

ORIGINAL ARTICLE

Medical specialties and life expectancy: An analysis of doctors' obituaries 1997–2019

Adam B. Brayne¹  | Ralph P. Brayne | Alexander J. Fowler²

¹ Department of Anaesthesia, North Devon District Hospital, Northern Devon Healthcare NHS Trust, Barnstaple, UK

² Critical Care and Peri-operative Medicine Research Group, Queen Mary University London, London, UK

Correspondence

Adam B. Brayne, Department of Anaesthesia, North Devon District Hospital, Northern Devon Healthcare NHS Trust, Raleigh Heights, Barnstaple EX31 4JB, UK.
Email: adam.brayne@nhs.net

An overview of this work was presented at the NHS-R programming language national meeting in November 2020.

Funding information

NIHR Doctoral Research Fellowship, Grant/Award Number: DRF-2018-11-ST2-062

Abstract

Background: Previous studies have found emergency medicine physicians have a reduced life expectancy compared to other doctors, using small subsets of data from the obituary section of the British Medical Journal. Technological advances now allow the entire catalogue of obituaries to be interrogated, which allows exploration of the relationship between medical specialty, age at death and cause of death in doctors.

Methods: Publicly available electronic records were obtained by web scraping and analysed with natural language processing algorithms. Obituaries published in the British Medical Journal between January 1997 and August 2019 were scraped and analysed for differences in age and cause of death and also relative survival analysis compared to the general U.K. population.

Results: Data were extracted from 8156 obituaries. The specialties with the oldest average age at death were general practitioners (80.3, *SD* = 12.5, *n* = 2508), surgeons (79.9, *SD* = 13.6, *n* = 853) and pathologists (79.8, *SD* = 13.8, *n* = 394). The specialties with the youngest average age at death were emergency physicians (58.7, *SD* = 23.6, *n* = 43), anaesthetists (75.5, *SD* = 16.1, *n* = 473) and radiologists (75.8, *SD* = 14.5, *n* = 172). Cancer was the most common cause of death and did not differ by specialty. Doctors on average have an older age at death than the general U.K. population.

Conclusions: A doctor's specialty has a significant association with their age at death, with general practitioners living the longest and emergency physicians the shortest, with proportionately more accidental deaths. Likely due to its recency as a separate specialty, the emergency physician group is the smallest, which may censor and falsely reduce this group's age at death. The observed increased life expectancy and the reduced cardiovascular disease in this cohort may be associated with lifestyle and socioeconomic factors.

KEYWORDS

epidemiology, life expectancy, life style

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Lifestyle Medicine* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Historically, doctors have been involved in epidemiology and public health not only as researchers but also participants.¹ As early as 1949, it was noted that the commonest causes of death amongst doctors were different from the general population.² However, there is little contemporary information regarding the life expectancy and cause of death of doctors in the United Kingdom.

The British Medical Journal's obituary column details the lives and deaths of doctors with a connection to the United Kingdom on a weekly basis, with its electronic archives containing over 20 years' worth of obituaries. This is a free service and deceased doctors need not be a member of the BMA or live or work in the United Kingdom. No obituaries are refused, and every submitted obituary is published online. Thus, the British Medical Journal (BMJ) website holds a collection of memories of doctors from a range of backgrounds. Within these memories is a wealth of epidemiological information which, although difficult to extract, provides interesting data. Subsets of these data have been analysed previously. Wright et al. used a sample of 572 obituaries, finding that doctors born in the Indian subcontinent die earlier than those born in the United Kingdom.³ Patel et al. used 3342 obituaries to describe that primary care doctors can expect to live 20 years longer than an emergency medicine (EM) doctor.⁴

Governments, institutions and companies are increasingly storing data online, allowing researchers easy access to collected data. Some websites facilitate access to this with an Application Program Interface (API); however, most do not. Web scraping is a specialist programming technique used in a variety of industries to automatically collect data from websites in bulk. Once bulk data are obtained, Natural Language Processing (NLP) can be applied to extract data variables. NLP refers to the method by which machines can be instructed on how to detect target variables stored in continuous prose rather than in a machine-friendly delimited format. Together, web scraping and natural language processing allow the procurement and interrogation of cumbersome online datasets.

This study aims to apply web scraping and natural language processing techniques to the BMJ obituary archives to produce a descriptive analysis of the data held in this large, publicly available dataset.

2 | METHODS

2.1 | Design and setting

This retrospective analysis of publicly available data was performed with the permission of the BMJ. Obituary pages were digitally 'scraped' on 7 August 2019 using an algorithm written in the Python (version 3.6) computing language.⁵

2.2 | Participants

Obituaries have been published by the BMJ since 1852. Entries are submitted by friends or family of the deceased within 1 year of the date

of death. Smaller numbers of obituaries about prominent members of the medical community are commissioned by the BMJ. The majority are U.K.-based doctors; however, occasional international and non-clinical submissions are published. Since 1997, obituaries have been available in a machine-readable format (in hypertext markup language). Prior to this, they were only available in print or as pdf documents. All obituaries published on the BMJ website in machine readable form until 7 August 2019 were eligible for inclusion.

2.3 | Variables

Natural language processing was used to automatically extract variables from each obituary entry. Extracted variables were as follows: gender, date of birth, cause of death, specialty, professional memberships and year and place of qualification of the deceased. In order to analyse these variables, they were categorised: Cause of death was assigned according to the NHS atlas of risk which has 18 categories. Specialties were put into 10 groups according to royal colleges: physicians, surgeons, GPs, anaesthetists, paediatricians, EM, obstetricians and gynaecologists, pathologists, psychiatrists and ophthalmologists. If multiple specialties were present in a given record, the main specialty was assigned based on a hierarchical model. First, any specialty preceded by 'consultant' was considered the main specialty, in accordance with U.K. practice. Secondly, if no consultant specialties were present, then we assumed the last recorded specialty to be the main specialty.

A sample of 1% of the automatically extracted variables were checked by a clinician and error rates for both incorrect and missed assignments were recorded.

2.4 | Outcomes

The primary outcome measure was age at death. As the exact date of death was inconsistently recorded, conservative age at deaths was used by subtracting 1 year from difference between the year of publication and the year of birth, given consistent reporting of year of birth.

2.5 | Statistical analysis

Demographic data were calculated and compared according to specialty. Averages were calculated as means with standard deviation unless otherwise stated. Analysis of variance (ANOVA) was used for mean age at death and Tukey's Honest Significant Difference (HSD) test was used to determine which groups had increased or decreased mean ages. Sensitivity analysis using multiple linear regression allowed the effect of assigning one main specialty for doctors with multiple specialties over their careers to be tested.

To avoid the pitfalls of performing risk analysis without a denominator, non-parametric relative survival analysis was estimated using the 'relsurv' package in R.⁶ Relative survival is defined as the ratio of the proportion of observed survivors in this cohort to the proportion of expected survivors in the general population. It is calculated by

TABLE 1 Completeness of automatic data extraction and manual check

Variable	Automatic extraction	False extraction	Missed extraction
Year of birth	8024/8156 (98.4%)	0/81 (0.0%)	0/81 (0.0%)
Year of publication	8156/8156 (100.0%)	0/81 (0.0%)	0/81 (0.0%)
Gender	8156/8156 (100.0%)	0/81 (0.0%)	0/81 (0.0%)
Year of qualification	7552/8156 (92.6%)	2/81 (2.5%)	2/81 (2.5%)
Specialty	7477/8156 (91.7%)	8/81 (9.9%)	2/81 (2.5%)
Cause of death	4345/8156 (53.3%)	9/81 (11.1%)	2/81 (2.5%)

Note: Completeness of automatic data extraction, and corresponding results from manual check of 1% of results. Details regarding cause of death were frequently omitted from obituaries. Automatic extraction refers to the proportion of cases where the algorithm assigned the variable. False extraction means that the automated process incorrectly assigned that variable. Missed extraction means that the variable was recorded in the obituary but was not extracted. Results are given as number/total (percentage).

dividing the percentage of the cohort who are still alive at the end of each period of time by the percentage of people in the U.K. general population of the same sex, age, and year of birth who are alive at the end of the same time period.⁷ The U.K. death table data were used as the reference cohort as the obituary section is intended for doctors with a U.K. connection, though not all of the deceased will have lived in the United Kingdom. The relative survival analysis shows whether being a doctor is associated with a different life expectancy compared to the general population. This approach is more usually applied to cancer registries; however, in this case instead of date of diagnosis, date of qualification is used to mark the start of the exposure. This approach considers the population mortality hazard of the baseline population, in which doctors make up an insignificant proportion. The resulting difference in mortality can be associated with becoming a doctor. Within the 'relSurv' package the 'rs.surv' function, the 'Ederer 2' model was used due to its more conservative estimates and robustness over long time periods.⁸

3 | RESULTS

A total of 8156 BMJ obituaries were obtained from the BMJ website spanning more than 22 years between 1997 and 2019. In this period, there were 6395 obituary webpages, of which we collected data from 6310; 85 of the webpages were not available to read on the date of our analysis. Missing data may be due to obituaries omitting details or failure of automatic data extraction logic (see Table 1 for breakdown).

There were significant differences in the mean age at death amongst different specialties (see Table 2). In a linear regression model (ANOVA and subsequent Tukey test), emergency physicians have a significantly reduced mean age at death compared to all other specialties. Sensitivity analysis using multiple linear regression showed this relationship persisted for doctors who had multiple specialties ascribed to them but had worked in EM during their career ($n = 89$). This cohort are also shown as a separate column in Table 2. Further subgroup analysis used the country of qualification to show comparable trends with U.K. qualified doctors only.

Relative survival analysis (Figure 1A) shows that, compared to U.K. contemporaries born in the same year, this cohort has a survival advan-

tage. This is evident 4 years after qualification with a tiny effect size (relative survival ratio 1.004 (95% confidence interval, 1.001–1.003)) but increases exponentially with time after qualification. This trend holds in all specialties (Figure 1B) except EM whose survival trend is below that of the reference population.

Cancer is the leading cause of death, causing 3191 of the 8156 (39.1%) deaths in this sample, and remains the leading cause regardless of sub-categorisation by specialty. The most common national cause of death, heart and circulatory disorders, is the second most common cause of death in all specialties, but EM for which it is non-transport accidents.

4 | DISCUSSION

There is a significant difference in age at death according to a doctor's specialty. Exposure to an emergency, anaesthetic, paediatric, radiology or psychiatry job is associated with earlier age at death, whereas pathology, surgery and GP were associated with an older age at death. The largest difference is seen in EM, with a younger mean age at death of 58.7 years and an increased proportion of accidental deaths than other specialties. The most common cause of death across all specialties according to the NHS risk atlas (cancer) differs from that of the general U.K. population (heart and circulatory disorders). On average, this cohort lives longer than the general population. Relative survival analysis demonstrates that this cohort shares the same risk as age- and sex-matched members of the U.K. population for the first 4 years of practice. Many years after qualification, there is a substantial survival advantage amongst all specialties aside from EM.

There are general lifestyle hazards associated with the profession relating to mental health and suicide, with suicide by self-poisoning and cutting more common than the general population.⁹ Furthermore specialty-specific hazards include the burden of complaints, radiation exposure, exposure to harmful gases and shift patterns.^{9–11} There is recent evidence that trainees in acute specialties, particularly EM, have very high rates of burnout.¹² The concerning risk of early death amongst EM doctors was first described a decade ago based on a much smaller subset of these data, containing only 17 EM doctors. Our study reconfirms this finding using our larger dataset, with 43 EM doctors.

TABLE 2 Demographic details by specialty

	n	Mean age at death in years (SD)	Causes of death (%)			%	
			#1	#2	#3	Female	Male
All specialties	8156	78.9 (14.1)	Cancer (39.1)	Heart and circulatory (26.7)	Nervous system (9.1)	15.5	84.5
Medical	1639	78.6 (14.2)	Cancer (38.7)	Heart and circulatory (26.1)	Nervous system (10.1)	11.5	88.5
Surgical	853	79.9 (13.6)	Cancer (39.4)	Heart and circulatory (27.4)	Infection (9.8)	3.4	96.6
Primary care	2508	80.3 (12.5)	Cancer (39.2)	Heart and circulatory (27.2)	Nervous system (9.6)	15.7	84.3
Anaesthesia	473	75.5 (16.1)	Cancer (41.2)	Heart and circulatory (27.6)	Infection (7.2)	18.6	81.4
Emergency medicine	43	58.7 (23.6)	Cancer (42.9)	Accidents (non-transport) (14.3)	Respiratory (14.3)	25.6	74.4
Obstetrics and gynaecology	396	78.7 (14.9)	Cancer (35.9)	Heart and circulatory (29.4)	Infection (10)	25.3	74.7
Paediatrics	397	76.1 (15.5)	Cancer (45.9)	Heart and circulatory (24.5)	Infection (7.7)	36.0	64.0
Radiology	172	75.8 (14.5)	Cancer (39.3)	Heart and circulatory (29.2)	Nervous system (11.2)	11.0	89.0
Pathology	394	79.8 (13.8)	Cancer (37.9)	Heart and circulatory (23.7)	Infection (10.6)	15.5	84.5
Psychiatry	460	76.5 (15.3)	Cancer (39.4)	Heart and circulatory (23.8)	Nervous system (8.6)	18.9	81.1
Ophthalmology	142	78.6 (14.5)	Cancer (44.2)	Heart and circulatory (26.7)	Infection (10.5)	10.6	89.4
Unknown specialty	679	80.9 (14.1)	Cancer (34.8)	Heart and circulatory (29.3)	Infection (9.1)	18.4	81.6
Emergency medicine at any time	89	71.3 (22.6)	Cancer (18)	Heart and circulatory (10.1)	Infection (6)	20.2	79.8

Note: Demographic details by specialty. Those in the emergency medicine specialty had the lowest mean age at death and a higher standard deviation. Cancer is the most common cause of death in this cohort, followed by heart and circulatory causes for all groups except emergency medicine for which it is accidental death (non-transport). The 679 people we were unable to assign a specialty to are included in the unknown specialty column and remain included in the 'all specialties' column. For comparison, a separate row is appended showing the demographic details for doctors who worked in emergency medicine at any point; this column duplicates cases held in other columns (e.g. it contains all doctors from the emergency medicine column plus 46 from the other columns). This group contains a higher than typical number of unknown causes of death.

However, these results should be interpreted with caution as EM as a specialty is only 50-year old, and so the population structure of emergency medics is different to that of other specialties. The fact that the specialty is new is the reason for the low numbers of emergency medics in the obituaries, and thus leads to censoring of older deaths, as they are yet to occur. For the other specialties with a reduced age at death, it can be reasonably assumed that the population structure is stable, given they are more established specialties. In relative survival analysis, which is compared to age matched U.K. death rates, emergency physicians are the only specialty to have a higher risk (reduced survival ratio) compared to the general population. For all other specialties 40 years after qualification relative survival compared to the general population increases dramatically, and this persists over time. The observed relative survival benefit for this cohort versus the general population does not imply causation. These outcomes are also likely to reflect socio-economic factors on health, as this cohort is likely to be wealthier and more educated than the general population.

Specialty-specific differences in life expectancy and cause of death are also evident in longer established specialties. The average age of death amongst surgeons is significantly higher than anaesthetists and radiologists. The exact reasons for this are unknown and it is particularly interesting given that surgeons and anaesthetists share

a similar work environment. Potential specialty-specific risks include exposure to volatile gases and the higher risk of substance misuse potentially due to ease of access to addictive medications in anaesthetists, for example opioids.¹³⁻¹⁴ Prior to 1950, radiologists were at higher risk of blood and potentially solid cancers due to exposure to high levels of ionising radiation.¹⁵ The life expectancy in radiologists is lower in our data; however, cause of death data in the radiology obituaries do not bear this out as a potential mechanism for reduced life expectancy compared to other specialties, with radiology actually having relatively lower cancer rates. Retirement age has historically differed by specialty, with primary care doctors more likely to retire earlier and less likely to return or continue part time than hospital specialists.¹⁶ Retirement facilitates lifestyle changes associated with reduced cardiovascular disease and therefore reduced mortality, for example no shift working, increased exercise and reduced stress.¹⁷ Therefore the relatively high life expectancy in primary care may be less to do with a difference in occupational hazards but more to do with an earlier and more complete retirement.

This study has a number of limitations. It is observational and cannot demonstrate causation; however, it is well documented that practicing medicine comes with occupational hazards. Furthermore, selection bias is a limitation due to the voluntary nature of submissions

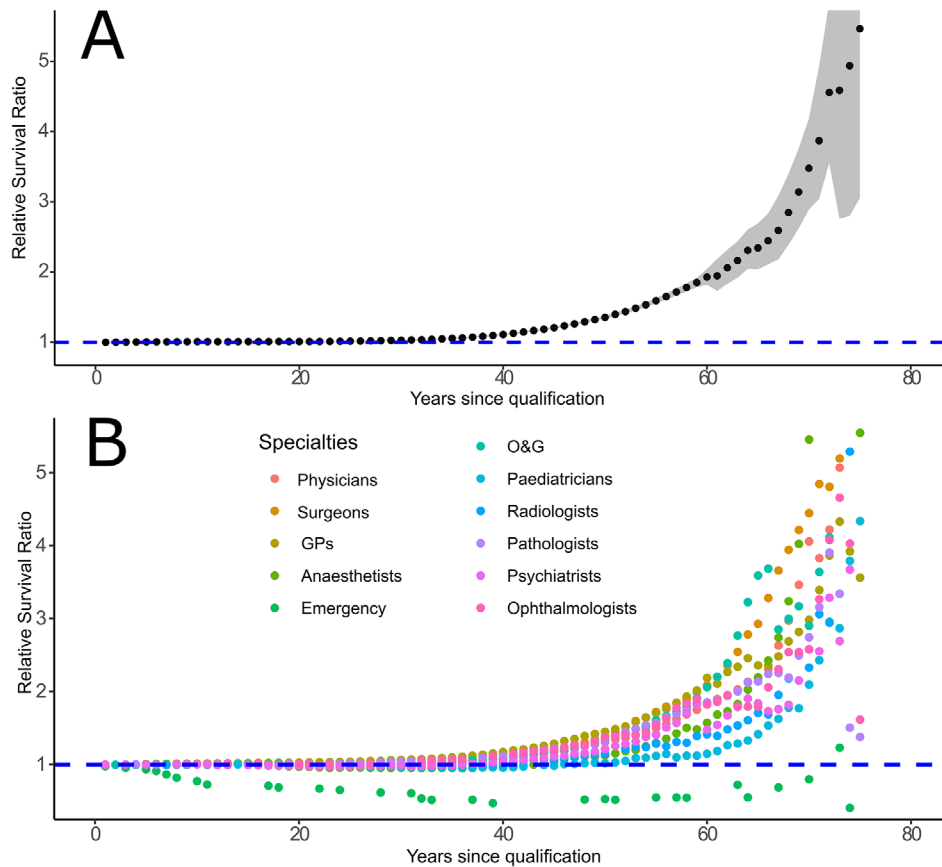


FIGURE 1 Relative survival ratio for the BMJ obituary cohort versus U.K. general population since qualification, corrected for age and year of birth (blue dashed line). (A) Relative survival ratio for all doctors with 95% confidence interval shaded in grey. (B) Relative survival ratio by specialty; note that emergency medicine has values below 1 and hence has a higher risk of death at that age

and open editorial policy, meaning that not every doctor who died in the United Kingdom has an obituary in the BMJ, and not every obituary in the BMJ is from a U.K.-based doctor. Given that the BMJ obituary data are not a formal data registry, the variables we sought to extract were inconsistently recorded. This is because the obituaries from which these data are extracted were not intended to be a repository of data, but a descriptive expression of memories, usually not written by the deceased and so prone to reporting errors. Simple demographic details were well recorded; however, specialty and cause of death were more poorly recorded. For example there is likely to be an underrepresentation of suicide in these data, as some obituaries for those who died by suicide are written in a way which suggests but does not explicitly state the cause of death. Even if fully recorded, our automated data extraction technique has an appreciable error rate in determining cause of death and specialty which should be considered when drawing conclusions.

This is the largest ever analysis of the causes of doctors' deaths in the United Kingdom. Using a novel approach of web scraping and natural language processing, we have demonstrated the value and limitations of automated data extraction. This approach gives benefits in terms of time efficiency, reproducibility and speed, but is not as accurate as manual extraction. As its application becomes more refined, and in combination with machine learning, its accuracy will improve

and it may become superior to manual data extraction in every way. Even now it is opening up the possibility to analyse datasets that were previously too unwieldy to consider and may make currently time-consuming tasks such as retrospective analyses of patient notes rapid and easy.

This study follows two previous notable manual analyses of the BMJ obituary columns, in which differences in life expectancy according to birth country and specialty were described.^{3,4} The data used by these studies form a subset of our data and add weight to the findings of Patel et al. regarding specialty-specific differences in age at death.⁴ At a time of increasing concern regarding doctors' welfare and the impact of the career on psychological and physical health, this study demonstrates that specialty-specific differences exist. A long-term registry of medical personnel, including lifestyle data, would allow these associations to be unpicked, with construction of a regression model to explore risk factors in more detail.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from BMJ. Restrictions apply to the availability of these data, which were used with permission for this study. Data are available from the authors with the permission of the BMJ.

ACKNOWLEDGEMENTS

The authors would like to thank the British Medical Journal (BMJ) for allowing us to scrape this information from their webpages.

AUTHOR CONTRIBUTIONS

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. ABB conceived the research concept, wrote analysis code, performed analyses and wrote the manuscript. RPB developed the research concept, wrote the scraping code, edited the manuscript and approved the final version for submission. AJF developed the research concept, wrote analysis code, edited the manuscript and approved the final version for submission. ABB affirms that the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

FUNDING INFORMATION

ABB and RPB had no financial support. AJF is supported by an NIHR Doctoral Research Fellowship (ref: DRF-2018-11-ST2-062). This work was completed in the authors' own time around existing commitments and no funding was received; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

CONFLICT OF INTEREST

The authors declare no conflict of interest. ABB and AJF are practising medical doctors in the United Kingdom.

ETHICAL APPROVAL

Ethical approval was not required for this study.

ORCID

Adam B. Brayne  <https://orcid.org/0000-0002-2000-9251>

REFERENCES

1. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ*. 2004;328(7455):1519.
2. Dickinson FG, Welker EL. The leading causes of death among physicians. *JAMA*. 1949;139(17):1129-1131.

3. Wright DJM, Roberts AP. Which doctors die first? Analysis of BMJ obituary columns. *BMJ*. 1996;313(7072):1581-1582.
4. Patel K, Yan YM, Patel K, et al. Lifespan and cardiology. *Br J Cardiol*. 2009;16(6):299-302.
5. Van Rossum G, Drake Jr FL. Centrum voor Wiskunde en Informatica Amsterdam. *Python Reference Manual*. Amsterdam: Centrum voor Wiskunde en Informatica; 1995.
6. Perme MP, Pavlic K. Nonparametric relative survival analysis with the R package relsurv. *J Stat Softw*. 2018;87(1):1-27.
7. Barbieri M, Wilmoth JR, Shkolnikov VM, et al. Data resource profile: the human mortality database (HMD). *Int J Epidemiol*. 2015;44(5):1549-1556.
8. Lambert PC, Dickman PW, Rutherford MJ. Comparison of different approaches to estimating age standardized net survival. *BMC Med Res Method*. 2015;15(1):64.
9. Hawton K, Clements A, Simkin S, Malmberg A. Doctors who kill themselves: a study of the methods used for suicide. *QJM*. 2000;93(6):351-357.
10. McNamee R, Keen RI, Corkill CM. Morbidity and early retirement among anaesthetists and other specialists. *Anaesthesia*. 1987;42(2):133-140.
11. Bourne T, Wynants L, Peters M, et al. The impact of complaints procedures on the welfare, health and clinical practise of 7926 doctors in the UK: a cross-sectional survey. *BMJ Open*. 2015;5(1), e006687.
12. Bailey J, Thomas C, McDowall A. 2018. 92 Grit and burnout in UK emergency medicine trainees. *BMJ Leader*. 2:A38.
13. Molina Aragonés JM, Ayora AA, Ribalta AB, et al. Occupational exposure to volatile anaesthetics: a systematic review. *Occup Med*. 2016;66(3):202-207.
14. Berry CB, Crome IB, Plant M, Plant M. Substance misuse amongst anaesthetists in the United Kingdom and Ireland: the results of a study commissioned by the Association of Anaesthetists of Great Britain and Ireland. *Anaesthesia*. 2000;55(10):946-952.
15. Yoshinaga S, Mabuchi K, Sigurdson AJ, Doody MM, Ron E. Cancer risks among radiologists and radiologic technologists: review of epidemiologic studies. *Radiology*. 2004;233(2):313-321.
16. Smith F, Goldacre MJ, Lambert TW. Retirement ages of senior UK doctors: national surveys of the medical graduates of 1974 and 1977. *BMJ Open*. 2018;8(6):e022475.
17. Steptoe A, Kivimäki M. Stress and cardiovascular disease. *Nat Rev Cardiol*. 2012;9(6):360-370.

How to cite this article: Brayne AB, Brayne RP, Fowler AJ. Medical specialties and life expectancy: An analysis of doctors' obituaries 1997-2019. *Lifestyle Med*. 2021;2:e23. <https://doi.org/10.1002/lim2.23>