Estimating Healthy Volunteer Bias Impact on Mortality in Observational Cohorts - a Projection for the German National Cohort

Paper submitted for presentation at PAA 2015

San Diego CA April 30 – May 2, 2015

Ulrich Mueller* & Wolfgang Hoffmann**

*Competence Center Mortality-Follow-Up of the German National Cohort

Philipps-University Marburg

mueller2@staff.uni-marburg.de

**Study Center Neubrandenburg of the German National Cohort

Ernst Moritz Arndt-University Greifswald

wolfgang.hoffmann@uni-greifswald.de

Abstract

Surveys participants are healthier and live longer than non-participants. Since health status influences all human activities, this healthy volunteer bias may compromise generalizability of survey findings to non-participants and the population of reference. Relevant are long-term prospects of the bias: grow, stabilize or diminish? Using data on subjective health in 13.226 respondents aged 20-69 in the German General Social Survey ALLBUS 2004-2012, we estimate the healthy volunteer bias impact on projected mortality and survival of the study protocol population of the German National Cohort (GNC) with n=200.000 aged 20-69, recruited 2014-2018. We show that paradoxically the healthy volunteer bias may increase with participation rates, but will have diminishing effects on generalizability of survey findings. Even with persistent survival differentials, any healthy volunteer bias in recruitment will leave only moderate traces in the mortality-follow-up. Empirically, survival differentials due to the healthy volunteer bias do not persist over long times anyway.

1. Objective

To assess the healthy volunteer bias's effect on survival and mortality with the German National Cohort, with a planned size of recruited n=200.000 one of the largest observational cohorts ever.

2. Introduction

In all kinds of surveys, where participation is voluntary or cannot be enforced totally, participants tend to be healthier, more health conscious, and more likely to survive the next years than non-participants. To the degree that healthy subjects behave differently in most fields of life, researchers may be concerned about how findings obtained from healthier participants may be generalized to non-participants or the population from which the sample has been drawn.

This healthy volunteer bias problem in health focussed observational cohorts raises at least four questions:

1. Which inferences about present health status can be made from participants to non-participants?
2. Which inferences about present health status can be made from participants to the population from which the sample was drawn – the sample frame of reference as well as the population total about which meaningful statements shall be made?
3. What are the long-term differential survival prospects for participants as compared to non-participants?
4. What are the long-term differential survival prospects for participants as compared to the population from which the sample was drawn – the sample frame of reference as well as the population total about which meaningful statements shall be made?

There is an abundance of reports on the healthy volunteer bias in various survey settings, and also on its effect on mortality:

Observation Studies

A „healthy volunteer bias“ for mortality in observational cohorts in Switzerland (Bopp et al. 2010)[1] (three waves of the MONICA study recruited between 1983-1992) and in a very large cohort in Austria (Klenk et al. 2009)[2], up to 40% Relative Risk Reduction within the first 1-2 decades has been reported - meaning that the survey participants population had a Standardized Mortality Ratio (SMR) of about 60% of the age and sex adjusted general population.

It is generally known, however, that the „healthy volunteer bias“ is maximal upon recruitment and tends to diminish over time, although there may be exceptions, depending on the cause of death.

In a large middle-aged observational cohort in Japan Hara et al. (2002)[3] found SMRs of about 50% for all causes of death, 70% for all cancers, even 37% for cerebrovascular and other circulatory causes in males, and 63% / 82% / 65% for females, in participants as compared to nonparticipants, but no such effect for ischemic heart disease. Also, the healthy volunteer effect disappeared for cancer mortality already 2 years after baseline, but remained stable over full 8 years observation for cerebrovascular causes.

Masters et al. (2013)[4] demonstrated in large datasets from the US National Health Interview Survey 1986-2004 linked to the National Death Index, that the weakening of the association between obesity and mortality with advancing age, that previous studies had found, is caused by confounding by disparate cohort mortality and age-related survey selection bias. In the obese population the healthy volunteer bias is even stronger than in the normal-weight population, and this effect grows with age. After statistical controlling for this, the association between obesity and mortality in fact grows with advancing age.

Screening Studies

In an Israeli cohort of industrial workers initially screened for cardiovascular disease in 1985 with 72% participants, Froom et al. (1999)[5] found that the SMR over 8 years follow-up for all cause mortality was 71% for participants and 99% for non-participants as compared to the age-adjusted general population – indicating a healthy volunteer effect on top of a – albeit small- healthy worker effect.

In a screening trial on prostate, lung, colorectal, and ovarian (PLCO) cancer, that had randomized 155.000 men and women 1993-2001, Pinsky et al. (2007)[6] found even after the exclusion of PLCO mortality an all cause SMR of 46% for men and 38% for women. SMRs increased from 31% at year 1 to 48% at year 7 of the study. For prostate cancer alone, Pinsky et al. (2012)[7] found 5 years later a prostate cancer specific SMR of 60% for the intervention and 55% of the control arm of the PLCO Cancer Screening Trial as compared to the general US population. SMR avoid biases of lead time and overdiagnosis, but are interpreted by the authors as the result of a healthy volunteer bias.

Intervention Studies

In a prospective study on mortality difference between participants and non-participants in a comprehensive health examination for the prevention of geriatric syndromes among 854 community-dwelling elderly people aged 70-84 in Tokyo, Iwasa et al. (2007)[8] after a three-years observation period found an all-cause Standardized Mortality Ratio (SMR) of 0.44 (95% CI 0.24-0.78) for participants, which authors attribute to self-selection of healthier individuals as well as an intervention effect of improved care as result of the examination.

Mere participation in a industry sponsored RCT may lead to a Relative Mortality Risk Reduction in mortality of more than 50% over 2 years observation, even with co-morbidity adjusted. Neither participating in the active arm vs. belonging to the controls nor the study drug made a difference in this respect (Baker at al. 2013)[9]. Jordan et al. (2013)[10] in an RCT identified deprived persons and current smokers as less likely to stay in the Follow-Up observation – with ORs of 0.29 and 0,20 at 6 months and =0.68 and 0.55 at 2 years.

Inclusion in a RCT, however, typically includes a check-up which in some cases also will lead to some therapeutic intervention that otherwise would not have occurred. Thus, the mortality risk difference may be caused not only by self selection but also by some improvement in health care.

3. Data and Methods

German General Social Survey (ALLBUS)

Subjective health status is an established predictor of the mortality risk (for example Fernández-Ruiz et al. 2013[11], Mavaddat et al. 2014[12]), outperforming even selected multi-biomarker panels, although the subjective health status predictive power may be further augmented by taking into account suitable biomarkers (for example Haring et al. 2011)[13]. However, the distribution of answer categories is known to vary between national settings.

Therefore, we use the distribution of the single-item self-rated health question in the German General Social Survey (ALLBUS) ("*Now I would like to ask some questions about your health. How would you describe your health in general? - A Very good - B Good - C Satisfactory - D Poor - E Bad?*") as it was obtained from respondents aged 20-69 in the general residential German population in 2004, 2006, 2008, 2010 and 2012 (n=13226, 6569 males and 6669 females). In the ALLBUS, there was no association between survey year and subjective health score distribution (gamma = .004, p=.569). Therefore, we may take the survey distribution as a good approximation of the score distribution to be expected among among participants in the first wave of the German National Cohort recruited 2014-2018, also aged 20-69.

This approximation will be used for assessing the likely short term effects of the healthy volunteer bias.

Sampling plan of the study protocol of the German National Cohort & cohort life tables of the German Federal Statistical Office

Based on the sampling plan of the study protocol of the German National Cohort (n=200.000 between 20 and 69 years of age, recruited 2014-2018), one of the largest observational studies ever, and the model population data in the cohort life tables of the German Federal Statistical Office, we deterministically predict total number of deaths of males and females from 2014 up to the last year 2110, when the youngest age group recruited – the birth cohort of 1998, recruited in 2018 - reaches age 112, the ultimate age in the cohort life tables used here. We predict the death counts by single calendar year that is approximately attributable to the birth cohorts of the five age groups at recruitment (20-29, 30-39, 40-49, 50-59, 60-69), namely 1945-1956, 1957-1966, 1967-1976, 1977-1986, 1987-1998. We differentiate by age group at death of 20-49, 20-64, 65-74, 75-84, 85-99, 100+, 105+, 110+. The health care quality parameter "avoidable deaths" or "mortality amenable to medical/health care" referring to deaths which by most definitions occur before age 49, 65 or 75, can be observed from survey onset until 2047, 2063 or 2073. On the other hand, the German National Cohort, from 2045 onwards, can be expected to generate centenarians (100+), of those from 2050 onwards semisupercentenarians (105+) and of those from 2055 onwards supercentenarians (110+). We differentiate between participants and non-participants in order to estimate the effects of the "healthy volunteer bias" under moderate and under strong assumptions.

For predicting death counts in the German National Cohort until the end of this century and beyond, there is no reliable information on the effects of a healthy volunteer bias at recruitment over such a long period of time. Even in the longest-running observational cohorts for which a healthy volunteer bias effect on mortality has been described (for example Bopp et al. 2010), the observation period is less than 30 years.

Instead of applying an analytical transformation (for example proportional hazard rates or accelerated/decelerated failure times) of general population mortality risks, we use the secular trend of increasing longevity expressed in cohort life tables: from 1935- to – mostly extrapolated – 2008 birth cohorts, life span increased about two and a half year every decade. Also, at age 51 (the median age of the recruitment sample of the German National Cohort) the Relative Risk Reduction in All-Cause Mortality is 15-20% in the observed mortality rates per decade. Thus we opt for modelling „*healthy volunteer bias*" by assuming a more recent birth cohort survival function for participants and an older one for non participants.

Since the study protocol assumes recruitment rates of about 40-50%, we may assume for non-participants a symmetrical negative RRR = Relative Risk Increase about equal in size.

Arbitrarily we chose a moderate assumption approach: participants' survival of a birth cohort 5 years younger, non-participants' survival of a birth cohort 5 years older than the actual birth year. For the oldest birth cohort 1945 recruited in 2014 only, this approach would result in a difference in remaining lifespan of 18,3 years for non-participating vs. 19,2 years for participating females (a difference of 0,9 years) and 15,4 years for non-participating vs. 16,2 years for participating males (a difference of 0,8 years). Fn the youngest birth cohort 1998 recruited in 2018 only, it would result in a difference in remaining lifespan of 71,6 years for non-participating vs.72,6 years for participating females (a difference of 1,0 year) and 67,0 for non-participating vs. 68,2 years for participating males (a difference of 1,2 years).

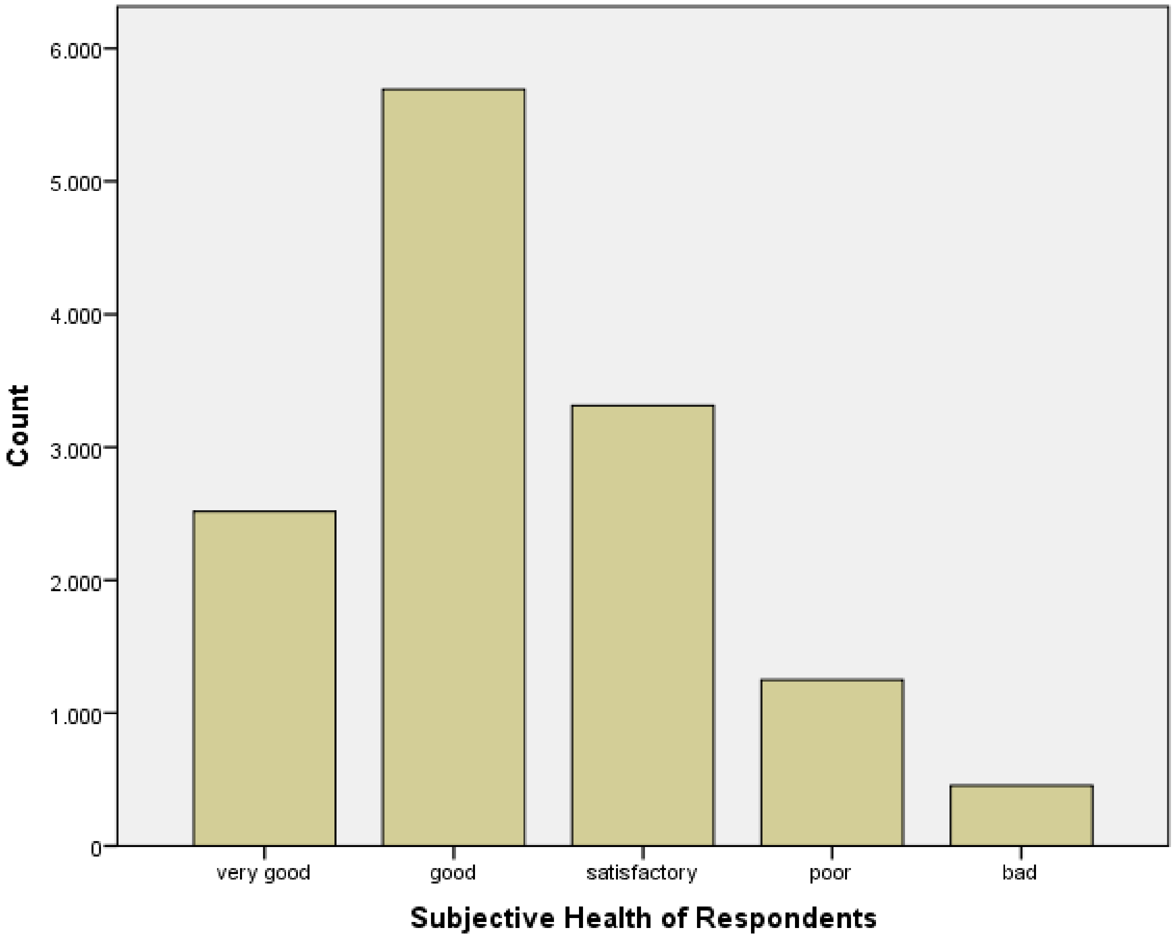
In addition, we chose a strong assumption approach: participants' survival of a birth cohort 10 years younger, non-participants' survival of a birth cohort 10 years older than the actual birth year.

For the oldest birth cohort 1945 recruited in 2014 only, this approach would result in a difference in remaining lifespan of 17,8 years for non-participating vs. 19,6 years for participating females (a difference of 1,8 years) and 14,9 years for non-participating vs. 16,7 years for participating males (a difference of 1,8 years). For the youngest birth cohort 1998, recruited in 2018 only, it would result in a difference in remaining lifespan of 71,0 years for non-participating vs. 73,0 years for participating females (a difference of 2,0 years) and 66,4 for non-participating vs. 68,8 years for participating males (a difference of 2,4 years).

4. Results

In the ALLBUS survey population, we get the distribution of subjective health scores as in Figure 1. Clearly, there was a substantial proportion of respondents who placed their health in the worst category offered, which means that bad health did not perfectly predict survey participation.

Figure 1



Distribution of subjective health scores in the ALLBUS sample

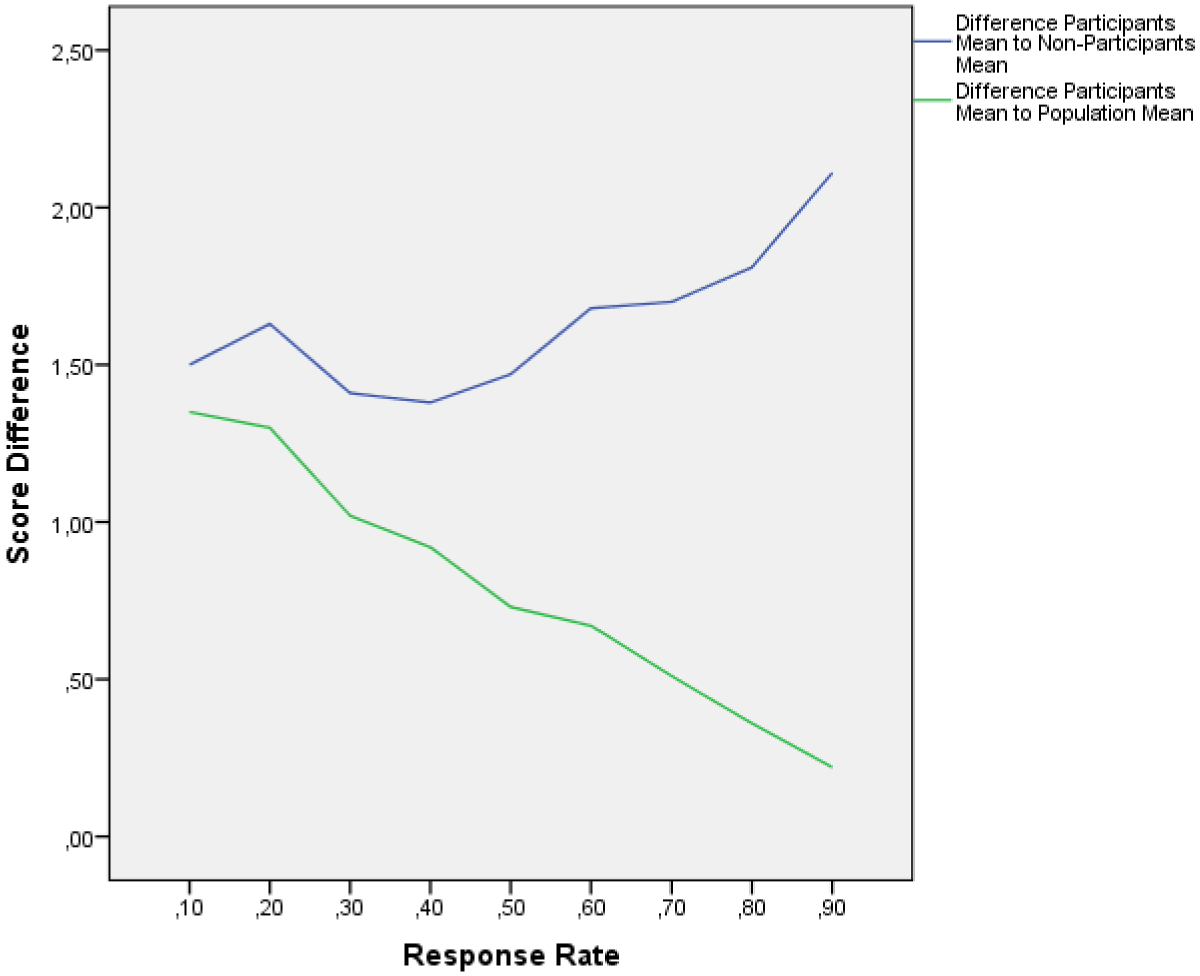
We do not know and cannot even make intelligent guesses about the subject health status of non-respondents in the ALLBUS surveys nor in the German National Cohort, and this statement of ignorance applies the more the worse the self rated health is among people who declined to participate.

However, we can make intelligent guesses about generalizability of the survey findings, given the skewed distribution of the subjective health scores, namely:

If participation strongly depends on Health Status, then with increasing response rate the Healthy Volunteer Bias – the difference in average health score between those we could measure and those we could not – will increase. With increasing response rate, on the other hand, generalizability of the sample findings to the population from which the sample was drawn – the sample frame of reference as well as the population total – will increase.

For a thought experiment, we assume that the distribution of subjective health scores in the population is as in Figure 1, but that response completely depends on subjective health status. Then, with increasing response rates, paradoxically, with the healthy volunteer bias, as measured by the difference between participants' score mean and non-participants' score mean will increase, but it will matter less and less for generalizing the survey findings to the population, as the difference between participants' score mean and total population score mean will get smaller and smaller, as shown in Figure 2

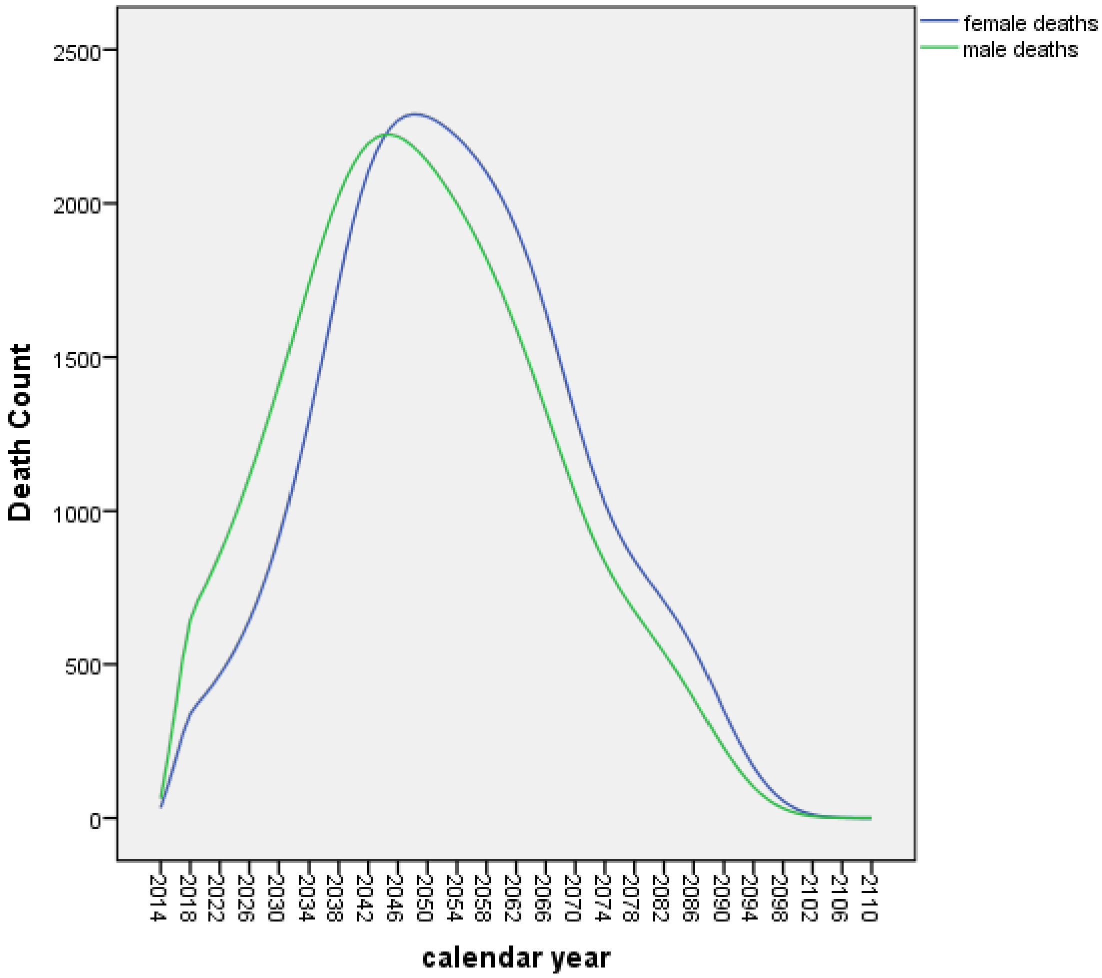
Figure 2



Next we deterministically predict total number of deaths of males and females from 2014 up to the last year 2110, when the youngest age group recruited – the birth cohort of 1998, recruited in 2018 - reaches age 112, the ultimate age in the cohort life tables used here.

Figure 3 shows the absolute death counts for males and females alike in the study protocol population of the German National Cohort.

Figure 3



Absolute death counts for males and females in the study protocol population of the German National Cohort

We differentiate between participants and non-participants in order to estimate the long-term effects of the “healthy volunteer bias”.

Predicting the size and time pattern of a „*healthy volunteer bias*“ over the lifetime of study participants, recruited between age 20 and 69, as in the German National Cohort, will be difficult. The reported follow-up times are typically shorter than a decade for intervention studies and may be somewhat longer for screening trials. For observational studies follow-up times reported in the literature often do not exceed 15-20 years.

Instead of applying an analytical transformation (for example proportional hazard rates or accelerated/decelerated failure times) of general population mortality risks, we use the secular trend of increasing longevity expressed in cohort life tables: From

1935- to – mostly extrapolated – 2008 birth cohorts, life span increased about two and a half year every decade.

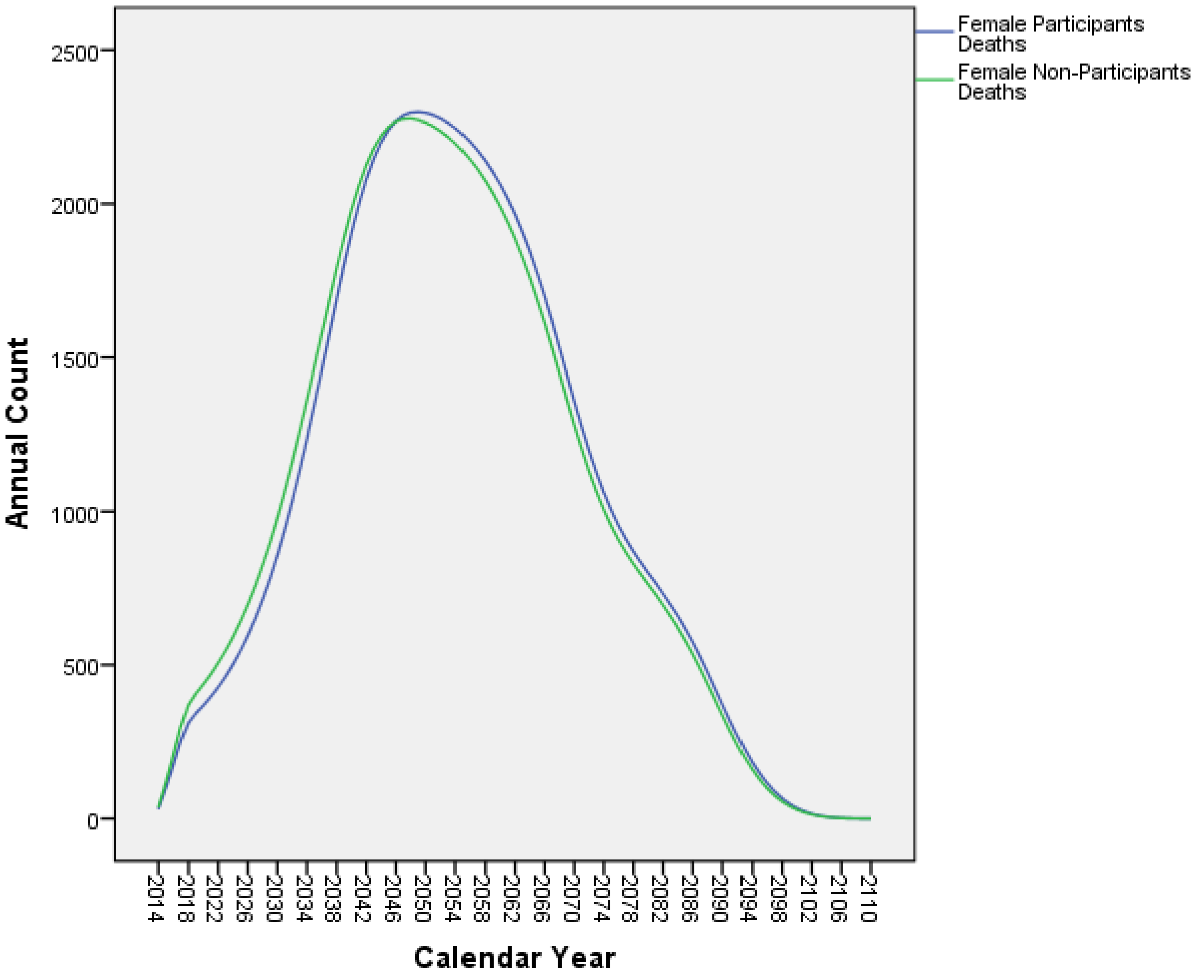
Also, at age 51 (the median age of the recruitment sample of the German National Cohort) the Relative Risk Reduction in All-Cause Mortality is 15%-20% in the observed mortality rates per decade.

We will test a moderate scenario, assuming that participants will have the survival of a cohort born 5 years later and non-participants of a cohort born 5 years earlier. Here, we would have a Relative Risk Reduction in All-Cause Mortality of 15%-20%.

Also, we test a strong scenario, assuming that participants will have the survival of a cohort born 10 years later and non-participants of a cohort born 10 years earlier. Here, we would have a Relative Risk Reduction in All-Cause Mortality of 30%-40%.

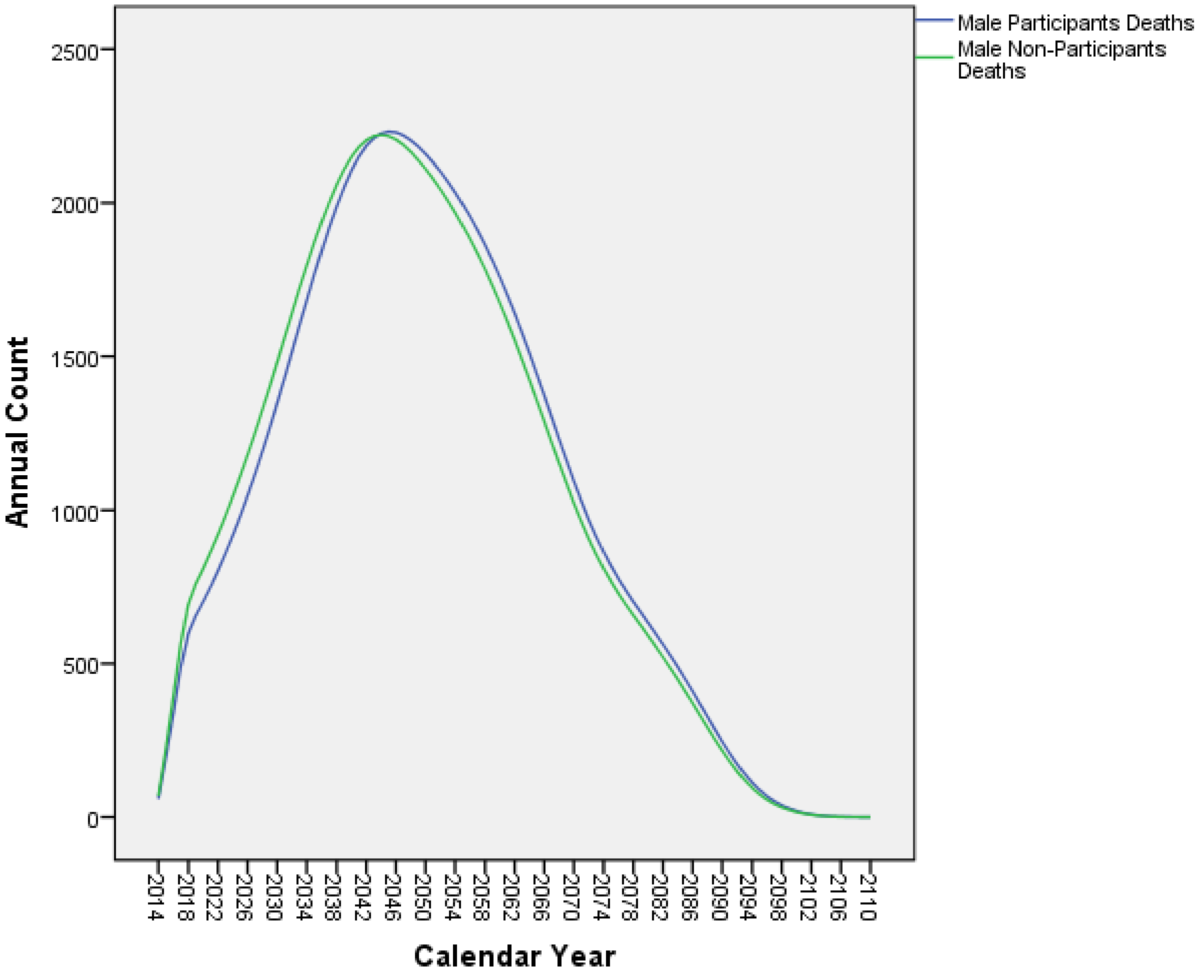
Figure 4 show the death counts for participating vs. non-participating females, Figure 5 for males in the moderate scenario.

Figure 4



Female death counts among participants vs. non-participants – moderate scenario.

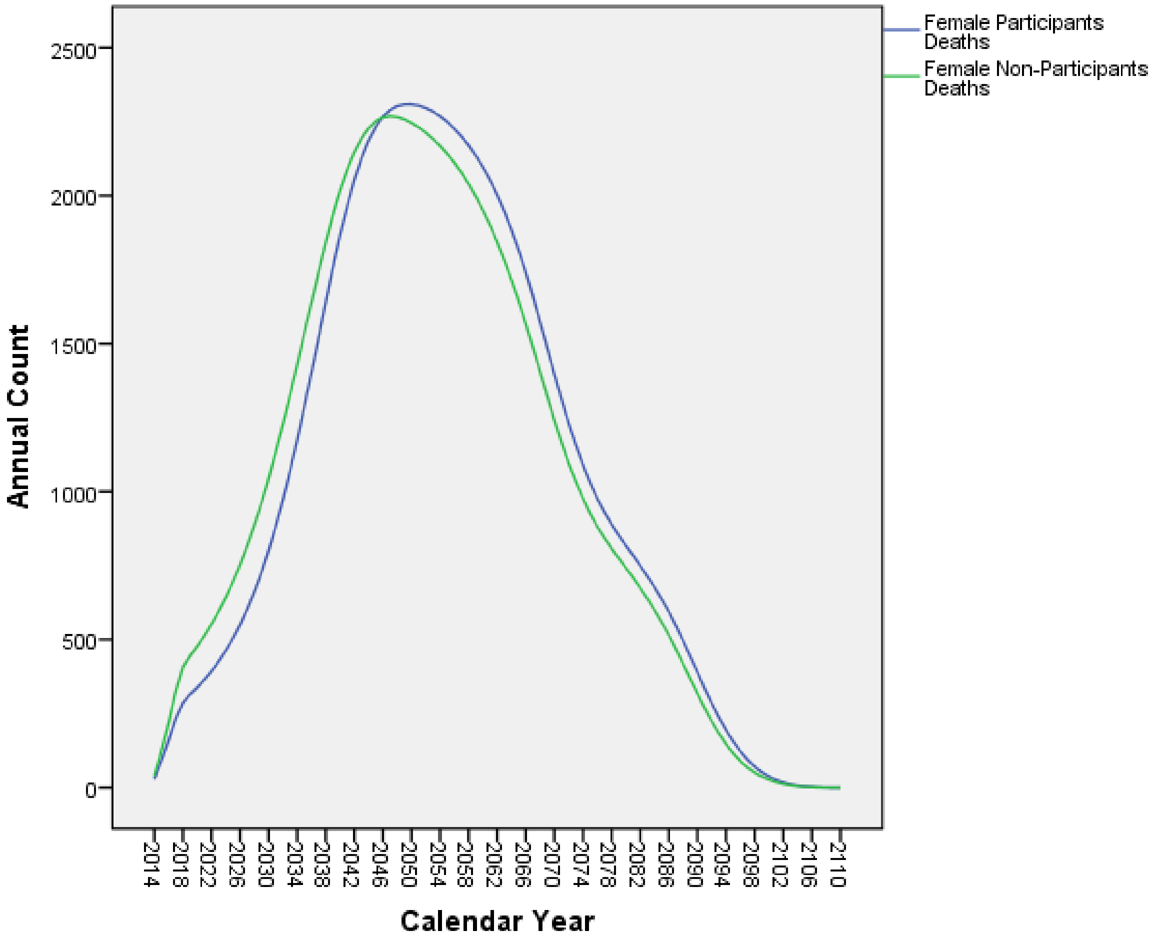
Figure 5



Male death counts among participants vs. non-participants – moderate scenario.

Figure 6 show the death counts for participating vs. non-participating females, Figure 7 for males in the strong scenario.

Figure 6

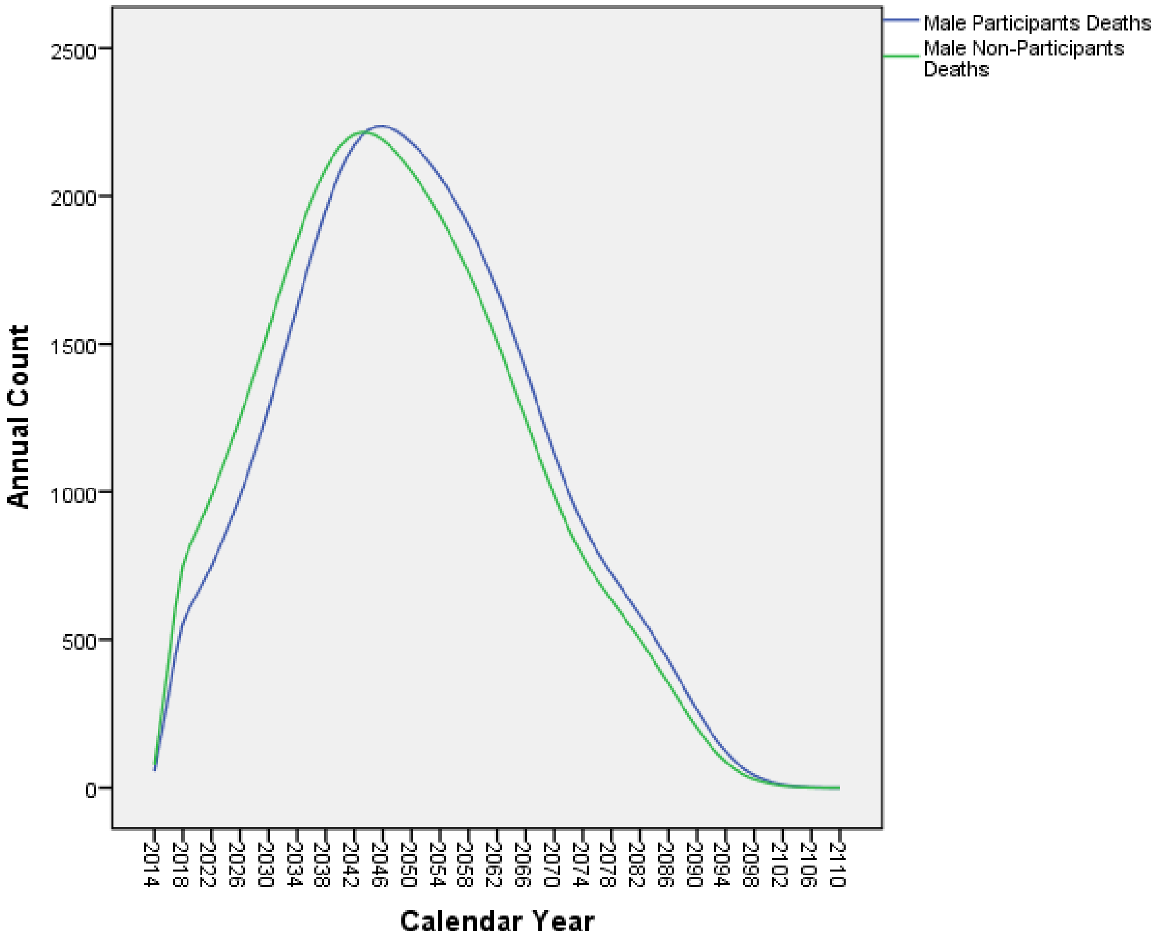


Female death counts among participants vs. non-participants – strong scenario.

In the moderate scenario, there will be a difference of just one year in the median death year for female as well as male participants vs. non-participants, respectively.

Even, however, if the Healthy Volunteer Bias Effect on Mortality in the German National Cohort is modelled as lifetime survival differential in the size of 20 years difference in birth cohort between participants and non-participants, there will be a difference of median year of death of 2 years between participating and non-participating females (2054 vs. 2052) and of 3 years between participating and non-participating males (2050 vs. 2047). If we compare participating subjects with the total population average, the Healthy Volunteer Bias Effect on Mortality shrinks to one year difference in median death year for females (2054 vs. 2053) and two for males (2050 vs. 2048).

Figure 7



Male death counts among participants vs. non-participants – strong scenario.

5. Discussion

In the short-term, generalizability of the sample findings despite a Healthy Volunteer Bias, may be only moderately affected, if we want infer from observed subjects to the general population from which the sample of recruited subjects is drawn.

In this respect, the healthy volunteer bias will be present, but will be not relevant for this purpose of the cohort.

In the long-term, in the moderate assumption scenario there is a difference in remaining life expectancy of about one year between participants and non-participants from the oldest to the youngest cohort. In the strong assumption scenario, this difference increases to 2 – 3 years. From all what is known about the “fading out” of a healthy volunteer bias, the strong scenario seems rather unlikely.

Furthermore, also in lifetime mortality, the difference between recruited subjects and the assumed population mean, will be even smaller.

However, a healthy volunteer bias in survival that persists over 25 – 30 years has never been described in reality.

Thus, the actual effects of a healthy volunteer bias probably may be even smaller than predicted by the moderate scenario, which predicts a difference of just one year in the median death year for female as well as male participants vs. non-participants, respectively.

It may come as a surprise, how small the effects of a “healthy volunteer bias” will be for the mortality distribution over the decades to come. Its effects can likely be neutralised by introducing a short time lag in the statistical models.

6. Conclusions:

1. A Healthy Volunteer Bias, as in all surveys, will be present also in the German National Cohort at recruitment, but by all experience will never be that extreme that contacted people in the worst subjective heath category will completely refrain from participating.
2. Given the skewed distribution of subjective health scores found in all general population samples, there is a paradoxical effect to be expected with for the response rate: the higher the response rate, the larger the Healthy Volunteer Bias, if participants are compared with non-participants, but the smaller the Healthy Volunteer Bias if participants are compared with the total population average – the sample frame of reference.
3. Since the Healthy Volunteer Bias is known to fade away already after the first years of follow-up, recruited subjects whose health worsens during the course of the study, may be not very different from contacted people, who for poor health reasons had decided against recruitment already in the first wave.
4. Even if the Healthy Volunteer Bias Effect on Mortality in the German National Cohort is modelled in a unrealistic strong scenario of a lifetime survival differential in the size of 20 years difference in birth cohort, there will be a difference of median year of death of 2 years between participating and non-participating females (2054 vs. 2052) and of 3 years between participating and non-participating males (2050 vs. 2047). If we compare participating subjects with the total population average, the Healthy Volunteer Bias Effect on Mortality shrinks to one year difference in median death year for females (2054 vs. 2053) and two for males (2050 vs. 2048). Lifetime survival differentials due to a Healthy Volunteer Bias at recruitment that persist over 30+ years after recruitment apparently have never been described.
5. The Healthy Volunteer Bias may be a serious problem for generalizing cohort subjects’ morbidity and mortality to the general population within the first years

after recruitment of the German National Cohort, but then will fade away fast. It's effects on generalizability of overall mortality findings within the cohort sample will be negligible.

6. Substantial efforts have to be made to maintain contact with recruited cohort subjects whose health worsens during the study observation time. They may serve as surrogates for people who for poor health reasons had decided against recruitment at all, and, thus, help to correct for this initial Healthy Volunteer Bias in the baseline data set.

7. Literature

[1] Bopp M, Braun J, Faeh D, Gutzwiller F; Swiss National Cohort Study Group.(2010): Establishing a follow-up of the Swiss MONICA participants (1984-1993): record linkage with census and mortality data. BMC Public Health. 2010 Sep 21;10:562

[2] Klenk J, Nagel G, Ulmer H, Strasak A, Concin H, Diem G, Rapp K; VHM&PP Study Group (2009): Body mass index and mortality: results of a cohort of 184,697 adults in Austria. Eur J Epidemiol. 2009;24(2):83-91

[3] Hara M, Sasaki S, Sobue T, Yamamoto S, Tsugane S. (2002): Comparison of cause-specific mortality between respondents and nonrespondents in a population-based prospective study: ten-year follow-up of JPHC Study Cohort I. Japan Public Health Center. J Clin Epidemiol. 2002 Feb;55(2):150-6

[4] Masters RK, Reither EN, Powers DA, Yang YC, Burger AE, Link BG.(2013): The impact of obesity on US mortality levels: the importance of age and cohort factors in population estimates. Am J Public Health. 2013 Oct;103(10):1895-901.

[5] Froom P, Melamed S, Kristal-Boneh E, Benbassat J, Ribak J. (1999): Healthy volunteer effect in industrial workers. J Clin Epidemiol. 52:731-5.

[6] Pinsky PF, Miller A, Kramer BS, Church T, Reding D, Prorok P, Gelmann E, Schoen RE, Buys S, Hayes RB, Berg CD (2007): Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. Am J Epidemiol. 165(8):874-81

[7] Pinsky PF, Black A, Parnes HL, Grubb R, David Crawford E, Miller A, Reding D, Andriole G. (2012): Prostate cancer specific survival in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Cancer Epidemiol. 36(6):e401-6

[8] Iwasa H, Yoshida H, Kim H, Yoshida Y, Kwon J, Sugiura M, Furuna T, Suzuki T. (2007): A mortality comparison of participants and non-participants in a comprehensive health examination among elderly people living in an urban Japanese community. Aging Clin Exp Res. 19:240-5.

[9] Baker JR, Vandal AC, Yeoh J, Zeng J, Wong S, Ryan SN (2013): Clinical trial participation improves outcome: a matched historical cohort study. Clin Trials 2013 Oct;10(5):735-43.

[10] Jordan S, Watkins A, Storey M, Allen SJ, Brooks CJ, Garaiova I, Heaven ML, Jones R, Plummer SF, Russell JT, Thornton CA, Morgan G (2013): Volunteer Bias in Recruitment, Retention, and Blood Sample Donation in a Randomised Controlled Trial Involving Mothers and Their Children at Six Months and Two Years: A Longitudinal Analysis. PLOS ONE 8 (7)

[11] Fernández-Ruiz M, Guerra-Vales JM, Trincado R, Fernández R, Medrano MJ, Villarejo A, Benito-León J, Bermejo-Pareja F. (2013): The ability of self-rated health to predict mortality among community-dwelling elderly individuals differs according to the specific cause of death: data from the NEDICES cohort. Gerontology. 2013;59(4)

[12] Mavaddat N, Parker RA, Sanderson S, Mant J, Kinmonth AL (2014): Relationship of self-rated health with fatal and non-fatal outcomes in cardiovascular disease: a systematic review and meta-analysis. PLoS One. 9(7)

[13] Haring R, Feng YS, Moock J, Völzke H, Dörr M, Nauck M, Wallaschofski H, Kohlmann T.(2011): Self-perceived quality of life predicts mortality risk better than a multi-biomarker panel, but the combination of both does best. BMC Med Res Methodol. 11:103