



Assessment of Relative Utility of Underlying vs Contributory Causes of Death

G. David Batty, PhD, DSc; Catharine R. Gale, PhD; Mika Kivimäki, FMedSci; Steven Bell, PhD

Abstract

IMPORTANCE In etiological research, investigators using death certificate data have traditionally extracted underlying cause of mortality alone. With multimorbidity being increasingly common, more than one condition is often compatible with the manner of death. Using contributory cause plus underlying cause would also have some analytical advantages, but their combined utility is largely untested.

OBJECTIVE To compare the relative utility of cause of death data extracted from the underlying cause field vs any location on the death certificate (underlying and contributing combined).

DESIGN, SETTING, AND PARTICIPANTS This study compares the association of 3 known risk factors (cigarette smoking, low educational attainment, and hypertension) with health outcomes based on where cause of death data appears on the death certificate in 2 prospective cohort study collaborations (UK Biobank [N = 502 655] and the Health Survey for England [15 studies] and the Scottish Health Surveys [3 studies] [HSE-SHS; N = 193 873]). Data were collected in UK Biobank from March 2006 to October 2010 and in HSE-SHS from January 1994 to December 2008. Data analysis began in June 2018 and concluded in June 2019.

MAIN OUTCOMES AND MEASURES Death from cardiovascular disease, cancer, dementia, and injury. For each risk factor–mortality end point combination, a ratio of hazard ratios (RHR) was computed by dividing the effect estimate for the underlying cause by the effect estimate for any mention.

RESULTS In UK Biobank, there were 14 421 deaths (2.9%) during a mean (SD) of 6.99 (1.03) years of follow up; in HSE-SHS, there were 21 314 deaths (11.0%) during a mean (SD) of 9.61 (4.44) years of mortality surveillance. Established associations of risk factors with death outcomes were essentially the same irrespective of placement of cause on the death certificate. Results from each study were mutually supportive. For having ever smoked cigarettes (vs never having smoked) in the UK Biobank, the RHR for cardiovascular disease was 0.98 (95% CI, 0.87-1.10; P value for difference = .69); for cancer, the RHR was 0.99 (95% CI, 0.93-1.05; P value for difference = .69). In the HSE-SHS, the RHR for cardiovascular disease was 0.94 (95% CI, 0.87-1.01; P value for difference = .09); for cancer, it was 1.01 (95% CI, 0.94-1.10; P value for difference = .75).

CONCLUSIONS AND RELEVANCE Risk factor–end point associations were not sensitive to the placement of data on the death certificate. This has implications for the examination of the association of risk factors with causes of death where there may be too few events to compute reliable effect estimates based on the underlying field alone.

JAMA Network Open. 2019;2(7):e198024. doi:10.1001/jamanetworkopen.2019.8024

Key Points

Question What is the relative utility for etiological research of using underlying cause of mortality from death certificates compared with using any field?

Findings In 2 large cohort studies with 696 528 participants combined, associations of known risk factors with an array of death outcomes were essentially the same, irrespective of the location of the death data on the certificate.

Meaning Our findings have implications for risk factor–end point computations in studies of rarer outcomes and those with smaller sample sizes.

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

JAMA Network Open. 2019;2(7):e198024. doi:10.1001/jamanetworkopen.2019.8024

July 31, 2019 1/11

Introduction

Death records have long been collected for the purposes of monitoring the health of populations,^{1,2} quantifying disease prognosis,^{3,4} and evaluating the impact of primary⁵ and secondary interventions.⁶ To examine the influence of environmental and genetic characteristics on disease and injury events, mortality records have also been extensively deployed in ecological,⁷ case-control,⁸ experimental,⁹ and, most frequently, prospective cohort studies.¹⁰⁻¹² The use of death records as a proxy for a health end point of interest is particularly important in contexts where linkages to other electronic health registries, such as hospital data or cancer records, are not viable, and clinical examination of study members is financially or logistically prohibitive. Linkage to death registers has the further advantage of having no additional burden on study participants. Scientific endeavors in which death data have been central to the understanding of disease etiology include the Framingham studies, where hypertension was first shown to be a risk factor for heart disease¹³ and stroke¹⁴; the original Whitehall study,¹⁵ where it was demonstrated that elevated blood glucose within the normal range was related to vascular events; and the British Doctors Study¹⁶ where, perhaps most famously, smoking was first prospectively linked to lung cancer. There are numerous other examples.¹⁷

To accord with the World Health Organization guidelines,¹⁸ death certificates are formulated in 2 parts. For the purposes of epidemiological research, the underlying (ie, immediate or direct) cause of death is almost exclusively extracted. Other diseases or injuries that contributed to the death but were not directly implicated appear in another section of the certificate. However, in practice, this contributory information is very rarely used.¹⁹ With multimorbidity being common in an era of effective treatments, more than one condition can be compatible with the manner of death.²⁰ Analyses that use only the underlying cause of death may therefore omit valuable information that is readily available.

In estimating burden of disease, reliance on underlying cause compared with incorporating contributory causes appears to lead to underestimates of the importance of several leading causes of death.²¹ However, the impact for etiological research is largely unknown. In the only study of which we are aware,²² investigators found the same predictive capacity for classic risk factors in analyses featuring cardiovascular disease deaths irrespective of placement on the death certificate. No such comparison was made for other important causes of death. Using data from the contributory field of death may have the analytical advantage of facilitating investigation of the determinants of rarer causes of death (eg, intentional injury²³ and dementia²⁴) where, particularly in smaller cohort studies, a reliance on underlying cause alone may result in too few events to facilitate statistical computations. The value to investigators of larger studies of commonly occurring conditions might be enhanced statistical precision.

Using data from 2 large cohort studies, we examined associations of 3 known risk factors with major causes of death, including cancer, cardiovascular disease, injury, and dementia. To provide findings of interest to a range of disciplines, we used physiological (hypertension²⁵), psychosocial (educational attainment²⁶), and behavioral (cigarette smoking^{27,28}) risk factors. Our aim was to examine if these risk factors had the same magnitude of association with cause-specific mortality when end point data were extracted from the underlying field alone vs the underlying and contributory fields combined (ie, any mention).

Methods

Included Cohort Studies

We used data from UK Biobank,²⁹ a prospective cohort study, and a pooling of 18 identical cohort studies from the Health Survey for England (15 studies) and the Scottish Health Surveys (3 studies) (HSE-SHS).³⁰⁻³² These studies were selected because they offer similar, standard processes for data collection. Participants in both studies gave full informed consent. In the UK Biobank, ethical approval was received from the North West Multi-center Research Ethics Committee, and the

research was carried out in accordance with the Declaration of Helsinki.³³ In HSE-SHS, ethical approval for data collection was granted by the London research ethics council or the local research ethics councils. This study analyzed existing anonymized data, and therefore, no further ethical approval was required. This report follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Baseline Data Collection

The sampling and protocols of these studies have been well described.²⁹⁻³² In brief, baseline data collection in the UK Biobank took place between March 2006 and October 2010 in research assessment centers across the United Kingdom, with data collected on 502 655 people aged 40 to 69 years (response proportion, 6%). Between January 1994 and December 2008, a total of 193 842 people aged 16 to 102 years (response proportion, 64%-78%) participated in HSE-SHS, with data collection taking place exclusively in the home (Table).

Responses to history of cigarette smoking habits (ie, ever smoker vs never) and highest attained educational qualification (ie, no university degree vs ≥university undergraduate degree) were collapsed into binary categories for the purposes of presentation brevity. In UK Biobank, systolic and diastolic blood pressure measurements were taken twice while the participant was seated using the Omron HEM-7015IT digital blood pressure monitor (Omron Healthcare).²⁰ Blood pressure in the present analyses was based on the average of the 2 measurements. In HSE-SHS, blood pressure was measured using the Dinamap 8100 automated device (GE Critikon).³⁴ Following a 5-minute seated rest, 3 readings of systolic and diastolic blood pressure were taken from the right arm at 1-minute intervals. Blood pressure used in the present analysis was based on the mean of the second and third measurement. We defined hypertension according to existing guidelines as systolic/diastolic blood pressure of at least 140/90 mm Hg and/or use of antihypertensive medication.³⁵

Ascertainment of Cause-Specific Mortality

Study participants were flagged using the procedures of the UK National Health Service (NHS) Central Registry. Underlying and contributing deaths were classified according to the *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)*¹⁸ as follows: attributable to (1) all cancers combined (codes C00-C97), (2) lung cancer (C34), (3) cardiovascular disease (I20-5, I50, I60-70, I73, and I74), (4) coronary heart disease (I20-5), (5) cerebrovascular disease (I60-9), (6) external causes (V01-Y99), and (7) dementia (F00-F02, F03, F05, F10, G30, G31, I67 and A81).²⁴ Where necessary, corresponding codes from earlier revisions of the *ICD* were used. The any mention category was a combination of the underlying and contributory cause of death fields.

Table. Characteristics of Study Members in UK Biobank and HSE-SHS

Characteristic	No. (%)	
	UK Biobank	HSE-SHS
Participants recruited, No. ^a	502 655	193 873
Women	273 472 (54.4)	106 469 (54.9)
Age at baseline, mean (SD) [range], y	56.5 (8.1) [38-73]	46.8 (18.5) [16-102]
Ever smoked	225 896 (45.2)	98 581 (51.1)
No university education	331 298 (67.3)	147 353 (76.0)
Hypertension	282 637 (57.2)	43 428 (36.9)
Duration of mortality surveillance, mean (SD), y	6.99 (1.03)	9.61 (4.44)
Death from any cause	14 421 (2.9)	21 314 (11.0)

Abbreviation: HSE-SHS, Health Survey for England and the Scottish Health Surveys.

^a Number of study members and events are higher in this Table than in the survival analyses, which have missing data.

Statistical Analysis

Hazard ratios (HRs) and accompanying 95% CIs were computed using Cox regression models³⁶ and adjusted for age and sex. In these survival analyses, we censored individuals according to the date of death or the end of follow-up (February 22, 2016, for UK Biobank, February 14, 2011, for HSE, and December 31, 2009, for SHS), whichever came first. For a summary measure of the difference between the HRs based on the 2 approaches to classifying mortality, we computed a ratio of the HRs (RHR) as has been used elsewhere.^{37,38} We also computed the *P* value for difference between underlying vs any mention HRs for the 2 studies. To do so, we used the Fisher *z* score measure to test for equality of HRs, with *z* calculated as the difference between the logarithms of the HRs (β_1 , β_2) divided by the square root of the sum of the square of their variances ($SE1^2$, $SE2^2$), where *z* follows a normal distribution: $z = (\beta_1 - \beta_2) / \sqrt{SE1^2 + SE2^2}$.³⁹ No prespecified level of statistical significance was set, and all tests were 2-tailed. Analyses were conducted using Stata version 15 (Stata Corp) and took place from June 2018 to June 2019.

Results

In the Table, we show the characteristics of study members in UK Biobank and HSE-SHS at baseline and their mortality experience. The proportion of women in the baseline sample was similar in UK Biobank (273 472 [54.4%]) and HSE-SHS (106 469 [54.9%]). The prevalence of people who reported ever having smoked and those without a university education was somewhat higher in HSE-SHS than UK Biobank (smoking: 98 581 [51.1%] vs 225 896 [45.2%]; no university degree: 147 353 [76.0%] vs 331 298 [67.3%]), while hypertension was less common (43 428 [36.9%] vs 282 637 [57.2%]).

In UK Biobank there were 14 421 deaths among 502 655 people (2.9%) during a mean (SD) follow-up period of 6.99 (1.03) years, while in the pooling of HSE-SHS, a mean (SD) follow-up period of 9.61 (4.44) years of mortality surveillance gave rise to 21 314 deaths among 193 873 individuals (11.0%). In **Figure 1**, **Figure 2**, and **Figure 3**, we show numbers of deaths for each major cause of mortality as retrieved from the underlying cause field as well as in combination with the contributing field (ie, any mention). The number of study participants and deaths in the sample used in the Figures is marginally lower than the full cohort owing to some missing data for the exposures of interest. The number of deaths in the any mention group was necessarily higher for all conditions, a differential that was least pronounced for cancer, which may reflect dissemination of the primary malignancy (ie, underlying cause).

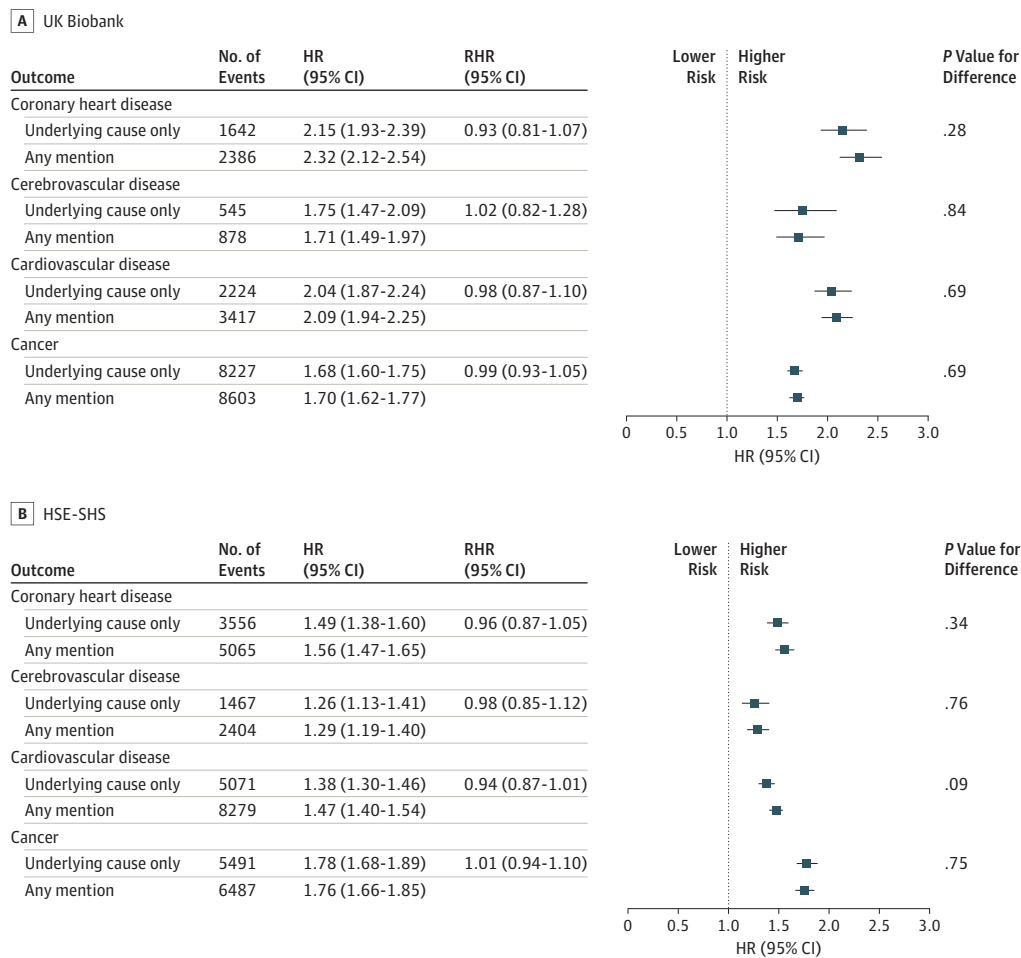
In Figure 1, we show the age-adjusted and sex-adjusted HRs for baseline cigarette smoking status associated with deaths from cardiovascular disease, its different presentations (ie, coronary heart disease and cerebrovascular disease), and all cancers combined. The expected associations were apparent, such that ever having smoked cigarettes was associated with an elevated rate of mortality from all conditions. Within studies, the size of the effect estimates was very similar, irrespective of whether the underlying or any mention field was used, so that the RHRs all hovered around unity (*P* value for difference $\geq .09$). For example, in the UK Biobank, the RHR for cardiovascular disease was 0.98 (95% CI, 0.87-1.10; *P* value for difference = .69); for cancer, the RHR was 0.99 (95% CI, 0.93-1.05; *P* value for difference = .69). In the HSE-SHS, the RHR for cardiovascular disease was 0.94 (95% CI, 0.87-1.01; *P* value for difference = .09); for cancer, it was 1.01 (95% CI, 0.94-1.10; *P* value for difference = .75). Similar observations were apparent for lung cancer, which, owing to the strong association with smoking in both UK Biobank (underlying cause: HR, 7.60; 95% CI, 6.50-8.88; any mention: HR, 7.84; 95% CI, 6.72-9.13) and HSE-SHS (underlying cause: HR, 10.38; 95% CI, 8.14-13.23; any mention: HR, 9.04; 95% CI, 7.33-11.15) could not be accommodated in the Figure alongside the expected weaker effect estimates for other outcomes. Owing to the higher numbers of deaths in the any mention group, statistical precision was somewhat higher, as evidenced by the tighter 95% CIs.

In Figure 2 we show HRs for the association of educational attainment with the same mortality outcomes featured in Figure 1, with the addition of dementia and external causes. As anticipated, a

more basic education, denoted by the absence of a university degree, was associated with an increased risk of mortality from all featured causes of death. The magnitude of these effect estimates was similar within the studies irrespective of the position of cause on the death certificate; thus, the RHRs ranged from 0.73 (95% CI, 0.45-1.18) for dementia in UK Biobank to 1.30 (95% CI, 0.70-2.44) for external causes in HSE-SHS, with a *P* value for difference > .20.

Last, in Figure 3 we depict the association of hypertension with mortality risk. People with hypertension experienced elevated rates of death from cardiovascular disease compared with their unaffected counterparts. Cross-study HRs were very similar with the RHRs all being close to unity and the *P* values for difference all being nonsignificant at conventional levels. Thus, the RHRs for cardiovascular disease in UK Biobank (1.04; 95% CI, 0.90-1.20; *P* value for difference = .60) and in HSE-SHS (1.00; 95% CI, 0.89-1.13; *P* value for difference = .95) were very similar, as they were for its subtypes of coronary heart disease (UK Biobank: RHR, 1.02; 95% CI, 0.86-1.20; *P* value for difference = .86; HSE-SHS: RHR, 0.96; 95% CI, 0.83-1.12; *P* value for difference = .59), and cerebrovascular disease (UK Biobank: RHR, 0.98; 95% CI, 0.75-1.29; *P* value for difference = .90; HSE-SHS: RHR, 0.90; 95% CI, 0.73-1.12; *P* value for difference = .36).

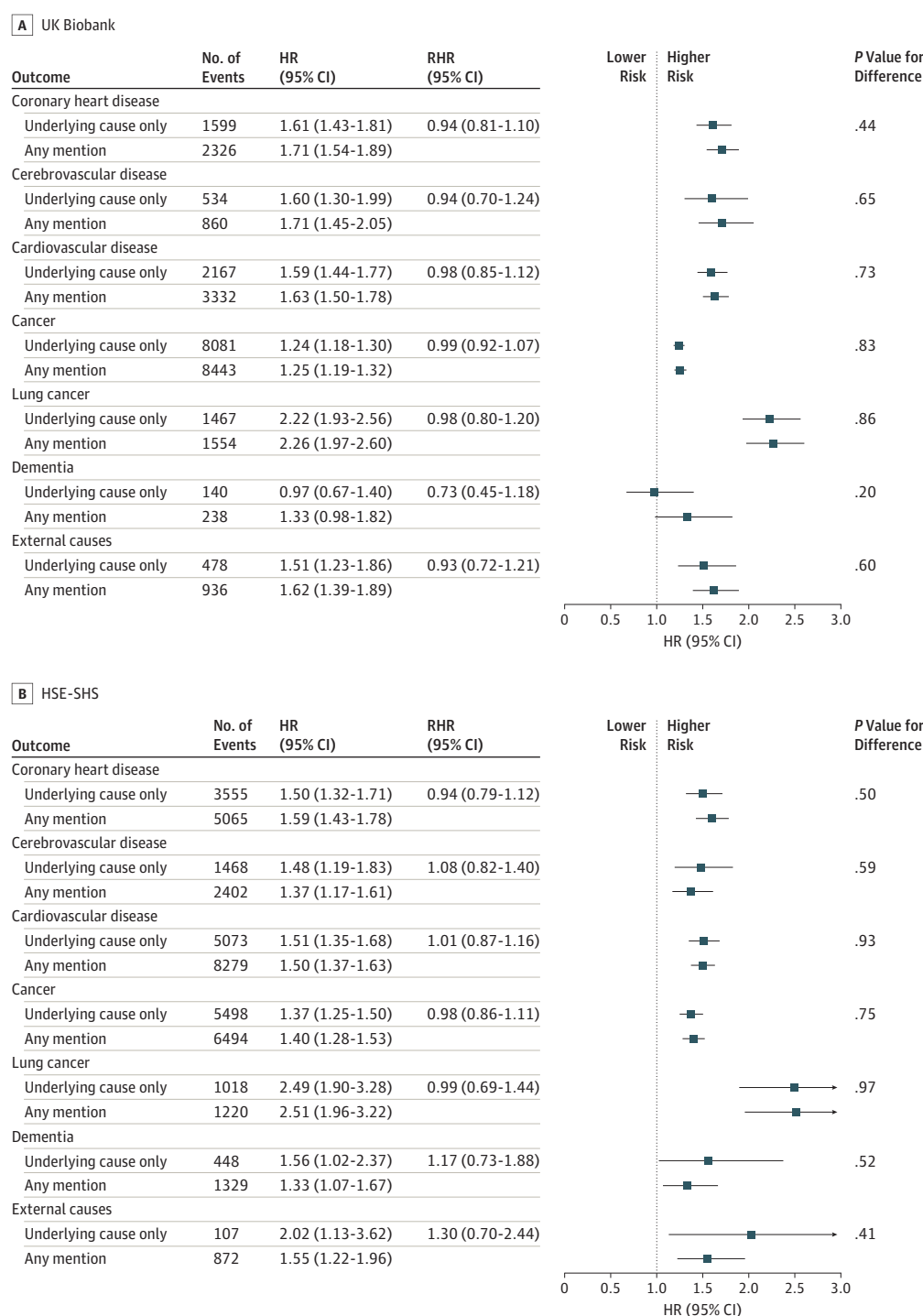
Figure 1. Association of Smoking Status With Cause-Specific Mortality in the UK Biobank (N = 499 701) and the Health Survey for England and Scottish Health Surveys (HSE-SHS; N = 193 037)



Shaded squares indicate the hazard ratios (HRs), and error bars denote the 95% CIs for the association of smoking status with the risk of death from a range of diseases. The reference group is never having smoked cigarettes. The ratio of hazard ratios (RHR) summarizes the difference, with underlying cause as the reference group, between the

effect estimate for the outcome as ascertained from different locations on the death certificate. The number of study participants and deaths in the sample used in this survival analysis is marginally lower than the full cohort owing to missing data for the exposure of interest.

Figure 2. Association of Educational Attainment with Cause-Specific Mortality in the UK Biobank (N = 492 513) and the Health Survey for England and Scottish Health Surveys (HSE-SHS; N = 193 702)



Shaded squares indicate the hazard ratios (HRs), and error bars denote the 95% CI for the association of educational attainment with the risk of death from a range of long-term diseases and injury. The reference group is having a university undergraduate degree or higher. The ratio of hazard ratios (RHR) summarizes the difference, with

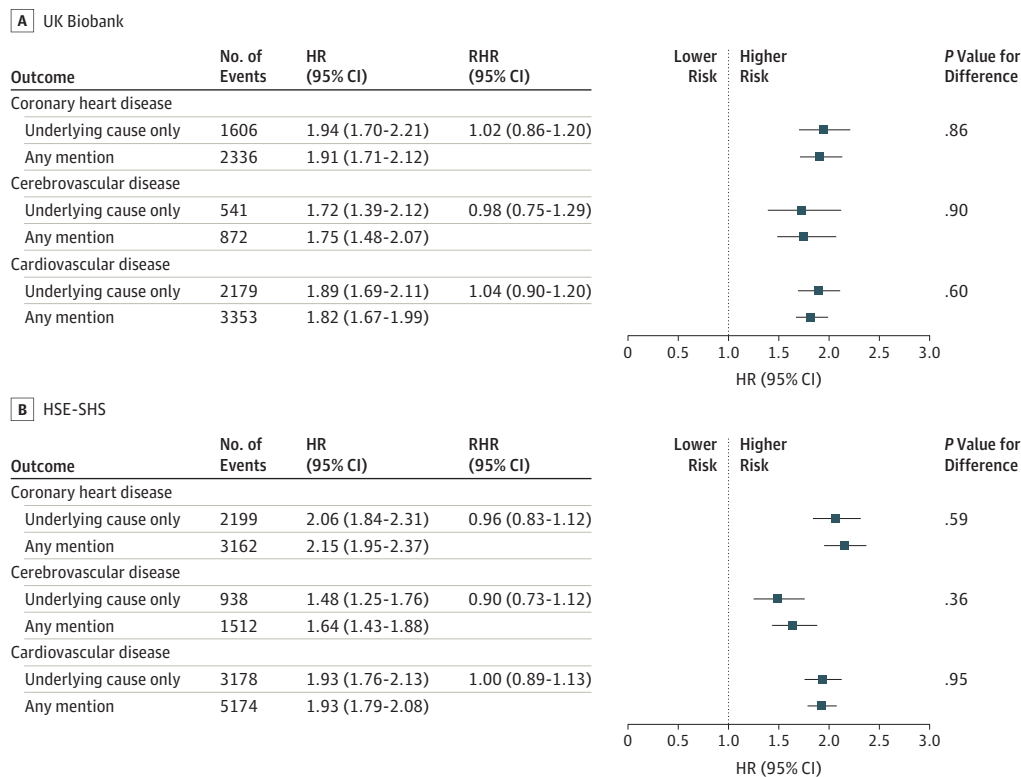
underlying cause as the reference group, between the effect estimate for the outcome as ascertained from different locations on the death certificate. The number of study participants and deaths in the sample used in this survival analysis is marginally lower than the full cohort owing to missing data for the exposure of interest.

Discussion

In this study, known associations of risk factors with an array of health end points were essentially the same irrespective of whether death data were drawn from the underlying cause field on the death certificate or a combination of underlying and contributory categories. These observations were confirmed in independent data sets. The any mention field is, as described, a combination of underlying and contributory fields. For outcomes where there is a small difference in absolute numbers of cases between these groups, such as cancer, the HRs based on analyses of each group will necessarily be nearly identical. More surprising is the similarity in effect estimates where discordance in the number of events is high, that is, for all other outcomes featured herein: cardiovascular disease (and the different presentations it comprises), dementia, and external causes. An implication of our findings is that using the contributory field alongside the underlying cause field may have the advantage of facilitating investigation of risk factors for the occurrence of rarer forms of death where, particularly in smaller cohort studies, there may be too few events to compute effect estimates using the underlying field alone. The value in larger studies, such as those used here, might be marginally improved statistical precision.

As described, we were able to identify only 1 other study that has systematically compared the utility for etiological research of using a combination of the underlying cause and contributory cause fields on death certificates with using the underlying cause filed alone.²² Using mortality records from the Western Electric Study, the 5 risk factors examined—age, systolic blood pressure, blood cholesterol, body mass index, and cigarette smoking status—revealed near-identical HRs for the

Figure 3. Association of Hypertension Status With Cardiovascular Disease Mortality in the UK Biobank (N = 492 513) and the Health Survey for England and Scottish Health Surveys (HSE-SHS; N = 117 606)



Shaded squares indicate the hazard ratios (HRs), and error bars denote the 95% CIs for the association of hypertension status with the risk of death from different presentations of cardiovascular disease. The reference group is not having hypertension. The ratio of hazard ratios (RHR) summarizes the difference, with underlying cause as the reference

group, between the effect estimate for the outcome as ascertained from different locations on the death certificate. The number of study participants and deaths in the sample used in this survival analysis is marginally lower than the full cohort owing to missing data for the exposure of interest.

association of these risk factors with cardiovascular disease mortality for the 2 sources of the mortality outcome. There are also examples of investigators who have followed this analytical process but reported their findings qualitatively only. Thus, in our previous work,⁴⁰ psychological distress was associated with an elevated risk of death as drawn from the major *ICD-10* chapters, whether extracted from underlying cause on the death certificate or in combination with contributory cause. Similar results were seen when the predictive capacity of alcohol intake and obesity for liver disease was assessed,⁴¹ when pulmonary function was associated with dementia death,⁴² and when we investigated the association of neuroticism with mortality from various causes.⁴³

Strengths and Limitations

While the present study has strengths—its relative novelty and the comparison of results across large, well-powered studies—there are inevitably some shortcomings. Biomedical data, while available for analyses in HSE-SHS, were not available in UK Biobank at the time of analyses. It has therefore not been possible to compare the association of cholesterol fractions, glycated hemoglobin, and inflammatory markers, all linked etiologically with cardiovascular disease,⁴⁴ with mortality across different placements of cause of death on the death certificate. Based on the present results, we think it is unlikely that these risk indices will yield results very different from the patterns described here. Second, our exposure variables are known to be associated with the major stroke subtypes (ie, ischemic and hemorrhagic); however, on UK death certificates stroke subtype is usually too ill defined to be useful.⁴⁵ Therefore, we were not able to run such analyses. Third, while UK Biobank is undoubtedly rare in its scale and broad in its content, it had an unconventionally low response to its baseline survey of approximately 6%. This has prompted debates about the generalizability of its findings.⁴⁶⁻⁴⁹ This notwithstanding, HSE-SHS, an independent data set, had response rates in the normal range (64%-78%).⁵⁰

Conclusions

Risk factor–end point associations were not sensitive to the placement of mortality data on the death certificate. Using cause of death positioned anywhere on a death certificate may have the advantage of facilitating investigation of risk factors for the occurrence of rarer forms of death where, particularly in smaller cohort studies, there may be too few events to facilitate effect estimate computations using the underlying field alone.

ARTICLE INFORMATION

Accepted for Publication: June 5, 2019.

Published: July 31, 2019. doi:[10.1001/jamanetworkopen.2019.8024](https://doi.org/10.1001/jamanetworkopen.2019.8024)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2019 Batty GD et al. *JAMA Network Open*.

Corresponding Author: G. David Batty, PhD, DSc, Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, United Kingdom (david.batty@ucl.ac.uk).

Author Affiliations: School of Biological and Population Health Sciences, Oregon State University, Corvallis (Batty); Department of Epidemiology and Public Health, University College London, London, United Kingdom (Batty, Kivimäki); Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom (Gale); Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom (Gale); Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom (Bell); The National Institute for Health Research Blood and Transplant Unit in Donor Health and Genomics, University of Cambridge, Cambridge, United Kingdom (Bell).

Author Contributions: Dr Gale had full access to the UK Biobank data, and Dr Bell had full access to the Health Survey for England and Scottish Health Survey data. Each person takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Batty.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Batty, Gale, Bell.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Gale, Bell.

Administrative, technical, or material support: Bell.

Conflict of Interest Disclosures: Dr Batty reported receiving grants from the UK Medical Research Council and the US National Institute on Aging during the conduct of the study. Dr Gale reported receiving grants from the UK Medical Research Council during the conduct of the study. Dr Kivimäki reported receiving grants from the UK Medical Research Council, the US National Institute on Aging, NordForsk, and the Academy of Finland and serving as the primary investigator of a research grant from the Institute of Life Science during the conduct of the study. Dr Bell reported receiving grants from the National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, the UK Medical Research Council, the British Heart Foundation, and National Institute of Health Research Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust during the conduct of the study.

Additional Information: This research has been conducted using the UK Biobank Resource under application 10279. Access to data from UK Biobank (<https://www.ukbiobank.ac.uk/>), the Health Survey for England (<https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000021>), and the Scottish Health Surveys (<https://www2.gov.scot/Topics/Statistics/Browse/Health/scottish-health-survey>) is available upon application.

REFERENCES

1. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1459-1544. doi:10.1016/S0140-6736(16)31012-1
2. Reid DD, Rose GA. Assessing the comparability of mortality statistics. *Br Med J*. 1964;2(5422):1437-1439. doi:10.1136/bmj.2.5422.1437
3. Allemani C, Matsuda T, Di Carlo V, et al; CONCORD Working Group. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet*. 2018;391(10125):1023-1075. doi:10.1016/S0140-6736(17)33326-3
4. Batty GD, Whitley E, Gale CR, Osborn D, Tynelius P, Rasmussen F. Impact of mental health problems on case fatality in male cancer patients. *Br J Cancer*. 2012;106(11):1842-1845. doi:10.1038/bjc.2012.150
5. Ebrahim S, Smith GD. Systematic review of randomised controlled trials of multiple risk factor interventions for preventing coronary heart disease. *BMJ*. 1997;314(7095):1666-1674. doi:10.1136/bmj.314.7095.1666
6. Davies C, Pan H, Godwin J, et al; Adjuvant Tamoxifen: Longer Against Shorter (ATLAS) Collaborative Group. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*. 2013;381(9869):805-816. doi:10.1016/S0140-6736(12)61963-1
7. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet*. 1986;1(8489):1077-1081. doi:10.1016/S0140-6736(86)91340-1
8. Leon DA, Saburova L, Tomkins S, et al. Hazardous alcohol drinking and premature mortality in Russia: a population based case-control study. *Lancet*. 2007;369(9578):2001-2009. doi:10.1016/S0140-6736(07)60941-6
9. Batty GD, Shipley MJ, Kivimäki M, Smith GD, West R. Impact of smoking cessation advice on future smoking behavior, morbidity, and mortality: up to 40 years of follow-up of the first randomized controlled trial of a general population sample. *Arch Intern Med*. 2011;171(21):1950-1951. doi:10.1001/archinternmed.2011.543
10. Hammond EC, Horn D. The relationship between human smoking habits and death rates: a follow-up study of 187,766 men. *J Am Med Assoc*. 1954;155(15):1316-1328. doi:10.1001/jama.1954.03690330020006
11. Reid DD, Hamilton PJ, McCartney P, Rose G, Jarrett RJ, Keen H. Smoking and other risk factors for coronary heart-disease in British civil servants. *Lancet*. 1976;2(7993):979-984. doi:10.1016/S0140-6736(76)90830-8
12. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet*. 2015;386(9993):533-540. doi:10.1016/S0140-6736(15)60175-1

13. Dawber TR, Moore FE, Mann GV. Coronary heart disease in the Framingham study. *Am J Public Health Nations Health*. 1957;47(4 Pt 2):4-24. doi:[10.2105/AJPH.47.4.Pt_2.4](https://doi.org/10.2105/AJPH.47.4.Pt_2.4)
14. Kannel WB, Dawber TR, Cohen ME, McNamara PM. Vascular disease of the brain: epidemiologic aspects: the Farmingham Study. *Am J Public Health Nations Health*. 1965;55:1355-1366. doi:[10.2105/AJPH.55.9.1355](https://doi.org/10.2105/AJPH.55.9.1355)
15. Fuller JH, Shipley MJ, Rose G, Jarrett RJ, Keen H. Coronary-heart-disease risk and impaired glucose tolerance: the Whitehall study. *Lancet*. 1980;1(8183):1373-1376. doi:[10.1016/S0140-6736\(80\)92651-3](https://doi.org/10.1016/S0140-6736(80)92651-3)
16. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report, 1954. *BMJ*. 2004;328(7455):1529-1533. doi:[10.1136/bmj.328.7455.1529](https://doi.org/10.1136/bmj.328.7455.1529)
17. Blackburn H. The origins and early evolution of epidemiologic research in cardiovascular diseases: a tabular record of cohort and case-control studies and preventive trials initiated from 1946 to 1976. *Am J Epidemiol*. 2019;188(1):1-8.
18. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems (10th Revision)*. Geneva, Switzerland: World Health Organization; 2010.
19. Batty GD, Shipley MJ, Marmot MG, Davey Smith G; Whitehall Study. Blood pressure and site-specific cancer mortality: evidence from the original Whitehall study. *Br J Cancer*. 2003;89(7):1243-1247. doi:[10.1038/sj.bjc.6601255](https://doi.org/10.1038/sj.bjc.6601255)
20. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012;380(9836):37-43. doi:[10.1016/S0140-6736\(12\)60240-2](https://doi.org/10.1016/S0140-6736(12)60240-2)
21. Redelings MD, Sorvillo F, Simon P. A comparison of underlying cause and multiple causes of death: US vital statistics, 2000-2001. *Epidemiology*. 2006;17(1):100-103. doi:[10.1097/01.ede.0000187177.96138.c6](https://doi.org/10.1097/01.ede.0000187177.96138.c6)
22. Crews DE, Stamler J, Dyer A. Conditions other than underlying cause of death listed on death certificates provide additional useful information for epidemiologic research. *Epidemiology*. 1991;2(4):271-275. doi:[10.1097/00001648-199107000-00006](https://doi.org/10.1097/00001648-199107000-00006)
23. Batty GD, Bell S, Stamatakis E, Kivimäki M. Association of systemic inflammation with risk of completed suicide in the general population. *JAMA Psychiatry*. 2016;73(9):993-995. doi:[10.1001/jamapsychiatry.2016.1805](https://doi.org/10.1001/jamapsychiatry.2016.1805)
24. Russ TC, Stamatakis E, Hamer M, Starr JM, Kivimäki M, Batty GD. Socioeconomic status as a risk factor for dementia death: individual participant meta-analysis of 86 508 men and women from the UK. *Br J Psychiatry*. 2013;203(1):10-17. doi:[10.1192/bjp.bp.112.119479](https://doi.org/10.1192/bjp.bp.112.119479)
25. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360(9349):1903-1913. doi:[10.1016/S0140-6736\(02\)11911-8](https://doi.org/10.1016/S0140-6736(02)11911-8)
26. Mackenbach JP, Stirbu I, Roskam AJ, et al; European Union Working Group on Socioeconomic Inequalities in Health. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. 2008;358(23):2468-2481. doi:[10.1056/NEJMsa0707519](https://doi.org/10.1056/NEJMsa0707519)
27. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum*. 2004;83:1-1438.
28. Batty GD, Kivimäki M, Gray L, Smith GD, Marmot MG, Shipley MJ. Cigarette smoking and site-specific cancer mortality: testing uncertain associations using extended follow-up of the original Whitehall study. *Ann Oncol*. 2008;19(5):996-1002. doi:[10.1093/annonc/mdm578](https://doi.org/10.1093/annonc/mdm578)
29. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. doi:[10.1371/journal.pmed.1001779](https://doi.org/10.1371/journal.pmed.1001779)
30. Mindell J, Biddulph JP, Hirani V, et al. Cohort profile: the health survey for England. *Int J Epidemiol*. 2012;41(6):1585-1593. doi:[10.1093/ije/dyr199](https://doi.org/10.1093/ije/dyr199)
31. Gray L, Batty GD, Craig P, et al. Cohort profile: the Scottish Health Surveys cohort: linkage of study participants to routinely collected records for mortality, hospital discharge, cancer and offspring birth characteristics in three nationwide studies. *Int J Epidemiol*. 2010;39(2):345-350. doi:[10.1093/ije/dyp155](https://doi.org/10.1093/ije/dyp155)
32. Batty GD, Russ TC, Stamatakis E, Kivimäki M. Psychological distress in relation to site specific cancer mortality: pooling of unpublished data from 16 prospective cohort studies. *BMJ*. 2017;356:j108.
33. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194. doi:[10.1001/jama.2013.281053](https://doi.org/10.1001/jama.2013.281053)
34. Bolling K. *The Dinamap 8100 Calibration Study*. London, United Kingdom: Stationery Office Books; 1994.

35. Chobanian AV, Bakris GL, Black HR, et al; National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003;289(19):2560-2572. doi:10.1001/jama.289.19.2560
36. Cox D. Regression models and life-tables. *J R Stat Soc [Ser B]*. 1972;34(2):187-220. doi:10.1111/j.2517-6161.1972.tb00899.x
37. Huxley R, Barzi F, Woodward M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ*. 2006;332(7533):73-78. doi:10.1136/bmj.38678.389583.7C
38. Huxley RR, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *Lancet*. 2011;378(9799):1297-1305. doi:10.1016/S0140-6736(11)60781-2
39. Paternoster R, Brame R, Mazerolle P, Piquero A. Using the correct statistical test for the equality of regression coefficients. *Criminology*. 1998;36(4):859-866. doi:10.1111/j.1745-9125.1998.tb01268.x
40. Russ TC, Stamatakis E, Hamer M, Starr JM, Kivimäki M, Batty GD. Association between psychological distress and mortality: individual participant pooled analysis of 10 prospective cohort studies. *BMJ*. 2012;345:e4933. doi:10.1136/bmj.e4933
41. Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *BMJ*. 2010;340:c1240. doi:10.1136/bmj.c1240
42. Russ TC, Starr JM, Stamatakis E, Kivimäki M, Batty GD. Pulmonary function as a risk factor for dementia death: an individual participant meta-analysis of six UK general population cohort studies. *J Epidemiol Community Health*. 2015;69(6):550-556. doi:10.1136/jech-2014-204959
43. Gale CR, Čukić I, Batty GD, McIntosh AM, Weiss A, Deary IJ. When is higher neuroticism protective against death? findings from UK Biobank. *Psychol Sci*. 2017;28(9):1345-1357. doi:10.1177/0956797617709813
44. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation*. 2001;104(22):2746-2753. doi:10.1161/hc4601.099487
45. Leon DA, Smith GD, Shipley M, Strachan D. Adult height and mortality in London: early life, socioeconomic confounding, or shrinkage? *J Epidemiol Community Health*. 1995;49(1):5-9. doi:10.1136/jech.49.1.5
46. Swanson JM. The UK Biobank and selection bias. *Lancet*. 2012;380(9837):110. doi:10.1016/S0140-6736(12)61179-9
47. Ebrahim S, Davey Smith G. Commentary: should we always deliberately be non-representative? *Int J Epidemiol*. 2013;42(4):1022-1026. doi:10.1093/ije/dyt105
48. Nohr EA, Olsen J. Commentary: epidemiologists have debated representativeness for more than 40 years: has the time come to move on? *Int J Epidemiol*. 2013;42(4):1016-1017. doi:10.1093/ije/dyt102
49. Richiardi L, Pizzi C, Pearce N. Commentary: representativeness is usually not necessary and often should be avoided. *Int J Epidemiol*. 2013;42(4):1018-1022. doi:10.1093/ije/dyt103
50. Russ TC, Kivimäki M, Starr JM, Stamatakis E, Batty GD. Height in relation to dementia death: individual participant meta-analysis of 18 UK prospective cohort studies. *Br J Psychiatry*. 2014;205(5):348-354. doi:10.1192/bjp.bp.113.142984