



Original Contribution

Assessing the Role of Selection Bias in the Protective Relationship Between Caregiving and Mortality

The Caregiver-Study of Osteoporotic Fractures

Meghan L. Smith, Timothy C. Heeren, Lysie R. Ranker, and Lisa Fredman*

* Correspondence to Dr. Lisa Fredman, Epidemiology Department, School of Public Health, Boston University, 715 Albany Street, Boston, MA 02118 (e-mail: lfredman@bu.edu).

Initially submitted December 2, 2018; accepted for publication July 23, 2019.

Caregivers have lower mortality rates than noncaregivers in population-based studies, which contradicts the caregiver-stress model and raises speculation about selection bias influencing these findings. We examined possible selection bias due to 1) sampling decisions and 2) selective participation among women (baseline mean age = 79 years) in the Caregiver-Study of Osteoporotic Fractures (Caregiver-SOF) (1999–2009), an ancillary study to the Study of Osteoporotic Fractures (SOF). Caregiver-SOF includes 1,069 SOF participants (35% caregivers) from 4 US geographical areas (Baltimore, Maryland; Minneapolis, Minnesota; the Monongahela Valley, Pennsylvania; and Portland, Oregon). Participants were identified by screening all SOF participants for caregiver status (1997–1999; $n = 4,036$; 23% caregivers) and rescreening a subset of caregivers and noncaregivers matched on sociodemographic factors 1–2 years later. Adjusted hazard ratios related caregiving to 10-year mortality in all women initially screened, subsamples representing key points in constructing Caregiver-SOF, and Caregiver-SOF. Caregivers had better functioning than noncaregivers at each screening. The association between caregiving and mortality among women invited to participate in Caregiver-SOF (41% died; adjusted hazard ratio (aHR) = 0.73, 95% confidence interval (CI): 0.61, 0.88) was slightly more protective than that in all initially screened women (37% died; aHR = 0.83, 95% CI: 0.73, 0.95), indicating little evidence of selection bias due to sampling decisions, and was similar to that in Caregiver-SOF (39% died; aHR = 0.71, 95% CI: 0.57, 0.89), indicating no participation bias. These results add to a body of evidence that informal caregiving may impart health benefits.

ancillary studies; caregiving; cohort studies; mortality; participation bias; selection bias; survival

Abbreviations: ADL, activities of daily living; aHR, adjusted hazard ratio; CI, confidence interval; IADL, instrumental activities of daily living; RHR, relative hazard ratio; SOF, Study of Osteoporotic Fractures.

Theories of stress predict higher rates of adverse health outcomes among caregivers versus noncaregivers due to the chronic stress of caregiving (1). However, caregivers have lower mortality rates than noncaregivers in most population-based studies (2, 3), including the Caregiver-Study of Osteoporotic Fractures (Caregiver-SOF) (4), an ancillary study to the Study of Osteoporotic Fractures (SOF) (5). These lower mortality rates may be explained by caregivers keeping physically (6) and cognitively (7) active and reaping psychological benefits from caregiving, such as feeling valued and having a purpose in life (8). Yet they also could be due to selection bias, which is a concern in prospective studies of aging-related outcomes (9–11). In the current

study, we examined whether selection bias explained the protective relationship observed between caregiving and mortality among older women participating in Caregiver-SOF.

In cohort studies, selection bias occurs when both exposure (or predictors of exposure) and predictors of outcome status influence the probability of a participant's being in the analytical sample (12). Selection bias can occur through sampling decisions and selection procedures, such as decisions about exposure definitions or eligibility criteria (11, 13), or through participation factors, such as refusal or inability to take part in a study (14–17). Significant differences in sociodemographic or health characteristics of study participants and nonparticipants do not necessarily

indicate selection bias, because these may have little or no impact on associations of interest (14–18).

Selection bias from sampling decisions could have occurred in Caregiver-SOF through inclusion and exclusion criteria and through sample construction. In addition, as in any observational study, participation factors could have induced selection bias. Older adults whose physical or cognitive health is declining may be more likely to refuse to participate, or be unable to participate, in prospective studies. If health status differentially influenced selection or participation among caregivers and noncaregivers, the observed caregiving-mortality association would be expected to differ from that in the source population. Although selection bias via these mechanisms is theoretically possible, if selection pressures occurred equally in caregivers and noncaregivers (i.e., the same proportions of caregivers and noncaregivers did not participate), we would expect the distribution of participants' characteristics to differ between the source population and the Caregiver-SOF sample, without the caregiving-mortality association shifting.

To assess selection bias, exposure and outcome information are needed for both nonparticipants and participants. Most cohort studies lack information on nonparticipants, and thus selection bias is generally examined in studies conducted within population registries (11, 15, 16, 19, 20). Studies created from existing prospective cohorts, such as Caregiver-SOF, provide an opportunity to evaluate selection bias. Data on health, sociodemographic characteristics, and mortality were available for all SOF participants screened for Caregiver-SOF, regardless of whether they were invited to participate or agreed to participate in the study. These data allowed us to compare associations between caregiving and mortality in the Caregiver-SOF sample with that in the population from which the sample was drawn (SOF) and in subsets of SOF representing key points in the creation of the Caregiver-SOF sample. Finally, reasons for noninclusion were recorded, allowing us to distinguish different sources of potential selection bias.

We hypothesized that the magnitude of the protective relationship between caregiving and mortality over 10 years among all initially screened SOF participants would be similar to that among those SOF participants invited to participate in Caregiver-SOF, that is, no selection bias due to sampling decisions. We also hypothesized that the magnitude of this relationship in the sample of invited participants (including those who declined to participate) would be similar to that in the Caregiver-SOF sample, that is, no participation bias. In exploratory analyses, we examined whether our matching procedure influenced results by comparing results from a propensity-score-matched sample of participants eligible for Caregiver-SOF with those observed in the Caregiver-SOF sample.

METHODS

Study sample

Participants were drawn from SOF (5). The initial SOF sample comprised 9,704 white women aged 65 years or older who were recruited between 1986 and 1988 from 4 geographical areas of the United States: Baltimore County, Maryland; Minneapolis, Minnesota; Portland, Oregon; and the Monongahela Valley, Pennsylvania. Women were excluded if they were unable to

walk without help or had a history of bilateral hip replacement. At SOF visit 6 (1997–1999), a cohort of 662 African-American women with functional characteristics similar to those of the baseline SOF sample was added. For all participants, follow-up examinations were conducted at approximately 2- to 4-year intervals.

Caregiver-SOF subsample

Initial screening. The Caregiver-SOF sample was identified from SOF in 2 stages: initial screening and telephone rescreening, described elsewhere (Figure 1) (21). The initial screening took place at SOF visit 6, during which a Caregiver Screening Questionnaire was administered in person to all SOF participants in order to classify caregiver status. The questionnaire asked participants if they were currently helping a relative or friend with any of 7 activities of daily living (ADL) (walking across a room, grooming, transferring from bed to chair, eating, dressing, bathing, or using the toilet) (22) or 7 instrumental activities of daily living (IADL) (using the telephone, getting to places farther than walking distance, shopping, preparing meals, managing medications, managing finances, or doing heavy housework) (23) because that person was physically, cognitively, or psychologically unable to perform the task independently. Of the 4,036 SOF participants initially screened for Caregiver-SOF, 930 (23%) reported helping 1 or more person with at least 1 ADL or IADL task without pay and were categorized as caregivers; 3,106 did not help anyone with ADL or IADL tasks and were categorized as noncaregivers.

Rescreening and Caregiver-SOF enrollment. Starting in 1999 (upon National Institutes of Health funding for Caregiver-SOF), all caregivers identified at the initial screening were rescreened by telephone to determine whether they were still caregivers; 841 (90%) of the 930 initial caregivers were rescreened. For each caregiver who was rescreened, a subset of noncaregivers identified at the initial screening was matched on SOF site, age, race (white or African-American), and zip code. Participants from each matched set were rescreened until 1–2 who were still noncaregivers agreed to participate in Caregiver-SOF. Of the 3,106 initial noncaregivers, 2,438 (79%) were included in matched sets and 1,890 (78% of those matched) were rescreened. Reasons why participants were not rescreened, such as death or refusal, were documented (Figure 1).

All participants who were classified as current caregivers at rescreening were invited to participate in Caregiver-SOF. This included 348 women who were caregivers at both initial screening and rescreening (41% of initial caregivers who were rescreened) and 142 initial noncaregivers who met the criteria for being caregivers at rescreening (8.6% of initial noncaregivers who were rescreened). Of the initial noncaregivers who were identified as a potential match and rescreened, 1,653 (87.5%) remained noncaregivers, of whom 694 were not invited to participate in Caregiver-SOF because another match had agreed to participate. To avoid potential residual effects of caregiving-related stress, noncaregivers were excluded if they had been caregivers in the past 2 years. Of the 1,449 women invited to participate, 375 caregivers and 694 matched noncaregivers (76.5% and 72.4% of caregivers and noncaregivers invited to participate, respectively) agreed to participate in Caregiver-SOF.

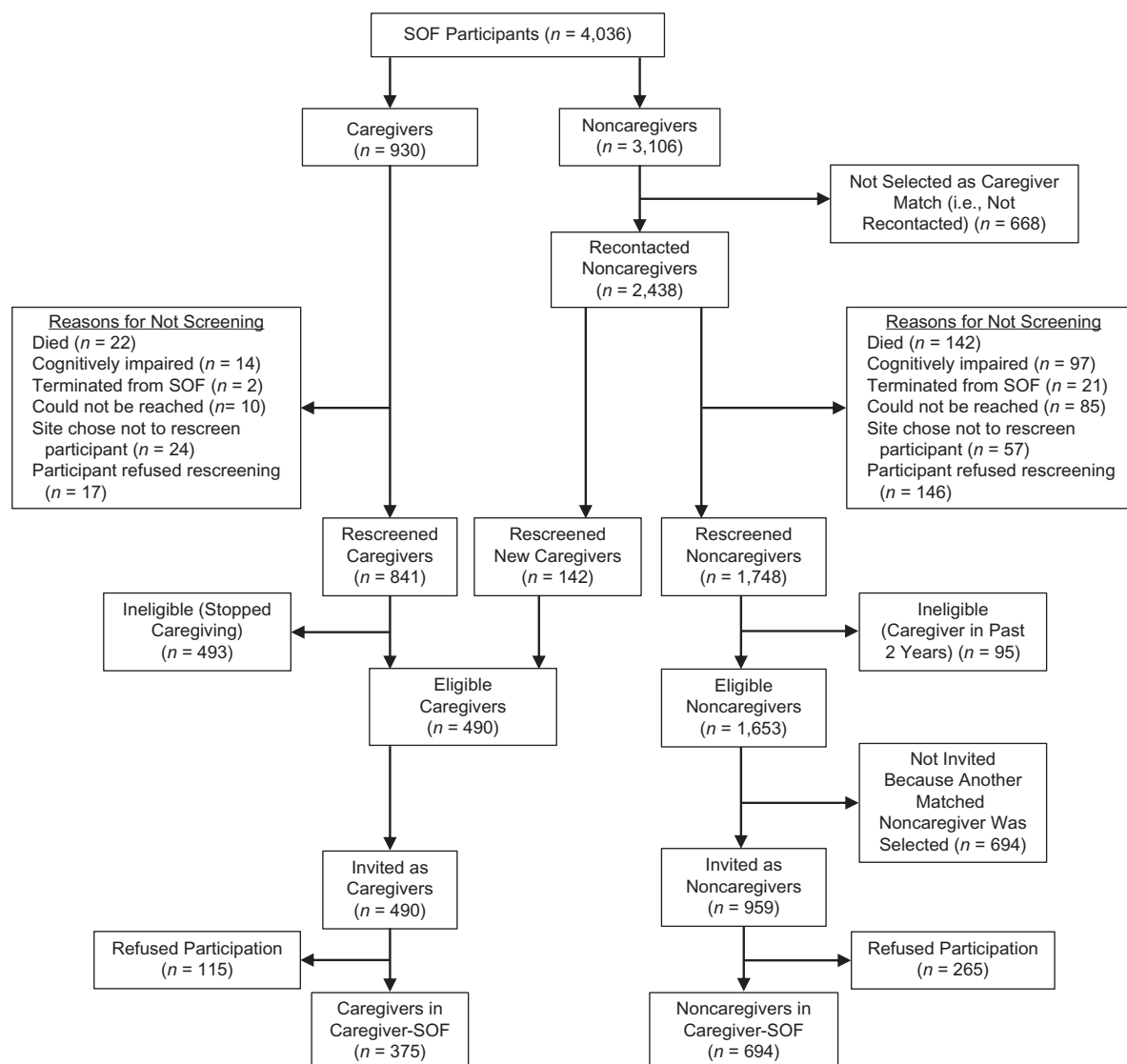


Figure 1. Creation of the Caregiver-Study of Osteoporotic Fractures (Caregiver-SOF) sample from participants in the Study of Osteoporotic Fractures (SOF), 4 US geographical areas (Baltimore, Maryland; Minneapolis, Minnesota; Monongahela Valley, Pennsylvania; and Portland, Oregon), 1997–2009.

Subset samples for analyses

We assessed multiple samples representing various stages in the creation of the Caregiver-SOF sample (Figure 1). These samples were: SOF participants who were initially screened ($n = 4,036$), participants who were selected for the telephone rescreening ($n = 3,368$; the “recontacted subset”), those who completed the rescreening ($n = 2,731$; the “rescreened subset”), those who met the eligibility criteria ($n = 2,143$; the “eligible subset”), those invited to participate in Caregiver-SOF ($n = 1,449$; the “invited subset”), and the Caregiver-SOF sample ($n = 1,069$).

SOF and Caregiver-SOF were approved by the institutional review board at each SOF site. Caregiver-SOF also received institutional review board approval from Boston University

Medical Center (Boston, Massachusetts). All participants provided written informed consent.

Measures

Measures came from SOF, the initial Caregiver Screening Questionnaire, and the telephone rescreening interview.

Caregiving status. At each stage in screening and sample creation, participants were categorized as caregivers if they currently helped one or more persons with at least 1 ADL or IADL task without pay, and as noncaregivers otherwise.

Mortality. All-cause mortality was assessed within 10 years of the initial screening date and (if applicable) the rescreening date. Per the SOF protocol, participants were contacted every

4 months. Death certificates were obtained by investigators at each SOF site for persons who had died, as described elsewhere (24). Follow-up for vital status was 99% complete.

Covariates. Covariates included SOF site, age at initial screening and rescreening, race (white or African-American), and educational level (attended college vs. less education, assessed at SOF visit 1). Marital status was assessed at SOF visit 4 (i.e., 2–4 years before the start of follow-up) for the original cohort and at visit 6 for African-American participants.

Health and functioning characteristics were measured at SOF visit 6. Women reported whether they had difficulty performing each of 5 IADLs (walking 2 or 3 blocks, climbing up 10 steps without stopping/resting, preparing meals, doing heavy housework, shopping for groceries or clothes, and walking down 10 steps) and were classified as having 1 or more IADL limitations versus none. Walking speed (m/second) was based on the time required to walk 6 m at a usual pace, averaged over 2 trials (25). Chair stand speed (seconds) was defined as the amount of time needed to rise from a chair 5 times without using one's arms. The grip strength of the dominant hand (kg) was measured as the average of 2 trials involving pressing a hand-held dynamometer. These performance-based items have high interrater reliability ($r = 0.70$, $r = 0.93$, and $r = 0.84$, respectively) (25). Interviewers also administered a Modified Mini-Mental State Examination, a brief assessment of general cognition with components for orientation, concentration, language, praxis, and immediate and delayed memory. Scores range from 0 to 26, with higher scores representing better cognitive functioning (26, 27). Physical and cognitive performance-based items were categorized as either unable to complete, below the median value, or at or above the median value.

Participants reported whether they had ever been diagnosed with the following conditions: stroke, diabetes, heart disease, Parkinson disease, chronic obstructive pulmonary disease/emphysema, and hypertension. Members of the original cohort reported whether they had ever been diagnosed with these conditions at visit 4 and whether they had been diagnosed since their last interview at visit 6. The African-American cohort reported at visit 6 whether they had ever been diagnosed with these conditions.

Analyses

We compared distributions of sociodemographic and health variables, using χ^2 tests for categorical variables and t tests for continuous variables. We used Cox proportional hazards modeling to calculate adjusted hazard ratios and 95% confidence intervals for death within 10 years among caregivers versus noncaregivers for all samples. For each sample, we analyzed 2 models. The first model adjusted for matching variables used to create the Caregiver-SOF sample, excluding zip code (SOF site, age, and race). The second (i.e., fully adjusted model) additionally controlled for education, physical and cognitive performance, IADL limitations, and the presence of one or more of 3 comorbid conditions that were associated with both caregiving status and mortality: chronic obstructive pulmonary disease/emphysema, Parkinson disease, and history of heart disease.

We calculated e-values using the “EValue” package (28) in R Studio, version 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria), correcting for mortality being a common

outcome (29). E-values indicate the magnitude of the association a hypothetical unmeasured confounder would need to have with *both* caregiver status and mortality in order to make their observed association null.

Alternate sample based on propensity score matching. We created an alternate sample using propensity score matching. The sample was drawn from the eligible subset ($n = 2,143$), augmented with 21 noncaregivers who were rescreened after enrollment ended but met eligibility criteria, and 580 of the 668 initially screened participants who were not recontacted because another noncaregiver from the same matched set agreed to participate. These 580 participants were added because SOF data indicated that they would have been alive, not cognitively impaired, and participating in SOF at the time of rescreening. We assumed that they would not have refused rescreening or changed caregiver status between initial screening and rescreening. The other 88 participants were excluded from these analyses because they died before the date on which they were added to a matched set ($n = 20$), their Modified Mini-Mental State Examination score at visit 6 was less than 20, indicating probable cognitive impairment ($n = 66$), or their last date of follow-up was before the date on which they were added to a matched set ($n = 2$). Adding these noncaregivers to the eligible subset resulted in 2,744 participants (490 caregivers and 2,254 noncaregivers).

For these 2,744 participants, we used logistic regression, including all variables in the fully adjusted model described above, to calculate propensity scores predicting the probability of being a caregiver. We applied this model separately for each SOF site. On the basis of propensity scores and caregiver status, all caregivers were then matched with 2 noncaregivers. Following previous studies, we ran a greedy propensity score matching algorithm (2, 30) 2 consecutive times using a macro (31) written for use in SAS (version 2.4.15; SAS Institute, Inc., Cary, North Carolina). We repeated the use of Cox proportional hazards models in our analyses of the propensity-score-matched sample.

Relative hazard ratios. We calculated relative hazard ratios and 95% confidence intervals (16, 20) to quantify the overall estimated amount of selection bias, the amount due to sampling decisions, and the amount due to participation factors. To assess selection bias overall and that due to sampling decisions, we compared the adjusted hazard ratios for the Caregiver-SOF sample and the invited subset, respectively, with the adjusted hazard ratio for the initially screened sample. To assess participation bias, we compared the adjusted hazard ratio for the Caregiver-SOF sample with that for the invited subset.

All analyses other than e-value calculation were conducted using SAS, version 9.4.

RESULTS

Among the 4,036 initially screened SOF participants, caregivers were slightly younger than noncaregivers (mean ages at initial screening were 79.0 years and 79.4 years, respectively) (Table 1). Among the 1,069 Caregiver-SOF participants, caregivers were slightly younger than noncaregivers at rescreening (mean ages were 81.3 years and 81.8 years, respectively). Approximately 90% of each sample was white. Caregivers comprised 23.0% of the initially screened sample, 33.8% of the invited subset, and 35.1% of the Caregiver-SOF sample.

Table 1. Characteristics (%) of Caregivers and Noncaregivers With Various Characteristics at Key Points in the Creation of the Caregiver-SOF Participant Sample, 4 US Geographical Areas (Baltimore, Maryland; Minneapolis, Minnesota; Monongahela Valley, Pennsylvania; and Portland, Oregon), 1997–2009

Characteristic	Initially Screened (n = 4,036) ^a		Invited Subset (n = 1,449) ^b		Caregiver-SOF Group (n = 1,069) ^b	
	Caregiver (n = 930)	Noncaregiver (n = 3,106)	Caregiver (n = 490)	Noncaregiver (n = 959)	Caregiver (n = 375)	Noncaregiver (n = 694)
Age, years ^c	79.0 (4.0)	79.4 (4.0)	81.3 (3.7)	81.9 (3.7)	81.3 (3.6)	81.8 (3.7)
Race						
White	88.2	90.8	87.1	88.6	88.0	88.5
African-American	11.8	9.2	12.9	11.4	12.0	11.5
>12 years of education	56.2	61.5	57.5	62.1	56.5	60.8
Married ^d	47.5	37.9	56.9	38.3	57.1	39.7
Needing help with ≥1 ADL or IADL	38.6	46.0	36.1	43.4	34.9	43.7
Functioning at or better than median ^e						
Chair stand time	45.6	47.7	44.1	43.9	43.1	43.8
Usual walking speed	56.2	48.4	61.2	54.0	62.4	54.3
Grip strength	55.3	46.9	57.1	50.0	57.6	49.0
Mental status	71.9	64.8	73.7	72.2	76.8	74.2
Usual walking speed ≥1 m/second ^f	35.1	30.6	38.8	34.1	40.8	34.9
Parkinson disease	1.0	1.1	0.8	0.6	0.5	0.4
Emphysema/chronic obstructive pulmonary disease	11.0	13.4	10.2	13.8	10.1	13.8
Heart disease	8.5	9.5	9.4	9.7	9.3	10.1

Abbreviations: ADL, activities of daily living; IADL, instrumental activities of daily living; SOF, Study of Osteoporotic Fractures.

^a In the initially screened participants, caregiver status and age are as measured at initial screening.

^b In the invited subset and the Caregiver-SOF sample, caregiver status and age are as measured at rescreening.

^c Values for age are expressed as mean (standard deviation).

^d Data on current marital status were missing for 13% of initially screened women, 11% of women in the invited subset, and 10% of women in the Caregiver-SOF sample. The statistics reported are for those participants for whom this information was available.

^e Participants who were unable to complete functional performance measures were considered to have below-median functioning.

^f Participants who were unable to complete the walking assessment were considered to have a walking speed of <1 m/second.

Across all samples, caregivers were slightly younger and less likely to be white and college-educated than noncaregivers. Caregivers had fewer IADL limitations and better functioning than noncaregivers on all performance-based measures and the Modified Mini-Mental State Examination.

Reasons why caregivers and noncaregivers identified at initial screening were not rescreened included death (2.4% and 5.8%, respectively), cognitive impairment (1.5% and 4.0%, respectively), and refusal to be rescreened (1.8% and 6.0%, respectively) (Figure 1). Among those rescreened, 53.0% of caregivers and 3.5% of noncaregivers were ineligible. Among eligible noncaregivers, 42.0% were not invited because of their matched caregiver not participating. Overall, 26.2% of all invited individuals declined to participate in Caregiver-SOF.

Caregiver status and mortality

In the initially screened sample, caregivers had a significantly lower hazard of mortality than noncaregivers, adjusting for matching variables and other covariables (adjusted hazard ratio (aHR) = 0.83, 95% confidence interval (CI): 0.73, 0.95) (Table 2). Adjusted hazard ratios were similar in the recontacted subset and the rescreened subset. However, this association was slightly more protective in the eligible subset, the invited subset, and the Caregiver-SOF subset (e.g., aHR = 0.71 (95% CI: 0.57, 0.89) in the Caregiver-SOF sample). The proportional hazards assumption was not violated in any models, based on a nonsignificant interaction of caregiver status with time in the adjusted models.

The propensity-score-matched sample contained 1,406 participants (34.4% caregivers). Confounders were balanced in

Table 2. Relationship Between Caregiving Status and Mortality Among SOF Participants at Different Stages in the Construction of the Caregiver-SOF Sample, 4 US Geographical Areas (Baltimore, Maryland; Minneapolis, Minnesota; Monongahela Valley, Pennsylvania; and Portland, Oregon), 1997–2009

Caregiver-SOF Sample Construction	No. in Sample ^a	% of Deaths ^b	Model Adjusting for Matching Variables ^c		Fully Adjusted Model ^d	
			aHR	95% CI	aHR	95% CI
Initially screened	4,036	36.6	0.75	0.65, 0.85	0.83	0.73, 0.95
Recontacted subset	3,368	35.9	0.75	0.66, 0.86	0.84	0.73, 0.96
Rescreened subset	2,731	42.0	0.80	0.70, 0.91	0.84	0.74, 0.96
Eligible subset	2,143	42.9	0.69	0.58, 0.81	0.73	0.61, 0.87
Invited subset	1,449	40.5	0.70	0.58, 0.84	0.73	0.61, 0.88
Caregiver-SOF sample	1,069	38.6	0.67	0.54, 0.83	0.71	0.57, 0.89
Propensity-score-matched sample ^e	1,406	36.6			0.77	0.64, 0.93

Abbreviations: aHR, adjusted hazard ratio; Caregiver-SOF, Caregiver-Study of Osteoporotic Fractures; CI, confidence interval; SOF, Study of Osteoporotic Fractures.

^a Between 0.6% and 1.2% of the samples were missing data on 1 or more covariates and were thus excluded from the analysis.

^b For the initially screened, the recontacted subset, and the rescreened subset, modeling of caregiving status at initial screening and death within 10 years of initial screening. For the eligible subset, the invited subset, the Caregiver-SOF sample, and the propensity-score-matched sample, modeling of caregiving status at rescreening and death within 10 years of rescreening.

^c Matching variables included age, race, and zip code.

^d Results were adjusted for age at the start of follow-up, race, educational level, SOF site, impairments in instrumental activities of daily living, physical and cognitive performance measures, and the presence of 1 or more of the following: chronic obstructive pulmonary disease/emphysema, Parkinson disease, and history of heart disease.

^e No additional covariates were adjusted for in the fully adjusted model.

this sample; thus, caregivers and noncaregivers did not differ meaningfully with regard to any covariates (see Web Table 1, available at <https://academic.oup.com/aje>). Caregivers had a lower hazard of mortality than noncaregivers (aHR = 0.77, 95% CI: 0.64, 0.93) (Table 2), falling between that of the initially screened sample and the Caregiver-SOF sample.

Based on the e-value results, an unmeasured confounder would need to be associated with both caregiving and mortality by a risk ratio of 1.51–1.85 (depending on the subsample) to increase the adjusted protective association to the null value of 1.00 (Web Table 2).

Quantification of potential selection bias

Relative hazard ratios comparing caregiver-mortality associations in the Caregiver-SOF sample with the initially screened sample indicated little evidence of overall selection bias (relative hazard ratio (RHR) = 0.86, 95% CI: 0.72, 1.02) (Table 3). The association in the invited subset was only slightly smaller than that in the initially screened sample (RHR = 0.88, 95% CI: 0.77, 1.00), indicating little evidence of selection bias due to sampling decisions. It was similar to that in the Caregiver-SOF sample (RHR = 0.97, 95% CI: 0.86, 1.10), indicating no participation bias.

DISCUSSION

Among older women participating in Caregiver-SOF, we found little evidence that selection bias accounted for lower rates of mortality in caregivers compared with noncaregivers.

The hazard ratio for invited participants was only slightly different from that in initially screened SOF participants, and the relative hazard ratio was not statistically significant, indicating little evidence of selection bias due to sampling decisions. Similar hazard ratios in analyses among women invited to participate in Caregiver-SOF and the final Caregiver-SOF sample indicated no bias due to participation factors. Thus, our hypothesis of no selection bias introduced at various critical points of study design and sample creation was largely supported.

One sampling decision that could have introduced bias was the exclusion of women who stopped caregiving between the initial screening and rescreening. In making this decision, we aimed to minimize residual health effects of caregiving-related stressors and to ensure correct classification of caregiver status in Caregiver-SOF. However, this resulted in our sample being predominantly a sample of long-term caregivers and it may have excluded noncaregivers who were healthy enough to perform caregiving tasks, thereby producing a slightly more protective hazard ratio. Additionally, matching noncaregivers to caregivers on age may have introduced bias. Because caregivers were generally younger than noncaregivers, this might have excluded the oldest noncaregivers, who had the highest risk of mortality.

Our results corroborate findings from other cohort studies that assessed possible participation bias, such as studies of cardiovascular risk factors (14) and the health effects of labor market activity (17). Our results extend these findings to assessment of selection bias due to sampling decisions. Previous studies (10, 32) documented reasons for noninclusion but did not evaluate whether these reasons biased study results.

Table 3. Relative Hazard Ratios Comparing Associations Between Caregiving Status and Mortality in 3 Samples of Participants From the Study of Osteoporotic Fractures, 4 US Geographical Areas (Baltimore, Maryland; Minneapolis, Minnesota; Monongahela Valley, Pennsylvania; and Portland, Oregon), 1997–2009

Type of Bias Assessed and Comparison	Model Adjusting for Matching Variables ^a		Fully Adjusted Model ^b	
	RHR	95% CI	RHR	95% CI
Overall selection bias				
Caregiver-SOF vs. initially screened	0.89	0.75, 1.07	0.86	0.72, 1.02
Selection bias due to sampling decisions				
Invited subset vs. initially screened	0.93	0.82, 1.06	0.88	0.77, 1.00
Participation bias				
Caregiver-SOF vs. invited subset	0.96	0.85, 1.08	0.97	0.86, 1.10

Abbreviations: Caregiver-SOF, Caregiver-Study of Osteoporotic Fractures; CI, confidence interval; RHR, relative hazard ratio; SOF, Study of Osteoporotic Fractures.

^a Matching variables included SOF site, age, race, and zip code.

^b Results were adjusted for age, race, educational level, SOF site, impairments in instrumental activities of daily living, and physical and cognitive performance measures.

The detailed tracking of SOF participants in the initially screened sample allowed us to differentiate sampling decisions from participation factors that might have led to selection bias and to create subsets for evaluating potential bias at each stage in this process. Such documentation could be useful in future cohort studies in which selection bias may be a concern.

The propensity-score-matched sample allowed us to evaluate whether matching procedures introduced bias. Propensity score matching is used to minimize bias and confounding in observational studies (2, 33). Similar results in the Caregiver-SOF sample and the propensity-score-matched sample provide support that our matching approach did not create selection bias.

These results are consistent with those of studies that found lower risk of mortality in older caregivers than in non-caregivers (34), and they support the “healthy caregiver” hypothesis (4). This hypothesis proposes that caregiving imparts health benefits, similar to the salutary health effects of volunteering (35, 36). Our results suggest that factors inherent to caregiving, not to selection of study participants, probably account for caregivers’ survival advantage.

Our conclusions must be considered in light of several limitations. First, selection factors may have influenced our source population. SOF participants who attended visit 6 might have been healthier and younger than those who were unable to engage in this visit. Comparing SOF visit 6 participants with the broader source population (all older community-dwelling women) would provide valuable insight, but we lacked data for this comparison. Second, our methods for quantifying selection bias did not account for potential residual confounding in the comparison populations (i.e., the denominator of each relative hazard ratio). Adjusting for age and physical and cognitive performance probably adjusted for the strongest confounders in each sample. However, we could not adjust for depression or marital status because these variables were not assessed at visit

6 for all participants. We determined that a hypothetical unobserved confounder would need to be moderately associated with both caregiving and mortality (changing each by 57%–77%, depending on the subsample) in order to nullify the protective associations observed. Moreover, we conducted a sensitivity analysis controlling for our marital status variable (which had a high degree of missingness and likely misclassification); the hazard ratios for caregiver status did not change in any sample. Third, despite having mortality and demographic information on nonparticipants, we lacked some information on the 668 noncaregivers who were not recontacted; for propensity score analyses, we used available SOF data to predict who would have been rescreened, but we lacked data with which to verify these predictions. Fourth, the SOF sample was comprised of primarily white and healthy older women; our results may have limited generalizability to other sociodemographic groups.

Nonetheless, our study had some methodological strengths. Neither caregiving status at the start of Caregiver-SOF nor mortality was likely to be misclassified. Participants were categorized as caregivers or noncaregivers on the basis of the same set of caregiving activities, rather than asking participants to self-identify as caregivers. In addition, misclassification due to change in caregiver status was minimized by assessing caregiving status at each point in creating the sample. Death certificates were used to assess mortality, making outcome misclassification unlikely. Additionally, we compared 2 methods of sample creation and quantified potential bias via relative hazard ratios.

In conclusion, our results underscore the value of ancillary studies for analyses of hard-to-measure biases like selection bias. By reducing concerns about possible selection bias, especially bias due to participation factors, our results add to a body of evidence that involvement in informal caregiving may impart health benefits for older adults. While the possibility of residual confounding in each estimate remains, validity is not likely to be affected by selection decisions.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, School of Public Health, Boston University, Boston, Massachusetts (Meghan L. Smith, Lynsie R. Ranker, Lisa Fredman); and Department of Biostatistics, School of Public Health, Boston University, Boston, Massachusetts (Timothy C. Heeren).

The Study of Osteoporotic Fractures (SOF) and Caregiver-SOF are supported by National Institutes of Health funding. This work was supported by National Institutes of Health grants R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, R01 AG027576, R01 AG18037, and R21 AG050428.

Conflict of interest: none declared.

REFERENCES

- Pearlin LI, Mullan JT, Semple SJ, et al. Caregiving and the stress process: an overview of concepts and their measures. *Gerontologist*. 1990;30(5):583–594.
- Roth DL, Haley WE, Hovater M, et al. Family caregiving and all-cause mortality: findings from a population-based propensity-matched analysis. *Am J Epidemiol*. 2013;178(10):1571–1578.
- Ramsay S, Grundy E, O'Reilly D. The relationship between informal caregiving and mortality: an analysis using the ONS Longitudinal Study of England and Wales. *J Epidemiol Community Health*. 2013;67(8):655–660.
- Fredman L, Lyons JG, Cauley JA, et al. The relationship between caregiving and mortality after accounting for time-varying caregiver status and addressing the healthy caregiver hypothesis. *J Gerontol A Biol Sci Med Sci*. 2015;70(9):1163–1168.
- Cummings SR, Black DM, Nevitt MC, et al. Appendicular bone density and age predict hip fracture in women. The Study of Osteoporotic Fractures Research Group. *JAMA*. 1990;263(5):665–668.
- Fredman L, Bertrand RM, Martire LM, et al. Leisure-time exercise and overall physical activity in older women caregivers and non-caregivers from the Caregiver-SOF Study. *Prev Med*. 2006;43(3):226–229.
- Bertrand RM, Saczynski JS, Mezzacappa C, et al. Caregiving and cognitive function in older women: evidence for the healthy caregiver hypothesis. *J Aging Health*. 2012;24(1):48–66.
- Gruenewald TL, Karlamangla AS, Greendale GA, et al. Feelings of usefulness to others, disability, and mortality in older adults: the MacArthur Study of Successful Aging. *J Gerontol B Psychol Sci Soc Sci*. 2007;62(1):P28–P37.
- Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology*. 2012;23(1):119–128.
- Knoll L, Felten MK, Ackermann D, et al. Non-response bias in a surveillance program for asbestos-related lung cancer. *J Occup Health*. 2011;53(1):16–22.
- Pendlebury ST, Chen PJ, Bull L, et al. Methodological factors in determining rates of dementia in transient ischemic attack and stroke: (I) impact of baseline selection bias. *Stroke*. 2015;46(3):641–646.
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615–625.
- Ness KK, Leisenring W, Goodman P, et al. Assessment of selection bias in clinic-based populations of childhood cancer survivors: a report from the Childhood Cancer Survivor Study. *Pediatr Blood Cancer*. 2009;52(3):379–386.
- Boshuizen HC, Viet AL, Picavet HS, et al. Non-response in a survey of cardiovascular risk factors in the Dutch population: determinants and resulting biases. *Public Health*. 2006;120(4):297–308.
- Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol*. 2009;23(6):597–608.
- Nohr EA, Frydenberg M, Henriksen TB, et al. Does low participation in cohort studies induce bias? *Epidemiology*. 2006;17(4):413–418.
- Carter KN, Imlach-Gunasakera F, McKenzie SK, et al. Differential loss of participants does not necessarily cause selection bias. *Aust N Z J Public Health*. 2012;36(3):218–222.
- Hatch EE, Hahn KA, Wise LA, et al. Evaluation of selection bias in an Internet-based study of pregnancy planners. *Epidemiology*. 2016;27(1):98–104.
- Jacobsen TN, Nohr EA, Frydenberg M. Selection by socioeconomic factors into the Danish National Birth Cohort. *Eur J Epidemiol*. 2010;25(5):349–355.
- Pizzi C, De Stavola BL, Pearce N, et al. Selection bias and patterns of confounding in cohort studies: the case of the NINFEA Web-based birth cohort. *J Epidemiol Community Health*. 2012;66(11):976–981.
- Fredman L, Tennstedt S, Smyth KA, et al. Pragmatic and internal validity issues in sampling in caregiver studies: a comparison of population-based, registry-based, and ancillary studies. *J Aging Health*. 2004;16(2):175–203.
- Katz S, Ford AB, Moskowitz RW, et al. Studies of Illness in the Aged. The Index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;185(12):914–919.
- George LK, Fillenbaum GC, Maddox GL. *Multidimensional Functional Assessment: the OARS Methodology*. Durham, NC: Duke University Center for the Study of Aging and Human Development; 1978.
- Ensrud KE, Ewing SK, Taylor BC, et al. Frailty and risk of falls, fracture, and mortality in older women: the Study of Osteoporotic Fractures. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):744–751.
- Nelson HD. Smoking, alcohol, and neuromuscular and physical function of older women. Study of Osteoporotic Fractures Research Group. *JAMA*. 1994;272(23):1825–1831.
- Teng EL, Chui HC. The Modified Mini-Mental State (3MS) Examination. *J Clin Psychiatry*. 1987;48(8):314–318.
- Yaffe K, Peltz CB, Ewing SK, et al. Long-term cognitive trajectories and mortality in older women. *J Gerontol A Biol Sci Med Sci*. 2016;71(8):1074–1080.
- VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167(4):268–274.
- VanderWeele TJ. On a square-root transformation of the odds ratio for a common outcome. *Epidemiology*. 2017;28(6):e58–e60.
- Parsons LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques. In: *Proceedings of the 26th Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc.; 2001:214–226.
- Rassen J, Doherty M, Huang W, et al. *Using the Pharmacoepidemiology Toolbox in SAS*. Boston, MA: Brigham and Women's Hospital and Harvard Medical School; 2013. http://www.drugapi.org/wp-content/uploads/2013/10/Using_the_Pharmacoepi_Toolbox_in_SAS_2.4.15.pdf. Accessed November 14, 2017.
- Clarisse B, Nikasinovic L, Poinard R, et al. The Paris prospective birth cohort study: which design and who participates? *Eur J Epidemiol*. 2007;22(3):203–210.

33. Ye Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of Alcoholics Anonymous in observational studies. *Drug Alcohol Depend.* 2010;104(1-2):56–64.
34. Roth DL, Fredman L, Haley WE. Informal caregiving and its impact on health: a reappraisal from population-based studies. *Gerontologist.* 2015;55(2):309–319.
35. Fried LP, Carlson MC, Freedman M, et al. A social model for health promotion for an aging population: initial evidence on the Experience Corps model. *J Urban Health.* 2004;81(1):64–78.
36. Varma VR, Tan EJ, Gross AL, et al. Effect of community volunteering on physical activity: a randomized controlled trial. *Am J Prev Med.* 2016;50(1):106–110.