Cancer Epidemiology 62 (2019) 101577

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Reporting and guidelines for mendelian randomization analysis: A systematic review of oncological studies

Gary C.Y. Lor[a], Harvey A. Risch[b], W.T. Fung[a], S.L. Au Yeung[a], Irene O.L. Wong[a], Wei Zheng[c], Herbert Pang[a,*]

[a] School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, G/F, Patrick Manson Building (North Wing), 7 Sassoon Road, Pokfulam, Hong Kong

[b] Department of Chronic Disease Epidemiology, Yale School of Public Health, Laboratory of Epidemiology and Public Health, 60 College St., New Haven, CT, CT 06510, United States

[c] Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University Medical Center, Vanderbilt Epidemiology Center Institute for Medicine and Public Health, 2525 West End Avenue, Suite 800, Nashville, TN, United States

ARTICLE INFO

Keywords:
Mendelian randomization
Reporting guideline
Cancers
Reporting quality
Systematic review

ABSTRACT

Background: Mendelian randomization (MR) analyses have been increasingly used to seek evidence of causal associations. This systematic review aims at characterizing and evaluating the reporting of MR analyses in oncological studies.

Methods: The PubMed database was searched to identify MR cancer studies until December 31, 2017. Two of the authors independently selected and evaluated reporting quality of the studies. Reporting quality in MR studies before 2016 and in 2016/17 was compared.

Results: Cancer studies with MR analyses in 2016 and 2017 accounted for 55.8% of the total number of studies identified. In the 77 eligible articles, 39 (50.6%) did not report subjects' characteristics, 53 (68.8%) did not conduct power estimation, 40 (51.9%) did not state all of the first three MR assumptions (i.e., genetic instrument is associated with exposure, is not associated with confounders, and acts on outcome only through exposure), and 31 (40.3%) did not exclude SNPs that diverged from Hardy-Weinberg equilibrium. More studies estimated power in 2016/2017 than before 2016 ($p = 0.028$).

Conclusions: Some MR cancer studies did not sufficiently report essential information, posing obstacles for critical appraisal. This study proposes for MR analysis a guideline/checklist for future publications in cancer and other biomedical research.

1. Introduction

Identification of causal relationships between exposures and cancers is the key to cancer prevention, early diagnosis and improved treatment. Although randomized controlled trials (RCTs) are often used to provide evidence for causal inference, such studies are often not feasible for ethical or practical reasons. Thus, observational studies are the major means by which associations have been found between exposures and cancers. Observational studies are susceptible to confounding and reverse causation in inferences about causation [1]. Therefore, analytic methods that have the potential to reduce bias and confounding in observational studies may provide stronger evidence for causal inference.

Mendelian randomization (MR) analysis is an analytic approach that seeks to obtain evidence for causation between an exposure and an observational outcome by using autosomal genetic variants as instrumental variables [1,2]. MR has been suggested to be a "natural RCT" because under certain assumptions, genetic variants from parents are randomly and independently distributed in their offspring, implying that the exposures determined by genetic variants are unlikely to be influenced by environmental confounding factors or biases or suffer from reverse causation [1,3]. Therefore, using a genetic instrument as a proxy to assess an association between an exposure and an outcome of interest can theoretically generate causal evidence of greater validity. For instance, the association between circulating polyunsaturated fatty acid (PUFA) level and prostate-cancer risk have been inconsistent in observational studies [4,5]. MR genetic instrumental variables for PUFAs were examined with respect to risk of prostate cancer [6] but no

* Corresponding author at: School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, School of Public Health Office, G/F, Patrick Manson Building (North Wing), 7 Sassoon Road, Pokfulam, Hong Kong.
E-mail address: herbpang@hku.hk (H. Pang).

https://doi.org/10.1016/j.canep.2019.101577
Received 2 April 2019; Received in revised form 16 July 2019; Accepted 22 July 2019
Available online 01 August 2019
1877-7821/

associations were found between PUFAs and prostate-cancer risk in MR analysis, providing evidence against the existence of a causal relationship. On the other hand, the observed associations between telomere length and hepatocellular carcinoma reported in observational studies have been evaluated with MR analysis and the relationship was confirmed [7,8].

However, the MR analytic technique also poses challenges to analysis and interpretation in cancer studies. A systematic review on reporting of Mendelian randomization demonstrated that publications have included inadequate reporting of the plausibility of the MR assumptions and statistical methods used in MR analyses [9]. In other types of instrumental variable analyses, reporting has often been a concern, since important information is often under-reported [10]. Inadequate reporting of methodologies, assumptions and sensitivity analyses may hinder the critical appraisal and utilization of the study data [11]. Further, many MR studies have overstated the degree of causal evidence of their results, because of the simplistic assumption that Mendelian independent inheritance of parental alleles vitiates all possible biases and confounding. These problems may also lead MR study authors to false-positive or false-negative conclusions [12]. To the best of our knowledge, no systematic reviews of reporting quality and trends of reporting over time in MR cancer studies have been published. Therefore, we attempt to evaluate the reporting quality of MR analyses in cancer studies and provide guidelines for reporting in MR biomedical research, for use in future publications.

2. Materials and methods

2.1. Search strategy and inclusion/exclusion criteria for existing studies

A systematic literature search of cancer studies with MR analyses was conducted in the PubMed database. The timeframe of the search was until December 31, 2017, without earlier date limits (the earliest date of the database was January 1, 1972). Search keywords included "Mendelian randomization", "Cancer" and their relevant or equivalent terms (as listed in Appendix 1). We included studies that (i) were published in English and (ii) used MR to investigate potential causal relationships between exposure(s) and cancer-related outcome(s). The study search excluded (i) conference abstracts, letters, commentaries, editorials, reviews, study proposals, book chapters, dissertations, purely methodological papers with MR analyses as examples, publications in health economics journals and purely theoretical papers, (ii) studies not applying MR analyses, and (iii) studies not aiming at identifying causation between exposure(s) and cancer related outcome(s).

2.2. Paper selection

Titles and abstracts resulting from the search keywords (Appendix 1) from PubMed were screened and selected independently by two of the authors (GCYL and JWF) according to the inclusion and exclusion criteria. When inconsistency in decision occurred, the two authors reviewed the full texts of the articles and arrived at consensus with the assistance from a third independent reviewer (HP).

2.3. Data extraction

The full texts of the eligible papers were examined to extract relevant information. For study characteristics, we recorded cancer type of interest, year of publication, impact factor of the publishing journal, number of studies included (for pooled data), number of subjects enrolled, number of genetic instruments, significance of results, and presence of traditional meta-analyses of exposure-outcome associations (for pooled data). Traditional meta-analysis in this study refers to combining data of multiple observational studies for summary estimates of association, but not involving MR analyses. Impact factors of journals publishing in 2016 were obtained from InCites Journal Citation Reports [13].

Items of reporting quality extracted from the published reports included provision of subjects' characteristics and cancer characteristics, application of sensitivity analyses, assessment of genetic variant-exposure correlations (i.e., strength of instrument), stating the fundamental assumptions of MR analysis, reporting of the methods to derive MR estimates, power estimation, consideration of heteroscedasticity in the calculation, discussion of pleiotropy, population stratification or linkage disequilibrium and report of between-study heterogeneity if data pooling was involved. Each item was classified either as "yes" or "no". We identified two types of characteristics reported: (i) cancer characteristics (e.g., stage of cancer at diagnosis, presence of metastasis, etc.) and (ii) subjects' characteristics, including socio-demographics, lifestyle, exposure status and genetic composition (e.g., age, gender, socio-economic status, smoking status, drinking habits, etc.). Articles reporting on these characteristics were classified as "yes", while the rest were classified as "no". Sensitivity analysis is defined as methods that determine how uncertainties in the output may be attributed to various uncertainties in the input of a model [14]. A study was classified as "yes" in application of sensitivity analysis if (i) the study mentioned that it had performed a sensitivity analysis of some type or, (ii) any kind of sensitivity analysis was identified during critical appraisal. The four fundamental assumptions of MR analysis are (i) the genetic instrument must be associated with the exposure, (ii) the genetic instrument must not be associated with confounders between the exposure and outcome, (iii) the genetic instrument must be associated with the outcome only through the exposure [2] and (iv), the parental alleles of the instrument must be sufficiently uncorrelated, as for example occurs under panmixia. Articles stating any of the four MR assumptions were classified as "yes" in the corresponding assumption. For the first assumption, assessment of genetic variant-exposure correlations refers to the estimation of the strength of genetic instruments from empirical data, instead of mere citation of fact of association in the literature. Articles that met this definition were classified as "yes" in "assessment of genetic instrument-exposure correlation (i.e., first MR assumption)" while those with citation of an article using a non-identical study population or without reporting of genetic variant-exposure correlation were classified as "no". For assessment of the second assumption, the article is classified as "yes" if it attempted to detect any correlations between common confounders and the genetic instrument. For the assessment of the third assumption, the article is classified as "yes" if MR-Egger regression or test for directional pleiotropy is used to detect pleiotropy. The fourth, though seldom explicitly stated assumption that choice of mating is uncorrelated with genotype, can be falsified if single nucleotide polymorphisms (SNPs) diverge from Hardy-Weinberg equilibrium (HWE) [15] so MR studies typically exclude such SNPs. Articles reporting Q-Q plots for SNP HWE p-values in the study or excluding SNPs divergent from HWE from the analysis were classified as "yes" on "assessment of the fourth assumption". Consideration of heteroscedasticity in the calculation referred to discussion of the influence of heteroscedasticity or the adjustment for its impact in calculations. Robust standard error (also known as sandwich estimator), nonparametric bootstraps, clustered robust standard errors, generalized estimating equations/generalized method of moments, were often reported to account for heteroscedasticity in instrumental variable analysis [10]. Discussion of impact or account for heteroscedasticity in calculation was classified as "yes" in the item "Consideration of risk of bias due to heteroscedasticity". If data pooling was involved, studies reporting between-study heterogeneity (e.g., inconsistency index) were classified as "yes" for the item "Report of between-study heterogeneity if data pooling is involved".

For a clearer landscape of MR studies in oncology, we also describe the types of subject characteristics reported, types of sensitivity analysis performed, and methods used in MR estimation. The categorization of sensitivity analysis was partially based on the publication of Burgess et al. [16] and other methods summarized by the authors (see Appendix

3).

In reviewing the publications, items of MR elements were independently extracted and verified by two authors (CGYL and JWF). Inter-rater agreement of 42 randomly selected articles was measured for these elements with Cohen's κ coefficient. According to McHugh, Cohen's κ coefficients of 0-0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80 and 0.81–1.00 correspond to level of agreement "slight", "fair", "moderate", "substantial" and "almost perfect" respectively [17]. For items with Cohen's score lower than 0.8, a third investigator (HP) reappraised the studies involved in inconsistency to arrive at a final decision.

2.4. Statistical methods

In order to detect changes in reporting quality over time, the major reporting items (provision of population's characteristics, stating the first three MR assumptions, and assessment of genetic instrument-exposure correlation) in 2016–2017 and before 2016 were compared with Fisher's exact test statistic. All reported p-values are two-sided and p-values of less than 0.05 are considered nominally statistically significant. Statistical analysis was performed using R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria).

3. Results

3.1. Study selection

The literature search identified 218 potentially relevant articles. Of these, 51 non-cancer papers, 33 studies not using MR analysis to assess causality, 29 reviews, 17 statistical, methodological or theoretical papers and 11 editorials, commentaries or letters were excluded, leaving 77 eligible articles for analysis (Fig. 1). Full texts were available for all eligible articles. This study closely followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist as attached in Supplementary Table 1 [18].

3.2. Trends of MR studies

The first eligible MR cancer study was published in 2005. The time trend of MR cancer studies is described in Fig. 2. The number of MR cancer studies has been accelerating since 2011, with published MR cancer studies in 2016 and 2017 accounting for 55.8% of the total of the published studies. Beyond the oncological field, MR analyses have also increased substantially. Similar to cancer studies, rising trends have been observed in studies of diabetes and coronary heart diseases (Supplementary Graph 1 and Supplementary Methods).

3.3. Characteristics of eligible studies

Characteristics of the included studies are presented in Table 1. Most of the articles (67.5%) studied the cause of one cancer type of interest while the rest (32.5%) targeted more than one cancer type. Among the studies of one cancer type, the most common cancers investigated with MR analysis were colorectal cancer (11 articles) and prostate cancer (10 articles). A majority (79.2%) of the reports pooled data from multiple studies to increase sample size. The median sample size was 27,640 subjects and the median number of studies included among the studies with data pooling was 11.5. There were 4 (5.2%) small sample size studies with fewer than 1000 subjects, 22 (28.6%) large sample size studies with 1000 to 9999 subjects and 51 (66.2%) very large sample size studies with 10,000 or more subjects. Among the studies, 40 (51.9%) constructed genetic risk scores from multiple SNPs as instrumental variables and 20 (26.0%) used multiple SNPs each as independent instrumental variables. Fifty-three (68.8%) studies reported positive or significant results. In twelve of the eligible articles, traditional meta-analysis was performed to estimate exposure-outcome associations. Comparing conclusions from traditional meta-analysis with MR analysis, fewer than half (33.3%) of the conclusions from traditional meta-analysis were consistent with MR analyses. The journals publishing MR cancer studies had a wide range of impact factors, from 1.3 to 28, with a median of 6.5.

3.4. Reporting quality of MR elements

Evaluation of reporting of MR cancer studies is presented in Table 2, and details of the reporting are given in Supplementary Table 2. Only 38 (49.4%) and 18 (23.4%) of the studies presented subjects' characteristics and cancer characteristics respectively. Six (7.8%) articles reported both. Fewer than half (48.1%) of the studies stated the three first MR assumptions in the full text. Only 1 (1.3%) article explicitly stated that random mating (i.e., panmixia) was an assumption of MR analysis. Forty-eight (62.3%), 29 (37.7%), 21 (27.3%), 46 (59.7%) articles assessed the first, second, third and fourth assumption respectively. Forty-six (59.7%) articles article excluded SNPs that diverge from HWE to assess the fourth assumption, but none of the studies reported Q-Q plots. Overall, 67 (87.0%) of the studies discussed at least one item of pleiotropy, population stratification or linkage disequilibrium, which were common issues in MR analyses and relevant to the MR assumptions.

In the 48 (62.3%) studies that assessed the correlation of the genetic instrument with the exposure of interest (i.e., strength of instrument, first assumption), five articles assessed the instrument strength and performed MR analysis with two different populations within the same study. Twelve out of 16 studies (75%) without data pooling assessed the strength of instrument but only 36 out of 61 studies (59.0%) with data pooling assessed it. The proportion of variance explained (R^2, 42 articles, 54.5%) and F-statistics (30 articles, 39.0%) were the two most commonly reported measures of instrument strength (Supplementary Table 4). Fifty-nine (76.6%) studies reported MR estimates of exposure-outcome associations, while the rest of the studies (23.4%) only reported associations between genetic variants and cancer outcomes. Among studies providing MR estimates, 51 (86.4%) reported clearly the method to calculate the estimates. The most commonly used method

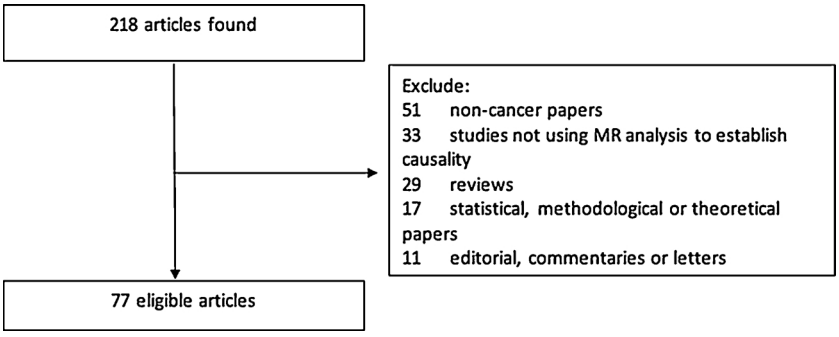

Fig. 1. Flowchart of selection process.

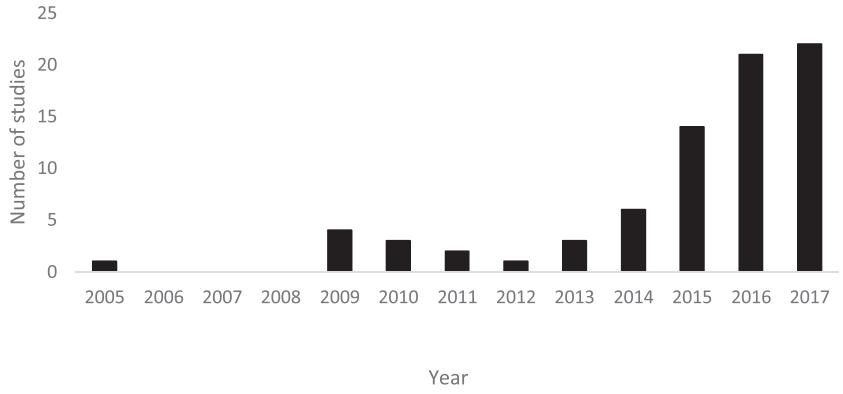

Fig. 2. Number of MR cancer studies published by year.

Table 1
Characteristics of eligible studies.

Characteristics	Number of studies (no. of studies/total no., %)
Cancer type (n = 77)	
One cancer type	52 (67.5)
Colorectal	11/52
Prostate	10/52
Breast	5/52
Other cancers	26/52
More than one cancer type	25 (32.5)
Exposure (n = 77)	
Obesity/BMI	14 (18.2)
Blood lipids	10 (13.0)
Vitamin D and its binding protein	8 (10.4)
Other	45 (58.4)
Data pooling (n = 77)	
Yes	61 (79.2)
No	16 (20.8)
Sample size (n = 77)	
Small (< 1000 subjects)	4 (5.2)
Large (1000-9999 subjects)	22 (28.6)
Very Large (10,000 subjects or more)	51 (66.2)
Number of included studies in studies with data pooling (n = 61)	
Median	11.5
Interquartile range (IQR)	21.75
Single gene/SNP as genetic instrument (n = 77)	
Yes, single gene/SNP was used as genetic instrument	17 (22.1)
No, multiple genes/SNPs were used	60 (77.9)
Genetic risk score was constructed as a single instrument	40/60
Individual SNPs/genes were assessed as independent instruments	20/60
Conducted traditional meta-analysis for exposure-outcome correlation (n = 77)	
Yes	12 (15.6)
Consistent with MR conclusion	4/12
Inconsistent with MR conclusion	8/12
No	65 (84.4)
Impact factor (n = 77)	
Range	1.3-28
Median	6.5
Interquartile range (IQR)	3.5
Positive Results (n = 77)	
Yes	53 (68.8)
No	24 (31.2)

Table 2
Evaluation of Reporting of MR cancer studies.

Reporting items	Number of studies (no. of studies/total no., %)
Provision of subjects' characteristics (n = 77)	
Yes	38 (49.4)
No	39 (50.6)
Provision of cancer characteristics (n = 77)	
Yes	18 (23.4)
No	59 (76.6)
Number of MR assumptions stated (n = 77)[a]	
None of the assumption stated	31 (40.3)
One assumption stated	7 (9.1)
Two assumptions stated	1 (1.3)
Three assumptions stated	38 (49.4)
All assumptions stated	0 (0)
Assessing the MR assumption (n = 77)[b]	
First assumption	
Yes	48 (62.3)
No	29 (37.7)
Second assumption	
Yes	29 (37.7)
No	48 (62.3)
Third assumption	
Yes	21 (27.3)
No	56 (72.7)
Fourth assumption	
Yes	46 (59.7)
No	31 (40.3)
Reporting MR estimate (n = 77)	
Yes, with calculation method	51 (66.2)
Yes, without calculation method	8 (10.4)
No	18 (23.4)
Power estimation (n = 77)	
Yes	24 (31.2)
No	53 (68.8)
Conducted sensitivity analysis (n = 77)	
Yes	61 (79.2)
No	16 (20.8)
Report of between-study heterogeneity if data pooling is involved (n = 61)[c]	
Yes	37 (60.7)
No	24 (39.3)

[a] 37 (48.1%) articles stated all the first three assumptions.

[b] First assumption: the genetic instrument must be associated with the exposure, Second assumption: the genetic instrument must not be associated with confounders between the exposure and outcome. Third assumption: the genetic instrument must be associated with the outcome only through the exposure. Fourth assumption: the parental alleles of the instrument must be sufficiently uncorrelated (e.g., under panmixia).

[c] only 61 studies involved meta-analysis of data from multiple studies.

was the Wald method (also called the ratio of coefficients method) (52.5%) [19]. Twenty-four (31.2%) studies conducted power estimation. Among the 77 included studies, 61 (79.2%) applied sensitivity analyses to examine robustness of the MR analysis. The most commonly used method was the calculation of associations between genetic instruments and common confounders to provide evidence against known confounding, which was done in 29 (37.7%) of the articles.

Thirty-seven studies (60.7%) reported between-study heterogeneity, mainly the inconsistency index (I^2). Only 11 (14.3%) studies considered

Table 3
Reporting quality of MR cancer studies by publication year (in 2016–2017 vs before 2016).

Variables	2016-2017 (%)	Before 2016 (%)	P value[a]
Provision of subjects' characteristics			
Yes	18 (41.9)	20 (58.8)	0.172
No	25 (58.1)	14 (41.2)	
Provision of cancer characteristics			
Yes	12 (27.9)	6 (17.6)	0.417
No	31 (72.1)	28 (79.4)	
State all the first three MR assumptions[b]			
Yes	25 (58.1)	12 (35.3)	0.066
No	18 (41.9)	22 (64.7)	
Assessment of genetic instrument-exposure correlation			
Yes	24 (55.8)	24 (70.6)	0.238
No	19 (44.2)	10 (29.4)	
Reporting MR estimation method (for studies with MR estimate reported) (n = 59)[c]			
Yes	31 (86.1)	20 (87.0)	1
No	5 (13.9)	3 (13.0)	
Power estimation			
Yes	18 (23.4)	6 (7.8)	0.028
No	25 (32.5)	28 (36.4)	

[a] Two-sided Fisher's exact test.
[b] The genetic instrument must be associated with the exposure, the genetic instrument must not be associated with confounders between the exposure and outcome, the genetic instrument must be associated with the outcome only through the exposure.
[c] only 59 studies reported MR estimates.

heteroscedasticity (i.e., varying variance) in the calculation and all of them adopted robust standard error estimates to account for it. The remainder of the studies assumed homoscedasticity (i.e., a constant variance). Cohen's κ coefficient of inter-rater agreement for "provision of subjects' characteristics", "assessment of genetic instrument-exposure correlation", and "state the first three MR assumptions" were 0.81 (almost perfect), 0.73 (substantial), and 1 (almost perfect) respectively (Supplementary Table 3), suggesting satisfactory agreement between the two independent investigators on the reporting items. Inconsistent studies for "assessment of genetic instrument-exposure correlation" were reappraised by HP. In order to detect possible changes in reporting among MR cancer studies over time, Fisher's exact test statistic was calculated for various reporting items in 2016–2017 versus before 2016 (Table 3). Significantly more studies conducted power estimation in 2016–2017 than before 2016 ($p = 0.028$). Other reporting items showed no significant difference between 2016–2017 and before 2016.

4. Discussion

Ongoing demand for stronger evidence of causation, increasing availability of Genome-wide Association Study (GWAS) databases, and creation of genetic instruments related to phenotype traits have fuelled the increasing trend of Mendelian randomization applications in biomedical research. Among the 12 oncological articles with both traditional meta-analysis and MR analysis in this review (Table 1), fewer than half (33.3%) of the conclusions from traditional meta-analysis were consistent with MR analyses. This may be due to potential biases and confounding of traditional observational studies which may jeopardize result validity. MR analysis has the potential to reduce the likelihood of reverse causation, as well as some degree of confounding by unmeasured factors. Inadequate or inconsistent reporting, however, may jeopardize the ability to appraise critically, reproduce or meta-analyse study results [20,21]. Such reporting may also reflect inadequate attention to underlying MR assumptions that may lead to greater risk of false-positive or false-negative findings in MR studies [12,22]. Our analysis demonstrated that reporting of MR analysis was heterogeneous among cancer studies and inadequate in some of them. Inadequate reporting of methodological details, fundamental assumptions and key statistics can compromise the ability to appraise critically a study and compare it with similar research. It can also hide reasons for incorrect study conclusions. Therefore, our findings provide useful reporting guidelines for MR analyses in cancer studies.

A fundamentally important aspect of observational epidemiology is the generalizability of study results to defined populations. This is perhaps the most unspoken limitation of RCTs: just because an apparently causal association can be shown in a small subset of a population does not mean that the association applies to the whole population in general. The representativeness of enrolled subjects for both a specified population and a disease condition enables generalization. Thus, provision of subject characteristics allows the comparison of the study participants with specific target populations. For instance, evidence from a study conducted primarily on patients with metastatic cancer may not be applicable to patients with earlier stages of the same cancer. Similarly, studies with smoking prevalence appreciably higher than the general population may not be useful for investigating causes of cancer in the general population. Although MR is regarded as a "natural RCT", this may not be true in some circumstances. Horizontal pleiotropy is a common problem of MR analysis in which a genetic instrument influences a variety of phenotypes on distinct pathways that can separately affect the outcome of interest [23,24]. If the genetic instrument influences the outcome through a pathway independent of the exposure of interest (i.e., horizontal pleiotropy exists and the third assumption is violated), the generalizability of an MR study conducted on a population that has appreciably different characteristics from the target population may be questionable. Our findings indicate that 50.6% and 76.7% of the MR cancer studies did not report subjects' characteristics and cancer characteristics, respectively. This lack of information of participant characteristics may pose challenges to critical appraisal, as the possibility of pleiotropy and other biases cannot be assessed.

Inadequate reporting and validation of the assumptions underlying MR may jeopardize apparent positive findings. Violations of these assumptions can result in biased or false conclusions. In our survey, 40 (51.9%) of the articles did not state all of the first three assumptions and none of the articles stated all four assumptions (Supplementary Table 2). Forty-eight (62.3%) articles assessed the first assumption by verifying the instrument strength with empirical data. Twenty-nine (37.7%) articles assessed the second assumption by detecting correlations between potential confounders and genetic instruments. Twenty-one (27.3%) of the articles assessed the third assumption by performing MR-Egger regression or tests for directional pleiotropy. Forty-six (59.7%) articles excluded SNPs that diverge from HWE to address the fourth assumption that the parental alleles of the instrument must be sufficiently uncorrelated (i.e., choice of spouse is uncorrelated with genotype) but none of the articles reported Q-Q plot to assess the assumption. According to Mendel's second law [25], the fourth assumption should hold true if mating is independent of genetic composition or under panmixia. In most studies, particularly in the US, however, common characteristics among participants (e.g., race and ethnicity, physical attributes, education, socio-economic status, risk-taking behavior) may be related to genotype [26] and violate the assumption. Although SNPs that diverge from HWE may have other causes, such as genotyping errors [15], detection of SNPs deviating from HWE should raise concern over the validity of the instrument and exclusion is recommended if possible. Chi-square test is commonly used to deviation from HWE [15]. There is no current consensus on cut-off for deviation from HWE, but in our samples, the cut-offs range from $p < 0.05$ to $p < 1 \times 10^{-7}$ (most commonly used $P < 0.05$, 8 studies). As a sensitivity analysis, a Q-Q plot for Hardy-Weinberg equilibrium is recommended to detect deviations from independent assortment. In any event, these procedures for assessing panmixia provide only a partial

picture, because the issue is not whether SNP deviations from HWE occur by chance in the study data, but how much deviation from HWE the observed frequencies represent, an issue of potential confounding. Magnitude, as opposed to statistical significance, of deviations from HWE has not generally been examined in MR studies. A recent analysis of spouse pairs in multiple generations of the Framingham study shows substantial spousal correlations (0.38-0.80) in ancestry-specific principal components [27], consistent with appreciable lack of panmixia in a typical area of the US. That study also clarified that adjustment for ancestry-specific principal components in MR studies is necessary to reduce biases underlying deviations of the fourth assumption. To our knowledge, MR studies have not generally included ancestry principal components adjustments in their instrumental variable regression models, and thus how well such analyses reflect true results is uncertain.

Other manifestations of assumption violation–pleiotropy, population stratification and linkage disequilibrium–are common technical problems for MR analyses. Ten (13.0%) of the articles, however, did not discuss any of these three generally encountered issues nor account for their possible impacts on the results. This may also pose challenges to critical appraisal, as possible directions and magnitudes of biases are not discussed. Since the validity of MR estimation rests primarily on true assumptions, and since the magnitudes of instrument association tend to be very low, detailed discussion is important to explain the assumptions, analyse the chance of assumption violation with existing knowledge and assess the potential impact of violation.

As it is generally impossible to verify empirically the MR assumption that the genetic instrument is not associated with confounders between exposure and outcome and that it is associated with the outcome only through the exposure [28], sensitivity analyses play an important role in understanding falsification of the assumptions if they are not true. This review found that a majority (79.2%) of the studies reported some types of sensitivity analyses. In light of the susceptibility of MR analysis to assumption violations, we recommend that all MR analyses conduct and report sensitivity analyses (e.g., MR-Egger regression, over-identification test statistics) to examine the robustness of estimates.

Assessment of the strength of the genetic instrument (i.e., first assumption) can be achieved empirically. Reliable MR analysis requires strong genetic instruments to reduce potential bias from unmeasured confounding for the instrument or from pleiotropic associations [29]. Although it is known that weak instruments are more likely to reflect biased estimation [30], there is no evidence-based consensus on a cut-off value for a strong instrument. F-statistic > 10 is arbitrarily adopted as an acceptable strength of instrument and it may be misleading [31,32]. Therefore, maximizing F-statistics with large sample sizes, proper instruments and genetic models, conducting sensitivity analyses and constructing a stronger instrument with multiple genotypes are important tools to confront weak instrument bias [32].We observed that 48 (62.3%) of the MR cancer studies reported assessment of the strength of the genetic instrument. The remainder (37.7%) of the studies merely cited the literature to support a relationship between their chosen genetic instruments and the exposures of interest. Among the 48 studies with assessment, 5 studies assessed the strength of instrument with an independent population that was not included in the MR analysis. Obtaining gene-exposure correlation (usually from GWAS) and gene-outcome correlation (usually from consortia) from two samples (which may have overlaps) is known as two-sample Mendelian Randomization, in contrast with the traditional one-sample MR in which both genetic instrument-exposure and genetic instrument-outcome relationship are obtained from the same population. When the instrument is weak, the conventional one-sample MR biases towards the confounded multivariable regression result, while the two-sample MR biases towards the null [33]. It is also important for authors of two-sample MR studies to ensure that the two samples are from the same underlying population (e.g., race) [34].

Another noteworthy finding is that a considerable number of studies (27 articles) reported the beta in linear regression as measure of instrument strength. Considered alone, the beta coefficient may overestimate the impact of a genotype with large magnitude of association but low population exposure frequency [31]. Therefore, the proportion of variance explained (i.e., R^2) is preferred in reporting of instrument strength in MR analysis. We found that 75% of studies without data pooling assessed the strength of the instrument but only 59.0% of the studies with data pooling reported such assessment. The combination of data from multiple studies may escalate complexity in the strength calculation, resulting in less empirical assessment but more likely citation of instrument strength from the literature. Since the strength of a genetic instrument is central to MR estimation in any methodology, proper assessment and calculation are important in all MR analyses. Ultimately, because many MR studies have obtained very low proportions of variance explained, authors have tended to limit conclusions to assertions of fact of causation, rather than to degree of causation. This results in the same problem as for RCTs, that unless the subjects, outcomes and exposures are representative of defined entities, the generalization of findings may be limited, as we have noted above. Low proportions of variance explained can also result from instrument component chance bias [31,34], a type of confounding frequently ignored in the assertion of causal MR conclusions.

MR estimation is usually based on an assumption of homoscedasticity; if the variance is heteroscedastic, resulting estimates may be biased [10]. Among all included MR cancer studies, only 11 articles (14.3%) reported robust standard errors to account for heteroscedasticity. In general, it is appropriate to examine heteroscedasticity within the dataset in order to verify the assumption of homoscedasticity. A robust standard error should be calculated if heteroscedasticity is detected. Following that, the MR estimate can be obtained through a range of methods. The Wald method (used in 52.5% of the studies) and the two-stage method (in 23.7%) [35] were the two most common methods of estimation in the reviewed studies. Eight studies (13.6%), however, did not clearly report the method of estimation. Power estimation is another reporting item of MR studies but only 24 (31.2%) of the studies reported estimated power. Power estimation provides readers with information about the capability of a study to have captured a true positive result if there were one. Significantly more studies with negative results (22/24) discussed the issue of power than the studies with positive results (18/53) ($p = 0.024$).

For better consistency and comprehensiveness of the reporting in MR analysis of cancer studies, we have proposed a set of recommendations and a checklist (Table 4). The full version of this is given in Supplementary Tables 5 and 6. These items were modified from STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) [36], STREGA (Strengthening the Reporting of Genetic Association Studies) [37] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [18] to adapt for use in MR studies. The major objectives of the guidelines are to report (i) subject characteristics, (ii) technical details of genotyping, quality control and imputation, (iii) selection of genetic instruments and evaluation of their strength, (iv) MR estimation methods and power estimation and, (v) sensitivity analyses to evaluate MR assumptions and the discussion of possible assumptions violation. Some of the specified items can be included in supplementary materials if manuscript length is limited. Although our recommendations originate from evaluation of MR cancer studies, they can be applied in a broader context to MR analyses in other fields.

This study has several strengths in its evaluation of MR cancer studies. First, it includes all MR analyses in cancer publications through December 31, 2017. Careful study selection and evaluation provides a comprehensive overview of the reporting quality of MR cancer studies. Second, the independent evaluation by two investigators and the report of Cohen's kappa coefficient help to reduce possible bias from subjective factors. This study also has some limitations worthy of note.

Table 4
Brief guidelines for reporting Mendelian randomization analysis (Combined version of Supplementary Table 5 and 6. Full versions in Supplementary File.).

	Item No	Recommendation	Check-box
Title and abstract	1	(a) Indicate the use of Mendelian randomization in the title or the abstract.	☐
Methods			
Quality control and Imputation	9	(For studies without data pooling) Clearly report details of quality control (e.g., method to address low call rates, excessive heterozygosity, incorrect sex assignments, and ambiguous familial relationship) and imputation.	☐
Selection of genetic instruments	10	(a) Describe in detail the strategy used to select genetic instrument. Specify methods to assess strength of genetic instrument in one-sample MR and possibility of linkage disequilibrium between instruments. For instrument strength assessment, specify from which population the strength is calculated. Test for and exclude SNPs that diverge from Hardy-Weinberg equilibrium.	☐
		(b) If genetic score is used, clear specify the calculation formula, selection of components and weights with explanations. Specify the method to assess its strength.	☐
Synthesis of results	11	(a) Describe the method used to calculate MR estimate if reported. Consider possible bias due to heteroscedasticity.	☐
		(b) If traditional meta-analysis is conducted, describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	☐
Bias	12	Describe how the potential biases due to MR assumption violation or other sources can be assessed or minimized.	☐
Statistical methods	15	(a) Describe all statistical methods, including those used to calculate MR estimate and control for confounding. Adjust for the top few principal components which represent the population's genetic structure.	☐
Sensitivity analysis	16	Describe how sensitivity analysis is applied to test for robustness of the MR assumptions.	☐
Results			
Descriptive data	18	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders. (may also be included in the Methods session or supplementary materials)	☐
Report of results	18	(a) Present results of MR analysis, including confidence intervals (if point estimate is reported) and measures of between-study heterogeneity.	☐
		(b) If traditional meta-analysis is also performed, reported confidence intervals and risk of bias across the studies.	☐
Report of bias	19	Present results of sensitivity analyses to assess biases due to assumption violation or other sources.	☐
Discussion			
MR assumptions	22	State clearly the four MR assumptions (if not yet stated in Introduction and Methodology Session). Interpret the results of sensitivity analysis and specify the assumptions assessed. Discuss the validity of MR assumptions in the study and potential impact of assumption violation.	☐
Strength of instrument	23	Discuss the strength of the genetic instrument and discuss potential impact due to weak instrument.	☐
Limitations	24	Discuss limitations of the study, taking into account sources of potential bias (e.g., pleiotropy, population stratification and linkage disequilibrium) or imprecision. Discuss both direction and magnitude of any potential bias.	☐

First, our survey focused only on cancer studies. As shown in Supplementary Fig. 1, cancer studies have substantially increased in the application of MR analyses, as compared with studies of diabetes and coronary heart disease. Trend in MR cancer studies can be a good representation of the reporting of recent MR analyses. Second, we only searched articles through the PubMed database. A systematic review reported that searching beyond PubMed had only modest effects on results [38]. We also did not search through non English-language publications. This limits our observations and generalizability to the English-language literature.

In conclusion, Mendelian randomization analysis has been increasingly applied with observational data in attempt to detect causal relationships between exposures and cancer outcomes. However, inadequate reporting and assumption validation in MR cancer studies has potentially compromised validity, reproducibility, transparency and consistency. Therefore, we propose a checklist to improve the quality of MR analyses in future publications. Possible future direction of research includes formulating a suggestion list of sensitivity analyses in different types of MR study designs.

Original publications

This manuscript contains original unpublished work and is not being submitted for publication elsewhere at the same time.

Ethics approval and consent to participate

This study does not involve human participants, human data or human tissue.

Availability of data and material

The data that support the findings of this study are openly available in Open Science Framework at http://osf.io.

Authors' contributions

GCYL and HP conceived the study and planned the analyses. GCYL, JWF and HP participated in study selection, data extraction and analysis. GCYL and HP drafted the manuscript. All authors (GCYL, HAR, JWF, SLAY, IOLW, WZ, and HP) participated in manuscript editing, data interpretation and final approval of the version to be published.

Funding

This study is supported by the University of Hong Kong Li Ka Shing Faculty of Medicine Research Internship Scheme.

Declaration of Competing Interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.canep.2019.101577.

References

[1] D.A. Lawlor, R.M. Harbord, J.A. Sterne, N. Timpson, G. Davey Smith, Mendelian randomization: using genes as instruments for making causal inferences in epidemiology, Stat. Med. 27 (8) (2008) 1133–1163.
[2] P. Sekula, M. Fabiola Del Greco, C. Pattaro, A. Köttgen, Mendelian randomization as an approach to assess causality using observational data, J. Am. Soc. Nephrol. 27 (11) (2016) 3253–3265.
[3] G. Davey Smith, S. Ebrahim, Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. 32 (1) (2003) 1–22.
[4] M. Carayol, P. Grosclaude, C. Delpierre, Prospective studies of dietary alpha-linolenic acid intake and prostate cancer risk: a meta-analysis, Cancer Causes Control 21 (3) (2010) 347–355.
[5] D.D. Alexander, J.K. Bassett, D.L. Weed, E.C. Barrett, H. Watson, W. Harris, Meta-

analysis of long-chain omega-3 polyunsaturated fatty acids (LCω-3PUFA) and prostate cancer, Nutr. Cancer 67 (4) (2015) 543–554.
[6] N.K. Khankari, H.J. Murff, C. Zeng, W. Wen, R.A. Eeles, D.F. Easton, Z. Kote-Jarai, A.A. Al Olama, S. Benlloch, K. Muir, Polyunsaturated fatty acids and prostate cancer risk: a Mendelian randomisation analysis from the PRACTICAL consortium, Br. J. Cancer 115 (5) (2016) 624.
[7] X. Fu, S. Wan, H.-W. Hann, R.E. Myers, R.S. Hann, J. Au, B. Chen, J. Xing, H. Yang, Relative telomere length: a novel non-invasive biomarker for the risk of non-cirrhotic hepatocellular carcinoma in patients with chronic hepatitis B infection, Eur. J. Cancer 48 (7) (2012) 1014–1022.
[8] Y. Cheng, C. Yu, M. Huang, F. Du, C. Song, Z. Ma, X. Zhai, Y. Yang, J. Liu, J.-X. Bei, Genetic association of telomere length with hepatocellular carcinoma risk: a Mendelian randomization analysis, Cancer Epidemiol. 50 (2017) 39–45.
[9] A.G. Boef, O.M. Dekkers, S. le Cessie, Mendelian randomization studies: a review of the approaches used and the quality of reporting, Int. J. Epidemiol. 44 (2) (2015) 496–511.
[10] N.M. Davies, G.D. Smith, F. Windmeijer, R.M. Martin, Issues in the reporting and conduct of instrumental variable studies: a systematic review, Epidemiology 24 (3) (2013) 363–369.
[11] R.M. Phelps, M.P. Dearing, J.L. Mulshine, Need for Uniformity in Collection and Reporting of Data in Cancer Clinical Trials, Oxford University Press, 1990.
[12] J.P. Ioannidis, The Proposal to Lower P Value Thresholds to. 005, JAMA (2018).
[13] InCites Journal Citation Reports (2017) (Accessed 16 Jan 2018), https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action?year=&edition=&journal=.
[14] A. Saltelli, Sensitivity analysis for importance assessment, Risk Anal. 22 (3) (2002) 579–590.
[15] S. Rodriguez, T.R. Gaunt, I.N. Day, Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies, Am. J. Epidemiol. 169 (4) (2009) 505–514.
[16] S. Burgess, J. Bowden, T. Fall, E. Ingelsson, S.G. Thompson, Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants, Epidemiology 28 (1) (2017) 30–42.
[17] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, biometrics (1977) 159–174.
[18] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, P. Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009) e1000097.
[19] A. Wald, The fitting of straight lines if both variables are subject to error, Ann. Math. Stat. 11 (3) (1940) 284–300.
[20] X.I. Yao, X. Wang, P.J. Speicher, E.S. Hwang, P. Cheng, D.H. Harpole, M.F. Berry, D. Schrag, H.H. Pang, Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies, JNCI 109 (8) (2017).
[21] A. MacCarthy, S. Kirtley, J.A. de Beyer, D.G. Altman, I. Simera, Reporting guidelines for oncology research: helping to maximise the impact of your research, Br. J. Cancer (2018).
[22] S. von Hinke, G.D. Smith, D.A. Lawlor, C. Propper, F. Windmeijer, Genetic markers as instrumental variables, J. Health Econ. 45 (2016) 131–148.
[23] M. Greco, F. Del, C. Minelli, N.A. Sheehan, J.R. Thompson, Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome, Stat. Med. 34 (21) (2015) 2926–2940.
[24] P.C. Haycock, S. Burgess, K.H. Wade, J. Bowden, C. Relton, G. Davey Smith, Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies, Am. J. Clin. Nutr. 103 (4) (2016) 965–978.
[25] A. MacLeod, A. Tweedie, S. McLellan, S. Taylor, A. Cooper, L. Sweeney, C.M.R. Turner, A. Tait, Allelic segregation and independent assortment in T. brucei crosses: proof that the genetic system is Mendelian and involves meiosis, Mol. Biochem. Parasitol. 143 (1) (2005) 12–19.
[26] H. Risch, The correlation between relatives under assortative mating for an X-linked and autosomal trait, Ann. Hum. Genet. 43 (2) (1979) 151–165.
[27] R. Sebro, G.M. Peloso, J. Dupuis, N.J. Risch, Structured mating: patterns and implications, PLoS Genet. 13 (4) (2017) e1006655.
[28] S.A. Swanson, M.A. Hernán, Commentary: how to report instrumental variable analyses (suggestions welcome), Epidemiology 24 (3) (2013) 370–374.
[29] M.A. Hernán, J.M. Robins, Instruments for causal inference: an epidemiologist's dream? Epidemiology 17 (4) (2006) 360–372.
[30] J. Bound, D.A. Jaeger, R.M. Baker, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, J. Am. Stat. Assoc. 90 (430) (1995) 443–450.
[31] D.I. Swerdlow, K.B. Kuchenbaecker, S. Shah, R. Sofat, M.V. Holmes, J. White, J.S. Mindell, M. Kivimaki, E.J. Brunner, J.C. Whittaker, Selecting instruments for Mendelian randomization in the wake of genome-wide association studies, Int. J. Epidemiol. 45 (5) (2016) 1600–1616.
[32] S. Burgess, S.G. Thompson, C.C.G. Collaboration, Avoiding bias from weak instruments in Mendelian randomization studies, Int. J. Epidemiol. 40 (3) (2011) 755–764.
[33] G. Davey Smith, G. Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, Hum. Mol. Genet. 23 (R1) (2014) R89–R98.
[34] D.A. Lawlor, Commentary: two-sample Mendelian randomization: opportunities and challenges, Int. J. Epidemiol. 45 (3) (2016) 908.
[35] J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression, Int. J. Epidemiol. 44 (2) (2015) 512–525.
[36] E. Von Elm, D.G. Altman, M. Egger, S.J. Pocock, P.C. Gøtzsche, J.P. Vandenbroucke, S. Initiative, The strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies, PLoS Med. 4 (10) (2007) e296.
[37] J. Little, J.P. Higgins, J.P. Ioannidis, D. Moher, F. Gagnon, E. Von Elm, M.J. Khoury, B. Cohen, G. Davey-Smith, J. Grimshaw, STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement, Hum. Genet. 125 (2) (2009) 131–151.
[38] C.W. Halladay, T.A. Trikalinos, I.T. Schmid, C.H. Schmid, I.J. Dahabreh, Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions, J. Clin. Epidemiol. 68 (9) (2015) 1076–1084.