MR JOHN ALEXANDER GERALD GIBSON (Orcid ID : 0000-0002-7944-8347)
MR THOMAS DOBBS (Orcid ID : 0000-0002-5176-7741)

Article type : Original Article

The Association of Smoking and Socioeconomic status on Cutaneous Melanoma: a population based, data linkage, case-control study

J.A.G. Gibson,[1,2*] T.D. Dobbs,[1,2] R. Griffiths,[3] J. Song,[3] A. Akbari,[3,4,5] S. Whitaker,[6] A. Watkins,[3,4] S.M. Langan[7,8], H.A. Hutchings,[4] R.A. Lyons[3,4,5] and I.S. Whitaker[1,2]

1. Reconstructive Surgery & Regenerative Medicine Research Group, Institute of Life Science, Swansea University Medical School, Swansea, UK.
2. The Welsh Centre for Burns and Plastic Surgery, Morriston Hospital, Swansea, UK
3. Health Data Research UK, Swansea University, Swansea, UK
4. Patient and Population Health and Informatics Research, Swansea University Medical School
5. Administrative Data Research Centre Wales, Swansea University Medical School, Swansea, UK.
6. Department of Dermatology, Singleton Hospital, Swansea, UK
7. Health Data Research UK, London, UK
8. London School of Hygiene & Tropical Medicine, London, UK

Corresponding author: John Gibson, email: johnaggibson@hotmail.com

Running head: The association of smoking and socioeconomic status on CMM
Conflict of interest: None
Funding: This work was supported by an iGrant from Tenovus Cancer Care. The ReconRegen research group is supported by the Royal College of Surgeons of England, the British Association of Plastic, Reconstructive and Aesthetic Surgeons and the Swansea Bay University Health Board. Health Data Research UK, which receives its funding from HDR UK Ltd (grant ref:NIWA1) funded by the UK Medical Research Council, Engineering and

Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust. This work was also supported by an ESRC award establishing the Administrative Data Research Centre Wales (ES/L007444/1). SML reports grants from Wellcome Senior Clinical Fellowship in Science (205039/Z/16/Z) during the conduct of the study.

Funders played no role in the study design, methods, data collection, data analysis, manuscript preparation or publication.

What is known about this subject?

- Previous studies have been contradictory with both negative and positive associations between smoking and the incidence of melanoma reported.
- Previous studies have either been limited by publication bias due to selective reporting or underpowered.

What does this study add?

- Our large study identified an inverse association between smoking status and melanoma incidence.
- Whilst smoking status was negatively associated with overall disease survival, no significant association was noted in melanoma-specific survival.
- Socioeconomic status remains closely associated with melanoma. Whilst higher socioeconomic populations are more likely to develop the disease, patients with lower socioeconomic status continue to have a worse prognosis.

Abstract

Background

Previous studies have identified an inverse association between melanoma and smoking; however data from population based studies are scarce.

Objective

To determine the association between smoking and socioeconomic status on the risk of development of melanoma. Furthermore, we sought to determine the implications of smoking and socioeconomic status on survival.

Methods

We conducted a population-based case-control study. Cases were identified from the Welsh Cancer Intelligence and Surveillance Unit (WCISU) during 2000-2015 and controls identified from the general population. Smoking and socioeconomic status were obtained from data linkage with other national databases. The association of smoking status and socioeconomic status on the incidence of melanoma were assessed using binary logistic regression. Multivariate survival analysis were performed on a melanoma cohort using Cox proportional hazard model using survival as the outcome.

Results

During 2000-2015, 9,636 patients developed melanoma. Smoking data were obtained for 7,124 (73.9%) of these patients. 26,408 controls were identified from the general population. Smoking was inversely associated with melanoma incidence (Odds Ratio (OR) 0.70 95% CI 0.65 -0.76). Smoking was associated with an increased overall mortality (Hazard Ratio (HR) 1.30 95% CI 1.09-1.55), but not associated with melanoma specific mortality. Patients with higher socioeconomic status had an increased association with melanoma incidence (OR 1.58 95% CI 1.44-1.73). Higher socioeconomic status was associated with an increased chance of both overall (HR 0.67 95% CI 0.56-0.81) and disease specific survival (HR 0.69 95% CI 0.53-0.90).

Conclusion

Our study has demonstrated that smoking appeared to be associated with reduced incidence of melanoma. Whilst smoking increases overall mortality, no association was observed with melanoma-specific mortality. Further work is required to determine if there is a biological mechanism underlying this relationship or an alternative explanation, such as survival bias.

Keywords:

Melanoma; Smoking; socioeconomic status; data-linkage; registry.

1. Introduction

Whilst there is a wealth of knowledge on the association of melanoma with risk factors such as ultraviolet light exposure, skin type and genetics[1], the relationship between tobacco smoke and melanoma is less clear. Tobacco smoke is a type 1 carcinogen, associated with 18 types of cancer[2]. Song et al[3] reported a moderate inverse association between melanoma and smoking in a meta-analysis of two cohort studies. This association was observed in both ex-smokers and current smokers in men, but not women. A larger meta-analysis, including 23 studies, reported a similar inverse association[4]. Both papers reported significant limitations, notably publication bias due to selective reporting in the published studies. Furthermore, confounding variables were not included in the analysis.

A recent, prospective cohort study has further explored the association. After adjusting for potential confounding factors, no association was observed between current smoking and melanoma (OR 1.01 95% CI 0.64 -1.61)[5]. Whilst the study addressed the aforementioned limitations by adjusting for confounding factors, the study was significantly underpowered; only a small proportion of the cohort developed melanoma and the average follow up duration was short (3.5 years).

The relationship between socioeconomic status and melanoma, on the other hand, is well established in the literature, with research dating back to the 1980s[6,7]. Those in higher income or higher educational groups are at an increased risk of developing melanoma, attributed to greater exposure to lifestyle factors, such as sun holidays and tanning bed use[8]. However, once diagnosed, those with a lower socioeconomic status have a worse prognosis, a finding seen across multiple jurisdictions with different health care systems[8]. Understanding and addressing this worsened prognosis is therefore a clear public health priority[9-11].

In this paper we describe the largest study investigating the association of smoking and melanoma published to date. We have used the power of routinely collected data to overcome limitations of previous studies and investigate the prognostic implications of smoking in this patient cohort. Furthermore, we sought to investigate the association of socioeconomic status on the incidence and survival of melanoma.

2. Methods

The described study has been reported in accordance with the Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement [12].

The study was conducted in two stages. In stage one; a case control study was performed to assess the relationship between smoking and the development of melanoma. In stage two, a cohort study was conducted to determine the association between smoking and survival within the melanoma cohort (Figure 1).

2.1 Overview of methods

Analysis of primary and secondary care National Health Service (NHS) data and national administrative data for 2000-2015 in Wales, UK (population 3.1 million) were performed. In instances where relevant data were unavailable from a single source, multiple datasets were linked. Data were retrieved from six national databases (Table 1). In Wales, population level de-identified person-based health and socio-economic administrative datasets are collated and linked within the Secure Anonymised Information Linkage (SAIL) Databank[13-15]. Robust policies, structures and controls are in place to protect privacy through a reliable matching and anonymization process, achieved in conjunction with the NHS Wales Informatics Service (NWIS) using a split file multiple encryption approach described in detail in previous published work[14].

Table 1 - List of databases used and their description

2.2 Cases

In Wales, all patients with a diagnosis of melanoma are recorded in the Welsh Cancer Intelligence and Surveillance Unit (WCISU) register. Cases were identified from WCISU using International Classification of Disease 10 (ICD-10) codes C43.0-C43.9 and morphology codes according to the International Classification of Diseases for Oncology (ICDO-3) 8720-8790[16]. Patients with melanoma in situ were not included in the study as either cases or controls. Demographic information was assessed at the diagnostic date. Melanoma specific variables (tumour location, stage and morphology) were assessed at the diagnostic date.

2.3 Controls

Four sets of general population controls were randomly selected from the Welsh Demographic Service Dataset (WDSD). Controls were not matched to cases. Both cases and controls needed to be alive and resident in Wales on the date of melanoma diagnosis. To increase the power of the study we aimed to have four controls for every case[17].

2.4 Smoking status

Self-reported smoking status, for cases and controls were obtained from the Welsh Longitudinal General Practice (WLGP) data, as recorded during patients' consultations with their General Practitioner in primary care, using Read codes that have been previously validated[18] (Appendix 1). Patients were defined as either a non-smoker (for lifelong non-smokers), ex-smoker (for those that had previously smoked) or current smokers. The smoking assessment window extended from the melanoma diagnosis date to six months prior. Where serial assessments were available, the smoking record most recent to the diagnosis was selected. Where "non-smoker" was recorded, the WLGP dataset was explored to establish whether the individual had previously been classified as a smoker. In such circumstances, the individual was classed as an ex-smoker.

2.5 Socioeconomic status

Socioeconomic status was measured using the Welsh Index of Multiple Deprivation (WIMD) version 2001, a measure based on the Index of Multiple Deprivation and used as the official measure of socioeconomic status for the Welsh Government[19]. Individual scores are based upon a person's postal address. Wales is divided into 1,896 Lower-Layer Super-Output Areas (LSOAs) following the 2001 Census, each consisting of approximately 1600 people. The WIMD scores for each LSOA are calculated from weighted scores from eight domains of socioeconomic status (income, employment, health, education, access to services, community safety, physical environment and housing socioeconomic status). Each LSOA in Wales has been ranked according to its WIMD score and grouped into quintiles, with quintile 5 being the highest socioeconomic status and 1 being the lowest.

2.6 Mortality data

Data relating to mortality, including cause of death, were obtained on the melanoma cohort from the Annual District Death Extract (ADDE) dataset, which contains the diagnostic codes listed on patient's death certificates, held within the SAIL Databank.

2.7 Charlson Co-morbidity Index

The Charlson co-morbidity index is a widely used measure of co-morbidity. An overall score is calculated from a list of conditions, each of which has been allocated a weight of between one and six based upon its adjusted relative risk of one-year mortality[20].

2.8 Ethical approval

Study approval was granted by the SAIL Databank independent Information Governance Review Panel (IGRP) (project 0593). Data held within the SAIL Databank are made available to researchers in an anonymised format and are therefore not subject to data protection legislation. SAIL follows all relevant legislative and regulatory frameworks in using population data for research.

2.9 Statistical analysis

Case – Control (Stage 1)

Descriptive statistics were used to characterise the melanoma cases and controls by smoking status and stage at diagnosis (cases only). An unconditional binary logistic regression model was used to calculate odds ratios with 95% confidence intervals for the association with melanoma. Sex, socioeconomic status and age at the time of diagnosis (as a continuous variable) were incorporated into the statistical model as confounders.

Cohort Study (Melanoma patients only) (Stage 2)

In this stage of the study on those with a diagnosis of melanoma were included (Figure1). Overall survival was calculated as the time from melanoma diagnosis to the time of death (outcome) or the end of the study (December 2018). Melanoma-specific survival was calculated as the time from melanoma diagnosis to the date of death from melanoma, or the end of the study for patients still alive (December 2018). Cases with missing variables were excluded from this aspect of the study.

Kaplan-Meier curves were generated for smoking status and socioeconomic status, with curves compared using the log-rank test. A Cox hazard proportional regression model was used to determine the association between smoking and mortality in the melanoma cohort. Sex, socioeconomic status, melanoma stage at diagnosis and age at diagnosis as a continuous variable were incorporated into the model as confounders. Both overall survival (deaths from any cause) and melanoma-specific survival (defined on their death registration held within ADDE) were analysed in the melanoma cohort. All data were analysed using IBM SPSS Statistics for Windows (IBM Corp. Released 2017. Version 25.0. Armonk, NY: IBM Corp). Statistical significance was assumed with a $p < 0.05$.

Results

Between 2000 and 2015, 9,636 patients were diagnosed with melanoma in Wales.

Stage 1 Case-Control study

Patient demographics and clinical characteristics of the cases and controls are outlined in Table 2. Data relating to smoking status were available for 7,124 (73.9%) of the melanoma cohort; 1,460 current smokers (20.6%), 3,065 (43.2%) ex-smokers and 2,599 (36.6%) non-smokers.

Smoking

After adjusting for sex, age and socioeconomic status, current smokers had 30% reduced odds for developing melanoma compared to non-smokers, (OR 0.70 95% CI 0.65-0.76) (Table 3). There was no association between being an ex-smoker or non-smokers and melanoma (OR 1.05 95% CI 0.98-1.12).

Socioeconomic status

We observed an inverse relationship between socioeconomic status and melanoma, whereby patients from higher socioeconomic WIMD quintiles were more likely to develop melanoma. Those in the highest socioeconomic quintile (WIMD 5) were 1.58 times more likely to develop melanoma as opposed to the lowest (HR 1.58 95% CI 1.44-1.73) (Table 3).

Table 3 Univariable logistic regression assessing risk factors for melanoma

Stage 2 Survival analysis of the melanoma cohort

Table 4 displays the demographics of the melanoma cohort.

Demographic data

Table 4 displays the demographics of the melanoma cohort. The median age at diagnosis was higher in non-smokers (66.7y) and ex smokers (64.5y) than in current smokers (62.4y). Socioeconomic status had significant variation amongst groups, with the current and ex-smokers being more likely to have lower socioeconomic status WIMD quintiles. Stage at diagnosis was not significantly different between smoking groups or socioeconomic status. No differences between the mean Charlson co-morbidity scores were noted between the smoking groups or between WIMD quintiles (Table 4).

Mortality

A total of 3,103 (32.2%) patients with melanoma died during the study period. Of these, 1,688 (54.4%) died from melanoma (melanoma listed as the primary cause of death on their death certificate) and 1,415 (45.6%) deaths were unrelated to melanoma. For patients who died from any cause, median time to death was 2.36 years. For patients who died of melanoma, median time to death was 1.73 years.

Univariate survival analysis

Median follow up duration of the entire cohort was 5.22 years (range: 0 – 18 years). Overall survival rates were different across the three smoking status groups, with ex-smokers having lower survival that current or non-smokers ($p<0.00$). In contrast, no difference was observed across the three smoking status groups for disease specific mortality ($p=0.88$). Overall and melanoma-specific survival rates by smoking status and socioeconomic status are shown in the supplementary figures. Figures 2 and 3 shows the overall and disease specific survival curves by smoking status.

Overall and disease specific survival rates differed significantly across the WIMD quintiles (Table 6 and 7).Figures 4 and 5 show the overall and disease survival curves by socioeconomic status.

Figure 2 Overall Survival by smoking status

Figure 3 Disease specific survival rates by smoking status

Figure 4 Overall Survival by socioeconomic status

Figure 5 Disease specific survival by socioeconomic status

Multivariable survival analysis

After adjusting for the aforementioned factors, current smokers had an increased overall risk of death as compared to non-smokers (HR 1.30 95% CI 1.09-1.55). There was no association between current smoking and melanoma-specific mortality. Increased odds of survival was noted in the highest socioeconomic WIMD quintile (quintile 5), compared to the lowest (quintile 1) (HR 0.67 95% CI 0.54-0.79). A similar trend was observed with disease specific mortality (HR 0.69 95% CI 0.56-0.81).

Males had an increased risk of overall and melanoma-specific death compared to females (Overall HR 1.28 95% CI 1.13-1.46) Disease specific (HR 1.35 95% CI 1.12-1.62). Tumour location was an important predictor of survival. For overall survival, tumours located on the upper limb were associated with increased survival compared to those on the trunk (HR 0.73 95% CI 0.61-0.88), with no association between tumours on the head and neck and lower limbs however. With regards to melanoma-specific mortality, tumours located on the trunk were associated with an increased risk of mortality when compared to those in other locations. Age was associated with a small increased risk of overall and melanoma-specific mortality (Overall HR 1.06 95% CI 1.05-1.06 $p < 0.00$; disease specific HR 1.02 95% CI 1.01-1.03 $p < 0.00$). Melanoma morphology was not associated with overall survival, however melanoma-specific mortality was increased in those with nodular melanoma (HR 1.23 95% CI 0.98-1.54) whereas those with lentigo maligna melanoma had improved survival (HR 0.43 95% CI 0.21-0.89). The Charlson co-morbidity index was not association with overall (HR 1.01 95% CI 1.00 -1.017) or melanoma-specific survival (HR 1.00 95% CI 0.99 -1.02).

Table 5 Cox model for overall and disease specific survival

Discussion

We found that smokers were less likely to develop melanoma in this population based, case-control study, but that their overall survival was reduced. After controlling for age, sex, socioeconomic status, tumour location, morphology and stage, the smoking group had an increased risk of death from all causes as compared to the non-smoking group. However, when investigating melanoma-specific mortality, no association was observed.

The mechanism responsible for the observed protective association of smoking on the risk of developing melanoma is not yet known, but several plausible hypotheses exist. Some authors hypothesize that the accumulation of nicotine in cells containing melanin suppresses the inflammatory response to UV-B[21-23]. Additionally, as smoking increases elastosis, it has been hypothesised that elastosis formation is protective of melanoma[24]. Alternative explanations include earlier deaths in current and ex-smokers leading to survival bias, whereby those exposed to smoking die before being at risk of developing melanoma.

Melanoma is not the only condition where smoking has shown to have a favourable association, such as in Parkinson's disease and ulcerative colitis[25,26]. The protective association in Parkinson's disease has been attributed to nicotine's ability to prevent brain damage and dopamine depletion. The depletion of dopamine occurs in the substantia nigra, an area of the brain populated by melanocytes. It is therefore plausible that Parkinson's disease and melanoma share similar pathogenesis[27]. Numerous studies have demonstrated an increased risk of melanoma in patients with Parkinson's disease and vice versa[28]. The inverse association of smoking and the risk of developing ulcerative colitis is well reported in the literature, however the pathogenesis is less well understood[29].

The relationship with smoking status has been investigated for Non Melanoma Skin Cancers (NMSC). In a prospective cohort study of over one million participants, current smokers were found to have a reduced risk of developing Basal Cell Carcinomas (BCC). Similar to our study, this "protective" association was not observed in ex-smokers. Squamous Cell Carcinomas (SCC) are conversely more common in smokers[30].

The Notch pathway, which functions broadly in specifying cell fates during embryogenesis and adult life, has a key role in linking the control of epidermal differentiation and proliferation[31]. Aberrant Notch signalling leads to skin cancer, although with different associations with different skin cancer types[31]. For melanoma, nodular and superficial BCC, Merkel Carcinoma and SCC in sun protected sites increased notched signalling has an oncogenic effect. Whilst for basosquamous BCC and SCC on sun exposed sites increased signalling has an oncosuppressive effects. The notch pathway has been found to be down regulated in smokers which could provide a further explanation on the protective association of smoking on melanoma and nodular BCC and the higher risk of SCC on sun exposed sites[31-34].

Whilst we observed that smokers appeared to be at reduced risk of melanoma, their overall survival was reduced. This finding is not surprising given the strong relationship between smoking and other life limiting conditions, such as the majority of cancers and cardio-respiratory disease. However, consistent with the potential protective influence of smoking on melanoma development, the risk of death from melanoma was not different between the smokers and non-smokers after adjusting for age, sex, stage of disease, morphology, socioeconomic status and tumour location. This might imply that smoking does not affect the disease progression of melanoma. This is however, not consistent with the work of Jones et al, who identified that at presentation, smokers had an increased risk of lymph node metastasis[35]. The discrepancy may be explained by the fact that the above study did not control for socioeconomic status. In addition, Jones et al reported an association between smoking status and Breslow thickness at presentation. Whilst in this study we did not have data on Breslow thickness, smoking status was not associated with stage at presentation.

Consistent with the published literature we found that the risk of developing melanoma was positively associated with socioeconomic status in this study[1]. The underlying explanation is poorly understood and likely to be complex and multifactorial. Socioeconomic status is closely linked with lifestyle factors such as travel, sunbed use and hobbies that are also associated with sunlight exposure, with the literature supporting the notion that those that are more affluent have greater exposure to lifestyle factors that increase melanoma incidence[1,8]. Our study also demonstrated that those in the highest socioeconomic status were less likely to smoke.

Despite higher socioeconomic status being associated with an increased risk of melanoma development, lower socioeconomic status is associated with poorer survival once diagnosed. This relationship was observed in both overall and disease specific survival rates. This is consistent with the broader health literature where it has been shown that lower socioeconomic status is associated with premature mortality from a number of conditions such as cardiovascular disease, respiratory disease and some malignancies[36]. In previous studies, low socioeconomic status has been associated with later stage of melanoma diagnosis, however this was not observed in this study. Our results may be explained by the measure used to classify socioeconomic status, the WIMD score. One of the seven domains used to determine the WIMD quintile is health, which is determined by the number of limiting long-term illnesses, all cause death rate, cancer incidence and birth weight. Patients within the low socioeconomic status group may therefore have other attributable factors influencing survival.

Limitations of this study included missing data, the lack of information available on ethnicity and UV light exposure. As with any population-based study, missing data prevented analysis on the total cohort. Data were missing for some of the cohort on smoking status and stage of disease. Smoking status was obtained from Welsh Longitudinal General Practice (WLGP), as recorded during patient's consultations with their GP. To date, the WLGP covers 80% of GP practices across Wales. Of the 2,512 patients for which smoking data were absent, 2,431 (96.7%) belonged to GP practices not contributing data to the SAIL Databank. It is therefore assumed that data for this variable were missing at random and would not bias the results. Additionally, information was not available on the quantity of tobacco smoked by participants. The Read codes listed in the appendix do indeed capture some information on the amount of smoking. In practice, these codes were rarely utilised by General Practitioners, with the majority simply recording 137R (Current smoker) and therefore we were unable to provide meaningful results. This is a substantial limitation as the cumulative exposure to tobacco was not assessed, thus it was not possible to calculate a dose response relationship.

When stage of melanoma was not recorded in the WCISU data and could not be obtained from other linked data, these data were missing. To assess the effect of this missingness, sensitivity analysis were performed. Missing data were incorporated into the regression model as a separate

category for stage. This was found not to affect the statistical significances outlined in the results section.

A further limitation of population-based studies using routinely collected data is incomplete control of confounding, that of data that are not specified, incompletely captured or misclassified, namely tumour location (relating to ICD 10 Code C43.9 melanoma unspecified) and tumour morphology (M7203 - MM NOS (melanoma – not otherwise specified)). The classification codes used to extract smoking status from GP data have shown to classify 8.6% ex-smokers as never smokers. Any misclassification would not significantly bias the results.

Ethnicity is only available on special request within the SAIL Databank and was therefore not incorporated into the statistical model. In Wales, population statistics reveal that 95% of the population are white and therefore the significance of ethnicity on the results would be minimal[37].

Conclusion

This is the largest study to date indicating that smoking has an inverse relationship on the risk of developing melanoma. Whilst the detrimental repercussions of smoking are well documented, further work is required to uncover the mechanism underlying this relationship, including further assessment about survival bias. If a biological association seems likely, this could lead to the development of novel prevention and treatment options, opening up a new wave of medical therapy for melanoma. Furthermore, this work reinforces the ongoing association between melanoma and socioeconomic status. Despite numerous public health strategies, higher socioeconomic groups continue to have a higher incidence of melanoma, however, lower socioeconomic status is related to poor survival once melanoma is diagnosed. The implications of these results, in a country such as the United Kingdom where healthcare is free to all, are significant. Further work is required to investigate how barriers to care may exist for the lowest socioeconomic status group so that policies can be implemented to prevent healthcare inequality and improve melanoma outcomes for all.

Declarations

Ethical approval and consent to participate

Study approval was granted by the SAIL Databank independent Information Governance Review Panel (IGRP) (project 0593). Data held within the SAIL Databank are made available to researchers in an anonymised format, and are therefore not subject to data protection legislation. SAIL follows all relevant legislative and regulatory frameworks in using population data for research.

Availability of data and materials

The data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK. All proposals to use SAIL data are subject to review by an independent Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been approved, it is gained through a privacy-protecting safe haven and remote access system referred to as the SAIL Gateway. SAIL has established an application process to be followed by anyone who would like to access data via SAIL https://www.saildatabank.com/application-process

Competing interests

The authors declare no competing interests,

Funding

No funding was sought for this study.

Authors Contributions

JAGG - Designed the study, performed statistical analysis, interpreted the results and wrote the early draft of the manuscript.

TDD - Involved in conception of the presented idea, assisted with study design, identification of data, supported analysis of the results, and editing of the manuscript.

RG + JS + AA - Sourced the data and assisted with manuscript preparation.

SW - Contributed relevant clinical evidence (dermatology) and editing the manuscript.

AW - Verified the statistical methods and edited the final manuscript.

HAH - Assisted with study design, statistical support and editing the manuscript.

SML - Contributed relevant clinical evidence (dermatology) and editing the manuscript.

RAL - Assisted with appropriate data retrieval, provided statistical support and edited the final manuscript.

ISW - Conceived the presented idea, encouraged JAGG to investigate the presented idea, contributed relevant clinical evidence (plastic surgery) and supervised manuscript preparation.

All authors discussed the results, provided a critical appraisal and contributed to the final manuscript.

Acknowledgments

This study makes use of the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research.

This work was supported by Health Data Research UK, which receives its funding from HDR UK Ltd (grant ref:NIWA1) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust.
This work was also supported by an ESRC award establishing the Administrative Data Research Centre Wales (ES/L007444/1).

References

1 Jiang AJ, Rambhatla PV, Eide MJ. Socioeconomic and lifestyle factors and melanoma: a systematic review. *British Journal of Dermatology* 2015; **172**: 885-915.

2 Secretan B, Straif K, Baan R *et al.* A review of human carcinogens;Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *The Lancet Oncology* 2009; **10**: 1033-4.

3 Song F, Qureshi AA, Gao X *et al.* Smoking and risk of skin cancer: a prospective analysis and a meta-analysis. *Int J Epidemiol* 2012; **41**: 1694-705.

4 Li Z, Wang Z, Yu Y *et al.* Smoking is inversely related to cutaneous malignant melanoma: results of a meta-analysis. *British Journal of Dermatology* 2015; **173**: 1540-3.

5 Dusingize JC, Olsen CM, Pandeya N *et al.* Smoking and Cutaneous Melanoma: Findings from the QSkin Sun and Health Cohort Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2018; **27**: 874-81.

6 Graham S, Marshall J, Haughey B *et al.* An inquiry into the epidemiology of melanoma. *Am J Epidemiol* 1985; **122**: 606-19.

7 Vagero D, Persson G. Risks, survival and trends of malignant melanoma among white and blue collar workers in Sweden. *Social science & medicine (1982)* 1984; **19**: 475-8.

8 Idorn LW, Wulf HC. Socioeconomic status and cutaneous malignant melanoma in Northern Europe. *British Journal of Dermatology* 2013; **170**: 787-93.

9 Buchanan L. Slip, slop, slap, seek, slide - is the message really getting across? *Dermatol Online J* 2013; **19**: 19258.

10 Makin JK, Warne CD, Dobbinson SJ *et al.* Population and age-group trends in weekend sun protection and sunburn over two decades of the SunSmart programme in Melbourne, Australia. *Br J Dermatol* 2013; **168**: 154-61.

11 Cancer Research UK. Be Clear on Cancer: Skin cancer campaign. In. 2014.

12 Benchimol EI, Smeeth L, Guttmann A *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS medicine* 2015; **12**: e1001885.

13 Ford DV, Jones KH, Verplancke JP *et al.* The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC health services research* 2009; **9**: 157.

14 Jones KH, Ford DV, Jones C *et al.* A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy-protecting remote access system for health-related research and evaluation. *Journal of biomedical informatics* 2014; **50**: 196-204.

15 Lyons RA, Jones KH, John G *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making* 2009; **9**: 3.

16 Fritz A, Percy C, Jack A *et al. International Classification of Diseases for Oncology*, 3rd edn. Geneva: World Health Organization. 2003.

17 Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. *Lancet* 2005; **365**: 1429-33.

18 Atkinson MD, Kennedy JI, John A *et al.* Development of an algorithm for determining smoking status and behaviour over the life course from UK electronic primary care records. In: *BMC Medical Informatics and Decision Making*, Vol. 17. 2017.

19 Welsh Government. What is the Welsh Index of Multiple Deprivation and how should it be used? In, Vol. 2018. 2014.

20 Charlson ME, Pompei P, Ales KL *et al.* A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases* 1987; **40**: 373-83.

21 Yerger VB, Malone RE. Melanin and nicotine: A review of the literature. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* 2006; **8**: 487-98.

22 Mills CM, Hill SA, Marks R. Transdermal nicotine suppresses cutaneous inflammation. *Arch Dermatol* 1997; **133**: 823-5.

23 Mills CM, Hill SA, Marks R. Altered inflammatory responses in smokers. *Bmj* 1993; **307**: 911.

24 Grant WB. Skin aging from ultraviolet irradiance and smoking reduces risk of melanoma: epidemiological evidence. *Anticancer Res* 2008; **28**: 4003-8.

25 Hernan MA, Takkouche B, Caamano-Isorna F *et al.* A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Annals of neurology* 2002; **52**: 276-84.

26 Thomas GAO, Rhodes J, Green JT *et al.* Role of smoking in inflammatory bowel disease: implications for therapy. *Postgraduate Medical Journal* 2000; **76**: 273.

27 Pan T, Li X, Jankovic J. The association between Parkinson's disease and melanoma. *Int J Cancer* 2011; **128**: 2251-60.

28 Huang P, Yang X-D, Chen S-D *et al.* The association between Parkinson's disease and melanoma: a systematic review and meta-analysis. *Translational neurodegeneration* 2015; **4**: 21-.

29 Parkes GC, Whelan K, Lindsay JO. Smoking in inflammatory bowel disease: Impact on disease course and insights into the aetiology of its effect. *Journal of Crohn's and Colitis* 2014; **8**: 717-25.

30 Pirie K, Beral V, Heath AK *et al.* Heterogeneous relationships of squamous and basal cell carcinomas of the skin with smoking: the UK Million Women Study and meta-analysis of prospective studies. *Br J Cancer* 2018; **119**: 114-20.

31 Panelos J, Massi D. Emerging role of Notch signaling in epidermal differentiation and skin cancer. *Cancer biology & therapy* 2009; **8**: 1986-93.

32 Balint K, Xiao M, Pinnix CC *et al.* Activation of Notch1 signaling is required for β-catenin–mediated human primary melanoma progression. *Journal of Clinical Investigation* 2005; **115**: 3166-76.

33 Nakayama K. Growth and progression of melanoma and non-melanoma skin cancers regulated by ubiquitination. *Pigment cell & melanoma research* 2010; **23**: 338-51.

34 Zhang M, Biswas S, Qin X *et al.* Does Notch play a tumor suppressor role across diverse squamous cell carcinomas? *Cancer medicine* 2016; **5**: 2048-60.

35 Jones MS, Jones PC, Stern SL *et al.* The Impact of Smoking on Sentinel Node Metastasis of Primary Cutaneous Melanoma. *Ann Surg Oncol* 2017; **24**: 2089-94.

36 Romeri E, Baker A, Griffiths C. Mortality by deprivation and cause of death in England and Wales, 1999-2003. *Health statistics quarterly* 2006: 19-34.

37 Stats Wales. Ethnicity by year and ethnic group. In: (Government W, ed), Vol. 2018. 2018.

Tables and Figures

Figure 1

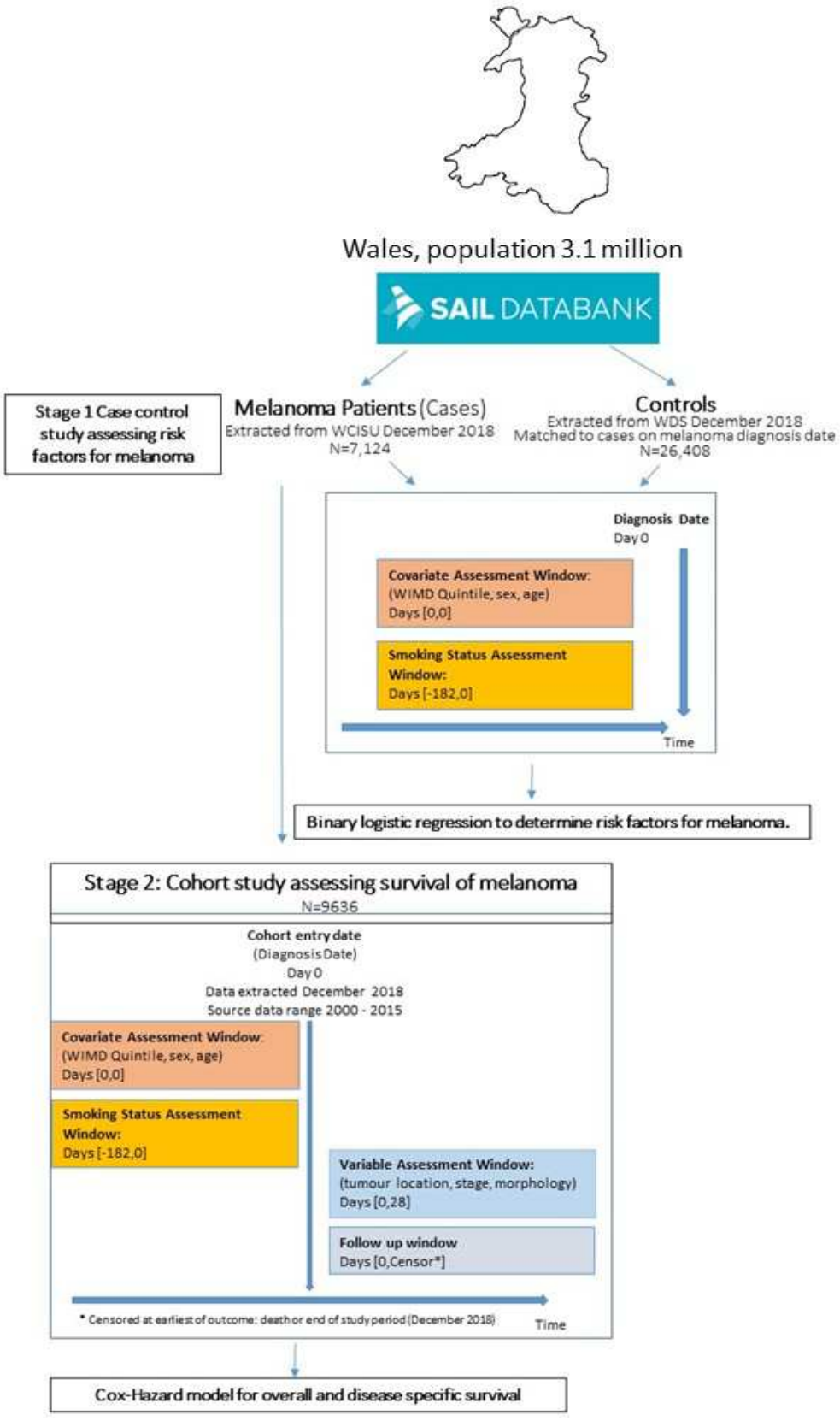


Table 1

Database	Description
Annual District Death Extract (ADDE)	Collected from the Office for National Statistics (ONS), containing death registration information, relating to Welsh residents including those who died outside of Wales.
Outpatient Dataset for Wales (OPDW)	Administrative and clinical data obtained from outpatient appointments in Wales.
Patient Episode Database for Wales (PEDW)	Administrative and clinical data for all hospital admissions, including diagnosis and operations performed.
Welsh Cancer Intelligence and Surveillance Unit (WCISU)	The national cancer registry for Wales. Captures all welsh melanoma patients from a number of sources; Multi-Disciplinary Team data, pathology data, other routine data sources in Wales and the English cancer registry.
Welsh Longitudinal General Practice (WLGP)	Administrative and clinical data from all patient visits to a General Practitioner.
Welsh Demographic Service Dataset (WDSD)	Administrative data about individuals resident or registered in Wales that have used National Health Service (NHS) services.

Table 2

Parameter	Cases (n=7,124)	Controls (n=24,608)	*P*-Value
Median (Interquartile range)	63.0 (50.0-74.0)	43.0 (26.0-60.0)	
Age Group, n (%)			
<20	46 (0.7)	3,980 (16.2)	0.00
20-29	262 (3.7)	3,866 (15.7)	
30-39	488 (6.9)	3,898 (15.8)	
40-49	833 (11.7)	4,230 (17.2)	
50-59	1,312 (18.4)	3,801 (15.5)	
60-69	1,582 (22.2)	3,180 (12.3)	
70-79	1,654 (23.2)	2,230 (9.1)	
80-89	974 (13.7)	1030 (4.2)	
>90	144 (2.0)	193 (0.8)	
Sex, n (%)			
Male	3,489 (49.0)	12,735 (51.8)	0.26
Female	3,635 (51.0)	1,3673 (55.6)	
WIMD Quintile, n (%)			
1	1,010 (14.18)	5502 (22.4)	0.00
2	1,202 (16.87)	5329 (21.7)	
3	1,464 (20.6)	5333 (21.7)	
4	1,446 (20.3)	4797 (19.5)	
5	1,996 (28.0)	5447 (22.1)	
Unspecified	6 (0.1)	0 (0.0)	
Smoking status			
Non-Smoker	2599 (36.5)	10,128 (41.2)	0.00
Ex-Smoker	3065 (43.0)	7,326 (29.8)	
Current Smoker	1460 (20.5)	8,954 (36.4)	

Table 3

Variable	*P*-Value	Odds Ratio (95% C.I.for Odds Ratio)
Age	0.00	1.04 (1.04 -1.05)
Non-Smokers	Reference	
Ex-Smokers	0.17	1.05 (0.98 - 1.12)
Smokers	0.00	0.70 (0.65 -0.76)
Male	0.26	0.97 (0.92 - 1.02)
WIMD Q1 (lowest socioeconomic status)	Reference	
WIMD Q2	0.09	1.09 (0.97 - 1.20)
WIMD Q3	0.00	1.20 (1.09 - 1.32)
WIMD Q4	0.00	1.30 (1.18 - 1.43)
WIMD Q5 (highest socioeconomic status)	0.00	1.58 (1.44 - 1.73)

Table 4

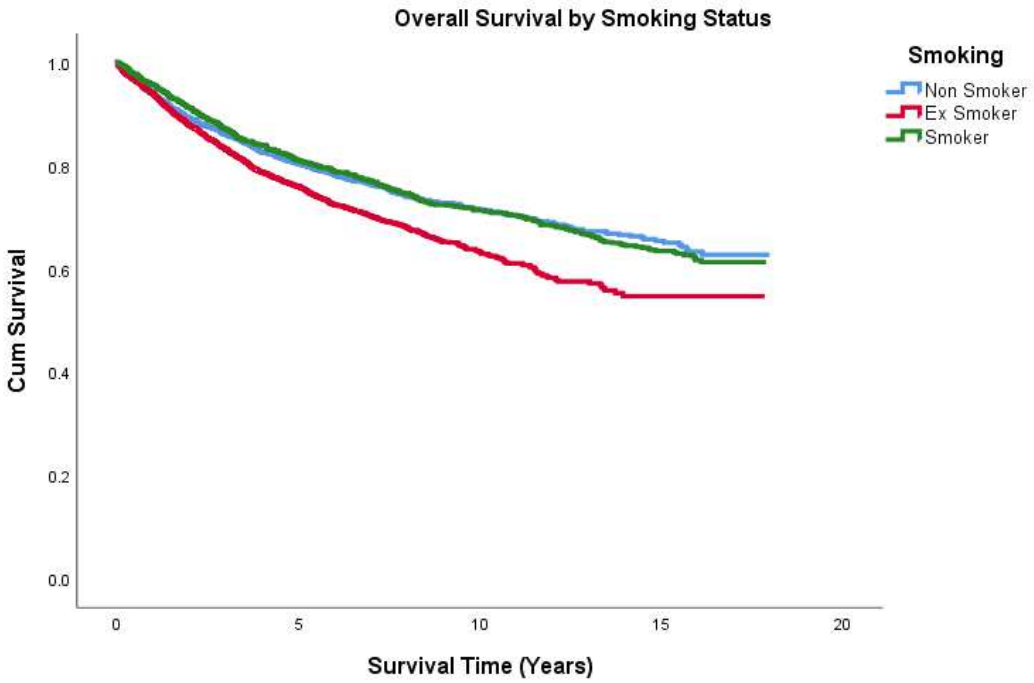
Characteristic	Total (n=9,636)	Unknown (n=2,512)	Non Smoker (n=2,599)	Ex-Smoker (n=3,065)	Current Smoker (n=1,460)	Chi-square P-Value
Median age (Interquartile range)	64.3 (50.5-75.5)	62.6 (48.9-75.0)	66.7 (51.9-77.0)	64.5 (51.4-75.4)	62.4(48.5-73.37)	
Age Group, *n* (%)						
0 - 9	<5* (0.1)	<5* (0.2)	0 (0)	0 (0)	0 (0)	
10-19	46 (0.5)	13 (0.5)	20 (0.8)	<5* (0.2)	9 (0.6)	
20-29	327 (3.4)	82 (3.3)	97 (3.7)	57 (1.9)	91 (6.2)	
30-39	726 (7.5)	180 (7.2)	221 (8.5)	154 (5.0)	171 (11.7)	
40-49	1,242 (12.9)	291 (11.6)	406 (15.6)	266 (8.7)	279 (19.1)	
50-59	1,615 (16.8)	385 (15.3)	480 (18.5)	417 (13.6)	333 (22.8)	
60-69	2,103 (21.8)	536 (21.3)	552 (21.2)	716 (23.4)	299 (20.5)	
70-79	2,085 (21.6)	571 (22.7)	471 (18.1)	850 (27.7)	193 (13.2)	
80-89	1,257 (13.0)	368 (14.6)	294 (11.3)	515 (16.8)	80 (5.5)	
90-99	230 (2.4)	82 (3.3)	57 (2.2)	86 (2.8)	5 (0.3)	
>100	<5*(0.1)	0 (0)	<5* (0.2)	0 (0)	0 (0)	
Sex, n (%)						
Male	4,750 (49.3)	1,261 (50.2)	1161 (44.7)	1661 (54.2)	667 (45.7)	<0.00
Female	4,886 (50.7)	1,251 (49.8)	1438 (55.3)	1404 (45.8)	793 (54.3)	
WIMD Quintile, n (%)						
1 (Lowest socioeconomic status)	1,300 (13.5)	290 (11.5)	269 (10.4)	450 (14.7)	291 (19.9)	<0.00

2	1,662 (17.2)	460 (18.3)	382 (14.7)	508 (16.6)	312 (21.4)	
3	1,951 (20.2)	487 (19.3	507 (19.5)	669 (21.8)	288 (19.7)	
4	2,169 (22.5)	723 (28.8)	558 (21.5)	606 (19.8)	282 (19.3)	
5 (highest socioeconomic status)	2,547 (26.4	551 (21.9)	881 (33.9)	828 (27.0)	287 (19.7)	
Unspecified	7 (0.1)	0 (0)	<5* (0.1)	<5* (0.2)	<5*	
Mean Charlson Co-morbidity Score	4.27	4.62	4.06	4.18	4.21	P=0.69
Location, n (%)						
Head & Neck	1,836 (19.1)	521 (37.9)	451 (17.4)	649 (21.2)	216 (14.8)	<0.00
Upper Limb	2,071 (21.5)	497 (19.8)	758 (29.2)	662 (21.6)	466 (31.9)	
Lower Limb	2,370 (24.6)	593 (23.6)	593 (22.8)	685 (22.3)	319 (21.8)	
Trunk	2,884 (29.9)	706 (28.1)	698 (26.9)	956 (31.2)	395 (27.1)	
Unspecified	476 (4.9)	195 (9.9)	99 (3.8)	113 (2.1)	64 (4.4)	
Stage, n (%)						
1	4,216 (43.8)	900 (35.8)	1220 (46.9)	1484 (48.4)	612 (41.9)	0.06
2	1,837 (19.1)	488 (19.4)	473 (18.2)	676 (22.1)	200 (13.7)	
3	319 (3.3)	100 (4.0)	82 (3.2)	95 (3.1)	42 (2.9)	
4	125 (1.3)	30 (1.2)	39 (1.5)	35 (1.1)	21 (1.4)	
Unspecified	3,139 (32.6)	994 (39.6)	785 (30.2)	775 (25.3)	585 (40.1)	
Morphology, n (%)						
MM NOS	3,122 (32.4)	954 (38.0)	798 (30.7)	844 (27.5)	526 (36.0)	<0.00
Superficial Spreading Melanoma	4,129 (42.8)	887 (35.3)	1,221 (47.0)	1,367 (44.6)	654 (44.8)	
Nodular Melanoma	1,578 (16.4)	436 (17.4)	387(14.9)	561 (18.3)	194 (13.3)	
MM in lentigo maligna	466 (4.8)	124 (4.9)	109 (4.2)	187 (6.1)	46 (3.2)	
Other[+]	347 (3.6)	111 (4.4)	84 (3.2)	106 (3.5)	40 (2.7)	

* = Results under 5 are not released from SAIL via disclosure control policies, to ensure privacy protection adherence.

+ = Balloon cell melanoma, Regressing melanoma, Amelanotic melanoma, MM in junctional naevus, Acral lentigous MM, Desmoplastic melanoma, MM in giant pigment naevus, mixed epithelial and spindle cell, Epitheliod cell, Spindle cell NOS, Spindle Cell type A

Figure 2



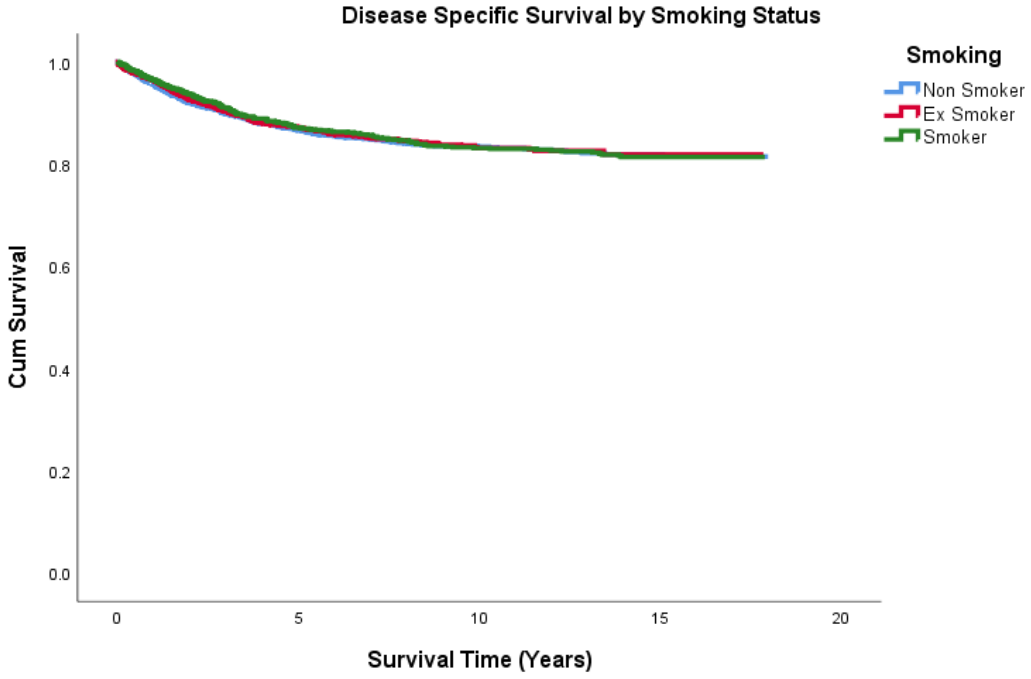


Figure 3

Figure 4

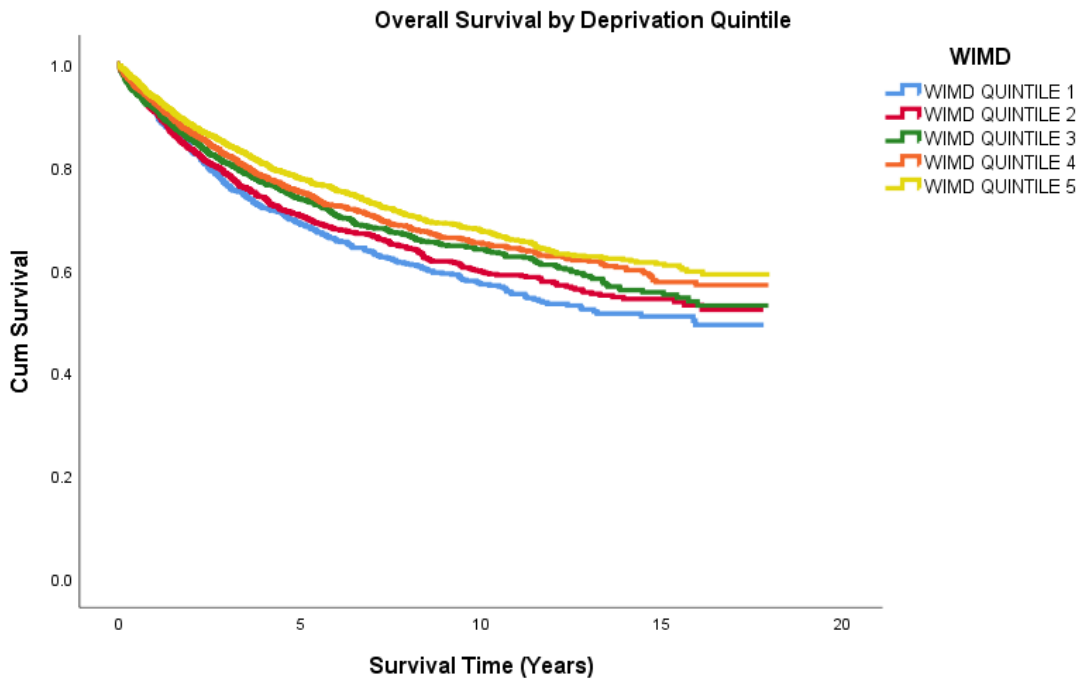


Figure 5

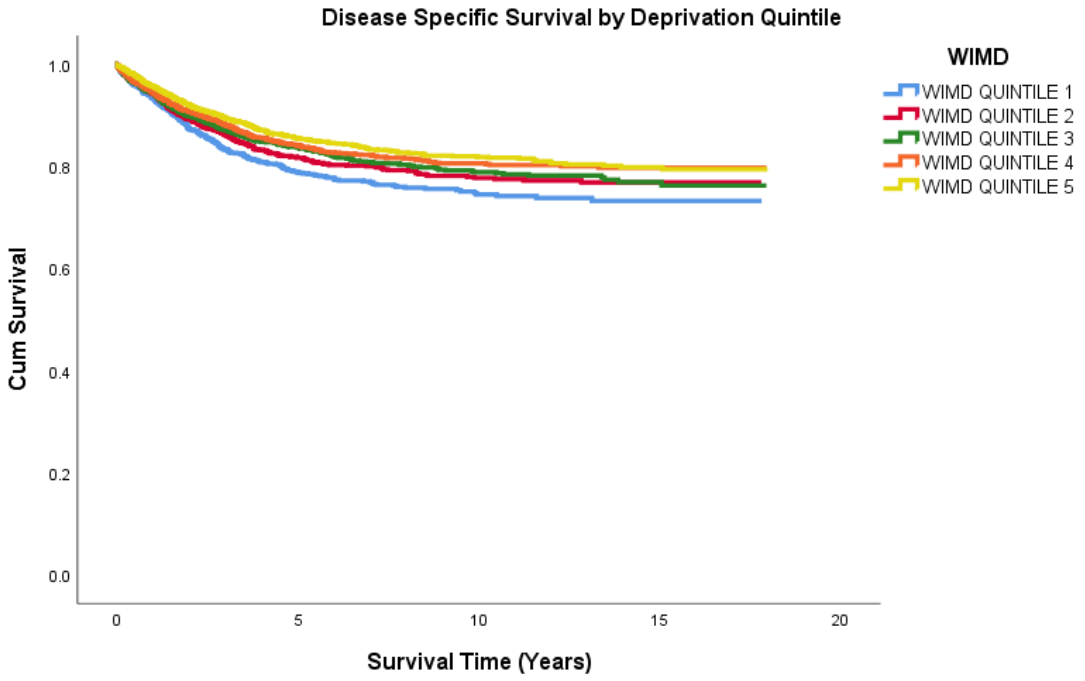


Table 5

Variable	Overall mortality			Disease specific mortality		
	P-Value	Hazard ratio (95% CI)		*P*-Value	Hazard Ratio (95% CI)	
Sex						
Female	*			*		
Male	0.00	1.28	(1.13 - 1.46)	0.01	1.35	(1.12 - 1.62)
Smoking status						
Non Smoker	*			*		
Ex-Smoker	0.93	1.00	(0.87 - 1.14)	0.20	0.88	(0.73 - 1.07)
Smoker	0.03	1.31	(1.09 - 1.55)	0.25	1.15	(0.91 - 1.45)
WIMD Quintile						
1 (Lowest socioeconomic status)	*			*		
2	0.75	0.97	(0.80 - 1.18)	0.93	0.99	(0.75 - 1.30)
3	0.01	0.78	(0.65 - 0.95)	0.09	0.79	(0.60 - 1.04)
4	0.04	0.75	(0.62 - 0.91)	0.08	0.78	(0.59 - 1.03)
5 (highest socioeconomic status)	0.00	0.67	(0.56 -0.81)	0.01	0.69	(0.53 – 0.90)
Charlson Co-morbidity Index	0.08	1.01	(1.00-1.02)	0.517	1.00	(0.99-1.02)
Location						
Trunk	*			*		
Lower Limb	0.10	0.86	(0.72 - 1.02)	0.00	0.79	(0.63 - 1.01)
Upper Limb	0.01	0.73	(0.61 - 0.88)	0.00	0.62	(0.48 - 0.79)
Head & Neck	0.48	0.94	(0.80 - 1.11)	0.06	0.80	(0.63 - 1.01)
Unspecified	0.28	1.21	(0.86 - 1.70)	0.83	1.05	(0.67 - 1.64)
Stage						
1	*			*		
2	0.00	2.48	(2.15 - 2.86)	0.00	6.24	(4.95 - 7.88)
3	0.00	3.65	(2.96 - 4.59)	0.00	11.48	(8.52 - 15.48)
4	0.00	11.78	(8.76 - 15.53)	0.00	32.55	(22.73 - 46.61)
Age**	0.00	1.06	(1.05 - 1.06)	0.00	1.02	(1.02 - 1.03)
Morphology						
Superficial Spreading Melanoma	*			*		
Nodular Melanoma	0.96	1.15	(0.98 - 1.35)	0.08	1.23	(0.98 - 1.54)
MM in lentigo maligna	0.70	1.05	(0.81 - 1.37)	0.02	0.43	(0.21 - 0.89)

Other[+]	0.12	1.25	(0.95 - 1.67)		0.50	1.16	(0.76 - 1.74)
Unspecified	0.01	1.24	(1.05 - 1.47)		0.04	1.28	(1.01 - 1.62)

* = Reference group

** = Age was included as a continuous variable in the model

Appendix

List of Read codes used for smoking status in this study.

Some codes (suffixed with %) have been presented as a set of codes under a wildcard.

Categories are S smoker, E ex-smoker, N never-smoker

Read codes	**Description**	**Category**	**Notes**
1371.	Never smoked tobacco	N	
9kn..	Non-smoker annual review - enhanced services administration	N	
137K.	Stopped smoking	E	
137L.	Current non-smoker	E	
137N.	Ex pipe smoker	E	
137O.	Ex cigar smoker	E	
137S.	Ex smoker	E	
137T.	Date ceased smoking	E	
1377.	Ex-trivial smoker (< 1 per day)	E	
1378.	Ex-light smoker (1 - 9 per day)	E	
1379.	Ex-moderate smoker (10 - 19 per day)	E	
137A.	Ex-heavy smoker (20 - 39 per day)	E	
137B.	Ex-very heavy smoker (40 + per day)	E	
137F.	Ex-smoker - amount unknown	E	
137i.	Ex tobacco chewer	E	
137j.	Ex-cigarette smoker	E	
137K0	Recently stopped smoking	E	
9km..	Ex-smoker annual review - enhanced services administration	E	
13p4.	Smoking free weeks	E	
137l.	Ex roll-up cigarette smoker	E	
745H%	(Various) Smoking cessation therapy	S	
du3%	(Various) Nicotine replacement therapy	S	
du6%	(Various) Bupropion	S	

du7%	(Various) additional nicotine replacement therapy	S	
du8%	(Various) Varenicline	S	
du9%	(Various) Nicotine withdrawal products	S	
E251%	(Various) tobacco dependence	S	
137..	Tobacco consumption	S	All with EVENT_VAL greater than 0
137Z	Tobacco consumption NOS	S	
137X.	Cigarette consumption	S	
137Y.	Cigar consumption	S	
137E.	Tobacco consumption unknown	S	
137g.	Cigarette pack years	S	
1372.	Trivial smoker - < 1 per day	S	
1373.	Light smoker - 1-9 per day	S	
1374.	Moderate smoker - 10-19 per day	S	
1375.	Heavy smoker - 20-39 per day	S	
1376.	Very heavy smoker - 20-39 per day	S	
137a.	Pipe tobacco consumption	S	
137b.	Ready to stop smoking	S	
137C.	Keeps trying to stop smoking	S	
137c.	Thinking about stopping smoking	S	
137e.	Smoking restarted	S	
137G.	Trying to give up smoking	S	
137H.	Pipe smoker	S	
137J.	Cigar smoker	S	
137M.	Rolls own cigarettes	S	
137P.	Cigarette smoker	S	
137Q.	Smoking started	S	
137R.	Current smoker	S	
137V.	Smoking reduced	S	
137D.	Admitted tobacco cons untrue?	S	
137d.	Not interested in stopping smoking	S	
137f.	Reason for restarting smoking	S	

137h.	Minutes from waking to first tobacco consumption	S	
6791.	Health ed. - smoking	S	
67910	Health education - parental smoking	S	
137m.	Failed attempt to stop smoking	S	
13p..	Smoking cessation milestones	S	
13p0.	Negotiated date for cessation of smoking	S	
13p8.	Lost to smoking cessation follow-up	S	
38DH.	Fagerstrom test for nicotine dependence	S	
67A3.	Pregnancy smoking advice	S	
67H1.	Lifestyle advice regarding smoking	S	
67H6.	Brief cessation for smoking cessation	S	
8B2B.	Nicotine replacement therapy	S	
8B3f.	Nicotine replacement therapy provided free	S	
8B3Y.	Over the counter nicotine replacement therapy	S	
8BP3.	Nicotine replacement therapy provided by community pharmacist	S	
8CAg.	Smoking cessation advice provided by community pharmacist	S	
8CAL.	Smoking cessation advice	S	
8CdB.	Stop smoking service opportunity signposted	S	
8H7i.	Referral to smoking cessation advisor	S	
8HBM.	Stop smoking face to face follow-up	S	
8HkQ.	Referral to NHS stop smoking service	S	
8HTK.	Referral to stop-smoking clinic	S	
8I2I.	Nicotine replacement therapy contraindicated	S	
8I2J.	Bupropion contraindicated	S	
8I39.	Nicotine replacement therapy refused	S	
8I3M.	Bupropion refused	S	
8I6H.	Smoking review not indicated	S	
8IAj.	Smoking cessation advice declined	S	
8IEK.	Smoking cessation program declined	S	

8IEM.	Smoking cessation drug therapy declined	S	
9hG..	Exception reporting: smoking quality indicators	S	
9hG0.	Excepted from smoking quality indicators: Patient unsuitable	S	
9hG1.	Excepted from smoking quality indicators: Informed dissent	S	
9kc..	Smoking cessation - enhanced services administration	S	
9kc0.	Smoking cessatn monitor template complet - enhanc serv admin	S	
9ko..	Current smoker annual review - enhanced service admin	S	
9N2k.	Seen by smoking cessation advisor	S	
9N4M.	DNA - did not attend smoking cessation clinic	S	
9Ndg.	Declined consent for follow-up by smoking cessation team	S	
9NdV.	Consent given follow-up after smoking cessation intervention	S	
9NdW.	Consent given for smoking cessation data sharing	S	
9NdY.	Declin cons follow-up evaluation after smoking cess interven	S	
9NdZ.	Declined consent for smoking cessation data sharing	S	
9NS02	Referral for smoking cessation service offered	S	
9OO..	Attends stop smoking monitor admin	S	
9OO1.	Attends stop smoking monitor	S	
9OO2.	Refuses stop smoking monitor	S	
9OO3.	Stop smoking monitor default	S	
9OO4.	Stop smoking monitor 1st lettr	S	
9OO5.	Stop smoking monitor 2nd lettr	S	
9OO6.	Stop smoking monitor 3rd lettr	S	
9OO7.	Stop smoking monitor verb.inv.	S	
9OO8.	Stop smoking monitor phone inv	S	

9OO9.	Stop smoking monitoring delete	S	
9OOA.	Stop smoking monitor check.done	S	
9OOB.	Stop smoking invitation short message service text message	S	
9OOB0	Stop smoking invitation first SMS text message	S	
9OOB1	Stop smoking invitation second SMS text message	S	
9OOB2	Stop smoking invitation third SMS text message	S	
9OOZ.	Stop smoking monitor admin.NOS	S	
E023.	Nicotine withdrawal	S	
J0364	Tobacco deposit on teeth	S	
SMC.	Toxic effect of tobacco and nicotine	S	
TJHy2	Adverse reaction to nicotine	S	
U6099	[X] Bupropion causing adverse effects in therapeutic use	S	
ZV4K0	[V] Tobacco use	S	
ZV6D8	[V] Tobacco abuse counselling	S	
13p5.	Smoking cessation programme start date	S	
9ko.	Current smoker annual review - enhanced service admin	S	

Appendix 2

The RECORD statement – checklist of items, extended from the STROBE statement, that should be repo routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items
Title and abstract				
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The typ should be specified in t abstract. When possibl the databases used sho RECORD 1.2: If applic geographic region and which the study took p reported in the title or a

				RECORD 1.3: If linka databases was conduct this should be clearly s or abstract.
Introduction				
Background rationale	2	Explain the scientific background and rationale for the investigation being reported		
Objectives	3	State specific objectives, including any prespecified hypotheses		
Methods				
Study Design	4	Present key elements of study design early in the paper		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection		

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 *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed		RECORD 6.1: The me population selection (s algorithms used to iden should be listed in deta possible, an explanatio provided. RECORD 6.2: Any val the codes or algorithms the population should b validation was conduct and not published elsev methods and results sh RECORD 6.3: If the st linkage of databases, c flow diagram or other g to demonstrate the data including the number o with linked data at eac

		Case-control study - For matched studies, give matching criteria and the number of controls per case		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A comp and algorithms used to exposures, outcomes, c effect modifiers should these cannot be reporte should be provided.
Data sources/ measurement	8	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		
Bias	9	Describe any efforts to address potential sources of bias		

Study size	10	Explain how the study size was arrived at		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If		

		applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses		
Data access and cleaning methods		..		RECORD 12.1: Autho describe the extent to v investigators had acces population used to crea population. RECORD 12.2: Autho information on the data methods used in the stu
Linkage		..		RECORD 12.3: State v

				included person-level, level, or other data link or more databases. The linkage and methods o evaluation should be pr
Results				
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Descri selection of the persons study (*i.e.,* study popul including filtering base quality, data availabilit The selection of includ be described in the text of the study flow diagr
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic,		

		clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures		

Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses		
Discussion				

Key results	18	Summarise key results with reference to study objectives		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discus implications of using d created or collected to specific research quest discussion of misclassi unmeasured confoundi and changing eligibilit they pertain to the stud
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence		
Generalisability	21	Discuss the generalisability (external validity) of the study		

		results		
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		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Autho information on how to supplemental informat study protocol, raw dat programming code.

Table 6 Overall and disease specific survival rates by smoking status

	Overall Survival			Disease Specific Survival		
	One year (%)	Five Year (%)	Ten year (%)	One year (%)	Five Year (%)	Ten year (%)
Non Smoker	94.1	79.8	70.5	95.6	86.3	85.9
Ex-Smoker	93.7	75.3	61.7	96.3	86.5	82.5
Current Smoker	95.5	80.7	70.7	96.4	86.9	82.9

Table 7 Overall and disease specific survival rates by socioeconomic status.

	Overall Survival			Disease Specific Survival		
	One year (%)	Five Year (%)	Ten year (%)	One year (%)	Five Year (%)	Ten year (%)
WIMD Quintile 1 (Lowest socioeconomic status)	90.5	68.3	56.1	93.2	78.3	73.9
WIMD Quintile 2	90.6	70	58.6	94.1	81.3	76.9
WIMD Quintile 3	91.1	73.2	62.9	94.2	83.3	78.3
WIMD Quintile 4	92.2	74.6	64.2	94.2	83.7	79.9
WIMD Quintile 5 (highest socioeconomic status)	93.4	77.2	66.6	95.5	85	81.1