefficient on the instrument were then used in the
second stage regressions with either costs or quality of life as the outcomes. F-statistics from
the first stage regressions were inspected to asse covariates. The predicted values of the coefficient on the instrument were then used in the second stage regressions with either costs or quality of life as the outcomes. F-statistics from the first stage regressions were the first stage regressions were inspected to assess instrument strength to assess whether the
PRS for each cancer was sufficiently associated with the observed cancers in the population
(23). the first stage regression were impressed to assess in strength to assess in the population
PRS for each cancer was sufficiently associated with the observed cancers in the population
Qur Mendelian randomization analysis e

profile).
Providely and the mean of the outcomes using
additive structural mean model (24–26), interpreted as the average change in each outcome
additive structural mean model (24–26), interpreted as the average change in `
Our M
addit
cause additive structural mean model (24–26), interpreted as the average change in each outcome
caused by the hypothetical average effect of having the cancer versus not across the population
(44). This assumes that a constant e caused by the hypothetical average effect of having the cancer versus not across the populati
(44). This assumes that a constant effect of genetic liability to each cancer on the cost and
quality of life outcomes. An alter (44). This assumes that a constant effect of genetic liability to each cancer on the cost and
quality of life outcomes. An alternative assumption is to interpret the results as a local average
treatment effect, which assu (44). This anality of life outcomes. An alternative assumption is to interpret the results as a local ave
treatment effect, which assumes a monotonic effect of genetic variants on each cancer
exposure (45). quality of the calculation alternative assumes in the representative as a local construction is to interpret
treatment effect, which assumes a monotonic effect of genetic variants on each cancer
exposure (45).
We compared

exposure (45).
We compared the Mendelian randomization estimates with estimates from convention:
separate multivariable linear regressions for QALYs and healthcare costs, with age, sex, expressively

We compared

separate multiv

genetic princip separate multivariable linear regressions for QALYs and healthcare costs, with age, sex, 40
genetic principal components, and recruitment centre as covariates. We performed Hausmar
endogeneity tests (27), in which a small genetic principal components, and recruitment centre as covariates. We performed Hausman
endogeneity tests (27), in which a small p-value indicates there was evidence the Mendelian
Randomization and multivariable effect es endogeneity tests (27), in which a small p-value indicates there was evidence the Mendelian
Randomization and multivariable effect estimates were different, although the power of these
tests is relatively low (46). endomization and multivariable effect estimates were different, although the power of thes
tests is relatively low (46).
2.3 Sensitivity analyses Randomization and multiplants of the community and the power of the power of these
Randomized estimates were different and the power of the
We conducted a variety of sens

2.3 Sensitivity analyses

2.3 Sensitivity analyses
We conducted a variety of sensitivity analyses. We performed a multivariable adjusted analysis with additional variables (Townsend Deprivation Index, household income, highest educational
qualification, ever smoked, and body mass index), to reduce the risk of confounding in the
multivariable adjusted analyses. with additional variables (Townsend Deprivation Index), to reduce the risk of confounding in the
qualification, ever smoked, and body mass index), to reduce the risk of confounding in the
multivariable adjusted analyses.
W

qualitivariable adjusted analyses.
The interpret smokes multivariable adjusted analysis in the re-ran the main Mendelian Randomization analysis stratified by sex and age group (less
Than 50 years, 50 to 59 years, and 60+ y mananaans argas caramayses.
We re-ran the main Mendelian R
than 50 years, 50 to 59 years, an
by when the participant was diag than 50 years, 50 to 59 years, and 60+ years), both separately and together. We also stratified
by when the participant was diagnosed with cancer, either before or after recruitment into UK
7 by when the participant was diagnosed with cancer, either before or after recruitment into UK by when the participant was diagnosed with cancer, either before or after records.
7

Biobank, as participants who received their cancer diagnosis before recruitment were more
likely to have smaller healthcare costs associated with diagnosis, and possibly curative
treatment, but more likely to have greater likely to have smaller healthcare costs and alleged the possible, analytical treatment, but more likely to have greater healthcare costs associated with managing progressing cancers, and vice versa for participants who rec trogressing cancers, and vice versa for participants who received their cancer diagnosi
recruitment.
We also examined potential violations of the exclusion restriction due to horizontal ple

progressing cancers, and vice versa for participants who received their cancer and α
recruitment.
This occurs when a variant influences the outcome through a channel other than the cancer o We also exan
This occurs w
interest. This This occurs when a variant influences the outcome through a channel other than the cancer of
interest. This may be via another disease, trait, behaviour, or health condition other than the
cancer specifically under investi

To test for the presence of pleiotropy, we estimated over-identified Mendelian Randomization interest. This may be via another another, trait, behaviour, or health condition other than the
cancer specifically under investigation.
To test for the presence of pleiotropy, we estimated over-identified Mendelian Random To test for the presence of pleiotropy, a
models comprising the individual SNPs
in the effect of these SNPs on the cost a models comprising the individual SNPs used in constructing cancer-specific PRSs. Heterogeneity
in the effect of these SNPs on the cost and quality of life outcomes could indicate a violation of
the exclusion restriction. in the effect of these SNPs on the cost and quality of life outcomes could indicate a violation of
the exclusion restriction.
Heterogeneity in effect estimates across SNPs can be assessed by comparing Cochran's Q

the exclusion restriction.
Heterogeneity in effect estimates across SNPs can be assessed by comparing Cochran's Q
statistic $Q = \sum_{j=1}^{J} \frac{1}{\sigma_{\rm v}^2} (\hat{\beta}_j - \hat{\beta}_{IVW})^2$ to the critical values of a chi-squared distributi Heterogeneity in effect expressed in the statistic $Q = \sum_{j=1}^{J} \frac{1}{\sigma_{Y_j}^2} (\hat{\beta}_j)$ statistic $Q = \sum_{j=1}^{J} \frac{1}{\sigma_{Y_j}^2} (\hat{\beta}_j - \hat{\beta}_{IVW})^2$ to the critical values of a chi-squared distribution. In the
formula, there are J total SNPs, $\hat{\beta}_{IVW}$ is the inverse variance weighted (IVW) effect calculat
for a statistic $Q = \sum_{j=1}^{\infty} \frac{1}{\sigma_{Y_j}^2} (p_j - p_{IVW})^{-1}$ to the critical values or a chi-squared distribution. In this
formula, there are J total SNPs, $\hat{\beta}_{IVW}$ is the inverse variance weighted (IVW) effect calculated
for a $=\sum_{j=1}^{J} \frac{1}{\sigma_{Y_j}^2}$ $\int_{j=1}^{J} \frac{1}{\sigma_{Y_i}^2} (\beta_j - \beta_{IVW})^2$ formula, there are J total SNPs, β_{IVW} is the inverse variance weighted (IVW) effect calculated
for all J SNPs, $\hat{\beta}_j$ is the effect estimate for a specific SNP j, and $\sigma_{Y_j}^2$ the variance of the SNP-
outcome asso for all J SNPs, β_j is the effect estimate for a specific SNP j, and $\sigma_{Y_j}^2$ the variance of the SNP-
outcome association. The intuition for this formula is that any $\hat{\beta}_j$ with a large impact on th
outcome that di $\overline{}$ outcome association. The intuition for this formula is that any β_j with a large impact on the outcome that different from the overall IVW effect of all SNPs on the outcome may suggest multiple pleiotropic channels of i multiple pleiotropic channels of influence of that SNP on the outcome not mediated by the
specific cancer exposure investigated.
We used consensus methods and a modelling method (47) for our sensitivity analyses, each

multiple pleis to provide a multiple channels of that SNP on the outcome in channels by the
specific cancer exposure investigated.
We used consensus methods and a modelling method (47) for our sensitivity analyses, each
wh specific cancer in the used consensus methods and a module which embodied different types of assex
exclusion restriction may be violated. which embodied different types of assumption about whether and in which respects the
exclusion restriction may be violated. We first estimated the IVW estimates needed to calculate
Cochran's Q statistic, which involved cal exclusion restriction may be violated. We first estimated the IVW estimates needed to ca
Cochran's Q statistic, which involved calculating instrumental variable estimates for each
separately, and then combining them in ran cochran's Q statistic, which involved calculating instrumental variable estimates for each SNP
separately, and then combining them in random-effects meta-analysis with weights determined
8 Separately, and then combining them in random-effects meta-analysis with weights determine
8 separately, and then combined then in random-effects meta-analysis with weights determined the separately in random-
Separately, with weights determined the separately in random-effects meta-analysis with weights determine by the precision of the association between the SNP and the outcomes (healthcare costs and
quality of life respectively). This assumes no violations of the exclusion restriction or no net
effect of any such violations on t

effect of any such violations on the point estimates.
We then compared this restrictive baseline model against more models that are robust to
different types of exclusion restriction violation. We implemented the Mendelian ve then compared this restrictive baseline model ag
different types of exclusion restriction violation. We
Randomization Egger estimator, which is consistent We different types of exclusion restriction violation. We implemented the Mendelian
Randomization Egger estimator, which is consistent even if all SNPs are pleiotropic provide
additional assumptions (relating to the associ MATE THE STEM CONDUCT THE MENDED TO METHEM THE METHEM AND MEDITION Randomization Egger estimator, which is consistent even if all SNPs are pleiotropic
additional assumptions (relating to the association between instrument additional assumptions (relating to the association between instrument strength and the dire
pleiotropic effect of SNPs) are met – see (48, 49). We also implemented a penalized weighted
median estimator, which is consisten pleiotropic effect of SNPs) are met – see (48, 49). We also implemented a penalized weighted
median estimator, which is consistent if at least 50% if the SNPs are valid instrumental variables
(50). We also implemented a we pleis are presented to substitute the state of the SNPs are valid instrumental variable
median estimator, which is consistent if at least 50% if the SNPs are valid instrumental variable
(50). We also implemented a weighted (50). We also implemented a weighted mode estimator, which is consistent if the largest
homogenous cluster of SNPs are valid even if more than 50% of SNPs are invalid as instrumental
variables (51). (50).
homogenous cluster of SNPs are valid even if more than 50% of SNPs are invalid as instru
variables (51).
Note that results from these estimators are on a different scale to the instrumental varial

homogenous cluster of SNPs are valid even if more almost a state of SNPs are included to the instrumental variable
Note that results from these estimators are on a different scale to the instrumental variable
estimates obt variables (* 1).
Note that resu
analyses reflec estimates obtained from the polygenic risk score models. The over-identified sensitivity
analyses reflect the measurement scales used for the outcomes in the respective source
genome-wide association studies (GWASs), which estimates summer in the polygens constants in the polygenic risk score
analyses reflect the measurement scales used for the outcomes in the respective source
genome-wide association studies (GWASs), which are based on logi analysing enome-wide association studies (GWASs), which are based on logistic regression. The source these outcomes therefore reflects a change in costs or quality of life per unit change in the odds of cancer status. Thes these outcomes therefore reflects a change in costs or quality of life per unit change in the log-
odds of cancer status. These results are on a relative rather than absolute scale, reflecting the
relative increase in gene estimates of genetically influenced changes in disease status in the analysis nonulation of cancer statust status in the cancer status change in an absolute scale, relative increasing
relative increase in genetic liability to cancer from increasing the number of risk-increasing
alleles. By contrast, the effect alleles. By contrast, the effect estimates obtained from the 2SLS just-identified models are
estimates of genetically influenced changes in disease status in the analysis population.
2.4 Modelling the cost-effectiveness of estimates of genetically influenced changes in disease status in the analysis population.
2.4 Modelling the cost-effectiveness of a novel anti-cancer interventior
We used our new estimates of the causal impact of cancer on

2.4 Modelling the cost-effectiveness of a novel anti-cancer intervention

estimates of generating interactioning of material changes in the analysis population.

2.4 Modelling the cost-effectiveness of a novel anti-cancer intervent

We used our new estimates of the causal impact of cancer on hea We used our new estimates of the causal impact of cancer on healthcare cost and QALYs to
simulate the impact of a new population-wide anti-cancer intervention. Zheng et al (52) used a
Mendelian Randomization study design t Mendelian Randomization study design to assess the impact of sodium-glucose cotransporter 2 (SGLT2) inhibitors (a type of oral glucose lowering drug) on prostate cancer risk. Their analysis
found that SGLT2 inhibition, equivalent to a 1.09% reduction in HbA1c levels, lowered the odds
9 found that SGLT2 inhibition, equivalent to a 1.09% reduction in HbA1c levels, lowered the odd:
9
9 found that SGLT2 inhibition, equivalent to a 1.099 reduction, lowered the odds in Hand levels, lowered the odd
SGLT2 in Hand levels, lowered the odds in Hand levels, lowered the odds in Hand levels, lowered the odds in H
S

in replication analysis (OR=0.51, 95% CI: 0.33 to 0.79).
In the absence of data on how such the hypothetical SGLT2 intervention would be dosed and costed for this new indication, we assumed that it would be offered in the in replication analysis (OR=0.5) CIX CI: 0.33 to 0.51, 95% In the absence of data on how such the hypothetical Secosed for this new indication, we assumed that it wou canagliflozin, a particular type of SGLT2 inhibitor. As costed for this new indication, we assumed that it would be offered in the form of 300mg of
canagliflozin, a particular type of SGLT2 inhibitor. As a monotherapy, a dose of 300mg of
canagliflozin had a broadly similar (53, canagliflozin, a particular type of SGLT2 inhibitor. As a monotherapy, a dose of 300mg of
canagliflozin had a broadly similar (53, 54) impact on HbA1c (approximately a 1% reduction) a
did the MR estimates in the Zheng et a canagliflozin had a broadly similar (53, 54) impact on HbA1c (approximately a 1% reductionally canagliflozin had a broadly similar (53, 54) impact on HbA1c (approximately a 1% reduction
did the MR estimates in the Zheng et did the MR estimates in the Zheng et al (52) paper. These comparisons are necessarily rather
coarse since the Zheng et al (52) estimates relate to a largely unselected population, while the
effectiveness estimates in the t coarse since the Zheng et al (52) estimates relate to a largely unselected population, while the
effectiveness estimates in the trials relate to effectiveness amongst diabetic patients, are
dependent (to some degree) on ba effectiveness estimates in the trials relate to effectiveness amongst diabetic patients, are
dependent (to some degree) on baseline HbA1c levels, and we assume in our analysis no side-
effects from prophylactic treatment w dependent (to some degree) on baseline HbA1c levels, and we assume in our analysis no s
effects from prophylactic treatment with these inhibitors.
We further note that the Zheng et al (52) estimates pertain to lifelong SGL

 m dependent (to some degree α) on example, the middle-to-early old age inen represented in α effects from prophylactic a summate with these inhibitors.
We further note that the Zheng et al (52) estimates pertain
initiation of daily canagliflozin by, for example, the middle
our UK Biobank cohort will likely underst initiation of daily canagliflozin by, for example, the middle to early old age men represented
our UK Biobank cohort will likely understate the cost-effectiveness of this intervention. Thes
are restrictive assumptions that our UK Biobank cohort will likely understate the cost-effectiveness of this intervention. These
are restrictive assumptions that may not apply to the use of these drugs in a "real-world"
deployment, but at least serve to i initial estimate of cost-effectiveness for a novel population-wide anti-cancer intervention. deployment, but at least serve to indicate how these methodologies can rapidly provide a
initial estimate of cost-effectiveness for a novel population-wide anti-cancer intervention.
Our hypothetical intervention therefore

deproyment, and at the process for a novel population-wide anti-cancer intervention.
Initial estimate of cost-effectiveness for a novel population-wide anti-cancer intervention.
Our hypothetical intervention therefore assu Our hypothetical intervention therefore assumed that every man in the analysis sample wa
offered and took daily SGLT2 inhibitors (and specifically 300mg of canagliflozin) from their
recruitment into the UK Biobank cohort. offered and took daily SGLT2 inhibitors (and specifically 300mg of canagliflozin) from their
recruitment into the UK Biobank cohort. To estimate the causal impact of prostate cancer on
healthcare cost and QALYs, we needed offerruitment into the UK Biobank cohort. To estimate the causal impact of prostate cancer
healthcare cost and QALYs, we needed to estimate 3 quantities: 1) the proportion of men
developed prostate cancer, 2) the proportio recruitment into the UK Biobana constant in the UK Biobana in proportion of men who
developed prostate cancer, 2) the proportion of men who wouldn't have developed prostate
cancer had they received the intervention, and 3) developed prostate cancer, 2) the proportion of men who wouldn't have developed prostate
cancer had they received the intervention, and 3) the total healthcare and QALY cost saved by
preventing those men from developing pr cancer had they received the intervention, and 3) the total healthcare and QALY cost saved by
preventing those men from developing prostate cancer. Each of these three quantities had
associated statistical imprecision. preventing those men from developing prostate cancer. Each of these three quantities had preventing these men from developing prostate cancer. Each of these three quantities had
associated statistical imprecision. associated statistical imprecision.

It is made available under a CC-BY 4.0 International license. **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2023.02.06.23285521;](https://doi.org/10.1101/2023.02.06.23285521) this version posted February 8, 2023. The copyright holder for this preprint

For the first quantity, we identified the proportion of all men in UK Biobank who were
diagnosed with prostate cancer, as well as within different age bands (less than 50 years, 50 to 59 years, and 60+ years), estimating the relevant standard errors (SE) of each proportion:

$$
SE = \sqrt{\frac{proportion \times (1-proportion)}{N}}
$$

where N is the total number of men (overall, or within each age band).

For the second quantity, we converted the odds ratios from Zheng et al (52) and the proportions estimated in quantity 1 into risk differences: proportions estimated in quantity 1 into risk differences:
Risk difference $= \frac{n \times 0R}{(N - n + n \times 0R)} - \frac{n}{N}$

Risk difference =
$$
\frac{n \times \text{OR}}{(N - n + n \times \text{OR})} - \frac{n}{N}
$$

where n is the number of men with prostate cancer, N is the total number of men, and OR is *N*
un
ba
et the odds ratio. For example, 6,155 of 144,032 men (4.3%) in UK Biobank were diagnosed with
prostate cancer up to 31 March 2017, so an OR of 0.29 from Zheng et al (52) gives a risk
difference of -3.0%.

For the third quantity, we used the relevant multivariable adjusted or Mendelian randomization prostate cancer a_proses and a right 2017, so an order to the 2019, so an _{(5.2}) gives a risk
difference of -3.0%.
For the third quantity, we used the relevant multivariable adjusted or Mendelian randon
results estimated For the third quantit
results estimated in
We multinlied the ri

For the third quantity, we used the relevant multipliers in galaxies or Mendelian randomization
results estimated in section 2.2, both for total annual healthcare costs and QALYs.
We multiplied the risk difference (quantit We multiplied the risk difference (quantity 2) by the relevant multivariable adjusted or
Mendelian randomization estimate (quantity 3) to estimate the annual total healthcare
and QALYs saved by reducing incidence of prosta We manapose and concurrence (quantity 3) to estimate the annual total healthcare
and QALYs saved by reducing incidence of prostate cancer with the intervention.
Because each quantity had associated uncertainty, we repeated

and QALYs saved by reducing incidence of prostate cancer with the intervention.
Because each quantity had associated uncertainty, we repeated the hypothetical intervention
10,000 times, with each quantity taken from a norm and QALIY save and a proceed by proceed the incorportation.
Because each quantity had associated uncertainty, we repeated the hypothetical
10,000 times, with each quantity taken from a normal distribution with a mean of
es Because each quantity taken from a normal distribution with a mean of the effect
estimate, and a standard deviation of the standard error of the effect estimate. We present the
median of the resulting estimates as the poin estimate, and a standard deviation of the standard error of the effect estimate. We present t
median of the resulting estimates as the point estimate for the effect of the intervention, wit
the 2.5 and 97.5 percentiles as estimate) and a standard a standard contract and the standard error of the standard in episonement
median of the resulting estimates as the point estimate for the effect of the intervention, with
the 2.5 and 97.5 percentil the 2.5 and 97.5 percentiles as the 95% confidence interval. This approach (implicitly) accounts
for survival up to a median of seven years of follow up.
11 for survival up to a median of seven years of follow up.
11 for survival up to a median of f survival up.

It is made available under a CC-BY 4.0 International license. **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2023.02.06.23285521;](https://doi.org/10.1101/2023.02.06.23285521) this version posted February 8, 2023. The copyright holder for this preprint

We estimated the cost-effectiveness of SGLT2 inhibition for population-wide prophylaxis of
incident prostate cancer. Using estimates from our conventional multivariable and Mendelian
Randomization estimates of the impact o Randomization estimates of the impact of prostate cancer on healthcare costs and qualityassuming that the cost of doing so was equivalent to the cost 300mg of daily canagliflozin.

We then compared the cost and effect of so doing to a cost effectiveness threshold or Ezo $_{j}$ assum of the cost of the cost and effect of so doing to a cost-effectiveness threshold of £20,
using a net monetary benefit calculation. This threshold is routinely used by the National
Institute of Health and Care Excelle using a net monetary benefit calculation. This threshold is routinely used by the National
Institute of Health and Care Excellence (NICE) to appraise the value of new medical
technologies, including pharmacotherapeutic age using a net institute of Health and Care Excellence (NICE) to appraise the value of new medical
technologies, including pharmacotherapeutic agents. Under this approach, interventions
cost-effective when their incremental b Institute of Health and Care Excellence (NICE) to approach care international technologies, including pharmacotherapeutic agents. Under this approach, interver
cost-effective when their incremental benefit (measured as the the cost-effective when their incremental benefit (measured as the difference between
intervention benefits and comparator benefits) scaled by the cost-effectiveness threshold
exceed incremental costs (measured as the diff intervention benefits and comparator benefits) scaled by the cost-effectiveness thre
exceed incremental costs (measured as the difference between intervention costs an
comparator costs). We measure benefits in the present intervention benefits and comparator benefits, contractly increased incremental costs (measured as the difference between intervention costs and
comparator costs). We measure benefits in the present context using quality-a comparator costs). We measure benefits in the present context using quality-adjusted l
years, and compared population-wide SGLT2 inhibition to a "do nothing" comparator, v
reflects the absence of any accepted pharmacothera comparator costs). We measure benefits in the present content along quality adjusted in the
years, and compared population-wide SGLT2 inhibition to a "do nothing" comparator, which
reflects the absence of any accepted phar reflects the absence of any accepted pharmacotherapeutic intervention for reducing
population-wide prostate cancer risk.
3 Data reflects the absence of any accepted pharmacotherapeutic intervention for reducing
population-wide prostate cancer risk.
3 Data

Data

population
3.1 Data
3.1 Study population 3.1 Study population

Who were recruited between the years 2006 and 2010 from 22 centres across the UK (6). We
used medical data from hospital episode statistics (HES) linked to all participants up to 31st
March 2015, and primary care (genera who were recruited between the year 200 km 2020 from 22 chinese and 2015 used medical data from hospital episode statistics (HES) linked to all participants up to 31st
March 2015, and primary care (general practice) data used medical data from hospital episode statistics (HES) linked to all participants up to 31"
March 2015, and primary care (general practice) data linked to approximately 31% of UK
Biobank participants registered with GP s March 2015, and primary care (general practice) data linked to approximately 31% of UK
Biobank participants registered with GP surgeries using EMIS Health (EMIS Web) and TPP
(SystmOne) software systems, up to 31st March (SystmOne) software systems, up to 31st March 2017. The study design, participants and control methods have been described in detail previously (7–9). UK Biobank received ethicap
approval from the Research Ethics Commit (SystmOne) software systems, up to 31" March 2017. The study design, participants and quality
control methods have been described in detail previously (7–9). UK Biobank received ethics
approval from the Research Ethics Com control methods have been described in detail previously (8–9). Understand termined three
approval from the Research Ethics Committee (REC reference for UK Biobank is 11/NW/038)
Further information on genotyping is availab Further information on genotyping is available in (10).
12 Further information on $\frac{1}{2}$ is available in (10).

We restricted the main analysis to unrelated individuals of white British ancestry (to avoid
confounding by population stratification) living in England or Wales at recruitment. In our
sample, 38% of participants had prima confounding by population stratification) in England or Material Conformation in Can

Sample, 38% of participants had primary care data.

3.2 Identification of cancers and creation of polygenic risk scores

We used UK Biob

3.2 Identification of cancers and creation of polygenic risk scores

We used UK Biobank as the source of our outcome data, the creation of which is described
below. To avoid biases from sample overlap (55), we searched for genome-wide association We used the source of our other the source of the source of our outcome the association
below. To avoid biases from sample overlap (55), we searched for genome-wide association
studies (GWASs) on any type of cancer not usi below. The contract of the Warret Studies (GWASs) on any type of cancer not using data from UK Biobank cohort. We searched
the MR Base (mrbase.org) database of GWASs and the Elizabeth Blackwell Institute GWAS
catalogue to stance (CONTECT) on any cype of cancer not using data from UNDERMINDENTIAL TO CONTENT THE UNITED THE STANDER T
the MR Base (mrbase.org) database of GWASs and the Elizabeth Blackwell Institute GWAS
catalogue to identify can catalogue to identify cancers with at least clumped 2 single nucleotide polymorphisms (SNPs)
associated with the cancer.

catalogue to identify cancers with specific international classification of disease
We also restricted our analysis to cancers with specific international classification of disease
(ICD) codes at one decimal place, for exa We also restricted our analysis to cancers with specific international classification of disease
(ICD) codes at one decimal place, for example, multiple myeloma is coded as "230.0" in ICD-9,
and "C90.0" in ICD-10. Where we (ICD) codes at one decimal place, for example, multiple myeloma is coded as "230.0" in ICD-5
and "C90.0" in ICD-10. Where we found multiple GWAS for the same cancer, we preferential
used the GWAS with the most participants and "C90.0" in ICD-10. Where we found multiple GWAS for the same cancer, we preferentially
used the GWAS with the most participants.
For each cancer, we clumped the genome-wide significant SNPs at an R^2 threshold of 0.

and "C90.0" in ICD-10. Where we found multiple 200.0" is the same cancer, μ preferentially
used the GWAS with the most participants.
For each cancer, we clumped the genome-wide significant SNPs at an R^2 threshold of For each cancer, we clumped the genome-v
within a 10,000 kilobase window. We search
the European subsample of 1,000 genomes For each cancer, we clumped the genome-wide signiticant SNPs at an R° threshold of 0.001
within a 10,000 kilobase window. We searched for proxies for all SNPs not in UK Biobank us
the European subsample of 1,000 genomes as the European subsample of 1,000 genomes as a reference panel (with a lower R² limit of 0.6)
(11). We used all included SNPs to construct PRS for each cancer, calculated as the weighted
sum of the SNP effect alleles for a the European subsample of 1,000 genomes as a reference panel (with a lower R° limit of 0.6)
(11). We used all included SNPs to construct PRS for each cancer, calculated as the weighted
sum of the SNP effect alleles for all (11). We used an included First Protein First Protein City, Christian as the Weighte
sum of the SNP effect alleles for all SNPs associated with each cancer, with each SNP weighte
by the regression coefficient from the corr

sum of the regression coefficient from the corresponding GWAS.
For some cancers, we found multiple GWASs which had subgroups of cancer as the outcome. In
these cases, we created PRSs for the main cancer (e.g. breast cancer For some cancers, we found multiple GWASs which had subg
these cases, we created PRSs for the main cancer (e.g. breast
subgroups (e.g. ER+ and ER- breast cancers). There was insufi These cases, we created PRSs for the main cancer (e.g. breast cancer) as well as the available
subgroups (e.g. ER+ and ER- breast cancers). There was insufficient detail in the ICD codes to
determine the subgroup of cancer subgroups (e.g. ER+ and ER- breast cancers). There was insufficient detail in the ICD codes to
determine the subgroup of cancer for all participants in UK Biobank, and as consequence the
exposure used in the main analyses determine the subgroup of cancer for all participants in UK Biobank, and as consequence the
exposure used in the main analyses and reported below was for the general (e.g. breast cance
exposure rather than its subtypes. In exposure used in the main analyses and reported below was for the general (e.g. breast cancer)
exposure rather than its subtypes. In Supplementary Table S3 we also report Mendelian
Randomization analysis using the availabl exposure rather than its subtypes. In Supplementary Table S3 we also report Mendelian exposure rather than its substitute rather than its substitution analysis using the available cancer subgroups.
13 R and R available cancer subset of available cancer subgroups.

Following this process, the following cancers were available to be analysed: bladder, breast, colorectal, lung, multiple myeloma, ovarian, prostate and thyroid. **Box 1** contains details for all PRS we generated for these colorectal, lung, multiple myeloma, ovarian, prostate and thyroid. Box 1 contains details for all
PRS we generated for these cancers, including the source GWAS, the number of SNPs in the
PRS, and the R² value of the PRS PRS, and the R² value of the PRS and the cancer from logistic regression with no covariates.
Supplementary Table 1 contains details for all SNPs used these PRSs. We created PRS for 8
cancers that had suitable GWAS, with breast cancer, 2 for lung cancer, 9 for ovarian cancer), see Box 1. Table 1 contains details for all SNPs used these PRSs. We created PRS for 8
d suitable GWAS, with an additional 13 PRS for subgroups of these cancers (2
2 for lung cancer, 9 for ovarian cancer), see **Box 1**. cancers that had suitable GWAS, with an additional 13 PRS for subgroups of these cancers (
breast cancer, 2 for lung cancer, 9 for ovarian cancer), see **Box 1**.
Box 1: Creation of cancer polygenic risk scores

	Box 1: Creation of cancer polygenic risk scores			
Cancer	Subgroup	GWAS	SNPs	R^2
Bladder	All	Figueroa et al (56) (2013)	10	0.57%
Breast	A	Michailidou et al (2017)(57)	139	2.54%
	ER-		105	0.56%
	ER+		36	2.38%
Colorectal	All	Schumacher et al (2015)(58)	8	0.30%
Lung	A	Wang et al (2014) (59)	5	0.08%
	Adenocarcinoma		$\overline{2}$	0.11%
	Squamous cell		$\overline{4}$	0.13%
Multiple myeloma	All	Mitchell et al (2016) (60)	13	0.72%
Ovarian	A	Phelan et al (2017) (61)	12	0.20%
	Clear cell		$\mathbf{1}$	0.01%
	Low malignant potential		4	0.10%
	Mucinous, invasive		3	0.002%
	Mucinous, invasive, low malignant potential		5	0.001%
	Mucinous, low malignant potential		4	0.002%
	Serous, high grade		15	0.24%
	Serous, high and low grade		15	0.25%
	Serous, low grade, low malignant potential		8	0.16%
	Serous, low malignant potential		3	0.10%
Prostate	A	Schumacher et al (2018)(62)	123	4.04%
Thyroid	All	Gudmundsson et al (2017)(63)	8	1 5 4 %

Prostate All (2018) (62) 123 4.04%

Thyroid All (2018) (62) 123 4.04%

Notes: GWAS: genome wide association study. SNP: Single nucleotide polymorphism. ER- and ER+: Estroge

negative (ER-) and estrogen receptor positive (Thyroid All All $(2017)(63)$ 8 1.54%

Notes: GWAS: genome wide association study. SNP: Single nucleotide polymorphism. ER- and ER+: Estroge

negative (ER-) and estrogen receptor positive (ER+). The "All" categories indicat Notes: GRAS: Graditional analyses, using GWAS for subgroups of cancer, although the exposure was all cancers in the respective cate negative and analyses, using GWAS for subgroups of cancer, although the exposure was all cancers in the respective
category. R² indicates the pseudo R² from a logistic regression of the cancer on its PRS, with no other

category. R² indicates the pseudo R² from a logistic regression of the cancer on its PRS, with no other included
covariates.
We used age, sex and UK Biobank recruitment centre reported at the UK Biobank bas
assessment www.ww
We used
assessme
control f \
c
c assessment as covariables, as well as 40 genetic principal components derived by UK Bioba
control for population stratification (35).
14 assessment as covariables, as well as 40 generic principal components derived by UK Biobank to Control for population stratification (35).
14 control for population stratification (35).

3.3 Estimation of quality-adjusted life years

We created health-related quality of life for all participants in structural continuation
until March 2017. The details of how these data were constructed are given in Harrison et al
(64). In brief, we used the data on qua until March 2017. In brief, we used the data on quality of life in Sullivan et al (65) for 240 health conditions
to create person-specific quality of life scores for every individual for all participants per day
UK Biobank (64). In the data on and the data on quality of life in the data on quality did participants per day in
UK Biobank from recruitment to 31 March 2017 or death, whichever came first. We multiplied
the results for QALYs by 10 UK Biobank from recruitment to 31 March 2017 or death, whichever came first. We multiplied
the results for QALYs by 100 to give the percentage of a QALY changed by having versus not
having the cancer. Note that the QALY es the results for QALYs by 100 to give the percentage of a QALY changed by having versus not the results for Caucasian average the cancer. Note that the QALY estimates represent the change in QALYs over an
average year of follow-up; that is, they are not cumulative over time.
These health conditions were recorded

having the cancer. Note that the QALY estimates represent the change in QALY estimates
average year of follow-up; that is, they are not cumulative over time.
These health conditions were recorded in primary care records, h are regely can extend to applicate typical procession.
These health conditions were recorded in primary care records, hospi
episode statistics) or self-reported in the baseline UK Biobank recruitr
data were averaged over y episode statistics) or self-reported in the baseline UK Biobank recruitment questionnaire. These
data were averaged over years of follow-up to create quality-adjusted life years (QALYs) for
each individual. The full list o episode statistical over years of follow-up to create quality-adjusted life years (QALYs) for
each individual. The full list of health conditions used to create these QALYs is reported in
Supplementary Table 2, alongside t data were averaged over years of following to create quality days create the quality for each individual. The full list of health conditions used to create these QALYs is reported in Supplementary Table 2, alongside the IC each individual. The can be condition of health conditions used to create these designed in Supplementary Table 2, alongside the ICD-9, ICD-10, Read version 2 and 3 codes (for primation care) associated with each condition Supplementary Table 2, along the ICD-9, along the ICD-9, increased the ICD-9, increased associated with each condition.
 S.4 Estimation of healthcare costs

We have previously reported our approach to creating healthcare care) associated with each condition.
 3.4 Estimation of healthcare costs

We have previously reported our approach to creating healthcare cost for the UK Biobank

Estimation of healthcare costs

cohort (22, 66). Briefly, we created healthcare costs from records of procedure and diagnosis
codes recorded in Hospital Episode Statistics linked to the UK Biobank cohort. This
encompassed all inpatient care (for which pa overnight) for publicly funded care in NHS hospitals from recruitment into UK Biobank to 31 encompassed all inpatient care (for which patients occupy a bed but do not necess
overnight) for publicly funded care in NHS hospitals from recruitment into UK Bioba
March 2015. We used these data to calculate inpatient ho encompassed all inpatient care (for unline patients occupy) all interaction at the transition, stay
overnight) for publicly funded care in NHS hospitals from recruitment into UK Biobank to 31
March 2015. We used these data March 2015. We used these data to calculate inpatient hospital costs per-patient per year of
follow-up. Primary care costs from recruitment into UK Biobank until 31 March 2017 were
estimated from the costs of primary care March 2021. We also have to calculate inpatient hospital cost per patient per year of
follow-up. Primary care costs from recruitment into UK Biobank until 31 March 2017 were
estimated from the costs of primary care consult estimated from the costs of primary care consultations and from the cost of prescribed dru
(as of November 2019). Further details on the construction of primary care costs are given
Harrison et al (64). (as of November 2019). Further details on the construction of primary care costs are given in
Harrison et al (64). (as of November 2019). Further details on the construction of primary care construction of given in
Harrison et al (64).
15 $\frac{15}{15}$

We combined annual primary and secondary care healthcare costs to create a total annual
healthcare cost figure per person for each year of follow up. This comprised costs related to
inpatient hospital care episodes, primar inpatient hospital care episodes, primary care appointments and primary care drug
prescriptions. Primary care data were available for 38% of the sample. We used multiple
imputation by chained equations to impute primary ca in the care of the sample. We used multipates in the set of the set of the sample. We used multimeutation by chained equations to impute primary care healthcare costs and other where data were missing. We created 100 imput prescriptions. Primary care and these are denoted to the samples the same inverped
imputation by chained equations to impute primary care healthcare costs and other outc
where data were missing. We created 100 imputed data where data were missing. We created 100 imputed datasets (16) and analysed these datasets
using Rubin's rules. Other elements of care, such as emergency care and private healthcare
undertaken in private facilities, were no where data were data were deniated as a series of the data weight of this conditional provide healthcare
undertaken in private facilities, were not available as outcomes for this cohort. Including these
costs would likely undertaken in private facilities, were not available as outcomes for this cohort. Including the
costs would likely increase the overall size of our total cost estimate per person but is very
unlikely to alter the magnitude costs would likely increase the overall size of our total cost estimate per person but is very
unlikely to alter the magnitude of the associations we estimate in conventional or Mendelian
Randomization models. condikely to alter the magnitude of the associations we estimate in conventional or Mendelia
Randomization models.
The price for 300mg of canagliflozin, used as an exemplar SGLT2 inhibitor to model the cos[.]

unlikely to alter the magnitude of the association of connect in conventions we estimate
Randomization models.
The price for 300mg of canagliflozin, used as an exemplar SGLT2 inhibitor to model the cost-
effectiveness of o The price for 300mg of c
effectiveness of offering
taken from the 2019 NH effectiveness of offering a population wide intervention to reduce prostate cancer risk, was
taken from the 2019 NHS drug tariff price (67). The monthly price for 30 canagliflozin tablets of
£39.20 was grossed up to give a

E39.20 was grossed up to give an annual cost figure of £470.
All costs used in our analysis were expressed in 2019 prices. Cost data not expressed in 2019
prices were inflated to this price level using the NHS cost inflat All costs used in our analysis were expressed in 2019 prices. Contains were inflated to this price level using the NHS cost inflated Ω All costs used in our analysis were expressed in 2019 prices. Cost and another suppressed in 2019
Prices were inflated to this price level using the NHS cost inflation index (68).
The empirical dataset will be archived wit

3.5 Data and code availability

prices were integrals of this prices were in the empirical dataset will be archived with UK Biobank and made available to
The empirical dataset will be archived with UK Biobank and made available to
obtain the necessary ne obtain the necessary permissions from the study's data access committees. The code used to
clean and analyse the data is available here: <u>https://github.com/sean-harrison-bristol/</u>
4 Results obtain and analyse the data is available here: <u>https://github.com/sean-harrison-bristol/</u>
 4 Results

4.1 Introduction

Results

clean and analysis are the data is available here: https://gitthe here: https://github.com/sean-harrison-britannical is available here: https://gitahere.org/
4.1 Introduction 4.1 Introduction

We analysed data on 310,913 unrelated individuals, of whom 166,981 (53.7%) were female.
Mean age at recruitment was 56.9 years (standard deviation: 8 years), and participants were
followed up for a mean of 8.1 years (stand followed up for a mean of 8.1 years (standard deviation 0.8 years). Median annual total followed up for a mean of 8.1 years (standard deviation 0.8 years). Median annual total total

healthcare costs per person were £601 (interquartile range: £212 to £1,217). The most
prevalent cancer we examined in UK Biobank was breast cancer (6.6% of women), followed by
prostate cancer (4.3% of men), colorectal canc prevalent cancer (4.3% of men), colorectal cancer (1.4%), bladder cancer (0.7%), lung cancer (0.7%), ovarian cancer (0.7% of women), multiple myeloma (0.2%), and thyroid cancer (0.2%). proses and the cancer (0.7% of women), multiple myeloma (0.2%), and thyroid cancer (0.2%), and cancer (0.2%), and cancer (0.2%), and cancer (0.2%), $\frac{1}{2}$ $(0.75, p = 0.75, p = 0.7$

Table 1 Summary of analysis sample

The Motes: SD: Standard deviation. IQR: Interquartile range. Results from imputed data, median & IQR are the medians of the

imputed medians/IQRs
 4.2 Results of main analysis *Notes: SD: Standard deviation. IQR: Interquartile range. Results from imputed data, median & IQR are the medians of the 100

4.2 Results of main analysis

4.2 Results
Estimates from co Estimate from conventional multiple analysis indicate and cancer analysis in
annual QALYs per person and increased annual healthcare costs per person (Table 2). The
quantitative impacts were material for all cancers; the l quantitative impacts were material for all cancers; the largest absolute impacts for QALYs
in relation to lung cancer and multiple myeloma for healthcare costs.
 in relation to lung cancer and multiple myeloma for healthcare costs. in relation to lung cancer and multiple myeloma for healthcare costs.

Notes
Tabl
outc
cost Total healthcare costs per year

yeality-adjusted life year.

ian Randomization estimates for the same cancer expos

ere relatively precise for breast cancer, showing a clear h Iuality-adjusted life year.

ian Randomization estimates for the same cancer experentatively precise for breast cancer, showing a clo

(£798 per year, 95% CI: £549 to £1,048 and -5.51%) Table 3 presents the Mendelian Randomization
outcomes. The estimates were relatively precis
cost and reduction in QALYs (£798 per year, 959 outcomes. The estimates were relatively precise for breast cancer, showing a clear healthcare
cost and reduction in QALYs (£798 per year, 95% CI: £549 to £1,048 and -5.51% QALYs per yea
95% CI: -9.32% to -1.70%), and for p cost and reduction in QALYs (£798 per year, 95% CI: £549 to £1,048 and -5.51% QALYs per year,
95% CI: -9.32% to -1.70%), and for prostate cancer, which had much lower healthcare costs and
change in QALYs (£134 per year, 95 95% CI: -9.32% to -1.70%), and for prostate cancer, which had much lower healthcare costs and change in QALYs (£134 per year, 95% CI: -£217 to £485 per year, and -2.68% QALYS per year,
95% CI: -7.48% to 2.12%).
Estimates were imprecise for all other cancers. The Mendelian randomization estimates are

change in QALYs (£134 per year, 95% CI: -7.48% to 2.12%).
95% CI: -7.48% to 2.12%).
Estimates were imprecise for all other cancers. The Mendelian randomization estimates are
necessarily less precise than the corresponding ,
Estimates were imprecise
necessarily less precise tha
function of the amount of Estimates were impressed in the corresponding conventional estimates, since their precision
function of the amount of variance in each cancer exposure explained by the respective
polygenic risk score. function of the amount of variance in each cancer exposure explained by the respective
polygenic risk score. function of the amount of variance in each cancer exposure expenses to your respective
polygenic risk score. polygenic risk score.

It is made available under a CC-BY 4.0 International license. **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2023.02.06.23285521;](https://doi.org/10.1101/2023.02.06.23285521) this version posted February 8, 2023. The copyright holder for this preprint

Notes

F sta

sugg

brea Total healthcare costs per year

Jotal healthcare costs per year

val; QALY: quality-adjusted life year.

est for lung (F-statistic = 21) and ovarian (F-statistic = 25) cancers

weak instrument bias, but very high for pros Pal; QALY: quality-adjusted life year.

Lest for lung (F-statistic = 21) and ovarian (F-statistic = 25) cance

Weak instrument bias, but very high for prostate (F-statistic = 1,

L,824) cancers, see Supplementary table S4 F statistics were lowest for lung (F-statistic = 21
suggesting possible weak instrument bias, but v
breast (F-statistic = 1,824) cancers, see Supplem F statistics were lowered to lung (F-statistic = 2, and ovarian (F-statistic = 2, anti-1-)
suggesting possible weak instrument bias, but very high for prostate (F-statistic = 1,72)
breast (F-statistic = 1,824) cancers, see suggest of the sast (F-statistic = 1,824) cancers, see Supplementary table S4. This suggests little evidence weak instrument bias for those exposures for which the most precise estimates were obtaine
in the main analysis. weak instrument bias for those exposures for which the most precise estimates were obtained
in the main analysis.
The multivariable adjusted analyses were generally consistent with the Mendelian

randomization analyses, with relatively little evidence from endogeneity tests for differences in the main analysis.
The multivariable adj
randomization analys
(except for thyroid ca randomization analyses, with relatively little evidence from endogeneity tests for d
(except for thyroid cancer), see Supplementary Table S6. Large cost impacts were a
multiple myeloma and thyroid cancer which differed dra randomization analyses, matematics, much stations from endogeneity sestimates were apparent fo
(except for thyroid cancer), see Supplementary Table S6. Large cost impacts were apparent fo
multiple myeloma and thyroid cance multiple myeloma and thyroid cancer which differed drastically from the multivariable adjusted
estimates for these cancers. Although no evidence of horizontal pleiotropy was observed for estimates for these cancers. Although no evidence of horizontal pleiotropy was observed for
either exposure (see Supplementary Table S3), it is possible that genetic variants for both
cancers indicate such large cost impac either exposure (see Supplementary Table S3), it is possible that genetic variants for both
cancers indicate such large cost impacts because they may be associated with a heritable fact
that is upstream of each cancer expo cancers indicate such large cost impacts because they may be associated with a heritable factor
that is upstream of each cancer exposure that itself increases costs.
20 can cancer in the such large costs.

that is upstream of each cancer exposure that itself increases costs.

20 t

4.3 Mendelian Randomization sensitivity analysis

We found little evidence of pleiotropy in the Mendelian Randomization estimates for breast
and prostate cancer, suggesting little evidence that the exclusion restriction was violated. The
p-values for SNP heterogeneity wer p-values for SNP heterogeneity were between 0.59 and 1.00 for Cochran's Q statistic, and
estimates from IVW, Egger, penalized weighted median, and weighted mode Mendelian
Randomization estimates were all broadly similar, s promates from IVW, Egger, penalized weighted median, and weighted mode Mendelian
Randomization estimates were all broadly similar, see Supplementary Table S3. There was
evidence of pleiotropy for the other exposures, altho Randomization estimates were all broadly similar, see Supplementary Table S3. There water-
Analysis for these other cancers was smaller than for prostate and breast cancer.
Analysis for these other cancers was smaller than Randomization estimates were all broadly similary significantly similarly similar environmentary
evidence of pleiotropy for the other exposures, although the number of SNPS available for
analysis for these other cancers wa

enalysis for these other cancers was smaller than for prostate and breast cancer.
We examined results stratified by sex (Supplementary Table S4). However, as the most pre
estimates were for breast and prostate cancer (each analysis for ansets and statistic materialism matrice prostate and also cancer.
We examined results stratified by sex (Supplementary Table S4). However, as the
estimates were for breast and prostate cancer (each affecting estimates were for breast and prostate cancer (each affecting predominantly one sex), there
was little evidence of differences between sexes for any other cancer. We also examined result
split by both age and diagnosis tim estimates were for a season process between sexes for any other cancer. We also examined result by both age and diagnosis time (pre- or post-baseline) see Supplementary Table S4. For breast cancer, the effect of breast can split by both age and diagnosis time (pre- or post-baseline) see Supplementary Table S4. For
breast cancer, the effect of breast cancer on QALYs per year was higher but more imprecise for
both cancers diagnosed pre-baselin split by both age and diagnosis time (pre- or post-baseline) see Supplementary Table S4. For
breast cancer, the effect of breast cancer on QALYs per year was higher but more imprecise for
both cancers diagnosed pre-baselin both cancers diagnosed pre-baseline and post-baseline, whereas total healthcare costs per year
were highest for cancers diagnosed post-baseline (Mendelian randomization estimate: £2,300
per year, 95% CI: £1,575 to £3,026). both cancers diagnosed post-baseline (Mendelian randomization estimate: £2,300
per year, 95% CI: £1,575 to £3,026).
For prostate cancer, there was little difference in the effects on healthcare costs and QALYs per

were mighted for cancers angulation post-baseline (Mendelian randomization estimates 22,300)
per year, 95% CI: £1,575 to £3,026).
year when split by either diagnosis time or age. There was little difference in the effects year when split by either diagnosis time or age. There was little difference in the effects on
healthcare costs and QALYs per year when split by age. For when split by either diagnosis time or age. There was little difference in the effects on
healthcare costs and QALYs per year when split by age.
4.4 Evaluating the impact of SGLT2 inhibition on prostate cancer risk year when splittly sumpliced and the effect of and the effect of splittle difference in the effect of splittle
4.4 Evaluating the impact of SGLT2 inhibition on prostate cancer risk
We used the Mendelian Randomization estim

4.4 Evaluating the impact of SGLT2 inhibition on prostate cancer risk

4.4 Evaluating the impact of SGLT2 inhibition on prostate cancer risk
We used the Mendelian Randomization estimates of the effect of genetic liability to prostate
cancer on healthcare costs and QALYs to model the impact Cancer on healthcare costs and QALYs to model the impact of SGLT2 inhibition. Our
hypothetical intervention assumed that every man in the analysis sample was offered daily
SGLT2 inhibitors from the start of follow-up in UK hypothetical intervention assumed that every man in the analysis sample was offere
SGLT2 inhibitors from the start of follow-up in UK Biobank. Since there are no popul
prophylactic pharmacotherapeutic agents which are know hypothetical intervention assumed that every manufactural intervention analysis samples was encounted that SGLT2 inhibitors from the start of follow-up in UK Biobank. Since there are no population-w
prophylactic pharmacoth SCLT2 inhibitors from the start of following in the start of the start of population inter-
prophylactic pharmacotherapeutic agents which are known to reduce prostate cancer
incidence, we assessed these effects against a d prophylactic pharmacotic pharmacotic pharmacotic pharmacotic pharmacotic pharmacotic pharmacotic pharmacotic pharmacotic prostate cancer of 0.29 (95% CI=0.13 to 0.65) and 0.51 (95% CI: 0.33 to 0.7
21 incidence of prostate cancer of 0.29 (95% CI=0.13 to 0.65) and 0.51 (95% CI: 0.33 to 0.79)
21 incidence of 0.33 (95% CI=0.13 to 0.51 to 0.51 (95% CI=0.13) and 0.51 (95% CI=0.133 to 0.79) and 0.51 (95% CI=
13 to 0.79) and 0.51 (95% CI=0.133 to 0.79) and 0.791 (95% CI=0.79) and 0.791 (95% CI=0.79) and 0.793 to 0.79

drawn from Zheng et al (52), we then applied our new conventional multivariable and
Mendelian Randomization estimates, accounting for associated uncertainty, of the effect of
genetic liability to prostate cancer on healthc genetic liability to prostate cancer on healthcare costs and QALYs to model cost-effectiveness of

genetic liability of the population wide SGLT2 inhibition in this cohort of men against a "do-nothing" comparator.
Results for the analyses using an OR of 0.29 (95% CI=0.13 to 0.65) for the intervention for total
healthcar population with a select minimization with servers of men against a "do-nothing" comparator.
Results for the analyses using an OR of 0.29 (95% CI=0.13 to 0.65) for the intervention for to
healthcare costs and QALYs for dif Result for the analysis for different age groups are shown in Table 4. As a sensitivity
analysis, we also repeated this analysis assuming the more conservative OR (0.51, 95% CI: 0.33
to 0.79 .of the replication analysis in analysis, we also repeated this analysis assuming the more conservative OR (0.51, 95% CI:
to 0.79 .of the replication analysis in Zheng et al (52) and report our findings in Supplement
Table S5. analysis in Zheng et al (52) and report our findings in Supplementary
Table S5.
Table 4: Results from the SGLT2 intervention analysis compared to a "do nothing" comparator Table S5.
Table 4: Results from the SGLT2 intervention analysis compared to a "do nothing" comparator
Age group | Outcome | Analysis Method | Effect estimate (95% CI)

Under the assumptions of this analysis, the provolution wide intervention dominates the "dependence"
Confidence in the sense that it lowers healthcare comences population wide intervention dominates the "do nothing" status quo for men at risk of provision
Cancer in the sense that it lowers healthcare costs and increases quality adjusted life years
Only if offered at zero cost (se population with an anti-time intervention dominate the "dominate" quotes with an intervention processes cancer
in the sense that it lowers healthcare costs and increases quality adjusted life years, but
only if offered at only if offered at zero cost (see Supplementary Table S6). This does not hold once the costs of
the drugs themselves are accounted for (Table 4). Net monetary benefit becomes positive only
if the SGLT2 inhibition drugs are only in the transference of the supplementary transference of the drugs themselves are accounted for (Table 4). Net monetary benefit becomes positive only if the SGLT2 inhibition drugs are offered at very modest prices; fo if the SGLT2 inhibition drugs are offered at very modest prices; for example, under the "All"
scenario above, the drug becomes just cost-effective (in the sense of resulting in a positive net
monetary benefit when accounti scenario above, the drug becomes just cost-effective (in the sense of resulting in a positive net
monetary benefit when accounting for other costs and the associated quality of life impacts) at
an annual cost of approximat monetary benefit when accounting for other costs and the associated quality of life impacts) at
an annual cost of approximately £32.
5 Discussion monetary benefit when an annual cost of approximately £32.
 S Discussion

Discussion

We developed the first Mendelian Randomization estimates of the impact of cancer status on
healthcare cost and on quality of life, the two outcomes necessary to undertake a Healthcare cost and on quality of life, the two outcomes necessary to undertake a
comprehensive economic evaluation of any intervention. We studied the impacts of bladder,
breast, colorectal, lung, multiple myeloma, ovaria healthcare cost and on quality of line, the current constrainty to and claims a
comprehensive economic evaluation of any intervention. We studied the impacts
breast, colorectal, lung, multiple myeloma, ovarian, prostate an breast, colorectal, lung, multiple myeloma, ovarian, prostate and thyroid cancers on these
outcomes.

breast, colorectal, and the metal, and to the Mendelian Randomization estimates are less likely to be biased by confounding from
1991 - The Mendelian Randomization estimates are less likely to be biased by confounding from The Mendelian Randomization estimates are less likely to be biased by confounding from
omitted variables and reverse causation than conventional analytic study designs such as t
based on cohort studies or decision-analytic The Mendelian Randomization estimates are less likely to be biased by conforming from

based on cohort studies or decision-analytic simulation models using observational effect

estimates (4). We used the largest available based on cohort studies or decision-analytic simulation models using observational effect
estimates (4). We used the largest available GWASs to identify cancer exposures without
restriction on site. Our outcome data were d based on consideration analytic simulation models analyses or decision-analytic simulation-assements of the de
testriction on site. Our outcome data were drawn from the large UK Biobank study which
comprehensive linked dat restriction on site. Our outcome data were drawn from the large UK Biobank study which
comprehensive linked data which permitted the calculation of patient-level healthcare co
quality of life and disease status over many y comprehensive linked data which permitted the calculation of patient-level healthcare costs,
quality of life and disease status over many years.
The estimates will be particularly relevant to cost-effectiveness models eval

comprehensive reading to the estimates will be particularly relevant to cost-effectiveness models evaluating the impact
The estimates will be particularly relevant to cost-effectiveness models evaluating the impact
cancer quality of life and disease statuties then, yearst
The estimates will be particularly relevant to cost-
.
provides an example of how these types of estima The estimated will be particularly relevant to constrain the constant to constanting the impact of
cancer on healthcare costs and on quality of life. The policy evaluation element of our paper
provides an example of how th provides an example of how these types of estimates may be used to simulate the long-term
23 provides an example of these types of these types of types of estimates may be used to simulate the long-term of the lon

impacts of anti-cancer intervention in the absence of, or in advance of, a randomized controlled
trial. These methods could also be used to prioritize future randomized controlled trials by
identifying therapies most likel

trial. These methods could also be cost-effective.
Identifying therapies most likely to be cost-effective.
Adjusted multivariable estimates indicated material impacts of cancer status on healthcare
costs and quality of lif idex of the stimates indicated material is
Costs and quality of life for all 8 included cancers. Esti
Models were much more imprecise, although genetic costs and quality of life for all 8 included cancers. Estimates from the Mendelian Randomization
models were much more imprecise, although genetic liability to breast and prostate cancers
was associated with reductions in costs and quality of the transition interests and transition to breast and prostate cancers
models were much more imprecise, although genetic liability to breast and prostate cancers
was associated with reductions in QALYs was associated with reductions in QALYs and increases in healthcare cost. There was little
evidence for heterogeneity and pleiotropy in violation of the exclusion restriction between
SNPs for these cancers. Summary Mendeli evidence for heterogeneity and pleiotropy in violation of the exclusion restriction betweer
SNPs for these cancers. Summary Mendelian randomization sensitivity estimates were sim
to each other and broadly consistent with t evidence for these cancers. Summary Mendelian randomization sensitivity estimates were simil
to each other and broadly consistent with the main PRS estimates, indicating the same caus
effect was likely being identified in to each other and broadly consistent with the main PRS estimates, indicating the same causal
effect was likely being identified in each model without gross bias from violations of the
exclusion restriction. effect was likely being identified in each model without gross bias from violations of the
exclusion restriction.
The results from the Mendelian Randomization analyses on cost and QALY outcomes are local

exclusion restriction.
Exclusion restriction.
The results from the Mendelian Randomization analyses on cost and QALY outcomes are
average treatment effect estimates that represent the effect of lifelong exposure to high The results from the I
average treatment ef
germline genetic liabi average treatment effect estimates that represent the effect of lifelong exposure to higher
germline genetic liability to each respective cancer. The results of the analysis cannot therefore
be interpreted as, for example, are rage treatment effect transfer that represent the streath in the represented to higher
germline genetic liability to each respective cancer. The results of the analysis cannot there
be interpreted as, for example, the germline interpreted as, for example, the impact on costs of reducing cancer risk at any particular
timepoint in the life course. Instead, they represent the impact on cost and QALY outcomes
from changing long-term germlin be interpreted as, for enample, the impact on the treating cancer risk and QALY outcomes
timepoint in the life course. Instead, they represent the impact on cost and QALY outcomes
from changing long-term germline risk in t from changing long-term germline risk in this population for each specific cancer. They may
correspond to the magnitude of changes in these outcomes that would be observed through
some other modification of cancer risk thr from changing long-term germine risk in these outcomes that would be observed through
correspond to the magnitude of changes in these outcomes that would be observed through
some other modification of cancer risk through, some other modification of cancer risk through, for example, manipulation of environmental
influences on cancer risk (69).

influences on cancer risk (69).
The cancer risks we modelled reflect the risks of incident disease. Our analysis includes SNPs
that may influence both incidence and progression. For some cancers, especially prostate influences on cancer risk, (22) .
The cancer risks we modelled
that may influence both incide
cancer, identifying aggressive That may influence both incidence and progression. For some cancers, especially prostate
cancer, identifying aggressive cancer is arguably more important than understanding incident
since most tumours in (for example) the that may cancer, identifying aggressive cancer is arguably more important than understanding incid
since most tumours in (for example) the prostate do not influence overall survival (70).
However, tumours that do not affec cancer, identifying aggressive cancer is arguad, inter important than understanding incidents
since most tumours in (for example) the prostate do not influence overall survival (70).
However, tumours that do not affect sur since most tumours that do not affect survival may still influence cost and quality of life a
influence will be reflected in our estimates. Nevertheless, there would be value in upda
influence will be reflected in our esti influence will be reflected in our estimates. Nevertheless, there would be value in updating this
24 influence will be reflected in our estimates. Nevertheless, there were word before words there were updated in
24

anade available.
Healthcare costs were estimated from inpatient care episodes, primary care consultations a
primary care drug prescriptions obtained or imputed from electronic primary care records. made available
Healthcare costs
primary care dru
Follow-up was t primary care drug prescriptions obtained or imputed from electronic primary care records.
Follow-up was two years shorter for secondary care costs than primary care costs but was
averaged over years of follow-up and should primary care in the UP of the Follow-up was two years shorter for secondary care costs than primary care costs but was
averaged over years of follow-up and should not have affected results. We did not capture
costs, includ Follow-up was two years of follow-up and should not have affected results. We did not capture
costs, including private care delivered in private healthcare facilities or outpatient care, ne
of which were available for this costs, including private care delivered in private healthcare facilities or outpatient care, neithe
of which were available for this cohort. This suggests our results may underestimate the costs
associated with cancer. of which were available for this cohort. This suggests our results may underestimate the costs
associated with cancer.

of which were available for this conduct with suggests our results may ansatz summary income
associated with cancer.
Despite its size, UK Biobank is not representative of the UK population as self-selected
participants ten Despite its size, UK Biobank is not representative of the UK population as self-selected
participants tend to be wealthier and healthier compared to wider population from whi
was recruited (31). It is possible, although we participants tend to be wealthier and healthier compared to wider population from wh
Was recruited (31). It is possible, although we cannot demonstrate, that our Mendelian
Randomization results have underestimated the cost participants tend to be wealthier and healthier compared to third population from minimizine
was recruited (31). It is possible, although we cannot demonstrate, that our Mendelian
Randomization results have underestimated Randomization results have underestimated the costs and quality of life impacts of canon
Since wealthier and healthier people may have more capacity and material resources to
ameliorate the detrimental effects on healthcar since wealthier and healthier people may have more capacity and material resources to
ameliorate the detrimental effects on healthcare cost and quality of life of increased geneti
liability to cancer. The scale of this typ ameliorate the detrimental effects on healthcare cost and quality of life of increased genetic
liability to cancer. The scale of this type of selection and any associated bias is unknown, but
may have less significance for liability to cancer. The scale of this type of selection and any associated bias is unknown, but manny of cancer the scale of this type of selection and any associated bias image of the
instrumental variable assumptions (71, 72), for which we separately report sensitivity analyse
and for more biologically distal expos matrumental variable assumptions (71, 72), for which we separately report sensitivity and
and for more biologically distal exposures (73) than we study. We controlled for genetic
principal components in our analysis; there instrumental variable assumptions (73), for antienal variable, purper sensitivity analysis
and for more biologically distal exposures (73) than we study. We controlled for genetic
principal components in our analysis; ther principal components in our analysis; there is evidence that this may not control for all
remaining population or geographic structure (74, 75) (33), both which could give rise to
residual biases. principal components in the maining population or geographic structure (74, 75) (33), both which could give rise to
residual biases.
Although Mendelian randomization is likely to be less affected by confounding and reve

residual biases.
Although Mendelian randomization is likely to be less affected by confounding and reverse
causality than conventional multivariable adjusted analyses, our analysis of a sample of residual
Although Mend
causality than co
unrelated indivi rausality than conventional multivariable adjusted analyses, our analysis of a sample of
unrelated individuals may give rise to further biases from dynastic effects and assortative
mating. Dynastic effects refer to the inf causality than contentional multimal multiparty and post-on-their protection provided individuals may give rise to further biases from dynastic effects and assortative mating. Dynastic effects refer to the influence of par mating. Dynastic effects refer to the influence of parental genotype on offspring phenotype.
Parents (and both their transmitted and untransmitted alleles) may influence the environmer
in which their children are raised an Parents (and both their transmitted and untransmitted alleles) may influence the environment in which their children are raised and could confound the SNP-cancer associations we estimate \sim 25

later in the life of these offspring. Given that health costs are measured in middle to early old
age, these impacts are probably modest, if present at all. Assortative mating refers to non-
random mating; that is, individ avoided in Mendelian Randomization. possessing alleles that increases the risk of cancer. This will tend to create clusters of similar
alleles in particular environments and may re-introduce environmental confounding otherwise
avoided in Mendelian Randomizat possessing alleles that increases the risk of the risk of the risk of cluster clusters of similar
alleles in particular environments and may re-introduce environmental confounding otherwi
avoided in Mendelian Randomization

avoided in Mendelian Randomization.
Within-family Mendelian randomization studies can account for some of these biases.
However, we did not have the statistical power to conduct these analyses since the binary Within-family Mendelian randomization studies can account for some of these biases.
However, we did not have the statistical power to conduct these analyses since the binary
cancer exposures we studied affected relatively However, we did not have the statistical power to conduct these analyses since the bir
Cancer exposures we studied affected relatively few people in UK Biobank (76). In any
there are as yet no large cohort of family trios However, we did note that the statistical power to conduct these analyses since the statistical cancer exposures we studied affected relatively few people in UK Biobank (76). In any ever
there are as yet no large cohort of charge are as yet no large cohort of family trios that have reached middle and early old age that
would have enabled this analysis for our UK Biobank cohort, and the magnitude of potential
biases may from these sources may would have enabled this analysis for our UK Biobank cohort, and the magnitude of potential
biases may from these sources may not be large for the cancer exposures that we study. For
more visible traits such as BMI (77, 78) would have enabled this analysis for the substitution, and the magnitude of potential
biases may from these sources may not be large for the cancer exposures that we study. For
more visible traits such as BMI (77, 78) and more visible traits such as BMI (77, 78) and biologically distal phenotypes such as education
(79) there is robust evidence of for family-related effects assortative mating and dynastic
effects (80), but it is less clear i (79) there is robust evidence of for family-related effects assortative mating and dynastic
effects (80), but it is less clear if these processes have any significance for the cancer exposu
studied here. (1998), but it is less clear if these processes have any significance for the cancer expo
studied here.
We measured the quality of life of UK Biobank participants by assigning quality of life

effects (80), but it is less clear if these processes have any significance for the cancer exposure
studied here.
decrements associated with 240 specific health conditions. This permitted the calculation and We measured the quality of life of UK Biobank participants by assigning quality of life decrements associated with 240 specific health conditions. This permitted the calcula
use of QALYs as an outcome. We note that, amongst these conditions, are some (but if
the cancers that we also study as exposures (specif use of QALYs as an outcome. We note that, amongst these conditions, are some (but not all) of
the cancers that we also study as exposures (specifically bladder, breast, colon, lung, and
prostate – see Supplementary Table S the cancers that we also study as exposures (specifically bladder, breast, colon, lung, and
prostate – see Supplementary Table S2). This creates a modest degree of circularity, but we
don't consider this as problematic bec prostate – see Supplementary Table S2). This creates a modest degree of circularity, but v
don't consider this as problematic because the cancers we study contribute relatively little
variance in QALYs across the full samp prostate – see Supplementary Table S2). This creates a modest degree of chromology during
don't consider this as problematic because the cancers we study contribute relatively little
variance in QALYs across the full sampl don't consider the as problematic as parts in the cancer the cancer of cancers we study the variation is st
sufficient to identify the causal effects of exposure to these cancer in the Mendelian
Randomization analysis. Mor sufficient to identify the causal effects of exposure to these cancer in the Mendelian
Randomization analysis. Moreover, these Mendelian Randomization estimates should also in
principle account for any other health conditi sufficient to identify the causal enters of exposure to these cancer in the Mendelian.
Randomization analysis. Moreover, these Mendelian Randomization estimates should
principle account for any other health conditions down principle account for any other health conditions downstream of the specific cancer exposures we study and which would otherwise be omitted from conventional multivariable analysis.
26 we study and which would otherwise be omitted from conventional multiparties be omitted from conventional multiparties and α

We used both conventional multivariable and Mendelian Randomization estimates to assess the potential cost-effectiveness of a hypothetical population-wide prophylactic intervention for
prostate cancer based on SGLT2 inhibition. Assuming the costs of this intervention would be as
for 300mg canagliflozin, this the procedulation of SGLT2 inhibition. Assuming the costs of this intervention would be as
for 300mg canagliflozin, this intervention was unlikely to be cost-effective at this price under
the assumptions of that analysis w pror 300mg canagliflozin, this intervention was unlikely to be cost-effective at this price under
the assumptions of that analysis which included (but were not limited) to reductions in the
odds of prostate cancer based on for 300 mg canagements, the intervention was unimity, to be constraint at this price under
the assumptions of that analysis which included (but were not limited) to reductions in the
odds of prostate cancer based on the fi odds of prostate cancer based on the findings of Zheng et al (52), and the absence of side
effects.

odder the carreling of the ffects.
Even given these limitations, we consider that these analyses demonstrate at least two
important conclusions. The first is to lend support for the conduct of a randomized contro Even given these limitations, we consider that these analyses demonstrate at least two Even given these intriductions, the sentimes that these analyses demonstrations as tensor the
important conclusions. The work of Zheng et al (52) indicated that they may have subs
impacts on the odds of prostate cancer, an trial of SGLT2 inhibitors. The work of Zheng et al (52) indicated that they may have substantial
impacts on the odds of prostate cancer, and are the sole pharmacotherapy to demonstrate
such an impact as a prophylactic for the paracts on the odds of prostate cancer, and are the sole pharmacotherapy to demonstrate
such an impact as a prophylactic for incident prostate cancer. Although the present work
demonstrates that such an intervention is impacts on the only control through the one pharmacother, the solution of such an impact as a prophylactic for incident prostate cancer. Although the present work
demonstrates that such an intervention is unlikely to be co demonstrates that such an intervention is unlikely to be cost-effective, this conclusion dep
on the price at which the rights holders of SGLT2 inhibitors would offer the drug for a nov
prostate cancer indication. demonstrates that such an intervention is minitely to be concerned, the conclusion a pends
on the price at which the rights holders of SGLT2 inhibitors would offer the drug for a novel
prostate cancer indication.
The secon

prostate cancer indication.
The second conclusion is that the methods we have developed and applied here for the first
time represent a feasible and rapid way of assessing the cost-effectiveness of both novel an prostate cancer indicated.
The second conclusion is th
time represent a feasible ar
established interventions. T time represent a feasible and rapid way of assessing the cost-effectiveness of both novel and
established interventions. These methods are an alternative but complementary approach to,
for example, Markov cohort simulation established interventions. These methods are an alternative but complementary approach to,
for example, Markov cohort simulations and related study designs that seek to model the long
term impact of exposures and intervent for example, Markov cohort simulations and related study designs that seek to model the long- \overline{RCTs} provided sufficient genetic evidence for the specific target of interest (see e.g. (81)) these methods as an alternative to fully articulated RCTs but they do offer considerable
promise as a means of rapidly and efficiently prioritising targets for investigation in subsequent
RCTs, provided sufficient genetic promise as a means of rapidly and efficiently prioritising targets for investigation in subs
RCTs, provided sufficient genetic evidence for the specific target of interest (see e.g. (81)
exists to enable robust Mendelian R promise as a means of rapidly and entirely processing ingerest investigation is an entropy and
RCTs, provided sufficient genetic evidence for the specific target of interest (see e.g. (81))
exists to enable robust Mendelia Exists to enable robust Mendelian Randomization analysis.
 6 Conclusion

Conclusion

exists to enable robust Mendelian Randomization analysis.
 6 Conclusion

We estimated the causal impact of liability to eight cancers on healthcare costs and on quality-

adjusted life years using a Mendelian Randomizati adjusted life years using a Mendelian Randomization study design using data drawn from large
27

genome-wide association studies and from the UK Biobank cohort. We identified impacts from
genetic liability to prostate cancer amongst men and breast cancer amongst women on both
healthcare costs and quality-adjusted life dealthcare costs and quality-adjusted life years. These results were concordant with those
obtained from pleiotropy-robust estimators and when stratifying by age and sex. Mendelian
Randomization results from less common ca obtained from pleiotropy-robust estimators and when stratifying by age and sex. Mendelia
Randomization results from less common cancers were imprecise.
We also used both conventional multivariable and Mendelian Randomizati

obtained from pleiotropy-robust estimators and measurement, m_{0} a y age and set in measurements.
Randomization results from less common cancers were imprecise.
We also used both conventional multivariable and Mendelia We also used both conventional multivariable and Mendelian Rand
the impact of hypothetical population-wide preventative intervent
prostate cancer based on the repurposing of SGLT2 inhibitors. Both the impact of hypothetical population-wide preventative intervention to reduce the risk of
prostate cancer based on the repurposing of SGLT2 inhibitors. Both sets of estimates indicated
that this intervention was unlikely prostate cancer based on the repurposing of SGLT2 inhibitors. Both sets of estimates indica
that this intervention was unlikely to be cost-effective if drugs were priced as for their curre
anti-diabetes indication. that this intervention was unlikely to be cost-effective if drugs were priced as for their current
anti-diabetes indication.

anti-diabetes indication.
Robust evidence of the causal impact of cancer exposures on health economic outcomes may
be used as inputs into decision analytic models of cancer interventions such as screening Robust evidence of the c
be used as inputs into de
programmes or simulatic Robust evidence of the causal impact of the captures on the cause of as screening
programmes or simulations of longer-term outcomes associated with therapies investigated in
RCTs with short follow-ups. We also demonstrated be used as increases associated with therapies investigation
Beta associated with therapies investigation
RCTs with short follow-ups. We also demonstrated a method to rapidly and efficiently est
the cost-effectiveness of a programmes or simulations or the cost-effectiveness of a hypothetical population-scale anti-cancer intervention to inform and
the cost-effectiveness of a hypothetical population-scale anti-cancer intervention to inform and the cost-effectiveness of a hypothetical population-scale anti-cancer intervention to inform and
complement other means of assessing long-term intervention cost-effectiveness. These the complement other means of assessing long-term intervention cost-effectiveness. These
estimates and methods will contribute to ongoing debates about managing the economic
impacts of cancer in the face of changes in canc estimates and methods will contribute to ongoing debates about managing the economic
impacts of cancer in the face of changes in cancer incidence, cancer survivorship, populati
structure, and ongoing debates about the cost impacts of cancer in the face of changes in cancer incidence, cancer survivorship, population
structure, and ongoing debates about the costs of cancer drug development and the
remuneration of anti-cancer therapies. structure, and ongoing debates about the costs of cancer drug development and the structure, and only contained the congress of containers about the costs of co remuneration of anti-cancer therapies.

It is made available under a CC-BY 4.0 International license. **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2023.02.06.23285521;](https://doi.org/10.1101/2023.02.06.23285521) this version posted February 8, 2023. The copyright holder for this preprint

Declarations

 $\begin{array}{c} \mathbf{I} \\ \mathbf{F} \\ \mathbf{I} \end{array}$ Funding statement: P.D., S.H., and R.M.M. received support from a cancer Research OK, Onted
Kingdom (C18281/A29019) program grant (the Integrative Cancer Epidemiology Programme).
RMM is a National Institute for Health Rese MM is a National Institute for Health Research Senior Investigator (NIHR202411). R.M.M is
also supported by the NIHR Bristol Biomedical Research Centre which is funded by the NIHR
(BRC-1215-20011) and is a partnership betw also supported by the NIHR Bristol Biomedical Research Centre which is funded by the NIHR
(BRC-1215-20011) and is a partnership between University Hospitals Bristol and Weston NHS
Foundation Trust and the University of Bri (BRC-1215-20011) and is a partnership between University Hospitals Bristol and Weston NHS
Foundation Trust and the University of Bristol. P.D., S.H., and R.M.M. are affiliated with the
Medical Research Council Integrative (BRC-1215) Foundation Trust and the University of Bristol. P.D., S.H., and R.M.M. are affiliated with the
Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is
supported by the Medica Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is
supported by the Medical Research Council (MC_UU_00011/1, MC_UU_00011/3,
MC_UU_00011/6, and MC_UU_00011/4) and the University of supported by the Medical Research Council (MC_UU_00011/1, MC_UU_00011/3,
MC_UU_00011/6, and MC_UU_00011/4) and the University of Bristol. PD acknowledges
support from a Medical Research Council Skills Development Fellowshi support from a Medical Research Council Skills Development Fellowship (MR/P014
Department of Health and Social Care disclaimer: The views expressed are those of
and not necessarily those of the NHS, the NIHR or the Departm support from a Medical Research Council Skills Development Fellowship (MR/P014259/1).
Department of Health and Social Care disclaimer: The views expressed are those of the aut
and not necessarily those of the NHS, the NIHR Department of Health and Social Care disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.
Conflict of interest statement: The authors declare no conflicts of interest.
Acknowledgments: This research was conducted using t

and in the Euronia, and distributed in the NHS. The Department of the AHS on the Department of the NHS.
Conflict of interest statement: The authors declare no conflicts of interest.
Application Number 29294. We are also ve $\frac{1}{2}$ Conflict of interest statement: The authors declare no conflicts of interest.
Acknowledgments: This research was conducted using the UK Biobank Reso
Application Number 29294. We are also very grateful to Jie Zheng for advi $\frac{1}{2}$ Acknowledgments: This research was conducted using the OK Biobank Resource under
Application Number 29294. We are also very grateful to Jie Zheng for advice on the mod
of SGLT2 inhibition. Application Number 2029
Application Number 2020
Application of SGLT2 inhibition.

 $\frac{1}{\sqrt{2}}$ $\frac{1}{2}$ Global, Reg
References
References
Close to Analysis for a
S
C Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years fo
Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study.
JAMA Oncology. 2017;3(4):524-48.

Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study.
JAMA Oncology. 2017;3(4):524-48.
2. Global Burden of Disease Cancer C. Global, Regional, and National Cancer Incidence,
Mortality, JAMA Oncology. 2017;3(4):524-48.
2. Global Burden of Disease Cancer C. Global, Regional, and National Cancer Incidence,
Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years fo
Canc 2. Global Burden of Disease Ca
Mortality, Years of Life Lost, Years L
Cancer Groups, 1990 to 2016: A Sys
JAMA Oncology. 2018;4(11):1553-6
3. Siegel RL, Miller KD, Fuchs H
2021:71(1):7-33 Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years fo
Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study.
JAMA Oncology. 2018;4(11):1553-68

Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study.
JAMA Oncology. 2018;4(11):1553-68.
3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin.
2021;71(1

surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records Siegel RL, Miller KD, Fuchs HE,
2021;71(1):7-33.
4. Allemani C, Matsuda T, Di Carle
surveillance of trends in cancer surviva
for 3705130025 patients diagnosed w 2021;71(1):7-33.
4. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global
5. surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual re
5. The statistics in 71 countri 2022, 2022, 2022, 2023

2025 pregistries in 71 component

2025 pregistries in 71 component

2025 pregistries in 71 component

2020 Mow JNCL Monos surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual refor 3725132025 patients diagnosed with one of 18 cancers from 322 population-based
registries in 71 countries. The Lancet. 2018;391(10 surveillance of the cancers from 322 population-based
for 3725132025 patients diagnosed with one of 18 cancers from 322 population-based
registries in 71 countries. The Lancet. 2018;391(10125):1023-75.
5. Halpern MT, Lipsc

For 31 Strategistrich angulation (18) 391(10125):1023-75.
For the signifies in 71 countries. The Lancet. 2018;391(10125):1023-75.
For the Fundmann MT, Lipscomb J, Yabroff KR. Cancer Health Economics Research: The Fundmann registries in 12 countries in 12 countries. The Lancer Health Econor
The Lancer Health Econor
Now. JNCl Monographs. 2022;2022(59):102-6.
6. Mariotto AB, Enewold L, Zhao J, Zeruto CA, Yabroff KR. Me
with Cancer Survivorship 5. Halpert Marketter Marketter Marketter Marketter Marketter Marketter New JNCl Monographs. 2022;2022(59):102-6.
1. Mariotto AB, Enewold L, Zhao J, Zeruto CA, Yabroff KR. Medical Care Costs Associated
1. With Cancer Surviv Mariotto AB, Enewold L, Zhao J, Zeruto
With Cancer Survivorship in the United States. (2020;29(7):1304-12.
2020;29(7):1304-12.
7. American Cancer S. Cancer facts & figure Sullivan R, Peppercorn J, Sikora K, Zalcb
affordabl with Cancer Survivorship in the United States. Cancer Epidemiol Biomarkers Prev.
2020;29(7):1304-12.
7. American Cancer S. Cancer facts & figures 2014: American Cancer Society; 2014.
8. Sullivan R, Peppercorn J, Sikora K,

2020;29(7):1304-12.
7. American Cancer S. Cancer facts & figures 2014: American Cancer Society;
8. Sullivan R, Peppercorn J, Sikora K, Zalcberg J, Meropol NJ, Amir E, et al. Del
affordable cancer care in high-income countr 2020;
2020; American Can
8. Sullivan R, Pe_l
affordable cancer can
9. Falzone L, Salı
turn of the third mille
10. Miller KD, No. 8. Sullivan R, Peppercorn J, Sikora K, Zalcberg J, Meropol NJ, Amir E, et al. Delivering
affordable cancer care in high-income countries. The Lancet Oncology. 2011;12(10):933-
9. Falzone L, Salomone S, Libra M. Evolution o

affordable cancer care in high-income countries. The Lancet Oncology. 2011;12(10):933-8

8. Falzone L, Salomone S, Libra M. Evolution of cancer pharmacological treatments a

10. Miller KD, Nogueira L, Devasia T, Mariotto A 9. Falzone L, Salomone S, Libra M. Evolution of cancer pharmacological treatments at the
turn of the third millennium. Front Pharmacol. 2018:1300.
10. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et

turn of the third millennium. Front Pharmacol. 2018:1300.
10. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et al. Cancer
treatment and survivorship statistics, 2022. CA Cancer J Clin. 2022;72(5):409-10. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabro
treatment and survivorship statistics, 2022. CA Cancer J Clin
11. Ekwueme DU, Halpern MT, Chesson HW, Ashok M,
Economics Research in Primary Prevention of Cancer: As treatment and survivorship statistics, 2022. CA Cancer J Clin. 2022;72(5):409-36.
11. Ekwueme DU, Halpern MT, Chesson HW, Ashok M, Drope J, Hong Y-R, et al. Hea
Economics Research in Primary Prevention of Cancer: Assessmen treatment and survively predictively and survively and survively equipment and the Economics Research in Primary Prevention of Cancer: Assessment, Current Challe
Future Directions. JNCl Monographs. 2022;2022(59):28-41.
12. Economics Research in Primary Prevention of Cancer: Assessment, Current Challenges, and

Future Directions. JNCl Monographs. 2022;2022(59):28-41.

12. Mailankody S, Prasad V. Five years of cancer drug approvals: Innovation

Economic Microsofterm in Prevention of Cancer Challenges, and Future Directions. JNCl Monographs. 2022;2022(59):28-41.

12. Mailankody S, Prasad V. Five years of cancer drug approvals: Innovation, efficacy, and

13. Prasad Future Directions. In the graphs. 2022, 2022

12. Mailankody S, Prasad V. Five years of cancer drug ap

costs. JAMA Oncology. 2015;1(4):539-40.

13. Prasad V, De Jesús K, Mailankody S. The high price of

implications, barr

22. May 2015, 1992 2015, 1993 2015, 1993 2016.

13. Prasad V, De Jesús K, Mailankody S. The high price of anticancer drugs: origins,

implications, barriers, solutions. Nature Reviews Clinical Oncology. 2017;14(6):381-90.
 THERTMAN PROTEGY-PEED, CHIRTMAN OF
13. Prasad V, De Jesús K, Mailankody S
implications, barriers, solutions. Nature R
14. Aggarwal A, Ginsburg O, Fojo T. Ca
debate? Perspectives from the EU, Canada
15. Fojo T, Mailankody S

13. Prasad V, Denemarko, Prasado, Prasado, Prasado, 2017; 14(6):381-90.
14. Aggarwal A, Ginsburg O, Fojo T. Cancer economics, policy and politics: What inf
15. Fojo T, Mailankody S, Lo A. Unintended Consequences of Expensi 14. Aggarwal A, Ginsburg O, Fojo T. Cancer economics, policy and politics: What information debate? Perspectives from the EU, Canada and US. Journal of Cancer Policy. 2014;2(1):
15. Fojo T, Mailankody S, Lo A. Unintended C debate? Perspectives from the EU, Canada and US. Journal of Cancer Policy. 2014;2(1):1-11.
15. Fojo T, Mailankody S, Lo A. Unintended Consequences of Expensive Cancer
Therapeutics—The Pursuit of Marginal Indications and a debate. Perspective From Therapeutics—The Pursuit of Marginal Indications and a Me-Too Mentality That Stifles
Therapeutics—The Pursuit of Marginal Indications and a Me-Too Mentality That Stifles
Innovation and Creativity: Therapeutics—The Pursuit of Marginal Indications and a Me-Too Mentality That Sti
Innovation and Creativity: The John Conley Lecture. JAMA Otolaryngology–Head &
Surgery. 2014;140(12):1225-36.
30 Innovation and Creativity: The John Conley Lecture. JAMA Otolaryngology–Head & Neck
Surgery. 2014;140(12):1225-36.
30 Surgery. 2014;140(12):1225-36.
30 $\frac{30}{140}$
30

16. Wait S, Han D, Muthu V, Oliver K, Chrostowski S, Florindi F, et al. Towards sustainable
cancer care: Reducing inefficiencies, improving outcomes—A policy report from the All.Can
initiative. Journal of Cancer Policy. 20 16.

initiative. Journal of Cancer Policy. 2017; 13:47-64.
17. Nicholson BD, Thompson MJ, Hobbs FDR, Nguyen M, McLellan J, Green B, et al.
Measured weight loss as a precursor to cancer diagnosis: retrospective cohort analysis o 17. Nicholson BD, Thompson MJ, Hobbs FDR, N
Measured weight loss as a precursor to cancer diag
43 302 primary care patients. Journal of Cachexia,
18. Doll R, Peto R, Boreham J, Sutherland I. Mo
observations on male British Measured weight loss as a precursor to cancer diagnosis: retrospective cohort analysis
43 302 primary care patients. Journal of Cachexia, Sarcopenia and Muscle. 2022;n/a(n/
18. Doll R, Peto R, Boreham J, Sutherland I. Mor

43 302 primary care patients. Journal of Cachexia, Sarcopenia and Muscle. 2022;n/a(n/a).
18. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years'
observations on male British doctors. BM 43 302 primary care pentent to alternate care patients. Journal of the Nation II Supersety
18. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years&
19. Prospective Studies C. Body-mass index

observations on male British doctors. BMJ. 2004;328(7455):1519.
19. Prospective Studies C. Body-mass index and cause-specific mortality in 9002000 adults:
collaborative analyses of 57 prospective studies. The Lancet. 2009; collaborative analyses of 57 prospective studies. The Lancet. 2009;373(9669):1083-96.
20. Fang Z, Song M, Lee DH, Giovannucci EL. The Role of Mendelian Randomization Studies
in Deciphering the Effect of Obesity on Cancer. 20. Prospective Studies. The Lancet. 2009;373(9669):1083-96.

20. Fang Z, Song M, Lee DH, Giovannucci EL. The Role of Mendelian Randomization Studies

20. Fang Z, Song M, Lee DH, Giovannucci EL. The Role of Mendelian Rando collaborative analysis of prospective studies. The Role of Mendelian Randomization
in Deciphering the Effect of Obesity on Cancer. JNCI: Journal of the National Cancer Inst
2022;114(3):361-71.
21. Zhou W, Liu G, Hung RJ, H

20. Fangle of the Effect of Obesity on Cancer. JNCI: Journal of the National Cancer Institute.
2022;114(3):361-71.
21. Zhou W, Liu G, Hung RJ, Haycock PC, Aldrich MC, Andrew AS, et al. Causal relationships
between body mas in Decipheria, 1991.
2022; 114(3):361-71.
21. Zhou W, Liu G, Hung RJ, Haycock PC, Aldrich MC, Andrew AS, et al. Causal relationship:
between body mass index, smoking and lung cancer: Univariable and multivariable Mendeliar 21. Zhou W, Liu G
between body mass i
randomization. Int J (
22. Dixon P, Hollii
Randomization analy
2020: 70: 102300

between body mass index, smoking and lung cancer: Univariable and multivariable Mendelian
randomization. Int J Cancer. 2021;148(5):1077-86.
22. Dixon P, Hollingworth W, Harrison S, Davies NM, Davey Smith G. Mendelian
R between body mass index, smooth and multiparty material in the multiparty of the candomization. Int J Cancer. 2021; 148(5): 1077-86.
22. Dixon P, Hollingworth W, Harrison S, Davies NM, Davey Smith G. Mendelian
Randomizatio 22. Dixon P, Hollingworth W, Harrison S, Davies
Randomization analysis of the causal effect of adip
2020;70:102300.
23. Hazewinkel A-D, Richmond RC, Wade KH, D
the causal impact of body mass index and waist-hip
Hum Biol.

Randomization analysis of the causal effect of adiposity on hospital costs. J Health E
2020;70:102300.
23. Hazewinkel A-D, Richmond RC, Wade KH, Dixon P. Mendelian randomization
the causal impact of body mass index and wai 2020;70:102300.
23. Hazewinkel A-D, Richmond RC, Wade KH, Dixon P. Mendelian randomization anal
the causal impact of body mass index and waist-hip ratio on rates of hospital admission.
19. Hum Biol. 2022;44:101088.
24. Dix 23. Hazewinke
the causal impact
Hum Biol. 2022;4
24. Dixon P, S.
on healthcare cos
25. Harrison S the causal impact of body mass index and waist-hip ratio on rates of hospital admission. Econ
Hum Biol. 2022;44:101088.
24. Dixon P, Sallis H, Munafo M, Smith GD, Howe L. The causal effect of cigarette smoking
on healthcar

the causal impacts of the causal effect of cigarette smoking

19 Hum Biol. 2022;44:101088.

24. Dixon P, Sallis H, Munafo M, Smith GD, Howe L. The causal effect of cigarette smoking

25. Harrison S, Davies AR, Dickson M, T 24. Dixon P, Sallis H, Mu
24. Dixon P, Sallis H, Mu
25. Harrison S, Davies Aleffects of health conditions
randomization in UK Bioban
26. Laudicella M, Walsh on healthcare costs. med Rxiv. 2022:2022.07.05.22277228.
25. Harrison S, Davies AR, Dickson M, Tyrrell J, Green MJ, Katikireddi SV, et al. The causal
effects of health conditions and risk factors on social and socioeconomi Marrison S, Davies AR, Dickson M, Tyrrell J, Green M
effects of health conditions and risk factors on social and sc
randomization in UK Biobank. Int J Epidemiol. 2020.
26. Laudicella M, Walsh B, Burns E, Smith PC. Cost of 25. Harrison States Arrison Social and Socioeconomic outcomes: Mendelia
25. Harrison in UK Biobank. Int J Epidemiol. 2020.
26. Laudicella M, Walsh B, Burns E, Smith PC. Cost of care for cancer patients in England:
27. Worl

randomization in UK Biobank. Int J Epidemiol. 2020.
26. Laudicella M, Walsh B, Burns E, Smith PC. Cost of care for cancer patients in England:
evidence from population-based patient-level data. Br J Cancer. 2016;114(randomization in UK Biobank. Int J Epidemiol. 2020.

26. Laudicella M, Walsh B, Burns E, Smith PC. Cost of care for cancer patients in England:

evidence from population-based patient-level data. Br J Cancer. 2016;114(11): 27. World Health O. Technical report: pricing of cancer. 2016;114(11):1286-92.
27. World Health O. Technical report: pricing of cancer medicines and its impacts: a
comprehensive technical report for the World Health Assemb 27. World Health O. Technical report: pricing of cancer medicines and its impacts: a
comprehensive technical report for the World Health Assembly Resolution 70.12: opera
paragraph 2.9 on pricing approaches and their impact comprehensive technical report for the World Health Assembly Resolution 70.12: opera
paragraph 2.9 on pricing approaches and their impacts on availability and affordability c
medicines for the prevention and treatment of c

paragraph 2.9 on pricing approaches and their impacts on availability and affordability of
medicines for the prevention and treatment of cancer. 2018.
28. Howard DH, Bach PB, Berndt ER, Conti RM. Pricing in the Market for

29. Workman P, Draetta GF, Schellens JHM, Bernards R. How Much Longer Will We Put Up
With \$100,000 Cancer Drugs? Cell. 2017;168(4):579-83.
30. Leighl NB. Nirmalakumar S. Ezeife DA. Gvawali B. An Arm and a Leg:

Leighl NB, Nirmalakumar S, Ezeife DA, Gyawali B. An Arm and a Leg: The Rising Cost of Econ Perspect. 2015;29(1):139-62.
29. Morkman P, Draetta GF, Schellens JHM, Bernards R. How Much Longer Will We Put Up
With \$100,000 Cancer Drugs? Cell. 2017;168(4):579-83.
30. Leighl NB, Nirmalakumar S, Ezeife DA, Gyawali 29. Vorkman P, Draetta GF, Sch
With \$100,000 Cancer Drugs? Cell.
30. Leighl NB, Nirmalakumar S,
Cancer Drugs and Impact on Access
2021(41):e1-e12.
31 Barnes TA Amir F Templet With \$100,000 Cancer Drugs? Cell. 2017;168(4):579-83.
30. Leighl NB, Nirmalakumar S, Ezeife DA, Gyawali B. An Arm and a Leg: The Rising Cost of
Cancer Drugs and Impact on Access. American Society of Clinical Oncology Educa Mini (110)

20. Leighl NB, Nirmalakumar S, Ezeife DA, Gyawali B.

Cancer Drugs and Impact on Access. American Society of

2021(41):e1-e12.

31. Barnes TA, Amir E, Templeton AJ, Gomez-Garcia

safety, tolerability and price

30. Leight Na, Nirmalakumar S, 2001, 2011 1973, 2014 1983. Cancer Drugs and Impact on Access. American Society of Clinical Oncology Educational Book.
2021(41):e1-e12.
31. Barnes TA, Amir E, Templeton AJ, Gomez-Garcia S, Na 2021(41):e1-e12.
31. Barnes TA, Amir E, Templeton AJ, Gomez-Garcia S, Navarro B, Seruga B, et al. Efficacy,
safety, tolerability and price of newly approved drugs in solid tumors. Cancer Treat Rev.
2017;56:1-7. 2022
31. Barnes TA,
safety, tolerability
2017;56:1-7.
31 31. Barnes Ta, Amir E, Templeton AJ, Gomez-Garcia S, Navarro B, Seringe B, Et al. 2017;
Safety, tolerability and price of newly approved drugs in solid tumors. Cancer Treat Rev.
2017;56:1-7.
31 safety, tolerability and price of newly approved drugs in solid tumors. Cancer Treat Rev. 2017;56:1-7.

32. Mitchell AP, Tabatabai SM, Dey P, Ohn JA, Curry MA, Bach PB. Association Between
Clinical Value and Financial Cost of Cancer Treatments: A Cross-Sectional Analysis. Journal of
the National Comprehensive Cancer Network 32.

the National Comprehensive Cancer Network J Natl Compr Canc Netw. 2020;18(10):1349-53.
33. Fu M, Naci H, Booth CM, Gyawali B, Cosgrove A, Toh S, et al. Real-world Use of and
Spending on New Oral Targeted Cancer Drugs in th Fu M, Naci H, Booth CM, Gyawali B, Cosgrove A, Toh S, et al. Real-world Use of and
Spending on New Oral Targeted Cancer Drugs in the US, 2011-2018. JAMA Internal Medicine.
2021;181(12):1596-604.
34. Salas-Vega S, Shearer E

Spending on New Oral Targeted Cancer Drugs in the US, 2011-2018. JAMA Internal Medicir
2021; 181(12): 1596-604.
34. Salas-Vega S, Shearer E, Mossialos E. Relationship between costs and clinical benefi
new cancer medicines Spending 112):1596-604.

Salas-Vega S, Shearer E, Mossialos E. Relationship between costs and clinical benefits onew cancer medicines in Australia, France, the UK, and the US. Soc Sci Med. 2020;258:113042

Siddiqui M, Rajk 2021;181(12):1596-604.

new cancer medicines in Australia, France, the UK, and the US. Soc Sci Med. 2020;258:113042.
35. Siddiqui M, Rajkumar SV. The High Cost of Cancer Drugs and What We Can Do About It.
Mayo Clin Proc. 2012;87(10):935-43.
36. G S. Siddiqui M, Rajkumar SV. The High Cost of Cancer Drugs and What We Can Do About It.
Mayo Clin Proc. 2012;87(10):935-43.
36. Goss PE, Lee BL, Badovinac-Crnjevic T, Strasser-Weippl K, Chavarri-Guerra Y, Louis JS, et
al. P Goss PE, Lee BL, Badovinac-Crnjevic T, Strasser-Weippl K, Chavarri-Guerra Y, Louis JS, et Mayo Chin Processory, 2012-2013

36. Goss PE, Lee BL, Badovinac-Cr

al. Planning cancer control in Latin Ar

2013;14(5):391-436.

37. Kelly RJ, Smith TJ. Delivering n

stakeholders in cancer care. The lance

38. Lenei K. P

37. Kelly RJ, Smith TJ. Delivering maximum clinical benefit at an affordable price: engaging
stakeholders in cancer care. The lancet oncology. 2014;15(3):e112-e8.
38. Jenei K, Prasad V, Lythgoe MP. High US drug prices have

38. Jenei K, Prasad V, Lythgoe MP. High US drug prices have global implications. BMJ.
2022;376:0693.
39. Richmond RC. Smith GD. Mendelian randomization: Concepts and scope. Cold Spr 37. Kelly RJ, Smith
stakeholders in cance
38. Jenei K, Prasa
2022;376:0693.
39. Richmond RC,
Harb Persnect Med. . state the lancet oncology. 2021;2022=2013
38. Lenei K, Prasad V, Lythgoe MP. High US drug prices have global
2022;376:0693.
39. Richmond RC, Smith GD. Mendelian randomization: Concepts a
40. Davey Smith G, Ebrahim S. 'Mend

39. Richmond RC, Smith GD. Mendelian randomization: Concepts and scope. Cold Spring
Harb Perspect Med. 2021:a040501.
40. Davev Smith G. Ebrahim S. 'Mendelian randomization': can genetic epidemiology

2022;376:0693.
39. Richmond RC, Smith GD. Mendelian randomization: Concepts and scope. Cold Spr
Harb Perspect Med. 2021:a040501.
40. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology
contribute Harb Perspect Med. 2021:a040501.
40. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology
contribute to understanding environmental determinants of disease? Int J Epidemiol.
2003;32(1):1-22.
41. Pi Harb Perspect Med. 2021:a040501.
40. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology
contribute to understanding environmental determinants of disease? Int J Epidemiol.
2003;32(1):1-22.
41. Pi contribute to understanding environmental determinants of disease? Int J Epidemiol.

40. Contribute to understanding environmental determinants of disease? Int J Epidemiol.

2003;32(1):1-22.

41. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetics.

2018;19(9):566-8 contribute to understanding entited the second to the second of the spherids.
2003;32(1):1-22.
41. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using
data to strengthen causal inference in data to strengthen causal inference in observational research. Nature Reviews Genetics.
2018;19(9):566-80.
42. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a
guide, glossary, and checklist

2018;19(9):566-80.
42. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studi
guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
43. Davey Smith G. Hemani G. Mendelian randomization: gen 2020, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 20
2018;19(9):566-80.000, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 20
2

42. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a
guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
43. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors f

guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
43. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal
inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89 guide, grown, and checklist condelian randomization: gen
inference in epidemiological studies. Hum Mol Genet. 2014;23(R
44. Dixon P, Harrison S, Hollingworth W, Davies NM, Smith Gl
of liability to disease on healthcare cos 144. Dixon P, Harrison S, Hollingworth W, Davies NM, Smith GD. Estimating the causa
44. Dixon P, Harrison S, Hollingworth W, Davies NM, Smith GD. Estimating the causa
6 liability to disease on healthcare costs using Mendel 44. Dixon P, Harrison S, Hollingworth W, Davies NM, Smith GD. Estimation
of liability to disease on healthcare costs using Mendelian Randomization. E
2022:101154. Qingyuan Z, Jingshu W, Gibran H, Jack B, Dylan SS. Statisti

of liability to disease on healthcare costs using Mendelian Randomization. Econ Hum Biol.
2022:101154.
45. Qingyuan Z, Jingshu W, Gibran H, Jack B, Dylan SS. Statistical inference in two-sample
summary-data Mendelian r of liability to disease on healthcare costs using the lines into the the lines 2022:101154.

2022:101154.

2022:101154.

2010;43 Mendelian randomization using robust adjusted profile score. The Annals of

2010;48 (3):1742-45. Qingyu
summary-data
Statistics. 2020
46. Burges
in Causal Estim
47 Slob F4 summary-data Mendelian randomization using robust adjusted profile score. The Annals of
Statistics. 2020;48(3):1742-69.
46. Burgess S, Thompson S. Mendelian Randomization: Methods for Using Genetic Varian
in Causal Estimat

statistics. 2020;48(3):1742-69.
Statistics. 2020;48(3):1742-69.
46. Burgess S, Thompson S. Mendelian Randomization: Methods for Using Genetic Varia
in Causal Estimation. Boca Raton, Florida: CRC Press; 2015.
47. Slob EAW, Statistics. 2020;48(3):1742-69.
46. Burgess S, Thompson S. Mendelian Randomization: Methods for Using Genetic Variants
in Causal Estimation. Boca Raton, Florida: CRC Press; 2015.
47. Slob EAW, Burgess S. A comparison of ro

in Causal Estimation. Boca Raton, Florida: CRC Press; 2015.
47. Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using
summary data. Genet Epidemiol. 2020;44(4):313-29.
48. Bowden J, Davey Smith in Caustin Chinamic Pression, Franch Press; 2017
47. Slob EAW, Burgess S. A comparison of robust Mende
summary data. Genet Epidemiol. 2020; 44(4):313-29.
48. Bowden J, Davey Smith G, Burgess S. Mendelian ran
effect estimat 47. Summary data. Genet Epidemiol. 2020;44(4):313-29.
48. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instrument:
49. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization Summary Commarks, Bowden J, Davey Smith G, Burgess S. Mendeliant effect estimation and bias detection through Egger re
25.
49. Burgess S, Thompson SG. Interpreting findings
MR-Egger method. Eur J Epidemiol. 2017;32(5):377-48. Bowden J, David President III and Diamonds and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-
25. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the
MR-Egger

25.
49. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the
MR-Egger method. Eur J Epidemiol. 2017;32(5):377-89.
32 ---
49.
MR
32 49. Burgess S, Thompson Statistics, Thompson Statistics, Thompson Statistics, The MR-Egger method. Eur J Epidemiol. 2017;32(5):377-89.
32 $\frac{1}{32}$

50.

50. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet
Epidemiol. 2016;40(4):304-14.
51. Hartwig FP, Dave Epidemiol. 2016;40(4):304-14.
51. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelia
randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985
52. Zhen Epidemic Hartwig FP, Davey Smith
Fandomization via the zero mo
52. Zheng J LJ, Cui B, Qi J, Y;
Ling CS, Yeung S, Dimopoulou (
Yuan J, Lewis S, Gaunt T, Davey
inhibition on prostate cancer ri randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-9
52. Zheng J LJ, Cui B, Qi J, Yang Q, Xu M, Xu Y, Wang T, Li M, Zhao Z, Zheng R, Wang S, Lin
Ling CS, Yeung S, Dimopoulou O, Dixon P randomization via the zero modal pleiotropy with the zero model pleiotropy as Zheng R, Wang S, Lin H,
Ling CS, Yeung S, Dimopoulou O, Dixon P, Harrison S, Liu Y, Robinson J, Yarmolinksy J, Haycock
Yuan J, Lewis S, Gaunt T, Fing CS, Yeung S, Dimopoulou O, Dixon P, Harrison S, Liu Y, Robinson J, Yarmolinksy J, Haycock F
Yuan J, Lewis S, Gaunt T, Davey Smith G, Martin RM. Effect of sodium-glucose cotransporter 2
inhibition on prostate cancer ri Yuan J, Lewis S, Gaunt T, Davey Smith G, Martin RM. Effect of sodium-glucose cotransporter 2
inhibition on prostate cancer risk: evidence from Mendelian randomization and electronic
health records. Submitted. 2022.
53. Pin

Yuan's) There y State Cypracty Chamber of Mendelian randomization and electronic

health records. Submitted. 2022.

S3. Pinto LC, Rados DV, Remonti LR, Kramer CK, Leitao CB, Gross JL. Efficacy of SGLT2

inhibitors in glyce inhibition on prostate cancer risk: 1022.
health records. Submitted. 2022.
53. Pinto LC, Rados DV, Remonti LR, Kramer CK, Leitao CB, Gross JL. Efficacy of SGLT2
inhibitors in glycemic control, weight loss and blood pressur mann records. Pinto LC, Rados DV, Remo

inhibitors in glycemic control, wei

and meta-analysis. Diabetol Meta

54. Lupsa BC, Inzucchi SE. Use

benefits. Diabetologia. 2018;61(1 53. Philibitors in glycemic control, weight loss and blood pressure reduction: a systematic reviewed and meta-analysis. Diabetol Metab Syndr. 2015;7(1):A58.
54. Lupsa BC, Inzucchi SE. Use of SGLT2 inhibitors in type 2 diab

and meta-analysis. Diabetol Metab Syndr. 2015;7(1):A58.
54. Lupsa BC, Inzucchi SE. Use of SGLT2 inhibitors in type 2 diabetes: weighing the risks and
benefits. Diabetologia. 2018;61(10):2118-25.
55. Burgess S, Davies NM, T 54. Lupsa BC, Inzucchi SE. Use of SGLT2 inhibitors in ty
benefits. Diabetologia. 2018;61(10):2118-25.
55. Burgess S, Davies NM, Thompson SG. Bias due to p
Mendelian randomization. Genet Epidemiol. 2016;40(7):5
56. Figueroa

benefits. Diabetologia. 2018;61(10):2118-25.
55. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample
Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608.
56. Figueroa JD, Ye Y, Siddiq benefits and protocologia. Diabeton SS.

Bendelian randomization. Genet Epidemiol. 2

S6. Figueroa JD, Ye Y, Siddiq A, Garcia-Clos

Genome-wide association study identifies mu

Hum Mol Genet. 2014;23(5):1387-98.

S7 Michai Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608.
56. Figueroa JD, Ye Y, Siddiq A, Garcia-Closas M, Chatterjee N, Prokunina-Olsson L, et al.
Genome-wide association study identifies multiple loci associated wit 56. Figueroa JD, Ye Y, Siddiq A, Garcia-Closas M, Chatterjee N,
Genome-wide association study identifies multiple loci associated
Hum Mol Genet. 2014;23(5):1387-98.
57. Michailidou K, Lindström S, Dennis J, Beesley J, Hui Genome-wide association study identifies multiple loci associated with bladder cancer risk.
Hum Mol Genet. 2014;23(5):1387-98.
57. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis

Hum Mol Genet. 2014;23(5):1387-98.
57. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysi
identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.
58. Schum Hum Mol Genet. 2014;23(5):1387-98.

57. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis

identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.

58. Schumacher FR identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.
58. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-wide
association study of colorectal cancer identifies six new suscept identifies 1918 in the state of the state of the state of 58.

Sassociation study of colorectal cancer identifies six new susceptibility lc

2015;6:7138.

Wang Y, Wei Y, Gaborieau V, Shi J, Han Y, Timofeeva MN, et al. I

f

association study of colorectal cancer identifies six new susceptibility loci. Nat Commun.
2015;6:7138.
59. Wang Y, Wei Y, Gaborieau V, Shi J, Han Y, Timofeeva MN, et al. Deciphering associat
for lung cancer risk through i association study of colorestal cancer in the colorest care precisely constant colores 2015;6:7138.

S9. Wang Y, Wei Y, Gaborieau V, Shi J, Han Y, Timofeeva MN, et al. Deciphering assoc

for lung cancer risk through imputa 2020, Wang
59. Wang
for lung cance
Hum Genet. 2
60. Mitche
association st
2016:7:12050

for lung cancer risk through imputation and analysis of 122316 cases and 162831 controls. Eur J
Hum Genet. 2015;23(12):1723-8.
60. Mitchell JS, Li N, Weinhold N, Försti A, Ali M, van Duin M, et al. Genome-wide
association For lung cancer 2015;23(12):1723-8.
Hum Genet. 2015;23(12):1723-8.
60. Mitchell JS, Li N, Weinhold N, Försti A, Ali M, van Duin M, et al. Genome-wide
association study identifies multiple susceptibility loci for multiple m Mitchell JS, Li N, Weinhold
60. Mitchell JS, Li N, Weinhold
association study identifies multip
2016;7:12050.
61. Phelan CM, Kuchenbaecke
Identification of 12 new susceptib
Nat Genet 2017:49(5):680-91

association study identifies multiple susceptibility loci for multiple myeloma. Nat Com
2016;7:12050.
61. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et a
Identification of 12 new susceptibility 2016;7:12050.
61. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al.
Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cance
Nat Genet. 2017;49(5):680-91 2020;
61. Phelan
Identification o
Nat Genet. 201
Association ana

Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian c
Nat Genet. 2017;49(5):680-91.
62. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et a
Association anal Nat Genet. 2017;49(5):680-91.
62. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al.
Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility
loci. Nat Genet Nat Genet. 2017;49(5):680-91.

62. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al.

Association analyses of more than 140,000 men identify 63 new prostate cancer suscepti

loci. Nat Genet. 2

Association analyses of more than 140,000 men identify 63 new prostate cancer suscepti
loci. Nat Genet. 2018;50(7):928-36.
63. Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjo
SA, et al. A 63. Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjonsson
SA, et al. A genome-wide association study yields five novel thyroid cancer risk loci. Nat
Commun. 2017;8:14517.
64. Harrison S, Di

SA, et al. A genome-wide association study yields five novel thyroid cancer risk loci. Nat Commun. 2017;8:14517.
64. Harrison S, Dixon P, Jones HE, Davies AR, Howe LD, Davies NM. Long-term cost-
effectiveness of interventi SA, et al. A genome-wide association study yields five novel thyroid cancer risk loci. Nat
Commun. 2017;8:14517.
64. Harrison S, Dixon P, Jones HE, Davies AR, Howe LD, Davies NM. Long-term cost-
effectiveness of interventi Commun. 2017;8:14517.
64. Harrison S, Dixon P, Jones HE, Davies AR, Howe LD, Davies NM. Long-term cost-
effectiveness of interventions for obesity: A mendelian randomisation study. PLoS Med.
2021;18(8):e1003725. Commun. 2017;8:14517. effectiveness of interventions for obesity; A mendelian randomisation study. PLOS Mendelian randomisation study
2021;18(8):e1003725.
33 233

It is made available under a CC-BY 4.0 International license. **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2023.02.06.23285521;](https://doi.org/10.1101/2023.02.06.23285521) this version posted February 8, 2023. The copyright holder for this preprint

65.

United Kingdom. Med Decis Making. 2011;31(6):800-4.
66. Dixon P, Davey Smith G, Hollingworth W. The Association Between Adiposity and
Inpatient Hospital Costs in the UK Biobank Cohort. Appl Health Econ Health Policy. 2018. 66. Dixon P, Davey Smith G, Hollingworth W. The As
Inpatient Hospital Costs in the UK Biobank Cohort. Appl
67. NHS. Durg tariff Various years, [Available from: <u>P</u>
gp-practices-and-appliance-contractors/drug-tariff.
68. Jo

Inpatient Hospital Costs in the UK Biobank Cohort. Appl Health Econ Health Policy. 2018.
67. NHS. Durg tariff Various years, [Available from: https://www.nhsbsa.nhs.uk/pharing-
gp-practices-and-appliance-contractors/drug-t Inpatient Materian Costs, [Available from: https://www.nhsbsa.nhs.uk/pharm

Inpaticulary epocasion Appliance-contractors/drug-tariff.

Indees K, Burns A. Unit Costs of Health and Social Care 2021. Kent: Personal Social

Se

GP PERTURE TREADED TO BE PERTURE 168.

68. Jones K, Burns A. Unit Costs of Health and Sc

Services Research Unit,; 2022.

69. Imbens GW, Rubin DB. Causal inference in st

Cambridge University Press; 2015.

70. Bhanji Y, Is Services Research Unit,; 2022.
69. Jmbens GW, Rubin DB. Causal inference in statistics, social, and biomedical scienc
Cambridge University Press; 2015.
70. Bhanji Y, Isaacs WB, Xu J, Cooney KA. Prostate cancer predispositi Services Research Unit,; 2022.

69. Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciene

Cambridge University Press; 2015.

70. Bhanji Y, Isaacs WB, Xu J, Cooney KA. Prostate cancer predispos

 $\frac{5}{10}$. Rhanii Yaracs WB Xull Cooney KA Prostate cancer predisposition Urologic Clinics

Cambridge University Press; 2015.

70. Bhanji Y, Isaacs WB, Xu J, Cooney KA. Prostate cancer predisposition. Urologic Clinics.

2021;48(3):283-96.

71. Hughes RA, Davies NM, Davey Smith G, Tilling K. Selection Bias When Es Famalinge Emilton, 1999, 2021; 2021; 48(3): 283-96.

2021; 48(3): 283-96.

71. Hughes RA, Davies NM, Dav

Average Treatment Effects Using O

2019; 30(3).

72. Gkatzionis A Burgess S. Cor 2021;48(3):283-96.

71. Hughes RA, Davies NM, Davey Smith G, Tilling K. Selection Bias When Estimating

Average Treatment Effects Using One-sample Instrumental Variable Analysis. Epidemiology.

2019;30(3).

72. Gkatzionis 2022, 2022, 2022, 2022, 2023, 2023, 2023, 2023, 2034, Average Treatment Effects Using One-sample Instrumental Variable Analysis. Epidemiolo
2019;30(3).
72. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomizatio
bad is it likely to be? Int J Epidemio

2019;30(3).

Average Treatment Effects Using Selection bias in Mendelian randomization: h

bad is it likely to be? Int J Epidemiol. 2018;48(3):691-701.

T3. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault J-B, Kutalik Z. 2020, 2019, 2019

2019 Schot

2019 Schot

2019 Schot

2020 Schot

2020 Haw bad is it likely to be? Int J Epidemiol. 2018;48(3):691-701.
73. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault J-B, Kutalik Z. Correction for
participation bias in the UK Biobank reveals non-negligible impact on geneti 73. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault J-
participation bias in the UK Biobank reveals non-negligible
downstream analyses. bioRxiv. 2022:2022.09.28.509845.
74. Haworth S, Mitchell R, Corbin L, Wade KH, Duddi

participation bias in the UK Biobank reveals non-negligible impact on genetic associa
downstream analyses. bioRxiv. 2022:2022.09.28.509845.
74. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. A
 participation analyses. bio Rxiv. 2022:2022.09.28.509845.

14. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Apparent

latent structure within the UK Biobank sample has implications for epidemi 74. Haworth S, Mitchell R, Corbin L, Wade KH, Duddin
latent structure within the UK Biobank sample has implica
Nature Communications. 2019;10(1):333.
75. Lawson DJ, Davies NM, Haworth S, Ashraf B, Howe
structure in the gen

Fatent structure within the UK Biobank sample has implications for epidemiological analysis.

Nature Communications. 2019;10(1):333.

75. Lawson DJ, Davies NM, Haworth S, Ashraf B, Howe L, Crawford A, et al. Is population
 lature Communications. 2019;10(1):333.

175. Lawson DJ, Davies NM, Haworth S, Ashraf B, Howe L, Crawford A, et al. Is population

structure in the genetic biobank era irrelevant, a challenge, or an opportunity? Hum Genet.
 75. Lawson DJ, Davies NM, Haworth S,
structure in the genetic biobank era irreley
2020;139(1):23-41.
76. Davies NM, Howe LJ, Brumpton B,
Mendelian randomization studies. Hum M
77. Aisley TA Angquist L Silventoinen

structure in the genetic biobank era irrelevant, a challenge, or an opportunity? Hum Genet.
2020;139(1):23-41.
76. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family
Mendelian randomizati

marriages by body mass index have increased simultaneously with the obesity epidemic. Front
Genet. 2012;3:125. Mendelian randomization studies. Hum Mol Genet. 2019;28(R2):R170-R9.

77. Ajslev TA, Angquist L, Silventoinen K, Gamborg M, Allison DB, Baker

marriages by body mass index have increased simultaneously with the obe

Genet. Mendelian randomization states from the second of the products and the marriages by body mass index have increased simultaneously with the obe
Genet. 2012;3:125.
T8. Howe LD, Kanayalal R, Harrison S, Beaumont RN, Davies AR

Howe LD, Kanayalal R, Harrison S, Beaumont RN, Davies AR, Frayling TM, et al. Effects of marriages by body mass index have increased simultaneously with the obesity epidemic. Front
Genet. 2012;3:125.
78. Howe LD, Kanayalal R, Harrison S, Beaumont RN, Davies AR, Frayling TM, et al. Effects of
body mass index on Genet. 2012;3:125.
78. Howe LD, Kanayalal R, Harrison S, Beaumont RN, Davies AR, Frayling TM, et al. Effects of
body mass index on relationship status, social contact and socio-economic position: Mendelian
randomization an France Encyclosum
T8. Howe LD, Ka
body mass index on
randomization and
T9. Kong A, Thoi
nature of nurture: E
R0. Howe LL Niv body mass index on relationship status, social contact and socio-economic position: Mendelian
randomization and within-sibling study in UK Biobank. Int J Epidemiol. 2019.
79. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson

randomization and within-sibling study in UK Biobank. Int J Epidemiol. 2019.
79. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young Al, Thorgeirsson TE, et al. The
nature of nurture: Effects of parental genotypes. S randomization and within-sibility of the Vilhialmsson BJ, Young Al, Thorgein
That with a September of nurture: Effects of parental genotypes. Science. 2018;359(6374):42
80. Howe LJ, Nivard MG, Morris TT, Hansen AF, Rasheed mature of nurture: Effects of parental genotypes. Science. 2018;359(6374):424.

20. Howe LJ, Nivard MG, Morris TT, Hansen AF, Rasheed H, Cho Y, et al. Within-sibship

20. Howe LJ, Nivard MG, Morris TT, Hansen AF, Rasheed H 1900. Howe LJ, Nivard MG, Morris TT, Hansen AF, Rasheed H, Cho Y, et al. With
genome-wide association analyses decrease bias in estimates of direct genetic of
Genet. 2022.
31. Burgess S, Mason AM, Grant AJ, Slob EAW, Gkat

80. However 2, Internation, Internation, Internation, Internation, Inc. 1, 2022.
81. Burgess S, Mason AM, Grant AJ, Slob EAW, Gkatzionis A, Zuber V, et al. Using genet
83. Burgess S, Mason AM, Grant AJ, Slob EAW, Gkatzioni genome-wide association analyses as a series bias in estimates of another genetic effects.

81. Burgess S, Mason AM, Grant AJ, Slob EAW, Gkatzionis A, Zuber V, et al. Using genetic

association data to guide drug discovery 81. Burge:
association da
The American
 81. Burgess Contained American American American Systems American Andre and Application Andre American Journal of Human Genetics. 2023;110(2):195-214.
The American Journal of Human Genetics. 2023;110(2):195-214. association data to guide drug discovery and development: Review of Martin applications.
The American Journal of Human Genetics. 2023;110(2):195-214.
34 $\frac{1}{34}$