HHS Public Access

Author manuscript

Ann Epidemiol. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

Ann Epidemiol. 2020 May ; 45: 40–46.e4. doi:10.1016/j.annepidem.2020.03.010.

Survival Advantage of Cohort Participation Attenuates Over Time: Results from Three Long-Standing Community-Based Studies

Zihe Zheng, **Casey M. Rebholz**, **Kunihiro Matsushita**, **Judith Hoffman-Bolton**, **Michael J. Blaha**, **Elizabeth Selvin**, **Lisa Wruck**, **A. Richey Sharrett**, **Josef Coresh**

Department of Biostatistics, Epidemiology, and Bioinformatics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, United States (Zihe Zheng, MHS); Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States (Casey M. Rebholz, PhD, MPH, MS, MNSP, Kunihiro Matsushita, MD, PhD, Judith Hoffman-Bolton, BS, Elizabeth Selvin, PhD, MPH, A. Richey Sharrett, MD, Josef Coresh, MD, PhD, MHS); Ciccarone Center for the Prevention of Heart Disease, Johns Hopkins School of Medicine, Baltimore, Maryland, United States (Michael J. Blaha, MD, MPH); Center for Preventive Medicine, Duke Clinical Research Institute, Durham, North Carolina, United States (Lisa Wruck, PhD). Ms. Kendra Hay assisted with editing the manuscript.

Abstract

Cohort participants usually have lower mortality rates than non-participants, but it is unclear if this survival advantage decreases or increases as cohort studies age. We used a 1975 private census of Washington County, Maryland to compare mortality among cohort participants to non-participants for three cohorts CLUE I, CLUE II, and ARIC, initiated in 1974, 1989, and 1986, respectively. We analyzed mortality risk using time-truncated Cox regression models. Participants had lower mortality risk in the first 10 years of follow-up compared to non-participants [fully adjusted average hazard ratio (95% CI) were 0.72 (0.68, 0.77) in CLUE I, 0.69 (0.65, 0.73) in CLUE II and 0.74 (0.63, 0.86) in ARIC], which persisted over 20 years of follow-up [0.81 (0.78, 0.84) in CLUE I, 0.87 (0.84, 0.91) in CLUE II, and 0.90 (0.83, 0.97) in ARIC]. This lower average hazard for mortality among participants compared to non-participants attenuated with longer follow-up [0.99 (0.96, 1.01) after 30+ years in CLUE I, 1.02 (0.99, 1.05) after 30 years in CLUE II, and 0.95 (0.89, 1.00) after 30+ years in ARIC]. In ARIC, participants who did not attend visits had higher mortality but those who did attend visits had similar mortality to the community. Our results suggest the volunteer selection for mortality in long-standing epidemiologic cohort studies often diminishes as the cohort ages.

Ms. Zheng, Dr. Rebholz, and Dr. Coresh designed the study and directed its implementation. Ms. Hoffman-Bolton supervised the field activities and data acquisition in the census and CLUE. Dr. Blaha provided study population's vital statistics linkage with the United States Social Security Administration Death Master File. Ms. Zheng performed the statistical analyses. Ms. Zheng and Dr. Coresh drafted the paper. All the authors have participated in reviewing and critically revising the manuscript.

Keywords

survival advantage; cohort study; selection bias; mortality

INTRODUCTION

Prospective epidemiologic cohort studies often strive to be representative of a target population and provide generalizable conclusions. However, cohort participants are usually not fully representative of the source population initially due to intentional or unintentional self-selection during the study enrollment process (e.g., ability or willingness to visit a research site). Indeed, previous studies reported that cohort participants usually have fewer health behavior-related risk factors, are better educated, and are older compared to non-participants [1–4]. Moreover, multiple cohort studies have observed consistently lower mortality rates among participants versus non-participants [5–16] but whether this difference persists, increases, or decreases with long follow-up is uncertain. This topic is increasingly relevant as many cohorts were established during 1970–2000 era and their follow-up continues.

A well-defined population of non-participants is the key to assess study external validity. Due to the lack of data on people who did not participate or were not invited, most cohort studies were not able to examine cohort representativeness through direct comparison of participants to all non-participants. To evaluate the generalizability of a research cohort, previous studies had used active non-respondents (individuals who were invited to be in the study but refused to participate) as an approximation for non-participants [18, 19] and compared the baseline characteristics between the two group. To get a better estimation, some studies chose to resurvey a subgroup of active non-respondents [3] or have linked the non-participants' or active non-respondents' identifiers to regional or registry data for outcome ascertainment [5–9, 11, 13–16,18–20]. However, the follow-up time of these previous studies was relatively short, ranging from 2.8–10 years, with only one study reaching 15 years of follow-up [11], and each of these studies only examined a single cohort [5–8].

We leveraged a private census of Washington County, Maryland with over 40 years of follow-up for the ascertainment of mortality by linkage to participation in multiple cohort studies. The census provided unbiased data on population characteristics and risk of mortality for participants and non-participants alike. A key assumption we made in this study was that the differences in mortality rates and risks for death between cohort participants and non-participants can be used as indicators to access cohort representativeness. The aim of our study was to examine whether cohort representativeness, indirectly reflected by survival differences between participants and non-participants, would change over decades of follow-up. Our main hypotheses were: 1) participants have a survival advantage (i.e. lower risk of death) compared to non-participants; and 2) this survival advantage attenuates over follow-up time.

METHODS

Study design and study population

We linked data from three sources: study cohorts, private census, and a registry of vital status from Washington County, Maryland and the United States Social Security Administration Death Master File (DMF). The 1975 private census covered over 92% of the Washington County population and was used to estimate the source population. We compared risk of all-cause mortality between the participants and the non-participants in three studies: CLUE I (Campaign Against Cancer and Stroke), CLUE II (Campaign Against Cancer and Heart Disease), and ARIC (Atherosclerosis Risk In Communities).

Details regarding the study design of the three cohorts are described elsewhere [21, 22]. Briefly, targeting the entire Washington County population (age range 2–98 years) at two different time points 15 years apart, CLUE I enrolled 23,951 male and female volunteers in 1974 and CLUE II enrolled 25,067 volunteers 1989. Both studies used mobile vans and general advertising to cover all areas of the county whose adult population was approximately 100,000. At the time of recruitment, socio-demographics, brief health histories, and biospecimens were collected from participants [21]. CLUE I and CLUE II participants were followed-up through questionnaires in 1996, 1998, 2000, 2003 and 2005. The ARIC study enrolled 4,020 men and women aged 45–64 years from Washington County during 1986–1989 [22]. Investigators collected participants' information on demographics, social factors, and medical conditions. ARIC participants were invited to attend follow-up study visits at approximately three-year intervals after the baseline visit (visit 2, 1990–1992; visit 3, 1993–1995; visit 4, 1996–1998) and followed subsequently for mortality and hospitalizations.

Two private censuses were conducted among Washington County residents in 1963 and 1975. Census staff distributed and collected a questionnaire asking for information on demographics, health behaviors (mainly smoking), and indoor environment at each household in Washington County. For the purpose of this analysis, we used the 1975 Washington County private census data, which was chronologically closer to the CLUE I study enrollment and preceded the other cohort studies. Census data was collected on 92% of county residents, which allowed us a unique opportunity to identify the vital status of individuals who were eligible for the three cohort studies but did not participate. We excluded non-white individuals (3%) and adolescents younger than 20 years of age in 1975.

Definition of participation

To identify cohort study participants from the census population, we created a data linkage algorithm by matching the following information: individuals' current last name, first name, middle name, and date of birth. For individuals matched with date of birth but mismatched with current name, we additionally used the name at birth (first, last, and middle name) to establish linkage. A total of 4,361 (18.2%) CLUE I participants, 8,807 (35.1%) CLUE II participants, and 1,096 (27.3%) ARIC participants were not linked with the census database and were excluded, likely due to migration into the county in the years between the census

and study enrollment. Male cohort participants were more likely to not link to the census database.

Participants in the CLUE I and CLUE II cohorts volunteered themselves to the study during the enrollment period while ARIC investigators actively approached and invited the eligible study population. The ARIC study used a known sampling frame randomly drawing age-eligible individuals from the private census and drivers' licenses registered at the local motor vehicle administration office. Eligible individuals were invited to participate in ARIC by study staff in person. Therefore, we were able to divide non-participants into invited non-respondents and uninvited community members. Among the 6,174 individuals in Washington County who were invited to the ARIC study, 2,154 refused to participate. Among these invited non-respondents, we identified 1,165 individuals (54.1%) in the census.

We used the first four consecutive ARIC visits (visit 1–4) to subdivide participants into full participants (individuals who attended all four visits) and partial participants (individuals who were alive at the end of ARIC visit 4 on January 30th, 1999 but missed at least one of the follow-up visits 2–4). We then analyzed mortality subsequent to visit 4.

Linkage of data sources

The 1975 private census in Washington County, MD included 90,261 individuals. We excluded 31,031 individuals younger than 20 years of age in 1975, 692 non-white individuals, and 1,442 individuals with missing covariates. After linking the data sources, we identified 17,100 CLUE I participants, 12,661 CLUE II participants, and 2,921 ARIC participants within the census dataset. In addition, we identified 38,324 eligible non-participants for CLUE I, and 33,548 eligible non-participants for CLUE II; and 1,165 invited non-respondents and 17,830 uninvited community members for ARIC (Figure 1).

Mortality ascertainment

We ascertained deaths through linkage of the census data to the State of Maryland Vital Records and United States Social Security Administration Death Master File (DMF). The above linkage was based on last name, first name, middle name, sex, and date of birth. Follow-up for death was applied uniformly to participants and non-participants in all studies (i.e. we used census mortality ascertainment rather than cohort mortality follow-up of participants). The administrative censoring date for our mortality ascertainment was August 31st, 2015. To access potential information bias due to death misclassification, we compared the census ascertained deaths against deaths ascertained in the ARIC study (gold standard) (Appendix A).

Measurement of other covariates

We used individual-level data on demographics (age, sex), social factors (education, marital status), and health characteristics (smoking status, cancer) that were available in the census, measured universally for both cohort participants and non-participants. We classified education level as less than 12 years, equal to 12 years, or more than 12 years. We defined marital status as currently married, never married, or divorced/separated/widowed. We

categorized smoking status as never smoker, current smoker, or former smoker. Self-reported cancer at baseline was collected through census survey questionnaire and was coded as a binary variable. We excluded 1,442 individuals (<2% of the census population) who had missing data for any of these covariates.

Statistical analysis

Descriptive statistics for each cohort are presented as means (standard deviation, SD) for continuous variables and as counts (%) for categorical variables. We compared baseline covariates by participation status using analysis of variance (ANOVA) and chi-square tests. Since the CLUE I study started 1 year before the census (1974), the time of entry for CLUE I survival analysis was set to 1975 to exclude immortal person time. To calculate age-standardized mortality rates from participants and non-participants, we used age of the whole census population as the standard population. For CLUE I and CLUE II, the standard population was the combination of the participants and the non-participants; and for ARIC, the standard population was the combination of the participants, the invited non-respondents, and the uninvited community members. We created Kaplan-Meier curves to present cumulative survival by participation status and performed log-rank tests. We used Cox regression models to estimate hazard ratios (HRs) for all-cause mortality by participation status for the three studies, stratified by 5-year (ARIC study) or 10-year (CLUE I and CLUE II studies) periods of follow-up. We used HRs for mortality during different time intervals as an indicator that reflected cohort representativeness at different point of follow-up time. We compared the unadjusted relative hazards (model 1) to the estimates adjusted for age and sex (model 2) and then further adjusted for education category, marital status, smoking status, and self-reported cancer status (model 3). Within ARIC, we also estimated average HRs over 10-year incremental follow-up periods after ARIC visit 4, comparing the participation subtypes of full participants, partial participants, and invited non-respondents to the reference group of uninvited non-participants. We performed the statistical analyses with Stata statistical software, version 15.0 (StataCorp, College Station, Texas).

RESULTS

Baseline characteristics

At baselines, the age range was 19–96 years in CLUE I (year 1974), 34–97 years in CLUE II (year 1989), and 45–65 years in ARIC (year 1986). The mean age was similar between participants and non-participants (Table 1). Across the three studies, we found that compared to non-participants, participants were more likely to be women, have a high level of education (12 years of education or more), be married, and not be a current smoker at the time of the census. Participants of CLUE I and II were more likely to report cancer history; while in ARIC, cancer status was not significantly different across participation types (Chi-square test $P > 0.05$).

Risk of death over time

The median follow-up time was 35 years for CLUE I, 26 years for CLUE II, and 28 years for ARIC. Participants had a lower mortality rate than non-participants during the first 10 years in all three studies (Table 2). The difference in age-standardized mortality rates

decreased over time and disappeared after 30 years of follow-up in CLUE I, during 10–20 years of follow-up in CLUE II, and after 10 years of follow-up in ARIC. After 20 years of follow-up, there were higher mortality rates in participants (30.8/1,000 person-years, 95% CI 29.3, 32.4) compared to non-participants (22.9/1,000 person-years, 95% CI 22.0, 23.9) in CLUE II. In ARIC, we consistently observed higher mortality rates among invited non-respondents compared to the groups of participants and uninvited community members. The survival curves of participants and non-participants converged in the three studies by the third to fourth decade after baseline, and the proportional hazard assumption was not met (Figure 2, log-rank test $P < 0.05$, proportional hazard test $P < 0.001$).

Adjusted relative hazards

We observed a lower hazard of all-cause mortality for participants vs. non-participants during the first 10 years of follow-up in all three cohorts (Table 3; Table 4). Adjusting for age, sex, education, marital status, smoking status, and cancer history in model 3 reduced but did not eliminate the lower risk of death seen among cohort participants relative to non-participants: 0.72 (0.68, 0.77) in CLUE I, 0.69 (0.65, 0.73) in CLUE II, and 0.80 (0.65, 1.00) in ARIC. With non-overlapping following up time stratified into 5- or 10-year periods, participants' survival advantage over non-participants decreased over time. With full adjustment, the survival advantage of participants over non-participants disappeared after 20+ years in CLUE I (20–30 year HR: 1.04, 95% CI 1.00, 1.09), after 10+ years of follow-up in CLUE II (10–20 year HR: 1.05, 95% CI 1.00, 1.10), and in ARIC (10–15 year HR: 1.01, 95% CI 0.88, 1.15).

ARIC invited non-respondents had a higher hazard of death during the entire follow-up time (model 3 HRs adjusted for age, sex, education, marital status, smoking status, and cancer history ranged from 1.14 to 1.44). We found that there was no consistent difference in risk of mortality for the enumerated individuals (combination of ARIC participants and invited non-respondents) compared to the uninvited community members.

The sensitivity analysis examining participants' survival advantage over non-participants, using cumulative follow-up time, confirmed the pattern found in the main analysis (Table S1 and Table S2).

Full vs. partial participation in ARIC

By the end of ARIC visit 4 (January 1999), 2,700 participants, 1,015 invited non-respondents, and 16,021 uninvited non-participants were alive. We further classified the 2,700 participants into 2,295 full participants (attended visits 1 through 4) and 405 partial participants (missing at least one of visits 2 to 4). Partial participants had the lowest cumulative survival followed by invited non-respondents and the uninvited non-participants (general community members) (Figure 2D; log-rank $P < 0.001$). ARIC full participants had a survival benefit compared to the uninvited general community members during the first 10 years of follow-up, but this advantage attenuated during the second decade of follow-up as mortality rates rose slightly faster among full participants (Table S3). The lower hazard of death comparing full participants to uninvited community members and the higher hazard of death comparing partial participants to uninvited community members attenuated towards

over time (Table S4 and Table S5). Invited non-respondents had greater hazards of death than uninvited community members during the entire study follow-up time. Adjustment for demographics, social factors, and health characteristics explained very little of these survival differences.

DISCUSSION

Using three community-based studies, we characterized two major patterns of longstanding cohorts. First, there existed initial survival difference between cohort participants and non-participants after cohort formation but this survival difference gradually diminished over follow-up time. We found that the relative lower mortality rate among cohort participants persisted for 10 to 20 years in the three cohorts. The lower age-standardized mortality rate of participants compared to non-participants disappeared after 10+ years of follow-up in CLUE II and ARIC and after 30 years of follow-up in CLUE I. Differences in baseline demographics, social factors, and health characteristics did not explain the initial survival differences observed in the first decade of follow-up. In ARIC, individuals who declined an invitation by study staff (the invited non-respondents) had a higher risk of mortality which persisted for two decades compared to both the uninvited community members and ARIC participants. Also, in ARIC, partial participants who did not attend all visits had a higher risk of mortality than full participants and uninvited community members. Interestingly, full participants' survival advantage compared to the uninvited community members diminished over the second decade of follow-up.

The survival advantage of healthy participants to research cohorts resembles in some aspects of healthy worker effects (HWE) in occupational studies [23–25]. In terms of mechanism, HWE has been viewed as an example of confounding due to unmeasured health status of employees and imperfect choice of the using general population as the comparison group. However, the nature of the healthy participants effect in our study and other cohort studies alike is a combination of confounding and population selection bias. Similar to the patterns observed in our study, the HWE of employees over general or unemployed population lasted for 3 to 15 years and then attenuated [23, 26, 27]. Some difference could be explained by measured confounders while some remained and is attributed to selection and unmeasured confounders.

Our results agree with and extend the findings from several other long-standing cohorts. An early Framingham Heart Study (FHS) publication [28], comparing the mortality rate of participants and non-participants after one decade of follow-up, found that the mortality rate was higher among non-participants and it varied according to the reasons for non-participation. The study concluded that the final cohort participants were not fully representative of the pre-defined study target population. The British Regional Heart Study (BRHS) [5], which enrolled middle-age men in the community, found a decrease in the survival difference of participants relative to non-participants during a median of 7.2 years of follow-up. Survival advantages were also observed in the study of Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA) [4] with data linkage to the 1990 census of the Swiss population. Our study extends these results by adjusting for differences

in demographics and education as well as exploring how participants' survival advantage changes when follow-up is longer than a decade.

To address the violation of proportional hazard assumption signaled by the crossing Kaplan-Meier curves (Figure 2), we used two sets of survival models. Using mutually exclusive intervals allows the survival modeling to dissect the difference between participants and non-participants in each interval. However, concerns have been raised that survival to the beginning of each interval influences the subsequent risk [29]. Therefore, we also conducted a comparison of cumulative intervals starting at baseline which allows for an overall comparison of the groups. Results from the two models were consistent.

The implications of our study are important for people conducting, funding and interpreting results from longstanding cohort studies. Longstanding cohorts are particularly valuable for conditions where cumulative exposure and long latency are important. For example, cumulative exposure to high blood pressure may take decades to result in significant cognitive decline, brain changes including amyloid deposition and dementia [30]. Likewise, cancer initiators require long latency periods to have their full effects, during which cohort participants were exposed to promoters and had sufficient time for tumor growth leading to clinical outcomes [31]. In addition, the consequences of changes in exposure, such as quitting smoking, are important to study for long-term outcomes [32]. Although it cannot be determined whether for a specific exposure (such as blood pressure or a risk factor of dementia) the population bias is less as a result of the participants' survival benefit diminishing. Our findings provided evidence that a well-designed long-standing cohort could have better external validity and be more representative of the older target population than a newly established cohort. With increasing availability of passive follow-up data from registry, census, and broader real time "big data," it is possible to extend our analyses to other settings. Given the growing difficulty and cost of establishing new cohorts, acknowledging the value of existing long-standing cohorts is noteworthy.

Due the nature of census data, we were only able to adjust for time-fixed baseline covariates to assess the change of the relative survival advantage between participants and non-participants. The lack of data on updated population characteristics restricted our ability to quantify changes in the characteristics of participants and non-participants to provide insight as to the mechanisms for the shrinking difference in mortality risks between the groups. However, for the purpose of examining cohort representativeness, our study does not rely on time-updated information for hypothesis testing. Either through active "change in health behavior" or "passive preferential deaths" of sicker individuals, the difference in the overall health profiles between participants and non-participants became smaller, which contributed to our testing the hypothesis of decreased survival advantage of cohort participants.

There are several other limitations to our study. First, our cohorts were recruited from a single county which limits generalizability to other research settings. Second, some cohort participants could not be linked to the 1974 census. This could be due to not being in the county during the census as a result of migration or non-linkage due to a name change. This incomplete linkage favoring most stable residents may have resulted in more conservative estimation of the initial survival advantage of participants over non-participants as well as

the rate of survival advantage decline over time. To maximize linkage between data sources, we additionally matched individuals with their name at birth. Finally, given that our vital statistics information was based on a local death registry reports and annually updated United States DMF, we may have had some misclassification of deaths, especially those occurring outside of the State of Maryland. If participants were less likely to move, as happened in ARIC, the underestimation of mortality risk due to migration would be more relevant to non-participants than to participants. Using the ARIC death ascertainment which captured both in-state and out-state deaths as the "gold standard", we found that the overall percentage of missing deaths by the census was less than 7% (Table S6), similar to the estimate of missed deaths among participants age 95+ years by the end of follow-up.

In summary, our study characterized the representativeness of a community-based cohort by comparing participants' risk of mortality to that of non-participants. We found that cohort participants had a survival advantage over non-participants initially, but this survival advantage in these cohorts decreased after the one to two decade of follow-up. Compared to non-participants or participants that missed follow-up visits, individuals who participated in all study visits had survival profiles that gradually converged to that of the source population. Thus, in terms of mortality risk long-standing cohort studies with high follow-up rates are expected to be more rather than less similar to the target population over time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank the staff and participants of the ARIC study for their important contributions. The Atherosclerosis Risk in Communities study has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under contract nos. HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, and HHSN268201700005I. Dr. Rebholz is supported by a mentored research scientist development award from the National Institute of Diabetes and Digestive and Kidney Diseases (K01 DK107782) and a grant from the National Heart, Lung, and Blood Institute (R21 HL143089).

REFERENCES

1. Comstock GW and Helsing KJ, Characteristics of respondents and nonrespondents to a questionnaire for estimating community mood. Am J Epidemiol, 1973 97(4): p. 233–9. [PubMed: 4697644]
2. Jackson R, et al., Differences between respondents and nonrespondents in a multicenter community-based study vary by gender and ethnicity. Journal of Clinical Epidemiology, 1996 49: p. 1441–1446. [PubMed: 8970495]
3. Batty GD and Gale CR, Impact of resurvey non-response on the associations between baseline risk factors and cardiovascular disease mortality: prospective cohort study. Journal of epidemiology and community health, 2009 63(11): 952–955. [PubMed: 19605367]
4. Bopp M, Braun J, and Faeh D, Variation in mortality patterns among the general population, study participants, and different types of nonparticipants: evidence from 25 years of follow-up. American journal of epidemiology, 2014 180(10): 1028–1035. [PubMed: 25344298]
5. Walker M, Shaper AG, and Cook DG, Non-participation and mortality in a prospective study of cardiovascular disease. Journal of Epidemiology and Community Health, 1987 41(4): 295–299. [PubMed: 3455423]

6. Bisgard KM, et al., Mortality and cancer rates in nonrespondents to a prospective study of older women: 5-year follow-up. American journal of epidemiology, 1994 139: 990–1000. [PubMed: 8178787]
7. Froom P, et al., Healthy volunteer effect in industrial workers. Journal of clinical epidemiology, 1999 52(8): 731–735. [PubMed: 10465317]
8. Barchielli A and Balzi D, Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-responders. International journal of epidemiology, 2002 31: 1038–1042. [PubMed: 12435781]
9. Otto SJ, Schröder FH, and de Koning HJ, Low all-cause mortality in the volunteer-based Rotterdam section of the European randomised study of screening for prostate cancer: self-selection bias? Journal of medical screening, 2004 11(2): 89–92. [PubMed: 15153324]
10. Drivsholm Thomas, et al. Representativeness in population-based studies: a detailed description of non-response in a Danish cohort study. Scandinavian journal of public health, 2006 34(6): 623–631. [PubMed: 17132596]
11. Ferrie JE, et al., Non-response to baseline, non-response to follow-up and mortality in the Whitehall II cohort. International Journal of Epidemiology, 2009 38: 831–837. [PubMed: 19264846]
12. Bosmans JC, et al., Survival of participating and nonparticipating limb amputees in prospective study: consequences for research. Journal of rehabilitation research and development, 2010 47(5): 457–464. [PubMed: 20803389]
13. Knudsen AK, et al., The health status of nonparticipants in a population-based health study. American Journal of Epidemiology, 2010 172: 1306–1314. [PubMed: 20843863]
14. Montgomery M, et al., Characteristics of non-participation and potential for selection bias in a prospective cohort study. American journal of industrial medicine, 2010 53(5): 486–496. [PubMed: 20017198]
15. Strandhagen E, et al., Selection bias in a population survey with registry linkage: Potential effect on socioeconomic gradient in cardiovascular risk. European Journal of Epidemiology, 2010 25: 163–172. [PubMed: 20127393]
16. Lindén-Boström Margareta, and Persson Carina. A selective follow-up study on a public health survey. The European Journal of Public Health, 2012 23(1): 152–157. [PubMed: 22253457]
17. Rothman KJ, Greenland S, and Lash TL, Modern epidemiology. 2008.
18. Heilbrun LK, Nomura a., and Stemmermann GN, The effects of nonresponse in a prospective study of cancer. American journal of epidemiology, 1982 116: 353–363. [PubMed: 7114044]
19. Baumann A, Stieber J, and Löwel H. Nonparticipation as a factor influencing the value of follow-up studies. Results of a telephone 5-year follow-up interview of 55–74-year-old participants of the Augsburg 1989/90 MONICA Survey. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)), 1997 59: 19–25. [PubMed: 9235124]
20. Strandberg TE, et al., Mortality in participants and non-participants of a multifactorial prevention study of cardiovascular diseases: a 28 year follow up of the Helsinki Businessmen Study. British heart journal, 1995 74: 449–454. [PubMed: 7488463]
21. Perneger Thomas V., et al. A prospective study of blood pressure and serum creatinine: results from the'Clue'study and the ARIC study. Jama, 1993 269(4): 488–493. [PubMed: 8419668]
22. Aric Investigators. The atherosclerosis risk in community (ARIC) study: Design and objectives. American journal of epidemiology, 1989 129(4): 687–702. [PubMed: 2646917]
23. Monson Richard R. "Observations on the healthy worker effect." Journal of occupational medicine.: official publication of the Industrial Medical Association 286 (1986): 425–433. [PubMed: 3723215]
24. Shah Divyang. "Healthy worker effect phenomenon." Indian journal of occupational and environmental medicine 132 (2009): 77. [PubMed: 20386623]
25. Kirkeleit Jorunn, et al. "The healthy worker effect in cancer incidence studies." American journal of epidemiology 17711 (2013): 1218–1224. [PubMed: 23595008]
26. Fox Anthony J., and Collier PF. "Low mortality rates in industrial cohort studies due to selection for work and survival in the industry." Journal of Epidemiology & Community Health 304 (1976): 225–230.

27. Thygesen Lau C., et al. “Quantification of the healthy worker effect: a nationwide cohort study among electricians in Denmark.” BMC Public Health 111 (2011): 571. [PubMed: 21767353]

28. Gordon T, et al., Some methodologic problems in the long-term study of cardiovascular disease: observations on the Framingham Study. Journal of Chronic Diseases, 1959 10(3): 186–206.

29. Hernán Miguel, A. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.), 2010 21(1): 13–15.

30. Gottesman Rebecca F., et al. “Association between midlife vascular risk factors and estimated brain amyloid deposition.” Jama 31714 (2017): 1443–1450. [PubMed: 28399252]

31. Aguirre-Ghiso Julio A. “Models, mechanisms and clinical evidence for cancer dormancy.” Nature Reviews Cancer 711 (2007): 834–846. [PubMed: 17957189]

32. Ding Ning, et al. “Cigarette smoking, smoking cessation, and long-term risk of 3 major atherosclerotic diseases.” Journal of the American College of Cardiology 744 (2019): 498–507. [PubMed: 31345423]

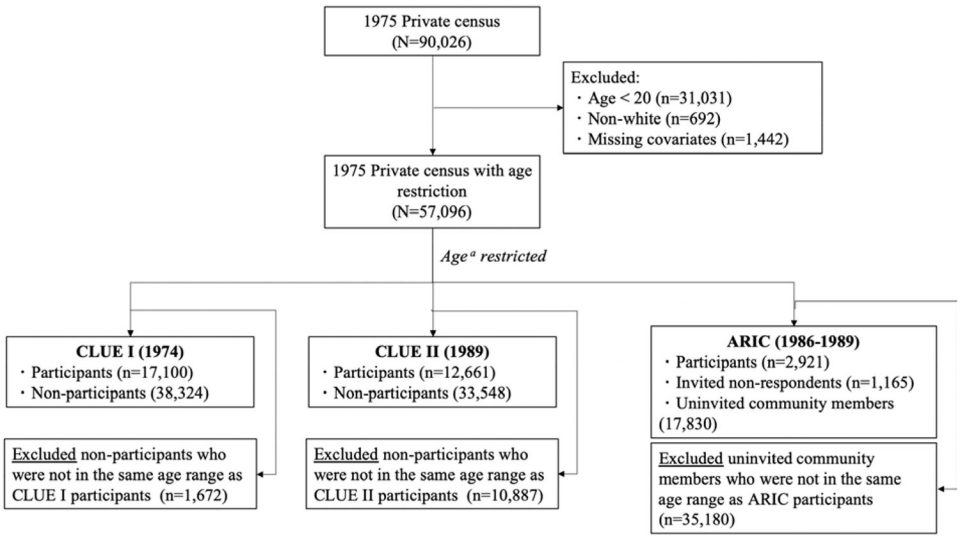

Figure 1.

Diagram of the study populations including cohort participants and comparison non-participant groups in the 1975 private census

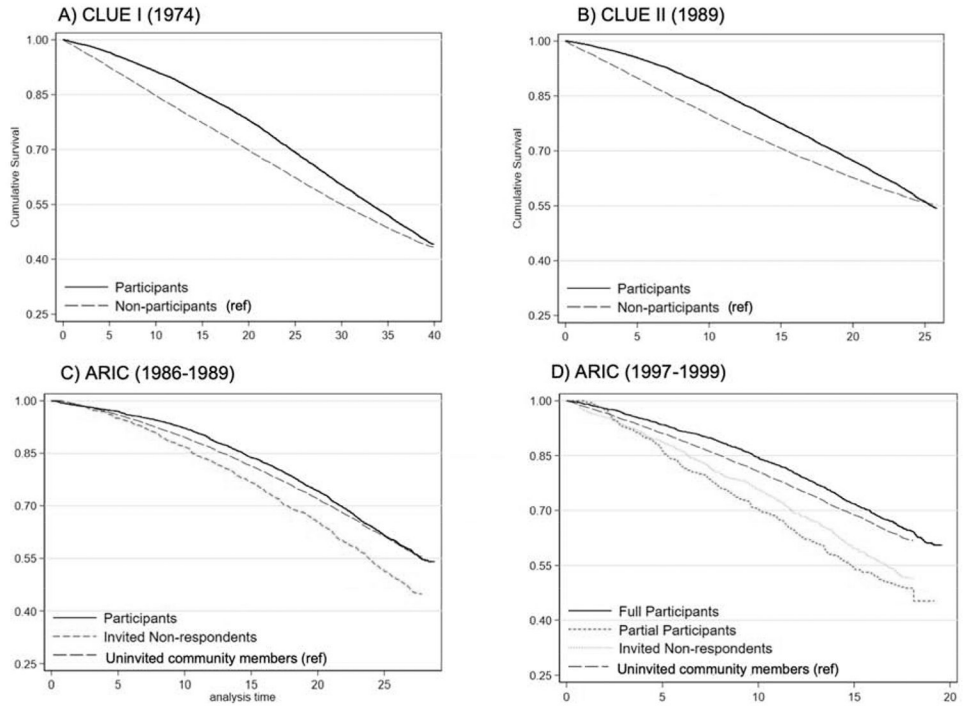

Figure 2.

Kaplan-Meier curves of all-cause mortality by participation in CLUE I (A), CLUE II (B), and ARIC (C & D).

Panel A: Cumulative survival of CLUE I participants and non-participants, followed from CLUE I baseline in 1974. Panel B: Cumulative survival of CLUE II participants and non-participants, followed from CLUE II baseline in 1989. Panel C: cumulative survival of ARIC participants, invited non-respondents, and uninvited non-participants, followed from ARIC baseline visit in 1986–1989; Panel D: cumulative survival of ARIC subtypes of participation-full participants, partial participants, invited non-respondents, and uninvited non-participants, followed from ARIC visit 4 in 1997–1999.

Table 1.

Characteristics of Participants and Non-Participants in CLUE and ARIC Studies, Washington County, Maryland, United States

	CLUE I		CLUE II		ARIC		
	Participants	**Non-participants**	**Participants**	**Non-participants**	**Participants**	**Invited non-respondents[b]**	**Uninvited community members[b]**
Counts	17,100	38,324	12,661	33,548	2,921	1,165	17,830
Age[a], years, mean (SD)	46.4 (14.7)	46.8 (17.7)	56.4 (12.7)	55.7 (15.1)	55.4 (5.5)	54.5 (5.9)	54.2 (6.9)
Male, n (%)	6,866 (40.2)	19,060 (49.7)	5,141 (40.6)	15,958 (47.6)	1,338 (45.8)	559 (48.0)	8,516 (47.8)
Education, %							
<12 years	5,800 (33.9)	17,334 (45.2)	3,583 (28.3)	12,554 (37.4)	936 (32.0)	574 (49.3)	6,474 (36.3)
=12 years	6,975 (40.8)	13,778 (36.0)	5,912 (46.7)	13,430 (40.0)	1,363 (46.7)	447 (38.4)	7,592 (42.6)
>12 years	4,325 (25.3)	7,212 (18.8)	3,166 (25.0)	7,564 (22.5)	622 (21.3)	144 (12.4)	3,764 (21.1)
Marital status, n (%)							
Currently married	13,791 (80.6)	28,680 (74.8)	10,609 (83.8)	25,656 (76.5)	2,710 (92.8)	1,052 (90.3)	15,504 (87.0)
Never married	1,098 (6.4)	3,729 (9.7)	988 (7.8)	4,051 (12.1)	65 (2.2)	49 (4.2)	790 (4.4)
Divorced/separated/widowed	2,211 (12.9)	5,915 (15.4)	1,064 (8.4)	3,841 (11.4)	146 (5.0)	64 (5.5)	1,536 (8.6)
Smoking, n (%)							
Never	7,524 (44.0)	15,499 (40.4)	5,878 (46.4)	13,601 (40.5)	1,190 (40.7)	395 (33.9)	6,318 (35.4)
Current	5,362 (31.4)	14,807 (38.6)	3,726 (29.4)	13,294 (39.6)	1,034 (35.4)	566 (48.6)	7,502 (42.1)
Former	4,214 (24.6)	8,018 (20.9)	3,057 (24.1)	6,653 (19.8)	697 (23.9)	204 (17.5)	4,010 (22.5)
Ever had cancer, n (%)	677 (4.0)	1148 (3.0)	379 (3.0)	717 (2.1)	67 (2.3)	34 (2.9)	418 (2.3)

Abbreviation: SD: standard deviation.

[a]Age was calculated from date of birth and baseline visit date for participants and median date of baseline for non-participants

[b]Invited non-respondents: participants of the 1975 private census who were invited to the ARIC study during cohort enumeration but refused to participate.

Uninvited non-participants (approximate study-define population): participants of the 1975 private census who were eligible but were not invited for ARIC study (e.g. age between 45–64 in 1987).

Table 2.

Standardized 10-Year All-Cause Mortality across Different Participation Groups

	CLUE I (Baseline: 1975)		CLUE II (Baseline: 1989)		ARIC (Baseline: 1987)		
	Participants	**Non-participants**	**Participants**	**Non-participants**	**Participants**	**Invited nonrespondents**	**Uninvited community members**
No. people at risk	17,100	38,324	12,661	33,548	2,921	1,165	17,830
No. of deaths	9,550	21,680	5,782	14,948	1,259	642	7,916
Age-standardized mortality rate per 1,000 person-years (95% confidence interval)							
0–10 years	10.4 (9.8, 11.1)	15.6 (15.2, 16.0)	13.9 (13.0, 14.8)	22.0 (21.5, 22.6)	7.8 (6.8, 8.8)	14.1 (11.6, 16.5)	11.4 (10.8, 11.9)
10–20 years	15.8 (15.1, 16.4)	19.7 (19.1, 20.2)	24.4 (22.5, 26.2)	26.2 (25.5, 26.8)	20.5 (18.7, 22.4)	30.0 (26.0, 34.1)	22.6 (21.8, 23.4)
20–30 years	22.6 (21.8, 23.5)	25.5 (24.8, 26.2)	30.8 (29.3, 32.4)	22.9 (22.0, 23.9)	35.6 (32.3, 38.9)	52.0 (44.8, 59.2)	34.8 (33.5, 36.1)
>30 years	26.6 (25.6, 27.6)	26.4 (25.6, 27.2)	NA	NA	NA	NA	NA

Abbreviation: NA: not applicable

Table 3.

Hazard ratio associated with cohort participation compared to non-participants across 10-year stratified periods of follow-up time in CLUE I and CLUE II

Models [b]	10-year period of follow-up	Hazard Ratio (95% CI)	
		CLUE I	CLUE II
Model 1	0–10 year	0.54 (0.51, 0.57)	0.59 (0.56, 0.62)
	10–20 year	0.82 (0.78, 0.86)	1.07 (1.02, 1.12)
	20–30 year	1.08 (1.03, 1.13)	1.73 (1.62, 1.84)
	>30 year	1.31 (1.24, 1.37)	NA
Model 2	0–10 year	0.69 (0.65, 0.73)	0.64 (0.60, 0.67)
	10–20 year	0.82 (0.78, 0.86)	0.98 (0.93, 1.03)
	20–30 year	0.98 (0.93, 1.02)	1.68 (1.57, 1.78)
	>30 year	1.24 (1.18, 1.30)	NA
Model 3	0–10 year	0.72 (0.68, 0.77)	0.69 (0.65, 0.73)
	10–20 year	0.89 (0.85, 0.93)	1.05 (1.00, 1.10)
	20–30 year	1.04 (1.00, 1.09)	1.75 (1.64, 1.87)
	>30 year	1.30 (1.24, 1.37)	NA

Abbreviation: CI: confidence interval; NA: not applicable

[a]Model 1: unadjusted; model 2: adjusted for baseline age and sex; model 3 adjusted for baseline age, sex, education level, marital status, smoking status, and cancer history.

Table 4.

Hazard ratio associated with cohort participation status compared to uninvited community members across 5-year stratified periods of follow-up time in ARIC

Models	Follow-up Period	Hazard Ratio (95% CI)		
		Enumerated individuals (n=4,086)	Enumerated Individuals	
			Participants (n=2,921)	Invited non-respondents (n=1,165)
Model 1	0–5 years	0.91 (0.76, 1.09)	0.78 (0.62, 0.97)	1.26 (0.96, 1.64)
	5–10 years	0.87 (0.76, 1.00)	0.72 (0.60, 0.86)	1.26 (1.02, 1.56)
	10–15 years	1.10 (0.98, 1.23)	1.01 (0.89, 1.16)	1.33 (1.11, 1.61)
	15–20 years	1.05 (0.94, 1.16)	0.96 (0.85, 1.09)	1.28 (1.08, 1.53)
	20–25 years	1.26 (1.15, 1.39)	1.18 (1.05, 1.32)	1.52 (1.29, 1.78)
	>25 years	1.15 (0.99, 1.33)	1.06 (0.88, 1.27)	1.35 (1.07, 1.71)
Model 2	0–5 years	0.88 (0.74, 1.05)	0.73 (0.59, 0.91)	1.29 (0.99, 1.69)
	5–10 years	0.83 (0.72, 0.95)	0.67 (0.56, 0.80)	1.27 (1.03, 1.57)
	10–15 years	1.04 (0.93, 1.17)	0.93 (0.82, 1.07)	1.35 (1.12, 1.63)
	15–20 years	0.97 (0.87, 1.08)	0.87 (0.77, 0.99)	1.30 (1.09, 1.55)
	20–25 years	1.17 (1.06, 1.29)	1.05 (0.94, 1.18)	1.55 (1.32, 1.82)
	>25 years	1.04 (0.90, 1.21)	0.92 (0.77, 1.10)	1.37 (1.08, 1.73)
Model 3	0–5 years	0.92 (0.77, 1.09)	0.80 (0.65, 1.00)	1.16 (0.89, 1.52)
	5–10 years	0.86 (0.74, 0.99)	0.73 (0.61, 0.88)	1.14 (0.93, 1.41)
	10–15 years	1.08 (0.96, 1.21)	1.01 (0.88, 1.15)	1.24 (1.03, 1.50)
	15–20 years	1.00 (0.89, 1.11)	0.93 (0.82, 1.05)	1.18 (0.99, 1.41)
	20–25 years	1.18 (1.07, 1.30)	1.09 (0.97, 1.22)	1.44 (1.22, 1.69)
	>25 years	1.04 (0.89, 1.21)	0.93 (0.78, 1.12)	1.29 (1.02, 1.64)

[a]Model 1: unadjusted; model 2: adjusted for baseline age and sex; model 3 adjusted for baseline age, sex, education level, marital status, smoking status, and cancer history.

[b]Enumerated individuals (n=4,086) includes 2,921 participants and 1,165 invited non-respondents.