
**Copyright:**

This is the peer reviewed version of the following article: [Sharp L, McDevitt J, Brown C, Comber H. Smoking at diagnosis significantly decreases 5-year cancer-specific survival in a population-based cohort of 18 166 colon cancer patients. *Alimentary Pharmacology and Therapeutics* 2017, 45(6), 788-800.], which has been published in final form at [https://doi.org/10.1111/apt.13944](https://doi.org/10.1111/apt.13944). This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

**DOI link to article:**

[https://doi.org/10.1111/apt.13944](https://doi.org/10.1111/apt.13944)

**Date deposited:**

15/06/2017

**Embargo release date:**

08 February 2018
Smoking at diagnosis significantly decreases 5-year cancer-specific survival in a population-based cohort of 18,166 colon cancer patients

Linda Sharp1*, Joseph McDevitt2, Christopher Brown2, Harry Comber2

1 Institute of Health & Society, Newcastle University, Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne, NE2 4AX, England;

2 National Cancer Registry Ireland, Building 6800, Cork Airport Business Park, Kinsale Road, Cork, Ireland.

* Corresponding author: Linda Sharp, Institute of Health & Society, Newcastle University, Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne, NE2 4AX, England. Email: linda.sharp@ncl.ac.uk; telephone: +44 (0)191 208 6275.

Running title: Smoking and colon cancer survival

Keywords: colon cancer, survival, mortality, smoking, tobacco

Word count: 3,717
ABSTRACT

Background: Accumulating evidence suggests smoking may adversely affect cancer patients’ outcomes. Previous studies of smoking and survival in colon cancer have been limited by size and/or lack of a population-basis and results have been inconsistent.

Aim: This large population-based cohort study investigated whether: smoking status at diagnosis is an independent prognostic factor for cancer-specific survival in colon cancer; and treatment modifies any impact of smoking.

Methods: Colon adenocarcinomas diagnosed 1994-2012 were abstracted from the National Cancer Registry Ireland, and classified by smoking status at diagnosis. Cancer-specific death rates over 5 years were compared in current, ex- and never smokers using multivariable Cox proportional hazards models, and subgroup analyses by treatment (combinations of cancer-directed surgery and chemotherapy) conducted.

Results: Of 18,166 colon cancers, 20% were current smokers, 23% ex-smokers and 57% never smokers. Compared to never smokers, current smokers had a significantly raised cancer death rate (multivariable hazard ratio (HR)=1.14, 95% CI: 1.07, 1.12). There was a significant interaction between treatment and smoking (P=0.03). In those who had cancer-directed surgery only, but not other groups, current smokers had a significantly increased cancer death rate compared to never smokers (HR=1.21, 95% CI: 1.09, 1.34).

Conclusions: Smoking at diagnosis is an independent prognostic factor for colon cancer. The limitation of the association to surgically-treated patients suggests that the underlying mechanism(s) may be related to surgery. While further research is needed to elucidate mechanisms, continued efforts to encourage smoking prevention and cessation may yield benefits in terms of improved survival from colon cancer.
INTRODUCTION

Each year worldwide more than 750,000 new cases of colorectal cancer are diagnosed,¹ around two-thirds of which arise in the colon. There is compelling evidence for the role of a range of lifestyle factors - including smoking, physical activity and sedentary behaviour, weight and weight change, and diet - in the aetiology of colon cancer in particular.²⁻⁶ Evidence is also starting to suggest that lifestyle factors may be associated with survival in patients with colon cancer (see, for example, ⁷⁻⁹). Given that many lifestyle factors are potentially modifiable, greater clarity on the role of these in colon cancer outcomes would be of considerable public health importance.

A 2014 report by the US Surgeon General concluded that smoking may be causally related to higher all-cause and cancer-specific mortality and increased risk of second primary cancers.¹⁰ A recent systematic review and meta-analysis of six studies of smoking and colorectal cancer survival estimated that, compared to never smokers, current smokers had significantly increased risk of death (hazard ratio=1.26, 95%CI 1.15-1.37),¹¹ but the result was based on all-cause mortality which is influenced by the effects of smoking of deaths from diseases other than cancer.¹² Moreover, no distinction was made between cancers of the colon and rectum. Four studies included in the systematic review reported on smoking and survival in colon cancer patients.¹³⁻¹⁶ All described higher risk of death, or shorter survival, among smokers (who were defined in various ways), but results only reached statistical significance in two of the studies.

The majority of colon cancer patients undergo surgery and suitable patients with stage II or III disease may receive adjuvant chemotherapy.¹⁷ Studies suggest that short-term post-operative morbidity and 30-day mortality are worse in colorectal cancer patients who smoke.¹⁸,¹⁹ One small study of 284 colorectal cancer patients with pathologically complete resection reported a significant raised hazard ratio in active smokers²⁰ but no studies appear to have investigated whether smoking is also associated with survival in non-surgical colon cancer patients. Nicotine inhibits apoptosis by
chemotherapeutic agents in colon cancer cells\textsuperscript{21} but associations between smoking and chemotherapy receipt in colon cancer patients have not been examined.

This large population-based cohort study aimed to investigate (1) whether smoking status at diagnosis is an independent prognostic factor for cancer-specific survival in colon cancer and (2) whether the association varies by whether or not patients undergo cancer-directed surgery or receive chemotherapy.

METHODS

Setting

The study was conducted in Ireland, which has a mixed public-private healthcare system. All residents are entitled to care within the public system. A population-based colorectal cancer screening programme commenced roll-out in 2013. Prior to that, screening was not generally available.

Data

The data source was the National Cancer Registry Ireland, which aims to record all cancers newly diagnosed in the population usually resident in Ireland, using an active registration process. Completeness of registration is estimated to be at least 97\%.\textsuperscript{22} Following standard protocols, trained tumour registration officers collect patient (e.g. date of birth, address at diagnosis), tumour (e.g. cancer site, date of diagnosis) and first course treatment (surgery, chemotherapy or radiotherapy within a year of diagnosis) details for each tumour. The information collected includes patients’ smoking status at diagnosis, as recorded in their hospital records. The Central Statistics Office notifies deaths to the Registry, where they are linked to registered cancers, using probabilistic matching methods; this allows dates and causes of death to be identified.
For this analysis, all cases of colon cancer diagnosed 1994-2012 (ICD10 C18, n=24,953) were abstracted from the Registry. Cases were excluded if they had another invasive cancer (other than non-melanoma skin cancer) diagnosed prior to, or following, the colon cancer (n=3,889), had a tumour morphology other than adenocarcinoma (n=2,842) or were identified from death certificates only (n=56). Ascertainment of deaths was complete to 31/12/2012. Colon cancer-specific deaths were defined as those for which the underlying cause of death was coded as: cancer at the same diagnosis site; cancer of the same body system; cancer of another specified site; or cancer of unknown site.  

In terms of treatment, each case was classified according to whether or not the patient received, within a year of diagnosis: (i) cancer–directed surgery (i.e. colectomy, hemicolecotomy limited excision with or without anastomosis, or endoscopic local excision/destruction) and (ii) at least one course of chemotherapy. Age, marital status and sex were available from Registry records. Cases were assigned to one of five deprivation categories based on address of residence at the time of diagnosis. Categories ranged from least (1) to most deprived (5) and were based on a score derived from 2002 census variables; the 2002 census was used as this was closest to the middle of the study period. In terms of clinical variables, summary stage at diagnosis was defined according to the UICC classification and grade as well differentiated, moderately differentiated or poorly differentiated/undifferentiated. Cases were categorised as being located in the proximal (ICD10 18.0-18.5) or distal colon (18.6-18.8) or colon not otherwise specified (NOS; 18.9).

Smoking status at diagnosis was classified as: never smoked; ex-smoker (had smoked at least once every month in the past but not in the previous year); and current smoker (had smoked at least once every month in the previous year). For 26% of patients, smoking at diagnosis was unknown. Information was incomplete or missing for tumour stage (4%) and grade (12%), which are important prognostic factors. In total, 36% of patients had missing values for one or more of smoking status, stage and grade. The missing data items in these field were populated by multiple
Multiple imputation is considered superior to other approaches to dealing with missing data because it produces less biased estimates than complete case analysis under a range of missing data scenarios and, under other scenarios, is more efficient than complete case analysis.\textsuperscript{30-32} Fifty datasets were generated, in which missing values for smoking status, grade and stage were imputed using multinomial logistic regression including, as covariates, socio-demographic and clinical variables, indicators of cancer-specific death and death from any cause, and time to cancer death (or censor date) and any death (or censor date) (Supporting table 1).

**Statistical analysis**

The primary analysis was based on the dataset containing the imputed data. As recommended, a complete case analysis (considered a sensitivity analysis) was conducted for comparison.\textsuperscript{33} Our interest was in associations between smoking and rate of death so cancer-specific survival was the primary outcome and non-cancer deaths were censored (as opposed to being considered competing risks). For each patient, survival time was computed from the date of diagnosis to date of death, 5-years of follow-up, or the censoring date (31\textsuperscript{st} December 2012), whichever occurred first. Chi-square tests were used to compare patient-related and clinical characteristics of never, ex- and current smokers at the time of diagnosis. Curves of cumulative incidence of cancer-specific death up to 5-years post-diagnosis were generated by smoking status and – to aid interpretation – similar curves for cumulative incidence of deaths due to other causes. Unadjusted and multivariable hazard ratios for cancer-specific death within 5-years by smoking status were computed using Cox proportional hazards regression. Hazard ratio estimates from the 50 imputed datasets were combined to provide a single estimate with the standard error adjusted for inter- and intra-imputation variance according to Rubin’s rules.\textsuperscript{34} Variables for inclusion in the multivariable model were selected using a backwards stepwise approach; variables were retained if the Wald test $P$-value was $<0.05$. For each covariate, a
test for proportionality of hazards was performed and, since the hazards for age and stage were non-proportional, these variables were fitted as strata in the model.

Initially the analysis was done for all cases then repeated for subgroups defined by treatment. Since most treatments are received shortly after diagnosis, four subgroups were created by combining variables for receipt of cancer-directed surgery (yes/no) and receipt of chemotherapy (yes/no). The interaction between smoking and treatment was tested by fitting a cross-product term. Those who die soon after diagnosis have no possibility of undergoing treatment and this complicates comparisons of treated and untreated groups; how this might influence the relationship between smoking and survival is unclear. Therefore, we repeated this analysis using a landmark approach, which involved restricting the analysis to patients who were alive at 6 months post-diagnosis (since almost all treatments had been delivered by 6 months; n=14,387).

In the sensitivity analysis, we analysed the dataset which included all patients without missing data for smoking status, stage or grade (“complete cases”, n=11,648). We repeated the descriptive analyses and re-ran the final multivariable model to estimate hazard ratios by smoking status. All analyses were done in Stata 11.0.

Review of previous studies

To set our findings in context, we searched Medline for other studies reporting some aspect of survival by smoking status in people with colon cancer. MeSH headings and text word searches were undertaken, combining terms for disease, smoking exposure, and outcome. To be eligible studies had to be published as full papers, in the English language, by 30th November 2016. Information on the study population, sample size, assessment of smoking status, smoking prevalence and hazard ratios for survival by smoking status was abstracted from the papers identified.

RESULTS
Primary analyses: multiple imputed dataset

The analysis included 18,166 incident colon cancers. After imputation, 57% of patients were classified as never smokers at diagnosis, 23% as ex-smokers and 20% as current smokers (Table 1). Slightly more than half of patients were male, 13% were aged under 55 at diagnosis and 35% were aged 75 or older. Cancer-directed surgery was received by 85% (n=15,502) and 38% (6,851) had chemotherapy. In terms of treatment combinations, 53% had cancer-directed surgery only, 33% had surgery and chemotherapy, 5% had chemotherapy only, and 10% had neither treatment.

There were significant associations between smoking status and all of the socio-demographic and clinical variables (chi-square $P<0.01$ throughout; Table 1). A higher proportion of current smokers were diagnosed under 55, than ex-smokers or never smokers. The proportion of current smokers who lived in the most deprived areas was higher than for ex- and never smokers. 27% of current smokers had stage IV disease at diagnosis compared to 24% each of ex-smokers and never smokers. The percentage of current smokers who had cancer-directed surgery was slightly lower (82%) than among ex- (85%) and never smokers (86%). In contrast, chemotherapy receipt was more common among current smokers (41%) than ex- or never smokers (both 37%).

During the 5-years post-diagnosis, there were 7,488 cancer specific deaths. The cumulative incidence of cancer-specific death was highest in the current smokers; the curves for ex-smokers and never smokers were almost identical (Supporting figure 1(a)). For non-cancer deaths, the cumulative incidence was consistently higher in current and ex-smokers than in never smokers (Supporting figure 1(b)).

Univariate and multivariable hazard ratios for cancer-specific death by smoking status for all patients are shown in Table 2. In the univariate analysis, smoking was a significant prognostic factor; compared to never smokers, current smokers had a significantly increased rate of cancer death, while ex-smokers did not. This pattern persisted after adjustment for socio-demographic and clinical
variables (sex, marital status, deprivation category, period of diagnosis, grade and tumour location). Compared to never smokers, the hazard ratio of current smokers was raised by 14% and this was statistically significant (multivariable HR=1.14, 95%CI 1.07-1.22). The hazard in ex-smokers did not differ from unity (HR=1.00, 95%CI 0.94-1.17).

Table 3 shows the univariate and multivariable hazard ratios for smoking status in the four treatment groups. There was a significant interaction between smoking and treatment (p=0.034). A significant association between smoking and cancer-specific death was apparent among patients who had surgery only (P<0.01); in adjusted analysis, comparing to never smokers, current smokers had a 21% higher rate of cancer death (HR=1.21, 95%CI 1.09-1.34) while the rate in ex-smokers was not raised. In the other three treatment groups, current smoking was not associated with a higher cancer death rate.

In the landmark analysis, the same pattern was evident although the test for interaction was no longer statistically significant (P=0.118). In the surgery only group, current smokers had a significantly higher rate of death (HR=1.22, 95%CI 1.09-1.34; P<0.01); the estimates for current versus never smokers in the other treatment groups did not differ from unity (surgery & chemotherapy: 1.09, 95%CI 0.97-1.22; chemotherapy only: 1.01, 95%CI 0.82-1.25; neither surgery nor chemotherapy: 0.99, 95% CI 0.84-1.17).

**Sensitivity analyses: complete case dataset**

Supporting table 2 summarises the characteristics of the 11,648 patients in the complete case analysis. The same pattern of association between smoking status and cumulative incidence of cancer-specific death was evident in this analysis as in the primary analysis based on the dataset with imputed values (data not shown). In univariate analyses current smokers had a 21% higher rate of cancer-specific death than never smokers; this attenuated slightly in the multivariable analysis but remained statistically significant (multivariable HR=1.17, 95%CI 1.08-1.26)(Supporting table 3).
hazard ratio for ex-smokers was 0.99 in both univariate and multivariable analysis. There was a significant interaction between treatment and smoking (\(P<0.01\)) and, as in the primary analysis, the association between current smoking and cancer specific death was found only among cases who had surgery without chemotherapy (multivariable HR=1.28, 95%CI 1.14-1.43) (Supporting table 4). The same pattern, and a significant interaction (\(P<0.01\)), was seen in the landmark analysis (data not shown).

**Review of previous studies**

Eight studies of smoking and survival in colon cancer were identified, six from the USA, one from Canada and one from Germany (Table 4).\(^7\)\(^{-}\)\(^{16}\)\(^{-}\)\(^{37}\)\(^{-}\)\(^{39}\) Sample sizes ranged from 424 to 4,213. Studies varied in how, and at what time point, they assessed smoking status. The prevalence of current smokers ranged from 7% to 18%. Five studies reported on cancer-specific survival, and four each on disease-free, recurrence-related and overall survival. For cancer-specific, disease-free or recurrence-free survival, all eight studies reported an increased hazard for current versus never- or non-smokers, but this only reached statistical significance in four studies.

**DISCUSSION**

In this population-based cohort study - which is, by far, the largest study of smoking at diagnosis and survival in colon cancer - current smokers (but not ex-smokers) had a statistically significant, 14% higher, rate of cancer death compared to never smokers.

Although all of the eight previous studies of colon cancer patients reported poorer survival (by at least by one measure cancer-specific, disease-free or recurrence-free survival) in smokers than non- or never smokers,\(^7\)\(^{-}\)\(^{16}\)\(^{-}\)\(^{37}\)\(^{-}\)\(^{39}\) the association was not statistically significant in four studies and, in one further study, there was a significant reduction in disease-free survival but not in cancer-specific survival. In part this is likely to be due to the relatively modest sample size and hence limited
statistical power of most of the studies. In the studies which reported significant findings, the hazard ratio for current smokers was around 1.5.\textsuperscript{7,15,16,39} One reason for the more modest association in the current study than these studies may that our data was from a population-based cancer registry; the analysis therefore included all (or, at least 97% of)\textsuperscript{22} colon adenocarcinomas diagnosed in Ireland during 1994-2012 eliminating any possibility of selection or participation biases. Of the previous studies, none was truly population-based; three of the eight studies sampled potential participants (who could then agree or decline to participate) from a population-based sampling frame\textsuperscript{13,15,37} and, of these, only one reported a significant increased hazard for smokers.\textsuperscript{15} Another possible explanation for the more modest association in our study is that we adjusted for deprivation. Deprivation is a marker of socio-economic status and past studies have reported positive associations between various markers of socio-economic status and colon cancer survival.\textsuperscript{40-42} Smoking prevalence tends to be higher in people of lower socio-economic status and/or in more deprived areas,\textsuperscript{43} and this is also seen in Ireland.\textsuperscript{44} Thus, adjustment for deprivation would be likely to attenuate observed associations between smoking and survival.

The observed interaction between smoking and treatment, such that the adverse effect of current smoking was limited to patients who had cancer-directed surgery only, is noteworthy. Several studies have convincingly shown more post-operative complications and higher post-operative mortality in colon (or colorectal) cancer patients who smoke.\textsuperscript{18,19} However, these studies mainly considered short-term outcomes (e.g. 30-day mortality). The current study therefore adds to the evidence-base by showing that smoking adversely affects 5-year survival in surgically-treated patients. Moreover, the landmark analysis, which included patients who had survived 6 months, suggests that the adverse effect of smoking is not limited to the period immediately post-surgery but extends longer-term. In addition, our results confirm – in a much larger series and at the population-level - the findings of Munro et al\textsuperscript{20} who reported that active smokers (but not ex-smokers) had a significantly increased hazard compared to non-smokers; and extend these to show
that there is no adverse effect of smoking in non-surgical patients. The consistent results in the landmark analysis indicate our results are not due to immortal time bias.\textsuperscript{45} Our findings, and those of Munro et al,\textsuperscript{20} suggest that the mechanism(s) by which smoking impacts on colon cancer survival are related, in some way, to surgery.

The lack of association between smoking and survival in surgical patients who also had chemotherapy is intriguing. It seems likely that those who receive chemotherapy are fitter, suggesting that the effect of smoking may be stronger (or only evident) in less fit patients. This might usefully be investigated in future studies.

A significant proportion of colon cancers presents as emergencies and require rapid surgical intervention.\textsuperscript{46,47} A Swedish study found that colon cancer patients with a lower level of education, or lower income, were more likely to present as emergencies\textsuperscript{48} and, in analyses of a subgroup of colon cancers in Ireland, 31% of current smokers presented as emergencies compared to 24% of ex-smokers and never smokers combined. Since cases which present as emergencies have higher mortality,\textsuperscript{46,49} this suggests that a higher frequency of emergency presentation among smokers could help explain the observed association. Positive margins are a risk factor for recurrence following resection for colon cancer\textsuperscript{17} and a single study has reported a slightly higher rate of radial margin positivity among current smokers (6.4%) than other colon cancer patients (5.1%).\textsuperscript{50} However, the overall rate of radial margin positivity appears too low to fully explain the survival differential between current and never smokers. Another possible explanation follows from the observations that smoking induces a systemic inflammatory response\textsuperscript{51} and that systemic inflammatory response is a prognostic factor following colorectal resection.\textsuperscript{52} Finally, smoking may be a marker for raised levels of the complex glycoprotein carcinoembryonic antigen (CEA), which is a strong prognostic factor in colorectal cancer.\textsuperscript{53} Studies among apparently healthy volunteers have found associations between smoking and higher CEA levels.\textsuperscript{54,55} Further investigations of these potential mechanisms would be of considerable value.
Like a few previous studies, we found that current smokers were more often diagnosed with more advanced disease. While this might point to a tumour promoting effect of smoking, it may also be due to different patterns of healthcare consultation in smokers and non-smokers. Further research to better understand this relationship might illuminate the mechanisms by which smoking influences cancer progression.

**Strengths and limitations**

Although our data was population-based, the fact that it was derived from a cancer registry meant that we did not have information on disease free survival, nor did we have any information on tobacco exposure other than smoking status at diagnosis. We also lacked information on other lifestyle factors some of which have been related to colon cancer survival and may be associated with smoking. Smoking status was derived from medical records and, as such, may be subject to misclassification. Data suggests that, in face-to-face interviews for research studies, few people in the general population misrepresent their use of tobacco products, but whether this holds for cancer patients in discussions with health professionals who are involved in their treatment is less certain. It seems most likely that – due to social desirability bias - any misclassification will be differential, such that current smokers are more likely to have been classified as non-smokers than vice versa; this would mean that the hazard ratios are biased towards the null and we have underestimated the true association between smoking and survival. We chose to base our analysis on cancer-specific survival rather than overall survival because the latter is likely influenced by the impact of smoking on other major causes of death, notably cardiovascular disease. Cancer-specific survival analyses are reliant on accurate information on cause of death. While it is well recognised that the cause of death recorded on some death certificates is inaccurate, there does not appear to be any published data to suggest that (in)accuracy is differential by smoking status. In addition, while the SEER classification of cancer-specific deaths includes cancers at sites outside the body system of interest, and in our study there were some deaths at sites other than the bowel, numbers were
small (~2% of deaths). Almost half were coded to the lung or liver so probably represent metastases from the primary tumour; others may represent cancers diagnosed before 1994, about which we had no information. Moreover, the rate of non-cancer deaths was higher among current smokers and ex-smokers than never smokers in our dataset, providing some reassurance about the accuracy of death certificates and suggesting that the poorer survival observed in smokers is unlikely to be due to non-cancer deaths being mis-coded as cancer deaths.

**Implications**

Although not specific to colon cancer, evidence suggests that some cancer patients give up smoking around the time of diagnosis and others want to stop,\textsuperscript{59,60} and that smoking cessation interventions can be successful among patients/survivors.\textsuperscript{61,62} A single trial of a smoking cessation intervention in patients undergoing colorectal resection found no difference in post-operative complication rates between arms,\textsuperscript{63} but the sample size was small (n=60). Further studies are needed to investigate whether smoking cessation around the time of diagnosis improves patients’ outcomes, either in the short or longer-term.

**Conclusions**

In this large, population-based, cohort study of newly diagnosed colon cancer patients, current smokers (but not ex-smokers) had a significantly increased rate of cancer death compared to never smokers. This effect was limited to those patients who had cancer-directed surgery, suggesting that the mechanism by which smoking impacts on survival is in some way related to surgery. Further research is needed to better elucidate these mechanisms but, meanwhile, continued efforts to encourage smoking prevention and cessation may bring benefits in terms of improved survival from colon cancer.
ACKNOWLEDGEMENTS

We are grateful to the tumour registration officers and the data processing staff in the National Cancer Registry Ireland who collected the data which formed the basis of this paper.

STATEMENT OF INTERESTS

Authors’ declaration of personal interests: None of the authors have any interests to declare.

Declaration of funding interests: This study was funded in part by the Health Research Board by grant number HRA/2012/30 (which provides support for CB). Data collection and processing by the National Cancer Registry is funded by the Department of Health.

AUTHORSHIP STATEMENT

Guarantor of article: LS is the guarantor.

Specific author contributions: LS had the idea for this study. HC supervised the acquisition of the data. JMcD undertook the statistical analysis and CB and LS provided statistical advice. All authors contributed to interpretation of the results. LS wrote the article and other authors contributed to the content. All authors approved the final version of the manuscript, including the authorship list.
REFERENCES


## STROBE Statement—checklist of items that should be included in reports of observational studies

<table>
<thead>
<tr>
<th>Item No</th>
<th>Recommendation</th>
<th>Details and/or location in manuscript</th>
</tr>
</thead>
</table>
| **Title and abstract** | 1  
(a) Indicate the study’s design with a commonly used term in the title or the abstract | Title contains description of design – “population-based cohort”; also included in abstract |
|  |  
(b) Provide in the abstract an informative and balanced summary of what was done and what was found | Abstract describes methods and results |
| **Introduction** |  |  |
| Background/rationale | 2  
Explain the scientific background and rationale for the investigation being reported | Review and critique of the past literature included in the Introduction (page 3-4) |
| Objectives | 3  
State specific objectives, including any prespecified hypotheses | Objectives described at end of Introduction (page 4) |
| **Methods** |  |  |
| Study design | 4  
Present key elements of study design early in the paper | Design stated in final paragraph of Introduction (page 4) |
| Setting | 5  
Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | Setting, population, sources of data, exposure and follow-up described in Methods (pages 4-5) |
| Participants | 6  
(a) **Cohort study**—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up  
Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls  
Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants | Eligibility criteria for inclusion in study population described on page 4; sources of participants described on page 4; follow-up described on pages 4 and 5 |
|  |  
(b) **Cohort study**—For matched studies, give matching criteria and number of exposed and unexposed  
Case-control study—For matched studies, give matching criteria and the number of controls | Design not matched |
<table>
<thead>
<tr>
<th>Variables</th>
<th>7</th>
<th>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sources/measurement</td>
<td>8*</td>
<td>For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group</td>
</tr>
<tr>
<td>Bias</td>
<td>9</td>
<td>Describe any efforts to address potential sources of bias</td>
</tr>
<tr>
<td>Study size</td>
<td>10</td>
<td>Explain how the study size was arrived at</td>
</tr>
<tr>
<td>Quantitative variables</td>
<td>11</td>
<td>Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12</td>
<td>(a) Describe all statistical methods, including those used to control for confounding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Describe any methods used to examine subgroups and interactions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) Explain how missing data were addressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(d) Cohort study—If applicable, explain how loss to follow-up was addressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case-control study—If applicable, explain how matching of cases and controls was addressed</td>
</tr>
</tbody>
</table>

Outcomes (definition of cancer and other deaths) described on page 5. Predictor (smoking) described on page 5. Potential confounders (socio-demographic and clinical characteristics) described on page 5.

Source and classification of variables of interest described in page 4 and 5.

Multiple imputation used to fill missing values in smoking status and key confounders – described on page 6. Done to deal with potential bias of complete case analysis. Full details provided in supporting table 1.

All primary colon adenocarcinomas registered in Ireland included; described on page 4

Classification of quantitative variables included on page 5 and table 1.

Full details of statistical analysis included on page 6 and 7.

Subgroup and interaction analysis (by treatment receipt) included on pages 6 & 7

Multiple imputation methods described on page 6

Deaths ascertained by linkage with national death records so loss to follow-up not a significant consideration.
Cross-sectional study—If applicable, describe analytical methods taking account of sampling
strategy

| (c) Describe any sensitivity analyses | Sensitivity analyses (complete case) described on page 7, results reported on page 9 and in supporting tables |

Continued on next page
### Results

<table>
<thead>
<tr>
<th>Participants</th>
<th>13*</th>
<th>(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed</th>
<th>Numbers of cases excluded – and reasons for this - described in methods (page 4); numbers included in primary and sensitivity analyses described in methods and tables 1-3 and supporting tables 2-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) Give reasons for non-participation at each stage</td>
<td></td>
<td></td>
<td>Reasons for exclusions reported on page 4.</td>
</tr>
<tr>
<td>(c) Consider use of a flow diagram</td>
<td></td>
<td></td>
<td>Definition of cohort straightforward so flow diagram not provided</td>
</tr>
<tr>
<td>Descriptive data</td>
<td>14*</td>
<td>(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders</td>
<td>Provided in table 1 (and supporting table 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Indicate number of participants with missing data for each variable of interest</td>
<td>Information provided in the methods (pages 4-7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) Cohort study—Summarise follow-up time (eg, average and total amount)</td>
<td>Cases followed-up for 5 years or to death; described in methods</td>
</tr>
<tr>
<td>Outcome data</td>
<td>15*</td>
<td>Cohort study—Report numbers of outcome events or summary measures over time</td>
<td>Numbers of deaths reported in Table 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case-control study—Report numbers in each exposure category, or summary measures of exposure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cross-sectional study—Report numbers of outcome events or summary measures</td>
<td></td>
</tr>
<tr>
<td>Main results</td>
<td>16</td>
<td>(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included</td>
<td>Unadjusted and adjusted estimates reported in tables 2 &amp; 3; rationale for adjustments described on pages 6 &amp; 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Report category boundaries when continuous variables were categorized</td>
<td>Categorises reported in Tables 1 &amp; 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</td>
<td>Results reported as hazard ratios</td>
</tr>
<tr>
<td>Other analyses</td>
<td>17</td>
<td>Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses</td>
<td>Subgroup analyses reported in Table 3; sensitivity analyses reported in supporting tables 2-4</td>
</tr>
</tbody>
</table>
**Discussion**

<table>
<thead>
<tr>
<th>Key results</th>
<th>18</th>
<th>Summarise key results with reference to study objectives</th>
<th>Results summarised and discussed pages 9-12 and repeated in conclusion (pages 12-13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limitations</td>
<td>19</td>
<td>Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias</td>
<td>Limitations discussed on page 12</td>
</tr>
<tr>
<td>Interpretation</td>
<td>20</td>
<td>Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence</td>
<td>Results interpreted in the context of previous findings on pages 9-11</td>
</tr>
<tr>
<td>Generalisability</td>
<td>21</td>
<td>Discuss the generalisability (external validity) of the study results</td>
<td>Generalisability discussed on page 10 (with reference to population-based nature of study)</td>
</tr>
</tbody>
</table>

**Other information**

| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | Information on funders provided in the Statement of Interests |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.