What can Mendelian randomization tell us about causes of cancer?

Daniela Mariosa, PhD[1]; Robert Carreras-Torres, PhD[1, 2]; Richard Martin, PhD[3, 4, 5]; Mattias Johansson, PhD[1]; Paul Brennan, PhD[1*]

Author Affiliations

[1] Section of Genetics, International Agency for Research on Cancer (IARC), Lyon, France

[2] ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Spain.

[3] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

[4] School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

[5] University Hospitals Bristol NHS Foundation Trust National Institute for Health Research Bristol, Nutrition Biomedical Research Unit, University of Bristol, Bristol, United Kingdom

Corresponding author: Paul Brennan

Head Section of Genetics, International Agency for Research on Cancer

150 Cours Albert Thomas, Lyon, France 69372 Cedex 08

Email: BrennanP@iarc.fr

Multiple important causes of cancer have been successfully identified over the past 70 years, including cigarette smoking, alcohol, obesity and UV light, as well as carcinogens in the occupational environment and different infections. However, despite these successes, about half of the cancer burden cannot be linked to known causes.[1] Difficulties in identifying causal factors for different cancers are due to a number of reasons including limitations in epidemiological study designs and the inherent problems of confounding and reverse causation, as well as inadequate statistical power to study relatively rare cancer types. Potential causes may also be relatively ubiquitous within populations, such as air pollution or water contaminants, and may only be relevant during time windows such as childhood or young adulthood.

Mendelian randomization (MR) offers an opportunity to overcome some of these limitations and further clarify underlying causes of cancer.[2] As discussed extensively in this issue, MR is an instrumental variable analysis that utilizes germline genetic variants, such as single nucleotide polymorphisms (SNPs), as proxies (‘instruments’) for the risk factor of interest. The advantage of MR over traditional observational studies that measure risk factors directly is that such genetic variants are randomly assigned at conception, and therefore bias due to reverse causation is avoided and the potential for confounding is more limited.

The availability of large-scale genome-wide data from large cohorts such as UK Biobank, and large genome-wide association study (GWAS) for individual cancers, along with important methodological advancements, now allow well-powered MR analysis on most cancer for a wide range of risk factors. MR analyses have for instance provided evidence that vitamin D and C-reactive protein are unlikely to have a strong causal effect on cancer, and longer telomeres increase the risk of several cancer types.[3-9] Diabetes is also an emerging cause of cancer for which there is still no conclusive evidence from traditional epidemiological studies and MR results have helped to clarify this relationship for multiple cancers including cancers of the kidney, pancreas, and lung.[10-12]

Similarly to any other study design, MR studies have their own weaknesses and it is important that the results are not interpreted in isolation, but should rather be seen as a complement to other study designs. With this background, we highlight different scenarios where MR can provide useful results (Fig 1). First, for putative risk factors that are either difficult or expensive to measure directly, MR has the potential to interrogate such risk factors at low cost and establish novel hypotheses. An example is provided by the relationship between circulating fasting insulin and renal cell carcinoma which had not previously been studied directly in a prospective setting. Cohort studies suggest that diabetes is associated with the risk of renal cell carcinoma, but a recent MR analysis identified that among the different factors that are related to diabetes, elevated insulin levels may be the causal factor driving the observed association.[12]

Many MR analyses are conducted for suspected causes of cancer where there is some evidence of association from observational epidemiological studies. Randomized trials to establish the causality of putative risk factor-disease associations are usually not feasible, and MR offers the opportunity to provide additional evidence for such associations using an independent study design whilst remaining in an observational setting. The concept of interrogating putative causes using independent methodologies is commonly referred as triangulation, the importance of which was recently highlighted in a commentary by Marcus R. Munafò and George Davey Smith.[13] The concept of triangulation implies the use of multiple study approaches with different and independent sources of bias to address the same underlying question. Risk factor-disease associations with concordant evidence from several different study designs would generally strengthen causal inference. A recent example of a study integrating multiple study approaches was conducted by Fanidi et al. who evaluated the relation between circulating vitamin B12 and lung cancer risk, where consistent associations with risk for elevated B12 were seen, both for directly measured blood levels in prospective case-control studies as in a parallel MR analysis.[3] In this example, residual confounding can exist for directly measured vitamin B12 due to imperfect

assessment of tobacco exposure, whereas the MR approach is unlikely to be affected by the same type of bias.

Because MR estimates are less likely to be affected by confounding and better reflect the exposure to the risk factor of interest across the lifespan, MR studies can also refine the knowledge on the actual effect size of a causal effect or identify that causal effects may be substantially different and relationships more complex than previously thought. Both scenarios have been reported, for instance when conducting MR for excessive body weight and risk of different cancer types.[10, 14]

To illustrate the importance of triangulation using MR for the identification of cancer causes we have chosen to present in detail the example of excessive body weight and obesity-related cancers.

Example of triangulation: the cancer burden of excessive body weight

In a comprehensive review of the observational and mechanistic data, thirteen cancer types were classified as demonstrating 'sufficient' evidence of an increased risk caused by excess body fatness.[15] Based on cancer-specific relative risk estimates observed in cohort studies, the population fraction of cancer that can be attributed to overweight and obesity has been estimated at 6%, a fraction that is likely to increase in the coming decades.[16] These estimates would also make excessive body fatness the second most important cause of cancer in high-income countries, after tobacco.

Whilst MR studies have confirmed the presence and direction of BMI-risk association for most of these cancers,[10, 12, 17-20] the case of body adiposity can also serve as an example of the use of MR in refining our understanding of the strength of effect by which specific risk factors influences a disease. In particular, risk estimates for body fatness and cancer are typically obtained using direct BMI measurements in large cohort studies. However, because such studies typically rely on measures of attained BMI at one time-point, which may poorly reflect lifetime body fatness, they may underestimate the impact of

elevated BMI on cancer risk. Similarly, reverse causation and residual confounding may influence the magnitude or direction of risk estimates for some cancers. In an MR framework, estimates for the cancer risk associated with BMI can be obtained by using genetic measures of BMI. Beyond the benefit of eliminating bias by reverse causation and its relative insensitivity of bias by confounding, MR may provide more accurate estimates of the underlying influence of obesity on cancer risk under the assumption that genetic proxies of BMI reflect differences in BMI across the lifespan, in contrast to direct BMI measures taken at one time-point.

To exemplify this concept we compared the genetic MR estimates of the association between BMI and cancer risk with that of classical cohort studies for eight common cancer types that have been linked to obesity.

Methods

An MR estimate for the effect of BMI on each cancer type was obtained using: (i) a genetic instrument for BMI based on a weighted panel of SNPs known to influence BMI; and (ii) results from a GWAS for each specific cancer.

The genetic instrument for BMI was obtained from a meta-analysis of the GIANT consortium and UK Biobank, comprising genome-wide data on 700 000 study participants with BMI measures, resulting in 714 independent genetic variants that were associated with BMI at a p-value of 5×10^{-8} (eMethods). A weighted score of these genetic variants explained 9.2% of BMI variance in the discovery sample combining the results of the GIANT consortium with that of UK Biobank.

Of the 13 cancer sites that have been linked to obesity, results from large GWAS were available for 7, including cancer of the colorectum, kidney, pancreas, ovary, endometrium, breast, and esophagus (eTable 1). These 7 cancer sites account for approximately 80% of all cancers that are linked to obesity in the UK. We also included lung cancer because a

potential association of elevated BMI with increased risk of lung cancer has been recently suggested.[11]

Based on the 714-SNP instrument for BMI (eTable 2), we applied two-sample MR (eMethods). Two-sample MR uses summary genetic association data from two independent samples, representing: a) the genetic variant-risk factor associations; and b) the genetic variant-outcome associations. We employed the likelihood-based approach as the main method of conducting the MR analysis. The major limiting issue of this method is the potential violation of the assumption of no horizontal pleiotropy. To investigate the validity of this and other MR assumptions we performed several sensitivity analyses, including the weighted median and mode estimators, the MR-Egger test, and a leave-one-out analysis (eMethods). For cancers of the breast, endometrium and esophageal adenocarcinoma, the existing GWAS data were not available but two-sample MR estimates based on these GWAS data and genetic instruments for BMI were instead retrieved from the latest published studies (eTable 1). We considered odds ratios as estimates of relative risks (RR).

The RR for BMI from observational studies were extracted from the World Cancer Research Fund (WCRF) Third Expert Report, except for lung cancer that used a large pooled analysis. [21, 22]

MR confirms and refines the causal effect of BMI on 6 obesity-related cancers

For 6 out of 7 evaluated cancer sites that were previously identified as being positively associated with BMI in the WCRF report, we also observed risk increases based on the MR analysis (Fig 2). However, for each of these 6 cancer sites, the MR estimate for a 5-unit increment in BMI was notably higher than the WCRF estimate, being approximately twofold higher for cancers of the kidney (e.g. MR RR_{5BMI}: 1.59, 95% CI: 1.45-1.74; Obs. RR_{5BMI}: 1.30, 95% CI: 1.25-1.35), endometrium, ovary and esophageal adenocarcinoma, and more than fourfold for pancreatic (e.g. MR RR_{5BMI}: 1.47, 95% CI: 1.31-1.66; Obs. RR_{5BMI}: 1.10, 95% CI: 1.07-1.14) and colorectal cancer. Sensitivity analysis for the MR results suggested estimates

even higher than main estimates from the likelihood-based approach, except for the weighted mode estimate for colorectal cancer, and both Egger regression and the leave-one-out analysis indicate that the MR results were unlikely to be biased due to pleiotropic effects (eTable 3, eTable 4 and eFigure 1A-1D).

The difference in relative risk estimates between the MR and traditional observational studies may to some extent be explained by the avoidance of bias by reverse causation and residual confounding in the MR analyses. More importantly, because the MR estimates rely on genetic proxies of BMI, the primary difference may be that it better reflects the risk implications of accumulated exposure to elevated BMI across the lifespan, thereby correcting for the exposure measurement error pertaining to single time-point BMI measurements.

Based on these MR-based results, the cancer burden explained by elevated BMI for 6 of the 8 cancers evaluated here alone, would be almost 8% in high income countries, as opposed to 3% based on previous estimates for these cancers, suggesting the cancer burden has been substantially underestimated for the majority of obesity-related cancers.

MR results contrast with observational associations of BMI with breast and lung cancers

For breast and lung cancer, the MR estimates of the association between BMI and risk were not consistent with previous observational studies (Fig 3). In particular, previous observational studies have reported a modest protective effect of elevated BMI for pre-menopausal breast cancer and a modest increase in risk for post-menopausal breast cancer. Conversely, the MR analysis suggested that elevated BMI would be associated with a decreased risk of both breast cancer subtypes (Fig 3 and eTable 5). Similarly, MR estimates for lung cancer overall and for two prominent subtypes (squamous cell and small cell) indicated an increased risk associated with high BMI, which contrasts with previous

observational evidence of an inverse association between overweight and lung cancer (Fig 3).

The positive association between BMI and lung cancer risk is in sharp contrast to that of traditional observational studies, and may in part be explained by an effect of BMI on smoking patterns. Indeed, recent genetic evidence indicates that increased BMI leads to a greater likelihood to initiate smoking, as well as a greater intensity of smoking behaviour.[23] This would imply that obesity is an unrecognised yet important causal factor in lung cancer, albeit mediated by its influence on smoking patterns.

For breast cancer, the MR analysis suggested a clear protective effect for increased BMI for both pre-menopausal and post-menopausal diseases, thus contrasting sharply to traditional observational studies that tend to show positive associations with post-menopausal breast cancer. These results suggest a more complex association between BMI and breast cancer than previously assumed. Emerging evidence of the importance of weight gain during adulthood, and not weight per se, as a determinant of breast cancer risk could provide a possible explanation for these discrepant results,[24] in that transitioning to being overweight or obese later in life may be an important risk factor for post-menopausal breast cancer. Several other MR studies, including the studies on breast size by Ooi et al. and on sex hormone binding globulin by Dimou et al. that are included in this issue of *International Journal of Epidemiology,* have investigated causality of breast cancer putative risk factors that add information about the potential role of BMI, though breast cancer represents a clear example where MR studies have highlighted a more complex relation with BMI that previously assumed, warranting a need for further research on this topic (Dimou and Ooi).

Conclusions

We provide examples where MR studies have advanced our understanding of cancer aetiology, primarily by complementing evidence attained through traditional observational studies. Examples include studies where MR have either provided confirmatory evidence for

putative risk factors or refined risk effect estimates, by highlighting more complex risk factor-disease relations, as well as identification of novel risk factors.

In particular, MR studies have proven instrumental in improving our understanding of obesity in cancer etiology. Results from large observational studies, which are usually based on a one-off measure of BMI, indicate that overweight and obesity account for approximately 6% of all cancers in high income countries. Our systematic analysis of genetic data for 6 of 7 obesity-linked cancers indicated that the estimates of cancer burden associated with BMI may be considerably underestimated. In addition, for lung and breast cancer the discrepancy between the MR and previous results clearly highlights a more complex relation that requires further elucidation.

It is clear that important knowledge gaps remain in our understanding of cancer aetiology in general, and changes in cancer incidence throughout the world indicate that many novel cancer risk factors remain to be established.[25] With important advances in available genetic data in large study populations, MR is likely to have an important role in furthering our understanding of cancer incidence in the years to come. Important contributions may come from the emerging availability of metabolomics and proteomics data with improved understanding of their genetic determinants. For instance, given the relatively simple genetic architectures of circulating proteins and metabolites compared to other more complex risk factors,[26, 27] two-sample MR may be a highly cost-efficient approach to agnostically identify metabolite and protein biomarkers that may have an etiological role in cancer development. Another largely unexploited application of MR methodology is in studies of cancer prognosis. Indeed the identification of prognostic factors may point, not only to important therapeutic targets, but also to new potential causes and differences in etiology between subgroups. There are examples of MR studies investigating outcomes such as cancer mortality[28] or survival time.[29] However, the contribution of these types of studies has been limited by lack of statistical power due to the intrinsic need of a larger pool of cancer cases for well-powered MR analyses. Efforts to fill the gap in collection of follow-up

data of cancer patients in genotyping studies have the potential to transform our understanding of cancer survival.

Many studies have already shown that MR is an important complement to traditional observational studies in evaluating cancer aetiology, offering opportunities to both identify new hypotheses and confirm the causal effects of putative risk factors. Future studies seeking to identify causes of cancers will undoubtedly benefit from the insight gained from MR.

Acknowledgments

Funding/Support: This study was supported by the World Cancer Research Fund International (grant number 2014/1193) and Cancer Research UK (grant number C18281/A19169).

Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Conflict of Interest: None declared.

References

1. Wild CP, Espina C, Bauld L, et al. Cancer Prevention Europe. Molecular oncology. 2019;13(3):528-34.
2. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology. 2003;32(1):1-22.
3. Fanidi A, Carreras-Torres R, Larose TL, et al. Is high vitamin B12 status a cause of lung cancer? International journal of cancer. 2018.
4. Pierce BL, Kraft P, Zhang C. Mendelian randomization studies of cancer risk: a literature review. Current epidemiology reports. 2018;5(2):184-96.
5. Dong J, Gharahkhani P, Chow WH, et al. No Association Between Vitamin D Status and Risk of Barrett's Esophagus or Esophageal Adenocarcinoma: A Mendelian Randomization Study. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2019.

6. He Y, Timofeeva M, Farrington SM, et al. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. BMC medicine. 2018;16(1):142.
7. Jiang X, Dimou NL, Al-Dabhani K, et al. Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study. International journal of epidemiology. 2018.
8. Kachuri L, Saarela O, Bojesen SE, et al. Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers. International journal of epidemiology. 2018.
9. Wang X, Dai JY, Albanes D, et al. Mendelian randomization analysis of C-reactive protein on colorectal cancer risk. International journal of epidemiology. 2018.
10. Carreras-Torres R, Johansson M, Gaborieau V, et al. The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study. Journal of the National Cancer Institute. 2017;109(9).
11. Carreras-Torres R, Johansson M, Haycock PC, et al. Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. PloS one. 2017;12(6):e0177875.
12. Johansson M, Carreras-Torres R, Scelo G, et al. The influence of obesity-related factors in the etiology of renal cell carcinoma-A mendelian randomization study. PLoS medicine. 2019;16(1):e1002724.
13. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International journal of epidemiology. 2016;45(6):1866-86.
14. Shu X, Wu L, Khankari NK, et al. Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. International journal of epidemiology. 2018.
15. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer--Viewpoint of the IARC Working Group. The New England journal of medicine. 2016;375(8):794-8.
16. (2018) CRU. When could overweight and obesity overtake smoking as the biggest cause of cancer in the UK.
17. Gao C, Patel CJ, Michailidou K, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. International journal of epidemiology. 2016;45(3):896-908.
18. Painter JN, O'Mara TA, Marquart L, et al. Genetic Risk Score Mendelian Randomization Shows that Obesity Measured as Body Mass Index, but not Waist:Hip Ratio, Is Causal for Endometrial Cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2016;25(11):1503-10.
19. Dixon SC, Nagle CM, Thrift AP, et al. Adult body mass index and risk of ovarian cancer by subtype: a Mendelian randomization study. International journal of epidemiology. 2016;45(3):884-95.
20. Thrift AP, Shaheen NJ, Gammon MD, et al. Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a Mendelian randomization study. Journal of the National Cancer Institute. 2014;106(11).
21. World Cancer Research Fund/American Institute for Cancer Research Continuous Update Project Expert Report 2018 Body fatness and weight gain and the risk of cancer Available at dietandcancerreportorg.
22. Yu D, Zheng W, Johansson M, et al. Overall and Central Obesity and Risk of Lung Cancer: A Pooled Analysis. Journal of the National Cancer Institute. 2018;110(8):831-42.
23. Carreras-Torres R, Johansson M, Haycock PC, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. Bmj. 2018;361:k1767.

24. Rosner B, Eliassen AH, Toriola AT, et al. Weight and weight changes in early adulthood and later breast cancer risk. International journal of cancer. 2017;140(9):2003-14.
25. Brennan P, Wild CP. Genomics of Cancer and a New Era for Cancer Prevention. PLoS genetics. 2015;11(11):e1005522-e.
26. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.
27. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. Nature genetics. 2014;46(6):543-50.
28. Ong JS, Gharahkhani P, An J, et al. Vitamin D and overall cancer risk and cancer mortality: a Mendelian randomization study. Human molecular genetics. 2018;27(24):4315-22.
29. Guo Q, Burgess S, Turman C, et al. Body mass index and breast cancer survival: a Mendelian randomization analysis. International journal of epidemiology. 2017;46(6):1814-22.

Figure legends

Fig 1. The contribution of Mendelian randomization (MR) to the identification of cancer causes. MR studies have the potential to contribute in different ways and examples of the different scenarios have been reported (white boxes).

Fig 2. WCRF (circles) and MR (squares) relative risks for the association between a 5-unit BMI increase and cancer risk by cancer site.

Fig 3. Observational (circles) and MR (squares) relative risks for the association between a 5-unit BMI increase and risk of breast and lung cancer.