Smoking, alcohol, diabetes, obesity, socioeconomic status and the risk of colorectal cancer in a population-based case-control study

Terry Boyle[1,2,3], Lin Fritschi[4], Seyed Mehdi Tabatabaei[5], Kathrin Ringwald[3], Jane S Heyworth[5]

1. Cancer Control Research, British Columbia Cancer Agency, Vancouver, British Columbia, Canada
2. School of Population and Public Health, The University of British Columbia, Vancouver, British Columbia, Canada
3. Harry Perkins Institute of Medical Research, The University of Western Australia, Perth, Western Australia, Australia
4. School of Public Health, Curtin University, Perth, Western Australia, Australia
5. School of Population Health, The University of Western Australia, Perth, Western Australia, Australia

Corresponding author:

Terry Boyle

Cancer Control Research

BC Cancer Research Centre

675 West 10th Avenue

Vancouver BC, V5Z 1L3, CANADA

Phone: +1 604-675-8066

E-mail: tboyle@bccrc.ca

Running Head: Smoking, alcohol, diabetes, obesity, socioeconomic status and the risk of colorectal cancer

Word Count: 3785

Keywords: case-control study; smoking; alcohol; diabetes; obesity; socioeconomic status; colorectal cancer; participation; response bias; survey methodology

ABSTRACT

Purpose: Although previous research has identified factors that may determine willingness to participate in research, relatively few studies have attempted to quantify the impact non-participation may have on exposure-disease associations. The aims of this study were to: (a) investigate the associations between smoking, alcohol, diabetes, obesity and socioeconomic status and the risk of colorectal cancer in a case-control study (59.7% and 47.2% response fractions among cases and controls respectively); and (b) perform sensitivity analyses to examine the possible influence of non-participation.

Methods: Logistic regression was used to estimate the exposure-disease associations. We then investigated the associations between various demographic and health factors and the likelihood that an individual would participate in the case-control study, and then performed two sensitivity analyses (sampling weights and multiple imputation) to examine whether non-participation bias may have influenced the exposure-disease associations.

Results: The exposures alcohol, smoking and diabetes were associated with an increased risk of colorectal cancer. We found some differences between cases and controls when examining the factors associated with participation in the study, and in the sensitivity analyses the exposure-disease associations were slightly attenuated when compared with those from the original analysis.

Conclusion: Non-participation may have biased the risk estimates away from the null, but generally not enough to change the conclusions of the study.

List of abbreviations: AOR, adjusted odds ratio; BMI, body mass index; CI, confidence interval; CRC, colorectal cancer; OR, odds ratio; RR, relative risk WA, Western Australia; WABOHS, Western Australian Bowel Health Study.

INTRODUCTION

Response rates in epidemiological studies have declined over the last few decades [1-3]. Non-participation in epidemiologic studies reduces precision, increases costs and may introduce selection bias, if the factors related to non-participation are directly or indirectly associated with both the exposure and the outcome. Although previous research has identified several factors that may determine willingness to participate in research studies, including sex, socioeconomic status and age [1], relatively few studies have attempted to quantify what impact non-participation may have on exposure-disease associations [4, 5].

In this study we used data from a case-control study and linked administrative datasets to: (a) determine the associations between demographic, socioeconomic and medical factors and the likelihood of participation in a case-control study of colorectal cancer (CRC); (b) determine the associations between various lifestyle-related exposures (alcohol intake, tobacco use, body mass index (BMI) and diabetes), socioeconomic status and the risk of CRC; and (c) undertake two sensitivity analyses (weighting and multiple imputation) to examine whether non-participation bias may have influenced the associations found. Diabetes, smoking, alcohol and obesity are established risk factors for CRC [6-12], while the evidence concerning socioeconomic status and CRC risk is inconsistent [13].

MATERIAL AND METHODS

The Western Australian Bowel Health Study (WABOHS)

The Western Australian Bowel Health Study (WABOHS) was a population-based case control study of CRC that was conducted in Western Australia (WA) in 2005-2007 [14, 15]. Cases were

males and females, aged between 40-79 years at the time of diagnosis, with a histopathologically confirmed incident CRC that was diagnosed and notified to the WA Cancer Registry between June 2005 and August 2007. Controls with no prior history of CRC were randomly selected from the WA electoral roll every three months to coincide with the recruitment of the cases (i.e., incidence-density sampling). They were frequency-matched for sex and five-year age group, based on the approximate distribution of five-year age-group and sex among incident cases of CRC in Western Australia in 2002. A total of 1538 eligible cases and 2163 eligible controls were invited to take part in the WABOHS, of whom 918 cases (59.7%) and 1021 controls (47.2%) participated. Ethics approval for the WABOHS was obtained from ethics committees at The University of WA and the WA Department of Health.

Exposure Measurement

Data from the cases and controls who participated in the WABOHS were collected via two self-administered questionnaires. Data on alcohol consumption and diet 10 years ago were obtained using a food frequency questionnaire [16]. Questions regarding alcohol intake included the number of glasses per day of total alcohol and frequency of consumption of beer, wine, fortified wines and spirits and liquors. Weekly intake of alcohol (grams) was calculated from these questions. Total alcohol intake was categorised as less than 1, 1 to 6.9, 7 to 20.9, and 21 or more standard drinks per week, where one standard drink is equivalent to 10 grams of alcohol. Full-strength beer intake and wine intake were classified according to frequency of consumption, and were categorised as less than one day per week, one to four days per week, and five or more days per week. The food frequency questionnaire also allowed estimates of energy intake (kilojoules per day, excluding alcohol) ten years ago. Participants were also asked to complete a lifetime recreational physical

activity questionnaire, from which we created variables for lifetime moderate-intensity physical activity, lifetime vigorous-intensity physical activity and lifetime total physical activity (moderate- and vigorous-intensity physical activity combined) [14].

Based on self-reported data, participants were categorized as having diabetes, high blood sugar level, or neither. Participants were also asked to record their height, as well as their weight one year ago, at ages 20, 40 and 60 years, and their maximum weight (excluding during pregnancy). Based on these data, BMI at ages 20 and 40 was calculated and categorized according to World Health Organization guidelines. For smoking, participants were asked if they had smoked more than 100 cigarettes, pipes or cigars in their life. If they answered yes, they were asked a series of questions, including the number of cigarettes per day they usually smoked, and the number of years in total that they had smoked. The following smoking metrics were used for analyses: never smoker/former smoker/current smoker at the time of completing the survey; and pack-years of smoking.

Information on Participants and Non-Participants

All eligible cases and controls were linked to the electoral roll and the hospital morbidity data system by Data Linkage Branch within the WA Department of Health to obtain five-year age-group, sex, residential postcode, and hospitalisations between 2000-2007. The probabilistic matching procedures used to link individuals have been estimated to be 99.9% accurate [17]. We were able to consider the effect of following factors on response to the case-control study: age; sex; geographic accessibility/remoteness; socioeconomic status; comorbidity; hospitalisation at the time of invitation; and hospitalisation in the last five years.

Socioeconomic status and accessibility/remoteness were both based on residential postcode. Participants were categorised into five groups of socioeconomic status using deciles of the Index of Relative Socio-Economic Disadvantage from the Socio-Economic Indexes for Areas [18]. Accessibility/remoteness was categorised as highly accessible, accessible or moderately accessible, and remote or very remote, based on the Accessibility and Remoteness Index of Australia [19]. Comorbidity was based on hospitalisation data and was determined using the Charlson Comorbidity Index [20]. For controls we included all hospitalisations (for any reason) in the past five years prior to the date of invitation to the study, while for cases we included all hospitalisations (for any reason) in the five years prior to three months before the date of CRC diagnosis. The Charlson Comorbidity Index was categorised as zero, one, two, or three or more. In addition, participants were classified as having been hospitalised for any reason or not during this five-year period, and they were also classified as having been in or not in hospital at or around the time they were invited to take part in the study (i.e., in the four months preceding the date the study invitation letter was sent). Finally, participants were classified as having been hospitalised or not in the five years prior to being invited to take part in the study (or five years prior to three months before the date of CRC diagnosis for cases) for each of cardiovascular disease, diabetes/renal disease, liver disease, cancer, or chronic obstructive pulmonary disease. These variables were only used in the imputation procedure in the second sensitivity analysis.

Histopathological information about the CRC of all eligible cases was obtained from the WA Cancer Registry. Cancer site was classified as proximal colon, distal colon, and rectum. Grade was categorised as low, moderate, high and unknown, while CRC surgery type was categorised as total/subtotal colectomy/proctocolectomy, rectosigmoidectomy/proctectomy, anterior

resection/hemicolectomy, limited excision, and other/unknown procedures. Stage data are not routinely collected so were unavailable.

Data analysis

Associations between the risk of CRC and smoking, alcohol intake, BMI, diabetes status, and socioeconomic status

For the original analysis, logistic regression was used to estimate the associations between the risk of CRC and smoking status, total alcohol intake, BMI at 20 years of age and 40 years of age, diabetes status, and socioeconomic status. All models were adjusted for age-group and sex because of the frequency matching. Smoking, alcohol, obesity, diabetes and socioeconomic status were all mutually adjusted for each other, and the multivariable model was additionally adjusted for lifetime vigorous physical activity and energy intake. Subsequent models replaced total alcohol intake with beer consumption and wine consumption, and smoking status with pack-years.

BMI at ages 20 and 40 years was missing for approximately 9% of participants, so these two variables were imputed using multiple imputation by chained equations [21]. The imputation procedure included height, all the weight variables listed above, and all other exposure, covariate and outcome variables in the original analysis. Ten datasets were added in the imputation procedure. There were no meaningful differences between the results of these imputed analyses and the results of analyses from a dataset containing only the 1618 participants with complete data (data not shown).

Factors associated with participation in the WABOHS

Modified Poisson regression was used to estimate the relative risk (RR) of participation associated with sex, age, socioeconomic status, remoteness/accessibility, hospitalization in the last five years or at the time of invite, comorbidity, and, for cases only, cancer grade, cancer site and surgery type [22]. Modified Poisson regression is an alternative to logistic regression when the outcome is common, and provides a direct estimate of relative risk [22]. Cases and controls were analysed separately. All variables were mutually adjusted. Trend tests were conducted by entering ordinal categorical variables into the model as continuous variables. We tested whether there were any significant interactions (i.e., $p<0.05$) between sex or age and the other variables, or between each other, however none were observed.

Sensitivity analyses to examine the possible influence of non-participation

We performed two sensitivity analyses - inverse probability weighting and multiple imputation - to investigate the possible influence that non-participation may have had on the associations between the risk of CRC and smoking, alcohol intake, BMI, diabetes status, and socioeconomic status. Both sensitivity analyses rely on the assumption that the missing data are missing at random (i.e., "any systematic difference between the missing values and the observed values can be explained by differences in observed data" [23].

In the first sensitivity analysis, the modified Poisson regression models outlined above were used to predict the likelihood that each eligible individual would take part in the WABOHS (separately for cases and controls). The inverse of the predicted likelihood of participation was then used to weight the data in the original logistic regression model examining the associations

between the risk of CRC and smoking, alcohol intake, BMI, diabetes status, and socioeconomic status. This analysis only involved individuals who took part in the case-control study, and participants who were predicted to be less likely to take part in the case-control study were weighted more heavily in the exposure-disease analysis.

The second sensitivity analysis involved using multiple imputation by chained equations to impute the 'missing' exposure and covariate data for the non-participants [21, 24, 25]. Multiple imputation involves the creation of multiple different plausible imputed datasets, thus allowing for uncertainty about the missing data, then combining the results from each dataset [23]. All the available participant/non-participant variables were included in the multiple imputation procedure. The outcome variable (case or control) was expanded to include CRC site (i.e., control, proximal colon cancer case, distal colon cancer case, rectal cancer case). We also included variables that indicated if a person had been hospitalised for each of cardiovascular disease, diabetes/renal disease, liver disease, cancer, and chronic obstructive pulmonary disease, as we thought that including these variables may result in better prediction of smoking, alcohol consumption, obesity and diabetes in the non-participants. The exposure and covariate data from the participating cases and controls were also included in the imputation procedure, as were total and moderate lifetime physical activity, height, weight (one year prior to the time of study completion, at age 60 years, and maximum), and education level. Fifty imputations were performed. Height, weight, alcohol intake, pack-years of smoking were imputed as continuous variables, and categorised (as body mass index for height and weight) following the imputation procedure. Energy intake was imputed as a continuous variable, lifetime physical activity was imputed as an ordinal variable, and diabetes status was imputed as a nominal variable. Following

the imputation procedure, logistic regression was used to investigate the exposure-disease associations in this imputed dataset, with the same model used in the original analysis.

As information on socioeconomic status was available for all participants, we also calculated the age- and sex-adjusted odds ratio for the association between socioeconomic status and CRC risk using only the participants included in the original analysis, and then again with the non-participants in addition to the participants included in the original analysis.

Linked data were not available for three cases so these participants were not included in any analyses, giving a total sample size of 1536 cases and 2163 controls for the analysis investigating the associations between demographic, socioeconomic and medical factors and the likelihood of participation in the case-control study. A total of 82 participants were excluded from the exposure-disease analyses as they reported very low energy intake (fewer than 500 kcal/day for women and 800 kcal/day for men) or very high energy intake (more than 3500 kcal/day for women and 4000 kcal /day for men) [26]. A further 10 participants were excluded from the exposure-disease analyses due to missing data on multiple covariates, giving a sample size of 1844 (872 cases and 972 controls) in the original analysis and the weighted sensitivity analysis and 3607 (1493 cases and 2114 controls) in the imputed sensitivity analysis. Stata 13.1 (StataCorp, Texas, USA) was used for all analyses.

RESULTS

Factors associated with participation

For cases, participation was more likely among males, people from areas with higher socioeconomic status, and people with no comorbidity (Table 1). For controls, participants aged 50 years and older and participants who had been hospitalised in the last five years were more likely to participate, while those from the most disadvantaged areas and with greater comorbidity were less likely to take part.

Associations between alcohol, smoking, BMI and diabetes and the risk of CRC

The distribution of the exposures and covariates in the participants in the case-control study are shown in the Table in Online Resource 1. For total alcohol consumption, consumption of 21 or more standard drinks per week ten years ago was associated with an elevated risk of CRC compared with consumption of less than one standard drink per week, although this was not statistically significant (Adjusted Odds Ratio (AOR)=1.25, 95% Confidence Interval (CI)=0.94-1.67) (Table 2). Participants who reported drinking beer on 5 or more days per week ten years ago had a significantly increased risk of CRC compared with those who reported drinking beer less than once per week (AOR=1.50, 95% CI=1.90-2.07). For wine consumption ten years ago, participants who drank wine on one or more days per week had a non-significant increased risk of CRC compared with participants who reported drinking wine less than once per week. The risk estimates from the weighted sensitivity analyses for total alcohol, beer consumption and wine consumption were similar to those from the original analysis, while the risk estimates for beer consumption and wine consumption attenuated by approximately 10% in the imputed sensitivity analysis.

Former smokers had a 24% higher risk of CRC than never smokers (AOR=1.24, 95% CI=1.01-1.53), but current smoking was not associated with CRC risk (Table 2). A significant dose-response relationship was found between pack-years and CRC risk (P_{trend}=0.049), with participants who had smoked for the equivalent of 20 or more pack-years having a 27% higher risk of CRC than never smokers (AOR=1.26, 95% CI=1.00-1.60). As with alcohol consumption, similar risk estimates were observed in the weighted sensitivity analysis, and attenuated (by approximately 10%) risk estimates were observed in the imputation sensitivity analysis.

BMI at ages 20 and 40 years were not significantly associated with CRC risk in the original, weighted or imputed analyses (Table 2).

Participants with diabetes had a 74% increased risk of CRC compared to participants without diabetes or high blood sugar levels (AOR=1.74, 95% CI=1.28-2.35) (Table 2). A significant association remained in the weighted and imputed sensitivity analyses, although the risk estimates decreased by 4% and 21% respectively.

In the age- and sex-adjusted original analysis, residing in the most disadvantaged areas was associated with non-significant 39% increased risk of CRC compared with participants in the least disadvantaged areas (OR=1.39, 95% CI=0.96-2.01) (Table 3). Inclusion of the non-participants led to a 17% decrease in the risk estimate (OR=1.15, 95% CI=0.79-1.68). The fully adjusted risk

estimates in the weighted and imputed sensitivity analyses were also much closer to the null than those from the age- and sex-adjusted and fully adjusted original analyses.

DISCUSSION

In this population-based case-control study we found that alcohol consumption (particularly beer consumption), smoking and diabetes were significantly associated with an increased risk of CRC. Although we found some differences between cases and controls when examining the factors associated with participation in the study, weighting the exposure-disease analysis according to the modelled likelihood that a person would take part in the study had minimal influence on the results, although the risk estimates were slightly attenuated. Imputing the 'missing' exposure and covariate data for the non-participants resulted in greater attenuation of the risk estimates, but generally not enough to change the conclusions of the study, with the exception of the results concerning wine consumption and tobacco use.

Our results concerning CRC risk and smoking, alcohol and diabetes are similar to previous findings, with recent meta-analyses providing strong evidence that these exposures are associated with increased risk [11, 27-29]. A possible association between socioeconomic status and CRC risk was found in our original analysis, but the risk estimate for the lowest versus highest socioeconomic status attenuated in the weighted and imputed sensitivity analyses, as well as in age- and sex-adjusted analyses including all the participants and non-participants. Previous research regarding socioeconomic status and CRC risk is inconsistent, with studies from North America generally finding a higher risk of CRC among lower socioeconomic groups and studies

from Europe and Australasia generally finding a lower risk among lower socioeconomic groups [13].

The results of the sensitivity analyses are broadly consistent with the results of other case-control studies that have investigated the influence of selection bias, which have generally found that non-participation does not have a large influence on risk estimates obtained in exposure-disease analyses. For example, two studies that used sampling weights based on data available for all participants and non-participants obtained similar risk estimates in weighted and unweighted analyses of the associations between renal cell carcinoma risk and smoking and hypertension respectively [30, 31]. Wigertz et al. linked non-participating cases and controls to central registries to obtain information about socioeconomic factors such as income level and working status, and their analyses indicated that non-participation did not have a large influence on the associations between these socioeconomic factors and the risk of brain tumour [32]. However Lopez et al. using sampling weights, found that non-participation may have biased some of the results in a case-control study of periodontitis [33]. Pandeya et al used data from a national health survey to impute smoking status and BMI for non-participating controls, and found modest changes in risk estimates for the association between these exposures and oesophageal and ovarian cancers [5]. Several other studies that have collected exposure data from a subset of non-participants to examine the possible influence of non-participation have also generally found that selection bias resulting from the variables considered was not likely to have had a large effect on risk estimates [34-37].

Although it is generally thought that females are more likely than males to take part in research studies [1], we found that male cases were more likely to participate in the case-control study than female cases. This is in contrast to previous studies of cancer cases, which have found no difference in response by sex [32, 38-40], or that female cancer cases are more likely to participate [41, 42]. One study also found that male cancer patients were more likely to participate than female cancer patients however [43], and a further study in colon cancer cases found a higher response among males than females [44]. Our finding that age did not influence the likelihood of participation among cancer cases is consistent with previous literature [32, 36, 39, 43, 45, 46], although some studies of cancer patients have found that younger age groups (younger than 30 to 40 years) and/or older age groups (older than 80 years) may be less likely to participate [38, 41, 42, 47]. In keeping with previous research, we found that higher socioeconomic status was associated with increased response in both cases and controls [1, 40, 41, 48].

This study had several limitations that should be taken into account. Recall bias is a possible explanation for the positive associations observed in this study, although our results are consistent with those of cohort studies, which do not suffer from this type of bias. Another limitation of this study was the potential for misclassification in the assessment of diabetes, BMI, alcohol consumption, smoking and other risk factors, as participants were asked to recall these data from 10 years earlier or at different ages over their lifetime.

With regards to the sensitivity analyses, although we were able to obtain comorbidity data for both participants and non-participants, this was limited to hospitalisations so did not include any illnesses that may have potentially influenced participation but not required hospitalisation. We

also lacked information on several factors that may influence participation in epidemiological studies, such as ethnicity, marital status and education [36, 40-42, 44], as well as information on topic-specific factors such as cancer stage, family history of CRC and cancer screening attendance. Not having this information limited our ability to determine what factors influenced participation in the study and to predict the likelihood of participation. None of the factors we investigated were strong predictors of participation, and this may be an explanation for the similar results obtained in the original and sensitivity analyses. Our ability to predict the missing exposure data is a further limitation. Although sex, comorbidity and socioeconomic status are all likely to be associated with the exposures investigated in the study, and we also included variables related to hospitalisations for various chronic diseases as we thought that these were likely to be associated with smoking, alcohol consumption, obesity and diabetes, it is possible that the data that were available for both the participants and non-participants did not allow for good prediction of the missing exposure variables, particularly as we were imputing large blocks of missing data (i.e., unit missing data) rather than imputing a small number of variables (i.e., item missing data). Finally, both of the sensitivity analyses rely on the assumption that the missing data are missing at random, however it is likely that people who did not participate were not a random subset of the source study population, so this assumption may not hold true. Although we included several variables that were associated with missingness, it is possible that the variables we included in the weighting and imputation procedures may not have accounted for any systematic difference between the missing values and the observed values. If this is the case, the missing at random assumption may not have been plausible and the results from the weighted and imputed analyses may not be valid.

In conclusion, in this study we found that the factors associated with participation differed between cases and controls, but the results of two sensitivity analyses indicate that any possible selection bias generally may not have a large influence on the associations between various lifestyle-related exposures, socioeconomic status and CRC risk.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Kieran McCaul for statistical advice, the people who generously gave their time to participate in this research, the clinicians who gave permissions to approach their patients, the WA Cancer Registry staff (Timothy Threlfall in particular) for assistance with case ascertainment and recruitment, and members of the WABOHS team (Cameron Platell, Kieran McCaul, David Crawford, Cassandra Clayforth, Clare Tran, Jenny Landrigan, Beatriz Cuesta-Briand) for study management and collection of data used in this study. This work was supported by the Australian National Health and Medical Research Council (Project Grant #353568, Fellowship #1072266 to TB, and Fellowship 37614900 to LF), the Canadian Institutes of Health Research (Fellowship #300068 to TB), the Michael Smith Foundation for Health Research (Fellowship #5553 to TB), and the Izaak Walton Killam Memorial Fund for Advanced Studies (Honorary UBC Killam Postdoctoral Research Fellowship to TB).

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Galea S, Tracy M. (2007) Participation rates in epidemiologic studies. Ann Epidemiol 17: 643-653

2. Hartge P. (2006) Participation in population studies. Epidemiology 17: 252-254

3. Morton LM, Cahill J, Hartge P. (2006) Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol 163: 197-203

4. Keeble C, Barber S, Law GR, Baxter PD. (2013) Participation bias assessment in three high-impact journals. SAGE Open 3: 1-5

5. Pandeya N, Williams GM, Green AC, Webb PM, Whiteman DC. (2009) Do low control response rates always affect the findings? Assessments of smoking and obesity in two Australian case-control studies of cancer. Aust N Z J Public Health 33: 312-319

6. Larsson SC, Orsini N, Wolk A. (2005) Diabetes mellitus and risk of colorectal cancer: a meta-analysis. J Natl Cancer Inst 97: 1679-1687

7. World Cancer Research Fund/American Institute for Cancer Research. (2007) Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR.

8. Cho E, Smith-Warner SA, Ritz J, et al. (2004) Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. Ann Intern Med 140: 603-613

9. Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M. (2009) The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer 125: 171-180

10. Moskal A, Norat T, Ferrari P, Riboli E. (2006) Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. Int J Cancer 120: 664-671

11. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. (2008) Smoking and colorectal cancer: a meta-analysis. JAMA 300: 2765-2778

12. Liang PS, Chen TY, Giovannucci E. (2009) Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. Int J Cancer 124: 2406-2415

13. Aarts MJ, Lemmens VEPP, Louwman MWJ, Kunst AE, Coebergh JWW. (2010) Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. Eur J Cancer 46: 2681-2695

14. Boyle T, Heyworth J, Bull F, McKerracher S, Platell C, Fritschi L. (2011) Timing and intensity of recreational physical activity and the risk of subsite-specific colorectal cancer. Cancer Causes Control 22: 1647-1658

15. Clapin HF, Fritschi L, Iacopetta B, Heyworth JS. (2012) Dietary and supplemental folate and the risk of left- and right-sided colorectal cancer. Nutr Cancer 64: 937-945

16. Hodge A, Patterson AJ, Brown WJ, Ireland P, Giles G. (2000) The Anti Cancer Council of Victoria FFQ: relative validity of nutrient intakes compared with weighed food records in young to middle-aged women in a study of iron supplementation. Aust N Z J Public Health 24: 576-583

17. Holman CDAJ, Bass AJ, Rouse IL, Hobbs MST. (1999) Population-based linkage of health records in Western Australia: Development of a health services research linked database. Aust N Z J Public Health 23: 453-459

18. Australian Bureau of Statistics. (2008) Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia. Catalogue Number 2033.0.55.001. Canberra, Australia: Australian Bureau of Statistics.

19. Department of Health and Aged Care and the University of Adelaide. (1999) Measuring remoteness: accessibility/remoteness index of Australia (ARIA). Occasional Papers: New Series No 6. Canberra: Commonwealth of Australia.

20. Charlson ME, Pompei P, Ales KL, MacKenzie CR. (1987) A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 40: 373-383

21. Royston P, White IR. (2011) Multiple imputation by chained equations (MICE): implementation in Stata. J Stat Softw 45: 1-20

22. Zou G. (2004) A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol 159: 702-706

23. Sterne JAC, White IR, Carlin JB, et al. (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 338:

24. The University of Wisconsin - Madison Social Science Computing Cooperative. (2013) Multiple Imputation in Stata. http://www.ssc.wisc.edu/sscc/pubs/stata_mi_intro.htm. Accessed September 14, 2013

25. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. (2011) Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 20: 40-49

26. Willett WC. (1998) Nutritional epidemiology. 2 ed. New York NY: Oxford University Press US.

27. Fedirko V, Tramacere I, Bagnardi V, et al. (2011) Alcohol drinking and colorectal cancer risk: an overall and dose–response meta-analysis of published studies. Ann Oncol 22: 1958-1972

28. Krämer HU, Schöttker B, Raum E, Brenner H. (2012) Type 2 diabetes mellitus and colorectal cancer: meta-analysis on sex-specific differences. Eur J Cancer 48: 1269-1282

29. Ma Y, Yang Y, Wang F, et al. (2013) Obesity and risk of colorectal cancer: a systematic review of prospective studies. PLoS ONE 8: e53916

30. Colt JS, Schwartz K, Graubard BI, et al. (2011) Hypertension and risk of renal cell carcinoma among white and black Americans. Epidemiology 22: 797-804

31. Kreiger N, Nishri ED. (1997) The effect of nonresponse on estimation of relative risk in a case-control study. Ann Epidemiol 7: 194-199

32. Wigertz A, Lönn S, Hall P, Feychting M. (2010) Non-participant characteristics and the association between socioeconomic factors and brain tumour risk. J Epidemiol Community Health 64: 736-743

33. López R, Frydenberg M, Baelum V. (2008) Non-participation and adjustment for bias in case–control studies of periodontitis. Eur J Oral Sci 116: 405-411

34. Hatch EE, Kleinerman RA, Linet MS, et al. (2000) Do confounding or selection factors of residential wiring codes and magnetic fields distort findings of electromagnetic fields studies? Epidemiology 11: 189-198

35. Ramos E, Lopes C, Barros H. (2004) Investigating the effect of nonparticipation using a population-based case–control study on myocardial infarction. Ann Epidemiol 14: 437-441

36. Madigan MP, Troisi R, Potischman N, et al. (2000) Characteristics of respondents and non-respondents from a case-control study of breast cancer in younger women. Int J Epidemiol 29: 793-798

37. Vrijheid M, Richardson L, Armstrong BK, et al. (2009) Quantifying the impact of selection bias caused by nonparticipation in a case–control study of mobile phone use. Ann Epidemiol 19: 33-41.e31

38. Hall AE, Sanson-Fisher RW, Lynagh MC, Threlfall T, D'Este CA. (2013) Format and readability of an enhanced invitation letter did not affect participation rates in a cancer registry-based study: a randomized controlled trial. J Clin Epidemiol 66: 85-94

39. Mols F, Aaronson NK, Vingerhoets AJJM, et al. (2007) Quality of life among long-term non-Hodgkin lymphoma survivors. Cancer 109: 1659-1667

40. Richiardi L, Boffetta P, Merletti F. (2002) Analysis of nonresponse bias in a population-based case–control study on lung cancer. J Clin Epidemiol 55: 1033-1040

41. Smith T, Stein KD, Mehta CC, et al. (2007) The rationale, design, and implementation of the American Cancer Society's studies of cancer survivors. Cancer 109: 1-12

42. Geller B, Mace J, Vacek P, Johnson A, Lamer C, Cranmer D. (2011) Are cancer survivors willing to participate in research? J Community Health 36: 772-778

43. Mols F, Oerlemans S, Denollet J, Roukema J-A, van de Poll-Franse LV. (2012) Type D personality is associated with increased comorbidity burden and health care utilization among 3080 cancer survivors. Gen Hosp Psychiatry 34: 352-359

44. Kelly B, Fraze T, Hornik R. (2010) Response rates to a mailed survey of a representative sample of cancer patients randomly drawn from the Pennsylvania Cancer Registry: a randomized trial of incentive and length effects. BMC Med Res Methodol 10: 65

45. Morabia A, Bemstein M, Héritier S, Khatchatrian N. (1996) Relation of breast cancer with passive and active exposure to tobacco smoke. Am J Epidemiol 143: 918-928

46. Wingo PA, Ory HW, Layde PM, Lee NC. (1988) The evaluation of the data collection process for a multicenter, population-based, case-control design. Am J Epidemiol 128: 206-217

47. Girgis A, Boyes A, Sanson-Fisher RW, Burrows S. (2000) Perceived needs of women diagnosed with breast cancer: rural versus urban location. Aust N Z J Public Health 24: 166-173

48. Maclure M, Hankinson S. (1990) Analysis of selection bias in a case-control study of renal adenocarcinoma. Epidemiology 1: 441-447

Table 1. Relative Risks for the Associations between Demographic and Medical Factors and Participation among Potential Cases and Controls in a Case-Control Study of Colorectal Cancer in Western Australia

	Controls (n=2163)				Cases (n=1536)			
	% of all Controls	Response Fraction (%)	Relative Risk[a]	95% CI	% of all Cases	Response Fraction (%)	Relative Risk[a]	95% CI
Sex								
Female	41.8	45.7	1.00		41.1	55.7	1.00	
Male	58.2	48.3	1.05	0.96, 1.15	58.9	61.4	1.11	1.02, 1.21
Age								
40 to 49 years	10.0	35.6	1.00		7.2	53.6	1.00	
50 to 59 years	24.0	43.5	1.20	0.98, 1.47	23.7	62.9	1.11	0.93, 1.34
60 to 69 years	31.3	53.4	1.48	1.22, 1.80	37.0	59.9	1.05	0.88, 1.26
70 to 79 years	34.7	47.5	1.33	1.09, 1.62	32.2	57.5	1.02	0.85, 1.23
$P_{trend\text{-}test}$				*<0.01*				*0.42*
Socioeconomic Status								
1 (least disadvantaged)	30.0	49.3	1.00		27.7	65.2	1.00	
2	14.7	47.0	0.94	0.82, 1.08	16.4	56.7	0.87	0.76, 0.99
3	25.0	50.2	1.01	0.91, 1.14	25.8	56.9	0.88	0.79, 0.98
4	21.1	45.6	0.88	0.77, 1.01	20.2	57.2	0.90	0.79, 1.02
5 (most disadvantaged)	9.2	36.2	0.71	0.58, 0.88	9.8	55.0	0.86	0.73, 1.01
$P_{trend\text{-}test}$				*<0.01*				*0.04*
Accessibility/Remoteness								
Highly accessible	85.1	46.9	1.00		84.5	59.1	1.00	

Accessible/ moderately accessible	10.9	51.3	1.19	1.04, 1.37	12.2	59.6	1.07	0.94, 1.23
Very remote/remote	4.0	43.0	1.10	0.86, 1.42	3.3	56.0	1.02	0.80, 1.31
$P_{trend\text{-}test}$				*0.05*				*0.46*
Hospitalised at Time of Invitation?								
No	86.8	47.4	1.00		41.7	60.2	1.00	
Yes	13.2	46.2	0.97	0.85, 1.12	58.3	58.3	0.96	0.89, 1.05
Hospitalised in Last Five Years?								
No	39.6	42.2	1.00		41.6	58.7	1.00	
Yes	60.4	50.5	1.25	1.13, 1.38	58.4	59.3	1.07	0.98, 1.17
Charlson Comorbidity Index								
0	83.6	48.5	1.00		80.4	60.4	1.00	
1	6.8	42.6	0.78	0.64, 0.95	9.4	51.7	0.83	0.70, 0.99
2	4.9	47.2	0.86	0.70, 1.06	5.8	57.3	0.98	0.82, 1.17
3 or more	4.6	30.0	0.56	0.42, 0.77	4.4	52.2	0.84	0.66, 1.06
$P_{trend\text{-}test}$				*<0.01*				*0.10*
Grade[b]								
1					15.0	62.2	1.00	
2					57.0	59.3	0.98	0.88, 1.10
3					11.8	54.1	0.94	0.80, 1.11
Unknown					16.3	58.8	1.01	0.87, 1.16
Surgery Type[b]								
Total/subtotal colectomy or proctocolectomy					6.1	55.3	1.00	

Rectosigmoidectomy or proctectomy	8.9	51.8	0.94	0.74, 1.20
Low or high anterior resection or hemicolectomy	67.8	62.1	1.10	0.92, 1.33
Limited excision/other excision	7.0	57.4	1.02	0.80, 1.31
Other/unknown	10.1	48.4	0.87	0.68, 1.11
Cancer Site[b]				
Proximal colon	31.1	61.4	1.00	
Distal colon	29.9	56.0	0.93	0.83, 1.04
Rectum	36.1	59.6	0.99	0.90, 1.10
Unknown	3.0	58.7	1.00	0.78, 1.27

Abbreviations: CI, Confidence Interval

[a] All variables are mutually adjusted

[b] Cases Only

Table 2. Associations between the Exposures Total Alcohol, Smoking, Diabetes and the Risk of Colorectal Cancer in a Case-Control Study in Western Australia

	Original Analysis (n=1844)		Weighted Analysis (n=1844)		Imputed Analysis (n=3607)	
	AOR[a]	**95% CI**	**AOR[a]**	**95% CI**	**AOR[a]**	**95% CI**
ALCOHOL INTAKE 10 YEARS AGO						
All Alcohol						
<1 standard drink/week	1.00		1.00		1.00	
1-7 standard drinks/week	1.09	0.83, 1.43	1.08	0.83, 1.42	1.05	0.82, 1.35
7-21 standard drinks/week	1.09	0.83, 1.43	1.05	0.80, 1.38	1.08	0.85, 1.36
21+ standard drinks/week	1.25	0.94, 1.67	1.21	0.90, 1.62	1.24	0.96, 1.61
P_{trend}		*0.155*		*0.275*		*0.113*
Beer						
Less than 1 day/week	1.00		1.00		1.00	
1-4 days week	0.99	0.74, 1.31	0.94	0.70, 1.26	1.05	0.81, 1.36
5-7 days/week	1.50	1.09, 2.07	1.46	1.06, 2.02	1.36	1.02, 1.80
P_{trend}		*0.033*		*0.061*		*0.055*
Wine						
Less than 1 day/week	1.00		1.00		1.00	
1-4 days week	1.22	0.96, 1.55	1.21	0.95, 1.54	1.08	0.86, 1.34
5-7 days/week	1.23	0.96, 1.58	1.25	0.97, 1.60	1.09	0.86, 1.39
P_{trend}		*0.065*		*0.05*		*0.406*
TOBACCO USE						
Smoking Status						
Never smoker	1.00		1.00		1.00	
Former smoker	1.24	1.01, 1.53	1.22	0.99, 1.51	1.11	0.91, 1.35
Current smoker	1.05	0.74, 1.48	1.01	0.71, 1.44	1.00	0.73, 1.38
Lifetime Pack-Years						
0	1.00		1.00		1.00	
0.1 to 19.9	1.16	0.92, 1.46	1.14	0.90, 1.45	1.06	0.86, 1.32

20 or more	1.26	1.00, 1.60	1.23	0.97, 1.56	1.12	0.89, 1.39
P_{trend}		*0.049*		*0.090*		*0.327*
BODY MASS INDEX						
20 years of age						
Normal weight	1.00		1.00		1.00	
Overweight	1.25	0.92, 1.71	1.26	0.92, 1.73	1.17	0.89, 1.52
Obese	0.89	0.44, 1.77	0.94	0.45, 1.95	1.00	0.54, 1.83
P_{trend}		*0.401*		*0.336*		*0.412*
40 years of age						
Normal weight	1.00		1.00		1.00	
Overweight	0.95	0.75, 1.21	0.94	0.74, 1.20	0.98	0.81, 1.20
Obese	1.19	0.80, 1.76	1.13	0.77, 1.67	1.06	0.78, 1.45
P_{trend}		*0.675*		*0.823*		*0.843*
DIABETES						
Neither	1.00		1.00		1.00	
High Blood Sugar	1.15	0.76, 1.75	1.10	0.72, 1.69	1.03	0.72, 1.47
Diabetes	1.73	1.27, 2.34	1.66	1.22, 2.25	1.37	1.08, 1.74

Abbreviations: AOR, Adjusted Odds Ratio; CI, Confidence Interval

[a] Adjusted for age-group, sex, socioeconomic status, energy intake, lifetime vigorous recreational physical activity, and all other exposures in the table

Table 3. Association between Socioeconomic Status and Risk of Colorectal Cancer in a Case-Control Study in Western Australia

	Original Analysis (n=1844)				Weighted Analysis (n=1844)		Full Dataset (n=3607)		Imputed Analysis (n=3607)	
	OR[a]	**95% CI**	**AOR[b]**	**95% CI**	**AOR[b]**	**95% CI**	**OR[a]**	**95% CI**	**AOR[b]**	**95% CI**
Socioeconomic status										
Group 1 (least disadvantaged**)**	1.00		1.00		1.00		1.00		1.00	
Group 2	1.05	0.79, 1.40	1.04	0.77, 1.39	1.17	0.87, 1.58	1.18	0.96, 1.46	1.17	0.95, 1.45
Group 3	0.97	0.76, 1.24	0.94	0.73, 1.21	1.08	0.84, 1.38	1.15	0.96, 1.38	1.14	0.95, 1.37
Group 4	1.01	0.77, 1.31	0.96	0.73, 1.26	0.96	0.73, 1.26	1.04	0.86, 1.26	1.02	0.84, 1.25
Group 5 (most disadvantaged)	1.39	0.96, 2.01	1.32	0.90, 1.92	1.15	0.79, 1.68	1.18	0.92, 1.51	1.14	0.88, 1.48
P_{trend}		*0.351*		*0.625*		*0.889*		*0.362*		*0.529*

Abbreviations: AOR, Adjusted Odds Ratio; CI, Confidence Interval; OR, Odds Ratio

[a] Adjusted for age-group and sex only

[b] Adjusted for age-group, sex, energy intake, lifetime vigorous recreational physical activity, and all the exposures in Table 2