Volunteering by Older Adults and Risk of Mortality: A Meta-Analysis

Morris A. Okun and Ellen WanHeung Yeung

Arizona State University

Stephanie Brown

University of Michigan and Stony Brook University Medical Center

Author Note

Morris A. Okun, Department of Psychology, Arizona State University; Ellen WanHeung Yeung, Department of Psychology, Arizona State University; Stephanie Brown, Center for Medical Humanities, Compassionate Care, and Bioethics, Stony Brook University Medical Center.

Correspondence concerning this article should be addressed to Morris A. Okun, Department of Psychology, Arizona State University, Tempe, AZ 85287-1104.

E-mail: okun@asu.edu

Abstract

Organizational volunteering has been touted as an effective strategy for older adults to help themselves while helping others. Extending previous reviews, we carried out a meta-analysis of the relation between organizational volunteering by late middle-aged and older adults (minimum age = 55 years old) and risk of mortality. We focused on unadjusted effect sizes (i.e., bivariate relations), adjusted effect sizes (i.e., controlling for other variables such as health), and interaction effect sizes (e.g., the joint effect of volunteering and religiosity). For unadjusted effect sizes, on average, volunteering reduced mortality risk by 47 percent with a 95% confidence interval ranging from 38% to 55%. For adjusted effect sizes, on average, volunteering reduced mortality risk by 24 percent with a 95% confidence interval ranging from 16% to 31%. For interaction effect sizes, we found preliminary support that as public religiosity increases, the inverse relation between volunteering and mortality risk becomes stronger. The discussion identified several unresolved issues and directions for future research.

Key words: volunteering, mortality, moderation

Word count: 11,472

Volunteering by Older Adults and Risk of Mortality: A Meta-Analysis

Using insights generated from evolutionary theory, Brown and her colleagues (Brown & Brown, 2006; Brown, Nesse, Vinokur, & Smith, 2003) advanced the hypothesis that helping behavior, whatever its effect on the recipient, promotes the psychological well-being and health of the *helper*. Providing assistance to another improves relationship satisfaction and enhances stress regulation (Post, 2007). Brown and her colleagues have shown that helping behavior among older adults is associated with accelerated recovery from depressive symptoms that accompany spousal loss (Brown, Brown, House, & Smith, 2008) and reduced mortality risk (Brown et al., 2003) even among caregivers (Brown, Smith, et al., 2009). An independent research team, using a sample of over a thousand older adults from New York City, reported similar findings for *morbidity* (Brown, Consedine, & Magai, 2005).

Prosocial behaviors refer to intentional efforts to provide assistance to another individual or communities. Planned prosocial activities include caregiving, providing support to others, contributing to other church-goers, and volunteering. Organizational or formal volunteering is an unpaid, voluntary activity that involves “. . . taking actions within an institutional framework that potentially provides some service to one or more other people or to the community at large” (Piliavin & Siegl, 2007, p. 454). In the current meta-analysis, we examine the relation between organizational volunteering and risk of mortality among adults 55 years old and older.

We chose to focus exclusively on organizational volunteering because, in contrast to helping familiar others and engaging in social activities with familiar others in informal social contexts, organizational volunteering entails helping unfamiliar others in an institutional context. Theoretically, we expect that social activities recruit different neural circuitry than helping others and, although, helping familiar and unfamiliar others should recruit similar neural circuitry under some conditions (Brown, Brown, & Preston, 2012), we reasoned that tests of this possibility should await an initial inquiry into whether there is indeed a reliable association between volunteering and reduced mortality risk.

We made the decision to exclude younger adults for both pragmatic and theoretical reasons. Pragmatically, the variability in mortality is much lower in younger than older adults, the causes of death differ for younger and older adults (Mathers, Boerma, & Fat, 2009), and the effects of volunteering on mortality via processes such as stress regulation are not likely to be observed until later life (Belloc & Breslow, 1972). From a theoretical perspective, aging is associated with life transitions that often involve role losses. Consequently, the role of volunteer may be especially important to the emotional and physical health of older adults (Van Willigen, 2000).

Because volunteering was a measured, rather than a manipulated, variable in the sources included in the current meta-analysis, it was important to take into account variables such as health, social interaction, and social connection that are positive selection factors for volunteering (Thoits & Hewitt, 2001). In examining, the adjusted relation between volunteering and mortality risk, the types of variables used as covariates included age, sex, physical health, socioeconomic status, health behaviors, marital status, religiosity/religious behavior, emotional health, social connection, social interaction, ethnicity, work status, cognitive functioning, and leisure activity.

Previous Reviews

We identified five reviews of studies examining the relation between organizational volunteering and mortality risk. In these reviews, the number of studies of the relation between volunteering and mortality risk ranged from five to eight. Brown and Okun (in press) focused on the bivariate relation between volunteering and mortality risk and reported that volunteering reduced mortality risk. The other reviewers examined adjusted relations in which sets of covariates were included in models examining the association between volunteering and mortality risk. These reviewers concluded that even when other factors are statistically controlled, individuals who volunteer are more likely to live longer (Grimm, Spring, and Dietz 2007; Harris & Thoresen, 2005; Oman, 2007; von Bonsdorff & Rantanen, 2011).

Although previous reviewers have drawn similar conclusions regarding the inverse relation between volunteering and mortality risk, several limitations in the methodologies of the individual studies suggest the need for meta-analytic techniques, which enhances our confidence in conclusions drawn from the results of several studies. Using these techniques we are able to provide more precise information on the distribution of effect sizes, including central tendency, variability, and confidence intervals. Based upon theoretical analyses (Brown & Brown, 2006) and the conclusions of previous reviewers (Grimm et al., 2007; Oman, 2007), we predicted that there would be (a) a significant (p < 05) inverse unadjusted (bivariate) association between volunteering and mortality risk and (b) a significant inverse adjusted association between volunteering and mortality risk. Inspection of the numerical estimates of the strength of the association between volunteering and mortality risk (Brown & Okun, in press) reveals that they vary substantially. Therefore, we predicted that there would be a significant ($p < .05$) amount of heterogeneity in the unadjusted and adjusted effect sizes.

Harris and Thoresen (2005) concluded that while the presence of covariates did not fully eliminate the association, it did substantially reduce the relation between volunteering and mortality risk. To examine whether part of the association is due to third party variables such as health and social interaction, we tested the hypothesis that there would be a significant (p < .05) reduction in the magnitude of the association between volunteering and mortality risk when adjusted effect sizes are compared with unadjusted effect sizes.

An unresolved issue in this literature pertains to the form of the relation between volunteering and morality risk. Researchers have reported that the relation between volunteering and mortality is (a) linear (Oman, Thoresen, & McMahon, 1999), (b) nonlinear, exhibiting a threshold effect (Luoh & Herzog, 2002), and (c) nonlinear, exhibiting a U-shaped effect (Musick, Herzog, & House, 1999). The threshold effect is based on the notion that a certain minimum amount of volunteering is required for older adults to obtain the health-related benefits. The curvilinear effect adds to the threshold effect the notion that volunteering beyond certain levels creates role strain which, in turn, nullifies the health-related benefits of volunteering. Therefore, we examined whether the data provide more support for depicting the relation between volunteering and mortality risk as linear or as nonlinear.

Another unresolved issue in this literature involves whether there are individual differences in the benefits that older adults derive from volunteering. Oman (2007) proposed two alternative hypotheses regarding how individual difference variables influence the association between volunteering and risk of mortality. According to the *compensatory* hypothesis, as the individual's resources (human, social, and cultural capital) decrease, the benefit of volunteering on mortality risk reduction increases. In contrast, according to the *complementary* hypothesis, as the individual's resources increase, the benefit of volunteering on mortality risk reduction increases. The assumption underlying the compensatory hypothesis is volunteering provides older adults with increased capital and a role that can offset the loss of other roles. In this case, volunteering should provide the greatest benefits to those with the fewest resources. According to the complementary hypothesis, volunteering by older adults taxes their limited reservoir of coping resources. Thus, the benefits of volunteering should be greatest for individuals who already have adequate amounts of funds or capital. To test between these alternative hypotheses, we carried out two types of analyses. We investigated whether the relation between volunteering and mortality risk differed in subsamples (e.g., older adults with weak and strong social ties). We also computed volunteering by moderator variable (e.g., volunteering by individual difference) interaction effect sizes.

The ability to statistically investigate whether the association between volunteering and mortality risk differs as a function of personal, social, situational and cultural influences is an advantage of using meta-analysis. Such analyses can shed light on why studies yield diverse effect sizes and suggest methodological and substantive boundary conditions on the relation between volunteering and mortality risk. Contingent upon finding that the effect sizes were heterogeneous and that a substantial proportion of the observed variation was not spurious, we sought to identify study-level moderator variables that might explain this variation including study focus, publication impact factor, country where volunteering took place, historical time of the study, age composition of the sample, and proportion of sample deceased.

Because of biases against publication of null effects, we expected that effect sizes would be stronger in studies explicitly focused on volunteering relative to studies with another focus. Using a similar rationale, we anticipated that effect sizes would be stronger in articles published in more as opposed to less prestigious journals. Because of cohort and/or period effects, the relation between volunteering and mortality risk may have shifted over historical time. Consequently, we used year of publication as a moderator variable. Due to differences in cultural norms regarding helping others via organizational volunteering, the relation between volunteering and mortality risk may vary between countries. In the current meta-analysis, we compared effect sizes derived from US samples with effect sizes derived from Israeli samples. Because role loss increases with age, we examined whether effect sizes are stronger in studies with older rather than younger minimum age requirements. As the death and volunteering rates deviate from .50, these variables have less variability and this may lead to smaller effect sizes. To determine whether variability in the death and volunteering rates were associated with effect size magnitude, we use percent deceased and percent volunteering as moderator variables.

Narrative and meta-analyses alike are potentially biased by the tendency for studies yielding statistically significant effects to be published whereas studies yielding non-statistically significant effects end up in the file draw. In the current meta-analysis, we examined the potential impact of publication bias using the trim and fill procedure (Duval & Tweedie, 2000).

Only 32 percent of the potential interaction effect sizes were retrieved from the studies. We posited that interaction effect sizes would be more likely to be missing when the tests of the interaction effects were not statistically significant. To test this notion, we investigated whether effect sizes were less likely to be reported when the *p* values associated with the test of the volunteer by moderator variable interaction effects were greater than .05 as opposed to less than .05.

Method

Literature Search Procedures

The processes of searching, selecting, and coding sources were carried out by the first author, a Ph.D., and the second author, a senior graduate student with extensive training in quantitative methods.

Inclusion criteria. To be included in this meta-analysis, (a) the source had to be published as a journal article or book chapter written in English; (b) the source had to report on empirical research; (c) the study had to include a measure of organizational volunteering and mortality had to be an outcome variable; (d) the design had to be prospective, that is volunteering had to be assessed prior to a mortality surveillance period; and (e) the unit of analysis had to be the individual.

Search strategies. We used multiple strategies to compile our list of studies. We searched the Medline and PsychINFO data bases on November 3rd 2011. The search strategy involved pairing volunteer, volunteerism, and volunteering in the document title with the keywords of mortality, death, longevity, or survival. For the Medline data base, the command line used for the search was:

((volunteer*.ti. and mortality.af. and english.lg.) or (volunteer*.ti. and death.af. and english.lg.) or (volunteer*.ti. and longevity.af. and english.lg.) or (volunteer*.ti. and survival.af. and english.lg.)). This search yielded 253 journal articles. The specific syntax used in the command line for the search of the PsychINFO data base was: (TI(volunteer*)) AND ((cabs(mortality) or cabs(death) or cabs(longevity) or cabs(survival))). This search yielded a total of 38 non-redundant sources. Next, we searched the reference lists of previous reviews of the relation between volunteering and mortality (Brown & Okun, in press; Harris & Thoresen, 2005; Grimm et al., 2007; Oman, 2007). This strategy netted us additional two sources.

Based upon a preliminary screening of the abstracts of the (291) sources, 13 sources were retrieved for coding. The low yield rate was due to the inclusion of the search term “volunteers”. This search term captured many studies in which the word “volunteers’ was used to describe how the sample was drawn and these studies had nothing to do with studying the effects of organizational volunteering. Eleven of these articles were included in the meta-analysis. One article was excluded because the unit of analysis was the neighborhood rather than the individual (Blakely, Atkinson, Ivory, Collings, Wilton, & Howden-Chapman, 2006) and a second article was excluded because the focus was on the survival times of terminally ill patients who did and did not receive support from volunteers (Herbst-Damm & Kulik, 2005). In an effort to obtain additional sources, we examined the reference lists of the article accepted for inclusion in the meta-analysis. This endeavor yielded another three sources raising the total number of sources included in the meta-analysis to 14.

Coding

Coding was guided by a codebook devised by the first author. A planning sheet was used to facilitate the coding process. The planning sheet enabled the coder to organize the information in the source with respect to whether effect sizes were extracted from the total sample and from subsamples, the measure(s) of volunteering, and the types of effect sizes extracted. Nine forms were developed for coding information regarding the source, the total sample, subsamples, volunteering measures, mortality measure, unadjusted effect sizes, adjusted effect sizes, interaction effect sizes, and interaction tests of statistical significance. The source form was filled out in its entirety except for the two sources that were rejected. For all accepted sources, the total sample form was completed. The remaining forms were used as many times as needed.

The first and second authors independently coded five of the accepted sources. The mean number of disagreements per 100 items coded was 4.50 (SD = 2.99). A total of 37 disagreements occurred including 8 disagreements over whether items needed to be coded (omission disagreements) and 29 regarding the values of items (commission disagreements). The first set of omission disagreements pertained to whether five statistical tests of interaction effects from one source should be coded. This set of disagreements arose because the authors of this source stated in their overview that they tested eight interaction effects but in the Results section they reported statistical tests for only three of the eight interaction effects (i.e., those that were statistically significant). To avoid this source of discrepancies, a sentence was added to the Codebook stating that coders should compare the overview of the statistical analyses with the entire set of analyses reported in the Results section. The second set of omission disagreements pertained to whether three interaction effect sizes from one source should be coded. This set of disagreements arose because the statistical tests of the interaction effects were carried out on subsamples and the Codebook did not provide guidance on what to do in this situation. To avoid this source of discrepancies, a sentence was added to the Codebook stating that two-way interaction effects should be coded only when the analyses were carried out on the total sample.

The item generating the most commission disagreements was “Number of types of covariates.” Discrepancies in coding this item arose for two reasons. First, one coder classified the covariates based on the labels provided by the authors whereas the other coder classified the covariates based upon the items used to assess them. Second, the definitions of the leisure, social connection, and social interaction covariates were blurred. To eliminate disagreements in coding number of types of covariates, we modified our definitions of the leisure, social connection, and social interaction covariates and coders were instructed to examine the survey items provided in the Method section as opposed to using the label provided by the authors.

The first author coded the remaining nine sources.

Effect Sizes

Effect sizes consisted of estimates of the relation between volunteering as measured via an item or items on a survey and mortality assessed after a period of time elapsed, referred to as the mortality surveillance period. We extracted hazard ratios (HR), odds ratios (OR), and relative risks (RR). We focused on three types of effect sizes. *Unadjusted* effect sizes assessed the magnitude of the relation between volunteering and mortality risk in the absence of covariates. In contrast, a*djusted* effect sizes assessed the magnitude of the relation between volunteering and mortality risk in the presence of covariates. Unadjusted and adjusted effect sizes were computed on the total sample and on independent sub-samples within a study (e.g., participants in poor health and participants in good health). *Interaction* effect sizes assessed the magnitude of the joint effect of volunteering and a moderator variable on mortality. These two-way interaction effects consisted of terms formed by multiplying scores on the volunteer variable by scores on a moderator variable. For example, Okun, August, Rook, and Newsom (2010) examined the joint effect of volunteering and functional health limitations on mortality risk. As indicated previously, two-way interaction effect sizes were extracted only from the total sample and not from subsamples. For volunteer by moderator variable interaction effects, we coded information reported in the article regarding the *p* value associated with the statistical test regardless of whether we were able to extract an effect size.

Complexities of Data Analysis

We established a common metric for effect sizes by converting RR estimates and OR estimates of effect sizes to HR estimates of effect sizes. In the current meta-analysis, the HR was an index of how often death occurred in a group of volunteers (or more frequent volunteers) compared to how often death occurred in a group of non-volunteers (or less frequent volunteers), over time. A RR and an OR were converted to HR using one or both of the following equations (Zhang & Yu, 1998):

$$RR = OR/[(1-r) + (r*OR)]$$ and

$$HR = \ln(1-RR*r)/\ln(1-r),$$

where *r* is the death rate for the reference group for volunteering.

When *r* was not provided in the source, we generated a predicted value for *r* using the proportion of deceased in the total sample as the predictor. The correlation between death rate in the total sample and *r* was .984, and the prediction equation was: $Y_i = 1.073X_i + .014$.

Prior to conducting the inferential analyses, the HR effect sizes were log transformed and weighted by the reciprocal of the conditional variance. For ease of interpretation, summary statistics were transformed back into HRs prior to presentation. When the conditional variance of the effect size was not available, we generated a predicted value using the sample size associated with the effect size as the predictor. The correlation between the sample size associated with the effect size and its conditional variance was -.511 and the prediction equation was: $Y_i = -.00000253X_i + .039$.

There are several sources of complexity that need to be taken into account in conducting meta-analytic analyses on effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009). In the present meta-analysis, we faced three sources of complexity—(a) effect sizes published in different journal articles from the *same* data set; (b) two or more unadjusted (or adjusted) effect sizes extracted from the *same* source; and (c) partitioning the sources of variability in effect sizes extracted from different studies.

In our meta-analysis, Lum and Lightfoot (2005) and Luoh and Herzog (2002) both reported analyses conducted using the Asset and Health Dynamics among the Oldest Old data set. Similarly, Harris and Thoresen (2005) and Sabin (1993) both reported analyses conducting using the Longitudinal Study of Aging. Our decision rule was to delete overlapping effect sizes from the source that yielded the fewest effect sizes. Consequently, we excluded the adjusted effect sizes extracted from the total samples from the Lum and Lightfoot (2005) and Sabin (1993) studies.

Two or more effect sizes can be extracted from a source when a researcher creates two or more volunteer-related predictor variables and examines their relations with mortality risk. For example, Okun et al. (2010) reported unadjusted and adjusted effect sizes in separate analyses of mortality risk in the total sample in which (a) volunteering was coded as a dummy variable and (b) volunteer frequency was coded as a continuous variable. In this case, we extracted two unadjusted effect sizes and two adjusted effect sizes from the source. Following the recommendation of Borenstein et al. (2009, pp. 227-230), we created a synthetic effect size and a synthetic variance for the unadjusted effect sizes and for the adjusted effect sizes.

Several indices of between-study variability in effect sizes have been developed. The oldest and most commonly used index of heterogeneity is the Q statistic. The Q statistic provides a test of the null hypothesis that all studies share a common effect size. The lack of a common effect size may indicate that each study has its own true population effect size or it may be due to sampling error. The main drawback of the Q statistic is that it does not partition the variability observed among studies into random error and "real" differences in the true effect sizes. To overcome this limitation, we report two additional statistics related to variability—Tau and I^2. Tau provides us with an estimation of the standard deviation of the true effect sizes which serves to contextualize the meaning of the estimate of the population effect size. I^2 tells us what proportion of the observed variance is due to differences in the true effect sizes. As I^2 increases, the proportion of the observed variance that is real increases. I^2 ranges from 0% to 100% and it has been suggested that 50% and 75% are benchmarks for moderate and high real variation, respectively (Borenstein et al., (2009, p. 119). Moderate and high values of I^2 indicate that it is worthwhile for researchers to search for study characteristics that account for the variation in effect sizes.

The meta-analytic analyses were conducted using Comprehensive Meta-Analysis, Version 2.0 (Borenstein, Hedges, Higgins, & Rothstein, 2006). Unless otherwise specified, we employed random effect models which take into account the amount of variance due to differences between studies as well as differences among participants within studies.

Results

Study Characteristics

Information describing the 14 studies is provided in Table 1. The articles were

---

Insert Table 1 here

---

published over a span exceeding 25 years. Four articles were published prior to 2000, seven articles were published between 2000 and 2009, and the remaining three articles were published between 2010 and 2012. Twelve different data sets were analyzed with the Longitudinal Study of Aging and the Asset and Health Dynamics among the Oldest Old each being analyzed in two studies. Nine of the studies used U.S. samples, and the remaining three studies employed Israeli (n = 2), and Taiwanese samples. The total sample sizes, which do not necessarily correspond to the sample sizes associated with the effect sizes, ranged from 868 to 15,938, with a median of 4,927.50. The minimum age of the participants ranged from 55 to 75 years old with a median of 66.50 years old. The mean and standard deviation for length of the mortality surveillance period in years were 5.94 and 1.86, respectively.

Measurement and Coding of Volunteering

Table 2 summarizes the measures and coding of volunteering as it pertained to the effect sizes extracted from the sources. Four types of volunteer predictor variables were used by researchers studying the relation between volunteering and mortality risk including (a) comparing non-volunteers with volunteers; (b) assessing individual differences in number of organizations volunteered for; (c) individual differences in hours volunteered; and (d) individual differences in frequency of volunteering.

Insert Table 2 here

Description of Unadjusted and Adjusted Effect Sizes

As can be seen in Table 3, we extracted 25 unadjusted effect sizes. Twenty-one of

Insert Table 3 here

the unadjusted effect sizes were derived from the total samples of nine studies (aggregate N = 49,320) and four of the unadjusted effect sizes were derived from sub-samples associated with two studies. The (unweighted) unadjusted effect sizes derived from total samples ranged from .31 to .96, with a median HR of .56. Unadjusted and adjusted HRs were coded such that values greater than 1.00 indicated that volunteering was associated with an *increased* risk of dying whereas values less than 1.00 indicated that volunteering was associated with a *decreased* risk of dying. A HR effect size of 1 indicated that volunteering was unrelated to risk of mortality. The variances associated with the unadjusted effect sizes derived from total samples ranged from .00 to .06, with a median of .03.

Excluding the adjusted effects from the total samples for the Lum and Lightfoot (2005) study and the Sabin (1993) study, we extracted 31 adjusted effect sizes. Twenty-five of the adjusted effect sizes were derived from the total samples of 11 studies (aggregate N = 49,400) and six of the adjusted effect sizes were derived from sub-samples associated with three studies. The (unweighted) adjusted effect sizes derived from total samples ranged from .40 to 1.11, with a median HR of .80. The variances associated with the adjusted effect sizes derived from total samples ranged from .00 to .06, with a median of .03.

Forest Plot of Unadjusted Effect Sizes

We began our inferential analyses by constructing a forest plot of the unadjusted effect sizes derived from total samples. In a forest plot, each study as well as the summary effect is depicted as a point estimate bounded by a confidence interval. As can be seen from Table 3, synthetic unadjusted effect sizes and variances were created for five of the studies--Harris and Thoresen (2005), Konrath, Fuhrel-Forbis, Lou, and Brown (2012), Musick et al. (1999), Okun et al. (2010), and Oman et al. (1999).

As can be seen in Figure 1, the confidence intervals for the nine unadjusted effect

Insert Figure 1 here

sizes were all below 1.00. The weighted mean of these effect sizes was 0.53 with a 95% confidence interval of 0.45 to 0.62. The *p* value associated with the weighted mean is less than .001. Thus, in the absence of control variables, the average effect size suggests that relative to non-volunteers, volunteers have a 47% decrease in the risk of death, with a 95% confidence interval of 38% to 55%.

Heterogeneity of the Unadjusted Effect Sizes

The *Q* test statistic with 8 degrees of freedom was 44.17, $p < .001$, indicating that the effect sizes are heterogeneous. Tau equals .22 which given a weighted average effect size of .53, means that the distribution of true effects is likely to include effect sizes ranging from .31 to .75. The value of I^2 was 82%, indicating that a large proportion of the observed variance reflects differences in the true effect sizes across studies.

Forest Plot of Adjusted Effect Sizes

Prior to analyzing the adjusted effect sizes, it is useful to note the frequency with which various types of variables were used as covariates. The percentage that each type of covariate was used ranged from 18% (leisure) to 100% (age, sex, and physical health). For the remaining type of covariates, the percentage of use was: socioeconomic status (91%), health behaviors (91%), marital status (73%), religiosity/religious behavior (64%), emotional health (64%), social connection (64%), social interaction (64%), ethnicity (55%), work status (55%), and cognitive functioning (27%). As can be seen from Table 3, synthetic adjusted effect sizes and variances were created for six of the studies-- Harris and Thoresen (2005), Konrath et al. (2012), Musick et al. (1999), Okun et al. (2010), Oman et al. (1999) and Shmotkin, Blumstein, and Modan (2003). As can be seen in Figure 2, the confidence intervals for seven of the 11 adjusted effect sizes derived from total samples were below

---

Insert Figure 2 here

1.00. The weighted mean of these effect sizes was 0.76 with a 95% confidence interval of 0.69 to 0.84. The *p* value associated with the weighted mean is less than .001. Thus, in the presence of control variables, the average effect size indicates that, relative to non-volunteers, volunteers have a 24% decrease in the risk of death, with a 95% confidence interval of 16% to 31%.

Heterogeneity of the Adjusted Effect Sizes

The *Q* test statistic with 10 degrees of freedom was 24.42, $p < .01$, indicating that the effect sizes are heterogeneous. Tau equals .11 which given a weighted average effect size of .76, means that the distribution of true effects is likely to include effect sizes ranging from .54 to .98. The value of I^2 was 59%, indicating that a moderate proportion of the observed variance is real rather than spurious.

Comparison of Adjusted and Unadjusted Effect Sizes

To examine the reduction in the relation between volunteering and mortality associated with the introduction of control variables, we used matched pairs of effect sizes from the nine studies that yielded both adjusted and unadjusted effect sizes. We created a difference score for each study by subtracting its unadjusted effect size from its adjusted effect size. For example, in the Rogers (1996) study, the adjusted and unadjusted effect sizes were .81 and .50, respectively. Thus, the difference score for this study was .31. Using the formula provided by Borenstein et al. (2009, p. 228), the variance of the differences in effect sizes was calculated for each study.

Forest plot of the difference between the adjusted and unadjusted effect sizes.

As can be seen in Figure 3, the difference between the adjusted and unadjusted effect sizes ranged from .07 to .32. The weighted mean of the difference between the adjusted and unadjusted effect sizes derived from total samples was .20 with a 95% confidence interval of

Insert Figure 3 here

0.16 to 0.25. The *p* value associated with the weighted mean is less than .001. Thus, the hazard ratio, on average, increases by .20 when the adjusted and unadjusted effect sizes are directly compared across the set of nine studies. Keeping in mind that for effect sizes below 1, larger values indicate smaller effects, the magnitude of the relation between volunteering and mortality risk is significantly ($p < .001$) reduced by the inclusion of covariates.

Heterogeneity of the difference between the adjusted and unadjusted effect sizes.

With 8 degrees of freedom, the *Q* test statistic was 54,693.49, $p < .001$, indicating that the difference between the adjusted and unadjusted effect sizes are heterogeneous. Tau equals .34. The value of I^2 was 99.9%, indicating that virtually all of the observed variance is real rather than spurious.

What is the Form of the Association between Volunteering and Mortality Risk?

To ascertain whether the association between volunteering and mortality risk is linear or nonlinear, we examined whether volunteer predictor variable was related to unadjusted and adjusted effect sizes. Because these studies focused on within-study differences in effect sizes as a function of volunteer predictor variable, fixed effects models were used. Table 4 summarizes the results of these analyses. Of the 10 comparisons, only two were statistically

---

Insert Table 4 here

---

significant. On the one hand, in the Musick et al. (1999) study, the adjusted effect size for volunteering for 1 organization (.60) was stronger than the adjusted effect size for volunteering for 2+ organizations (1.11), suggesting a curvilinear relation. On the other hand, in the Oman et al. (1999) study, the unadjusted effect size for volunteering for 1 organization (.74) was weaker than the unadjusted effect size for volunteering for 2+ organizations (.37), suggesting a linear relation. The results of these analyses do not provide a warrant for drawing a firm conclusion regarding whether the volunteering-mortality risk association is linear or curvilinear.

Do Adjusted Effect Sizes Vary with Moderator Variables?

We examined the relation between two categorical moderator variables—study focus and country-- and effect sizes using the Q_B statistic. We focused on adjusted effect sizes from total samples (N = 11) because these estimates should be more precise than estimates based on unadjusted effect sizes. Nine of the studies focused on the volunteering-mortality risk association and the remaining two studies did not. Also, nine of the studies were conducted in the U.S. whereas the remaining two studies were conducted in Israel. Neither test was significant, lowest $p > .39$.

We examined the relation between five quantitative moderator variables and adjusted effect sizes using meta-regression with a mixed-effects model estimated using the method of moments. The five moderator variables were (a) journal impact factor ($M = 2.48$; $SD = 0.81$), (b) year of publication (*Median* = 2005, $SD = 5.37$), (c) minimum age of sample ($M = 65.73$, $SD = 6.92$, (d) percentage of sample deceased ($M = 24.50$, $SD = 15.70$), and (e) percentage of sample volunteering ($M = 24.56$, $SD = 13.02$). None of the regression coefficients were statistically significant (all *p*s > .05).

Comparison of Effect Sizes Derived from Independent Subsamples within Studies

For independent sub-samples *within* studies, we used the *Q* test with fixed effects to compare five pairs of effect sizes. As can be seen in Table 5, three comparisons were

Insert Table 5 here

statistically significant (p < .01). In the Lee, Steinman, and Tan (2011) study, the unadjusted relation between volunteering and mortality risk was stronger among non-drivers/limited drivers than regular drivers. In the Hsu (2007) study, the adjusted relation between volunteering and mortality risk was inverse among Taiwanese males (.81) but positive and stronger among Taiwanese females (2.28). In the Konrath et al. (2012) study, the unadjusted relation between volunteering and mortality risk was stronger among participants who were primarily motivated to volunteer by concerns for others as opposed to concerns for self.

Volunteering by Moderator Variable Interaction Effect Sizes

In five of the studies, researchers tested for volunteering by moderator variable interaction effects. In tests of interaction effects, covariates were included in the model, along with the main effects of the volunteering and moderator variable, and one or more interaction terms. Table 6 provides a summary of the measurement and coding of the moderator variables. Of the 34 tests of volunteering by moderator variable interaction effects, four tests involved measures of leisure, four tests involved measures of religiosity, seven tests used measures of health, eight tests used measures of social connection, nine tests employed measures of social interaction, and sex was the moderator variable in the remaining two tests.

Insert Table 6 here

As can been seen in Table 7, we extracted 11 volunteering by moderator interaction effect sizes from four of the five studies (aggregate N = 5226). Interaction effect sizes were coded such that HR values greater than 1.00 indicated that the relation between volunteering and

---

Insert Table 7 here

---

mortality risk increased as the moderator variable *decreased* whereas HR values less than 1.00 indicated that the relation between volunteering and mortality risk increased as the moderator variable *increased*. A HR interaction effect size of 1.00 indicated that the relation between volunteering and mortality risk did not vary across levels of the moderator variable. Interaction effect sizes greater than 1.00 are consistent with the compensation hypothesis whereas interaction effect sizes less than 1.00 are consistent with the complementary hypothesis.

The (unweighted) interaction effect sizes ranged from .37 to 2.44. Seven of the volunteering by moderator variable interaction effect sizes were below 1.00 and the remaining four were above 1.00. The variances associated with the volunteering by moderator variable effect sizes ranged from .00 to .13, with a median of .03. Because sufficient information to extract effect sizes was reported for only 11 of the 34 volunteering by moderator variable interaction effects and because the moderator variables were diverse, we did not carry out inferential meta-analytic statistical techniques on the summary effect size.

Confidence intervals were generated for each of the 11 interaction effect sizes. As indicated in Table 7, the 11 confidence intervals were significant ($p < .05$). For all three volunteering by religiosity interaction effect sizes and for all three volunteering by social connection interaction effect sizes, the entire confidence intervals were below 1, which supports the complementary hypothesis. Similarly, for the social interaction effect size, the entire confidence interval was below 1. In contrast, for the two volunteering by health interaction effect sizes and for the volunteering by social interaction effect size, all three confidence intervals were entirely above 1 supporting the compensatory hypothesis. For the two effect sizes involving leisure as the moderator variable, one confidence interval was entirely above 1 whereas the other confidence interval was entirely below 1. Finally, regardless of whether we were able to extract a volunteering by moderator variable interaction effect size, we extracted information provided in the source regarding the *p* value associated with the statistical test of the volunteering by moderator variable interaction effect. When researchers did not report exact *p* values, we included information from the source regarding whether the *p* values were less than a specified value or greater than a specified value. Of the 34 *p* values, 10 were less than .05. Only two of the *p* values less than .05 were associated with tests of interaction effects that did not yield effect sizes. In the Harris and Thoresen (2005) study, religiosity and one of the social interaction variables both amplified the inverse relation between volunteering and mortality risk.

The Robustness of Conclusions Regarding Sets of Effect Sizes

We examined the robustness of the conclusions drawn regarding the unadjusted effect sizes and the adjusted effect sizes using the Duval and Tweedie (2000) trim and fill procedure which provides an estimate of the unbiased mean effect size. For the interaction effect sizes, we used the Fischer Exact Test (Agresti, 1992) to determine whether researchers were less likely to report effect sizes when the *p* values associated with the statistical test was greater than .05.

Unadjusted and adjusted effect sizes. Application of the trim and fill procedure separately for the unadjusted effect sizes and the adjusted effect sizes revealed that the means of the distributions of effect sizes did not change indicating the absence of publication bias.

Interaction effect sizes. The Fischer Exact Test revealed that interaction effect sizes were significantly (p = .001) more likely to be reported when the p values associated with the statistical tests were less than .05 (80%) as compared to when the p values associated with the statistical tests were greater than .05 (12.5%). This association indicates that researchers are biased toward reporting interaction effect sizes that achieve conventional levels of statistical significance. Thus, the volunteering by moderator variable interaction effect sizes included in the current meta-analysis overestimate the magnitude of the joint effect of volunteering and moderator variables on mortality risk.

Discussion

This is the first meta-analysis of the volunteering-mortality association and as such provides strong evidence in favor of the growing consensus that helping others yields health benefits for the helper. Across 11 studies, volunteerism appeared to reduce mortality risk by almost half in unadjusted models, when variables that likely mediate the effect are not first removed from the analysis. When the more conservative test is applied, one that controls for covariates such as age, sex, ethnicity, socioeconomic status, work status, marital status, religiosity, emotional health, health behaviors, social connection, social interaction, and physical health, the adjusted effect size remains substantial, predicting a 25% reduction in the risk of death. Furthermore, we detected no evidence of publication bias for our estimates of the means of the unadjusted and adjusted effect sizes.

However, researchers did exhibit a bias toward reporting interaction effect sizes only when the tests of the interaction effects reached conventional levels of statistical significance. Thus, it is premature to draw conclusions regarding the merits of the complementary and compensatory hypotheses. Keeping this caveat in mind, our analyses revealed that religious involvement appears to amplify the association between volunteering and mortality risk.

Consistent with the complementary hypothesis, the greater resources derived from religious involvement enhance the health-related benefits of volunteering. In fact, all four estimates of the volunteering by public religiosity interaction effect on mortality risk (see Table 7) were significant (highest $p < .05$). In trying to understand this effect further, we think it is reasonable to consider whether religiosity confers cultural capital, or reflects more altruistic values.

Wilson and Musick (1997) identified altruistic values as a resource that contributes to cultural capital. Konrath et al. (2012) found that volunteering reduced mortality risk only among older adults motivated primarily by a concern for others rather than a concern for oneself. Furthermore, Pargament (1997) posits that involvement in public religious activities is associated with stronger motivation to engage in actions that benefit humanity. Thus, publically religious older adults may benefit more from volunteering in terms of reduction of mortality risk than their non-publically religious peers because they are more motivated to volunteer by other-oriented motives, which have been theorized to be beneficial for physical health (Brown et al., 2012).

Our results also reveal substantial differences among studies, reflected in the heterogeneity of the unadjusted and adjusted effect sizes. In other words, the volunteering-mortality risk effect sizes reflect systematic differences among the studies and not just sampling error. Rather than conceiving the effect sizes extracted from the studies as variation around a single (common) population effect size, each study should be viewed as having its own population effect size which gives rise to a distribution of population effect sizes.

Unfortunately, we were unable to account for variation in effect size magnitude. One possibility is that the analyses of predictors of effect sizes magnitude were statistically under-powered. Thus, we cannot draw firm conclusions from our meta-regressions, including the form of the relation between amount of volunteering and mortality risk. Additional research is warranted that tests the notion that high levels of volunteering are detrimental to the health of the volunteers (Musick et al., 1999).

Limitations and Guidelines for Future Studies

Our review has several limitations. First, it was limited by the small, published literature and a relatively few “file draw” studies with null findings would negate the associations that we observed between volunteering and mortality risk. Second, the studies used non-experimental designs. Although researchers, on average, controlled for over nine types of covariates, these efforts do not permit us to draw conclusions regarding the causal impact of volunteering on mortality (von Bonsdorff & Rantanen, 2011). The third limitation of studies examining the volunteering-mortality relation has been the lack of standardization of volunteer predictor variables. For example, the lack of consistency in assessing and coding frequency of volunteering and hours volunteered makes it more difficult to establish the form of the relation between volunteering and mortality risk. Finally, some researchers did not (a) report both unadjusted and adjusted effect sizes; (b) test for interaction effects when reporting results for independent subsamples; and (c) provide effect sizes when testing interaction effects.

To rectify these limitations, in future studies of the relation between volunteering and mortality risk, researchers should (a) employ experimental designs; (b) assess frequency of volunteering and hours volunteered using clearly specified anchors (i.e., once a month rather than occasionally) and collect objective as well as self-report data; (c) report unadjusted as well as adjusted effects from non-experimental studies; (d) test interaction terms instead of carrying out separate tests of the effects of volunteering within each subsample; and (e) provide confidence intervals and hazard ratios (or odds ratios) for all main and joint effects that are tested.

Agenda for Future Research

We hope that the current meta-analysis inspires a new generation of research on the volunteering-mortality risk association that (a) explores individual differences in volunteer-related variables, (b) unpacks and tests causal mechanisms, and (c) expands the range of prosocial behaviors.

Individual differences in volunteer-related variables. The limited research conducted on who benefits the most from volunteering in terms of mortality risk reduction has largely ignored volunteer-related variables. Konrath et al. (2012) found that volunteers who were primarily motivated by self-oriented reasons did not live longer than non-volunteers. We believe that additional volunteer-related individual differences variables should be examined as moderator variables. More specifically, the positive impact of volunteering on health outcomes may vary with variables such as volunteer work autonomy, efficacy, and mattering. In other words, the health-related benefits of volunteering may be negated when volunteers do not derive a sense of control, competence, and making a difference from their unpaid work.

Unpacking and testing causal mechanisms. To make progress in unraveling the mystery of how volunteering reduces mortality risk, it will be important to take advantage of integrative, theoretical models that have been advanced in related research. Specifically, a "caregiving system" model that is grounded in evolutionary biology, neuroscience, social psychology, and attachment theory has recently been advanced which proposes mechanisms that link prosocial behavior with mortality risk (Brown et al., 2012; Brown & Preston, 2012). This framework integrates animal models of parenting (Numan, 2006) with human neuroimaging studies of parental responses to specify the triggers of prosocial behavior and conditions that favor beneficial versus harmful effects of prosocial behavior. This model suggests that perceptions of another's need in combination with the ability to meet the need trigger the motivation to help, which in turn activates neural circuits related to parenting that release hormones such as oxytocin and progesterone, both of which regulate stress and down-regulate inflammation. Critically, situational (recipient, interpersonal, organizational, and cultural) characteristics such as authenticity of need or interdependence with the recipient are hypothesized to interact with personal resources to generate either intrinsic motives to help (i.e., mediated by hypothalamic processes, Numan, 2006), or to generate extrinsic motives to help that by-pass other-regarding emotions and lengthen potential exposure to harmful levels of chronic stress and inflammation.

Thus, the caregiving system model suggests that volunteering can be mediated by neural circuitry that activates natural tendencies we all have to be caring toward others. Whether volunteering will produce health benefits is thought to depend not entirely on whether resources exist to give, but also the extent to which the signals for need are authentic—that is, they occur in the context of a trusting relationship, trusted organization, or cultural norms that minimize the possible threat of exploitation.

Expanding the range of prosocial behaviors. Finally, we advocate that investigators examine the health consequences of other types of prosocial behavior in addition to organizational volunteering. The majority of studies that demonstrate morality benefits associated with prosocial behavior were not initially designed to examine the health effects of providing support, but were instead designed to investigate the health effects of receiving support (Brown, et al., 2003). As such, systematic investigations into the health consequences of helping others are rare. In the absence of these efforts, researchers have reanalyzed existing data, which is supportive, but does not lend itself to meta-analytic techniques, tests of mechanisms, tests of causal relationships, or tests of boundary conditions. Given the strong evolutionary biological theoretical underpinnings of integrative approaches to prosocial and caregiving behavior (Brown et al., 2012), attempts to formally and systematically examine the health consequences of helping behaviors within close relationships are likely to be informative, relevant, and may suggest important caveats for translating basic research on volunteerism to health policy.

Implications for Public Health

The baby boomers pose a major challenge and innovative changes are required to sustain our system of public health including finding ways to keep them as healthy as possible and to integrate them into the fabric of our communities (Knickman & Snell, 2002). Volunteering has been described as a win-win activity because of the benefits derived by both the recipients (Wheeler, Gorey, & Greenblatt, 1998) and the providers (Post & Neimark, 2007).

The results of this meta-analysis suggest that it is no longer a question of whether volunteering is predictive of reduced mortality risk. Rather our results suggest that the volunteering-mortality association is reliable, and that the magnitude of the relationship is sizable. The findings of the current study are bolstered by research using a true experimental design to investigate the health-related benefits of volunteering. In a comparison of Experience Corps program volunteers with wait-listed controls, Fried et al. (2004) showed that whereas the control group exhibited decreases in their strength, the volunteering group exhibited increases in their strength. In a more recent study using cognitively at risk volunteers in the Experience Corps program, Carlson et al. (2009) demonstrated that relative to participants in the control group, participants in the intervention group who received training in general literacy support, library support, and conflict resolution exhibited more cognitive activity in the left pre-frontal cortex and anterior cingulate cortex (related to empathy) during a selective attention task.

At the same time that the health-related benefits of volunteering have been documented, forecasts suggest that there will be a severe shortage of volunteers (Gottlieb & Gillespie, 2008). Given these circumstances, strategies should be identified to encourage older adults to volunteer. For example, online volunteering activities are expanding the range of opportunities available to healthy older adults as well as to older adults with functional limitations (Cravens, 2003) although there may be boundary conditions on whom and under what circumstances volunteering has a salutary effect on health. In addition to interventions which focus on volunteering, it is also possible to leverage its benefits by incorporating volunteering into psychosocial interventions with other foci. For example, interventions that target family members who are caregivers can provide opportunities for them to serve as peer mentors for novice caregivers (Pillemer, Suitor, Landreneau, Henderson, & Brangman, 2000).

Ultimately, the possibility that volunteering reduces mortality risk is exciting and a mystery. Our results suggest that it is now permissible, desirable, and even necessary for researchers to begin to delve into this mystery. What we discover may do more than inform health policy and volunteerism. The complex and intricate systems of the body that either produce a volunteering—mortality association or account for it suggest that what we discover may tell us something bigger, about disease, the aging process itself, and/or how behavior, perception, and motivation are instantiated in the body.

References

References marked with a single asterisk indicate studies included in the meta-analysis that yielded effect sizes.

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, *7*, 131-153. doi: 10.1214/ss/1177011454

*Ayalon, L. (2008). Volunteering as a predictor of all-cause mortality: What aspects of volunteering really matter? *International Psychogeriatrics*, *20*, 1-14. doi: 10.1017/S1041610208007096

Belloc, N. B., & Breslow, L. Relationship of physical health status and health practices. *Preventive Medicine*, *1*, 409-421. doi: 10.1016/0091-7435(72)90014-X

Borenstein, M., Hedge, L. V., Higgins, J. P. T., & Rothstein, H R. (2006). *Comprehensive meta-analysis* (Version 2.0) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedge, L. V., Higgins, J. P. T., & Rothstein, H R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: John Wiley& Sons, Ltd.

Brown, S. L. & Brown, R. M. (2006). Selective Investment Theory: Recasting the functional significance of close relationships. *Psychological Inquiry* , *17*, 1-29. doi: 10.1207/s15327965pli1701_01

Brown, S. L., Brown, R. M., House, J., & Smith, D. (2008). Coping with spousal loss: The potential buffering effects of self-reported helping behavior. *Personality & Social Psychology Bulletin, 34*, 849-861. doi: 10.1177/0146167208314972

Brown, S. L., Brown, R. M., & Preston, S. (2012). The human caregiving system: A neuroscience model of compassionate motivation and behavior. In S. Brown, R. Brown, & L. Penner (Eds.), *Moving beyond self-interest: Perspectives from* *evolutionary biology, neuroscience, and the social sciences* (pp. 75-88). New York: Oxford University Press.

Brown, S. L., Nesse, R. M., Vinokur, A., & Smith, D. (2003). Providing social support may be more beneficial than receiving it: Results from a prospective study of mortality. *Psychological Science*, *14*, 320-327. doi: 10.1111/1467-9280.14461

Brown, S., & Okun, M. A. (in press). Using the caregiver system model to explain the resilience-related benefits that older adults derive from volunteering. In M. Kent, M. Davis, & J. Reich (Eds.), *The resilience handbook: Approaches to stress and trauma*. New York: Routledge.

Brown, S. L., & Preston, S. (2012). Mechanisms linking prosocial behavior to physical health. *Perpsectives on Psychological Science* (revise and resubmit).

Brown, S. L. Smith, D. M., Schulz, R. Kabeto, M., Ubel, P., Yee, J., . . . Langa, K. (2009). Caregiving and decreased mortality in a national sample of older adults. *Psychological Science*, *20*, 488-494. doi: 10.1111/j.1467-9280.2009.02323.x

Brown, W. M., Consedine, N. S., & Magai, C. (2005). Altruism relates to health in an ethnically diverse sample of older adults. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences*, *60B*, 143-152. doi: 10.1093/geronb/60.3.P143

Carlson, M. C., Erickson, K. I., Kramer, A. F., Voss, M. W., Bolea, N., Mielke, M., . . . Fried, L. P. (2009). Evidence of neuroplasticity in at–risk older adults: The Experience Corps Program. *The Journals of Gerontology: Series A: Biological and Medical Sciences*, *64*, 1275-1282. doi: 10.1093/Gerona/glp117

Cravens, J. (2003). Online mentoring: Program and suggested practices as of February 2001. *Journal of Technology in Human Services*, *21*, 85-109. doi: 10.1300/J017v21n01_05

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-463. doi: http://www.jstor.org/stable/2676988

Fried, L. P., Carlson, M. C., Freedman, M., Frick, K. D., Glass, T. A., Hill, J., . . . Zeger, S. (2004). A social model for health promotion for an aging population: Initial evidence on the Experience Corps model. *Journal of Urban Health*, *81*, 64-78. doi: 10.1093/jurban/jth094

Gottlieb, B., & Gillespie, A. A. (2008). Volunteerism, health, and civic engagement among older adults. *Canadian Journal on Aging*, *27*, 399-406. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.3138/cja.27.4.399

Grimm, R., Spring, K., & Dietz, N. (2007). *The health benefits of volunteering: A review of recent research*. New York: Corporation for National & Community Service.

*Gruenewald, T. L., Karlamangla, A. S., Greendale, G. A., Singer, B. H., & Seeman, T. E. (2007). Feelings of usefulness to others, disability, and mortality in older adults: The MacArthur Study of Successful Aging. *Journal of* Gerontology*: Psychological Sciences*, *62B*, P28-P37. doi: 10.1093/geronb/62.1.P28

*Harris, A. H., & Thoresen, C. E. (2005). Volunteering is associated with delayed mortality in older people: Analysis of the longitudinal study of aging. *Journal of Health Psychology*, *10*, 739-752. doi: 10.1177/1359105305057310

*Hsu, H. C. (2007). Does social participation by the elderly reduce mortality and cognitive impairment? *Aging & Mental Health*, *11*, 699-707. doi: 10.1080/13607860701366335

Knickman, J. R., Snell, E. K. (2002). The 2030 problem: Caring for aging baby boomers. *Health Services Research*, *37*, 849-884. doi: 10.1034/j.1600-0560.2002.56.x

*Konrath, S, Fuhrel-Forbis, A., Lou, A., & Brown, S. (2012). Motives for volunteering are associated with mortality risk in older adults. *Health Psychology*, *31*, 87-96. doi: 10.1037/a0025226

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*, 673-676. doi:10.1038/nature03701

*Lee, S. J., Steinman, M. A., & Tan, E. J. (2011). Volunteering, driving status, and mortality in U.S. retirees. *Journal of the American Geriatrics Society*, *59*, 274-280. doi: 10.1111/j.1532-5415.2010.03265.x

*Lum, T. Y., & Lightfoot, E. (2005). The effects of volunteering on the physical and mental health of older people, *Research on Aging*, *27*, 31-55. doi: 10.1177/0164027504271349

*Luoh, M. C., & Herzog, A. R. (2002). Individual consequences of volunteer and paid work in old age: Health and mortality. *Journal of Health and Social Behavior*, *43*, 490-509. doi: http://www.jstor.org/stable/3090239

Mathers, C. D., Boerma,T., & Fat, D. M. (2009). Global and regional causes of death. *British Medical Bulletin*, *92*, 7-32. doi: 10.1093/bmb/ldp028

*Musick, M. A., Herzog, A. R., & House, J. S. (1999). Volunteering and mortality among older adults: Findings from a national sample. *Journal of Gerontology*, *54B*, S173-S180. doi: 10.1093/geronb/54B.3.S173

Numan, M. (2006). Hypothalamic neural circuits regulating maternal responsiveness toward infants. *Behavioral Cognitive Neuroscience Review*, *5*, 163-190. doi: 10.1177/1534582306288790

*Okun, M. A., August, K. J, Rook, K. S., & Newsom, J. T. (2010). Does volunteering moderate the relation between functional limitations and mortality?. *Social Science & Medicine*, *71*, 1662-1668. doi:10.1016/j.socscimed.2010.07.034

Oman, D. (2007). Does volunteering foster physical health and longevity? In S. G. Post (Ed.), *Altruism and health: Perspective from empirical research* (pp. 15-32). New York: Oxford University Press.

*Oman, D., Thoresen, C., & McMahon, K. (1999). Volunteerism and mortality among the community-dwelling elderly. *Journal of Health Psychology, 4*, 301-316. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.1177/135910539900400301

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.2307/1164923

Pargament, K. I. (1997). *The psychology of religion and coping: Theory, research and practice*. New York: Guilford.

Piliavin, J. A., & Siegl, E. (2007). Health benefits of volunteering in the Wisconsin Longitudinal Study. *Journal of Health and Social Behavior*, *48*, 450-464. doi: 10.1177/002214650704800408

Pillemer, K., Suitor, J. J., Landreneau, L. T., Henderson, C. R., & Brangman, S. (2000). Peer support for Alzheimer's caregivers: Lessons from an intervention study. In Pillemer, K., Suitor, J. J., Landreneau, L. T., Henderson, C. R., & Brangman, S. (Eds.), *Social integration in the second half of life* (pp. 265-286). Baltimore, MD: John Hopkins University Press.

Post, S. G., Ed. (2007). *Altruism and Health: Perspectives from empirical research*. New York, Oxford University Press

Post, S. G, & Neimark, J. (2007). *Why good things happen to good people*. New York: Broadway Books.

*Rogers, R. (1996). The effects of family composition, health, and social support linkages on mortality. *Journal of Health and Social Behavior*, *37*, 328-329. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.2307/2137260

*Sabin, E. P. (1993). Social relationships and mortality among the elderly. *Journal of Applied Gerontology*, *12,* 44-60. doi: 10.1177/073346489301200105

*Shmotkin, D., Blumstein, T., & Modan, B. (2003). Beyond keeping active: Concomitants of being a volunteer in old-old age. *Psychology and Aging*, *18*, 602-607. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.1037/0882-7974.18.3.602

Thoits, P. A., & Hewitt, L. N. (2001). Volunteer work and well-being. *Journal of Health and Social Behavior*, *42,* 115-131. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.2307/3090173

Van Willigen, M. (2000). Differential benefits of volunteering across the life course. *The Journals of Gerontology: Psychological Sciences and Social Sciences*, *55,* S308-S318. doi: 10.1093/geronb/55.5.S308

Von Bonsdorff, M. B., & Rantanen, T. (2011). Benefits of formal voluntary work among older people. A review. *Aging Clinical and Experimental Research*, *23*, 162-169. doi: 10.3275/7200

Wheeler, J. A., Gorey, K. M., & Greenblatt, B. (1998). The beneficial effects of volunteering of older volunteers and the people they serve: A meta-analysis. *International Journal of Aging and Human Development*, *47*, 69-79. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.2190/VUMP-XCMF-FQYU-V0JH

Wilson, J., & Musick, M. (1997). Who cares? Toward an integrated theory of volunteer work. *American Sociological Review*, *62,* 694-713. doi: http://dx.doi.org.ezproxy1.lib.asu.edu/10.2307/2657355

Zhang, J., & Yu, K. F. (1998). What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association, 280*, 1690-1691. doi: 10.1001/jama.280.19.1690

Table 1. Description of Study Characteristics

First Author	Year of Publication	Data Set	Country	Total N	Minimum Age
Ayalon	2008	Israeli Census Bureau Survey	Israel	5,055	60
Gruenewald	2007	MacArthur Study of Successful Aging	USA	1,030	70
Harris	2005	Longitudinal Study of Aging	USA	7,496	70
Hsu	2007	Survey of Health and Living Status of the Elderly	Taiwan	2,825	60
Konrath	2012	Wisconsin Longitudinal Study	USA	10,317	68
Lee	2011	Health and Retirement Survey	USA	6,408	65
Lum	2005	Asset and Health Dynamics Among the Oldest Old	USA	7,322	70
Luoh	2002	Asset and Health Dynamics Among the Oldest Old	USA	4,860	75
Musick	1986	American’s Changing Lives	USA	1,211	65
Okun	2010	Later Life Study of Social Exchanges	USA	868	65
Oman	1999	Marin County	USA	1,972	55
Rogers	1996	National Health Interview Survey Supplement on Aging	USA	15,938	55
Sabin	1993	Longitudinal Study of Aging	USA	7,485	70
Shmotkin	2003	Cross-Sectional and Longitudinal Aging Study	Israel	1,343	75

Table 2. Measurement and Coding of Volunteering Relevant to Effect Sizes

First Author	Measure of Volunteering	Coding of Volunteer Variables
Ayalon	Volunteering within an organization.	Yes versus no
Okun	How often volunteered in the past month. Response options ranged from "never or almost never" (coded 0) to "daily" (coded 5)	Frequency of volunteering was treated as a continuous variable and as a binary variable (never or almost never versus all other response options combined)
Harris	How often did volunteer work. Response options included never, rarely, sometimes and frequently	Three dummy variables were formed (rarely, sometimes, and frequently) with never as the reference group
Hsu	Did volunteer work	Yes versus no
Lum	Hours volunteered in the past year	0 to 99 hours versus 100 or more hours
Rogers	Did volunteer work in the community	Yes versus no
Gruenewald	Volunteered in the past year	Yes versus no
Shmotkin	Volunteered with an organization and frequency of volunteering. Response options included several times a week, several times a month, less than several times a month, and did not answer frequency question	Yes versus no plus four dummy variables (several times a week, several times a month, less than several times a month and did not answer frequency question with non-volunteer as the reference group.
Musick	Volunteered in the past year for religious, school, political, senior citizen and "other" organizations and hours volunteered.	Two sets of dummy variables (1 organization and 2+ organizations versus non-volunteer and less than 40 hours and 40 or more hours versus non-volunteer
Oman	Number of organizations involved with as a volunteer and hours volunteered per week.	Two sets of dummy variables (1 organization and 2+ organizations versus non-volunteer and less than 4 hours and 4 or more hours versus non-volunteer

Table 2 Continued

First Author	Measures of Volunteering	Coding of Volunteer Variables
Lee	Spent time doing volunteer work for religious, educational, health-related, or other charitable organization	Yes versus no
Luoh	Hours volunteered in the past year	0 to 99 hours versus 100 or more hours
Sabin	Did volunteer work in the past 12 months	Yes versus no
Konrath	Volunteered in the past 10 years, regularity of volunteering in the past 10 years (0=not at all to 3=volunteered regularly the whole time), and hours volunteered per month during the past year	Yes versus no for volunteering in past 10 years, continuous measures of regularity of volunteering, and hours volunteered

Table 3. Unadjusted and Adjusted Effect Sizes and Variances

First Author	Unadjusted Effect Sizes (and Variances)	Adjusted Effect Sizes (and Variances)
Ayalon	.50 (.03)	.77 (.04)
Gruenewald		.72 (.05)
Harris	.59 (.04)	1.01 (.04)
	.58 (.01)	.71 (.00)
	.47 (.01)	.81 (.01)
Hsu		.81* (.03)
		2.28* (.03)
Konrath	.53 (.03)	.63 (.05)
	.74 (.01)	.97 (.00)
	.96 (.00)	.84 (.01)
	.92* (.07)	
	.37* (.04)	
Lee	.41 (.01)	.68 (.01)
	.41* (.01)	.65* (.01)
	.75* (.02)	.91* (.02)
Lum		.67[a] (.00)
Luoh	.31 (.03)	.40 (.03)
Musick	.40 (.03)	.60 (.03)
	.65 (.03)	1.11 (.03)
	.46 (.03)	.70 (.03)
	.58 (.03)	.93 (.03)
Okun	.86 (.00)	1.00 (.00)
	.56 (.03)	.82 (.03)

Table 3 Continued

First Author	Unadjusted Effect Sizes (and Variances)	Adjusted Effect Sizes (and Variances)
Oman	.58 (.02)	.80 (.02)
	.74 (.02)	.94 (.02)
	.37 (.06)	.56 (.06)
	.69 (.03)	
	.49 (.04)	
Rogers	.50 (.02)	.81 (.00)
Sabin		.58[b] (.02)
		.53* (.03)
		.80* (.03)
Shmotkin		.67 (.02)
		.62 (.03)
		.60 (.03)
		.86 (.03)
		.96 (.03)

*Extracted from sub-sample.
[a] Excluded from analysis because of overlap with adjusted effect size extracted from Luoh and Herzog (2002).
[b] Excluded from analysis because of overlap with adjusted effect size extracted from Harris and Thoresen (2005).

Table 4. The Relation between Volunteer Predictor Variable and Effect Sizes

First Author	Volunteer Predictor	Effect Size[a]	*df*	Q
Harris	Rarely	.59 (U)	2	2.95
	Sometimes	.58 (U)		
	Frequently	.47 (U)		
Harris	Rarely	1.01 (A)	2	3.97
	Sometimes	.71 (A)		
	Frequently	.81 (A)		
Shmotkin	< Several times a month	.60 (A)	2	3.46
	Several times a month	.86 (A)		
	Several times a week	.96 (A)		
Musick	1 Organization	.40 (U)	1	3.39
	2+ Organizations	.65 (U)		
Musick	1 Organization	.60 (A)	1	5.44*
	2+ Organizations	1.11 (A)		
Musick	< 40 hours per week	.46 (U)	1	0.77
	40+ hours per week	.58 (U)		
Musick	< 40 hours per week	.70 (A)	1	1.16
	40+ hours per week	.93 (A)		
Oman	1 Organization	.74 (U)	1	6.21*
	2+ Organizations	.37 (U)		
Oman	1 Organization	.94 (A)	1	3.36
	2+ Organizations	.56 (A)		
Oman	1-3 hours per week	.69 (U)	1	1.79
	4+ hours per week	.49 (U)		

[a](A) = Adjusted, (U) = Unadjusted
*p < .05

Table 5. Comparison of Effect Sizes from Within-Study Independent Samples

First Author	Independent Samples	Effect Size[a]	*df*	Q
Lee	Non-drivers/Limited drivers	.41 (U)	1	11.78***
	Regular drivers	.75 (U)		
Lee	Non-drivers/Limited drivers	.65 (A)	1	2.85
	Regular drivers	.91 (A)		
Sabin	Poor health	.80 (A)	1	2.98
	Good health	.53 (A)		
Hsu	Females	2.28 (A)	1	15.46***
	Males	0.81 (A)		
Konrath	Motivated by concern for others	.37(U)	1	7.76**
	Motivated by concern for self	.92 (U)		

[a](A) = Adjusted, (U) = Unadjusted
**$p < .01$, ***$p < .001$

Table 6. Measurement and Coding of Moderator Variables

First Author	Measure of Moderator	Coding of Moderator
Harris	Attended sporting or other event	Yes versus no
	Attended religious services	Yes versus no
	Sex	Male versus female
	Living alone	Yes versus no
	Living with spouse	Yes versus no
	Visited senior center	Yes versus no
	Visited with friends/neighbors	Yes versus no
	Visited with family	Yes versus no
Musick	Living alone	Yes versus no
	Frequency of talking with friends, neighbors or relatives and frequency of getting together with friends and relatives	Talking on scale from 1 (never) to 6 (more than once a day) and Getting together on scale from 1 (never) to 6 (once a week)
Okun	Functional health limitations	Limitations on scale from 0 (not at all) to 3 (very difficult),
	Number of health conditions	Number of health conditions out of 12
	Self-related health	Self-related health from 0 (poor) to 1 (excellent)
Oman	Number of leisure activities out of eight	1 = 3 or more leisure activities; 0 = 2 or fewer leisure activities
	Attended religious services weekly	Yes versus no
	Attended religious services at all	Yes versus no
	Attended other religious group activities monthly	Yes versus no
	Sex	Male versus female
	Feels close to 3 or more friends, feels close to 3 or more relatives, and see 3 close friends or relatives	0 to 3
	Living with others	Yes versus no
	Get out of house everyday	Yes versus no
	Participate in organizational group activities	Yes versus no
Shmotkin	Frequency of physical activities including walking, gardening, and any sport.	1 = not at all to 4 = 3 or more times a week
	Have a hobby	Yes versus no
	Frequency of passive activities (e.g., watching TV), talking with family and friends, going out to do something, and playing cards or another game	0 = never to 3 = every day

Table 7. Interaction Effect Sizes and *p* Values

First Author	Moderator Variable	N	Effect Size (and Variance)	Hypothesis Supported	*p****
Harris	Leisure	7,496			>.05
	Religiosity	7,496		Complementary	<.05
	Sex	7,496			>.05
	Social Connection	7,496			>.05
	Social Connection	7,496			>.05
	Social Interaction	7,496			>.05
	Social Interaction	7,496		Complementary	<.05
	Social Interaction	7,496			>.05
Musick	Social Connection	1,211	.51 (.03)*	Complementary	<.10
	Social Connection	1,211	.55 (.03)*	Complementary	<.10
	Social Connection	1,211			>.05
	Social Connection	1,211			>.05
	Social Interaction	1,211	1.88 (.03)**	Compensation	<.05
	Social Interaction	1,211			>.05
	Social Interaction	1,211			>.05
	Social Interaction	1,211			>.05
Okun	Health	868	1.05 (.00)**	Compensation	=.06
	Health	868	2.44 (.10)**	Compensation	<.01
	Health	868			>.12
	Health	868			>.12
	Health	868			>.12
	Health	868			>.12
Oman	Leisure	1,973	1.75 (.07)**	Compensation	<.05
	Religiosity	1,973	.40 (.10)*	Complementary	=.01
	Religiosity	1,973	.37 (.13)*	Complementary	=.01
	Religiosity	1,973	.66 (.04)*	Complementary	<.05
	Sex	1,973			=.15
	Social Connection	1,973	.69 (.03)*	Complementary	<.05
	Social Connection	1,973			>.05
	Social Interaction	1,973			>.05
	Social Interaction	1,973			>.05
Shmotkin	Health	1,174			>.05
	Leisure	1,174			>.05
	Leisure	1,174	.56 (.03)*	Complementary	=.04

*Entire confidence interval is below 1.
**Entire confidence interval is above 1.
****p* refers to the probability associated with a test of statistical significance for an interaction effect that was provided in the article by the authors.

Study name	Statistics for each study				
	Hazard ratio	Lower limit	Upper limit	Z-Value	p-Value
Ayalon, 2008	0.496	0.356	0.690	-4.159	0.000
Okun, 2010	0.711	0.562	0.901	-2.829	0.005
Harris, 2005	0.547	0.475	0.630	-8.374	0.000
Rogers, 1996	0.502	0.385	0.654	-5.085	0.000
Musick, 1986	0.523	0.420	0.651	-5.806	0.000
Oman, 1999	0.574	0.450	0.731	-4.488	0.000
Lee, 2011	0.407	0.347	0.477	-11.065	0.000
Luoh, 2002	0.309	0.214	0.445	-6.301	0.000
Konrath, 2012	0.744	0.639	0.865	-3.826	0.000
	0.528	0.447	0.624	-7.512	0.000

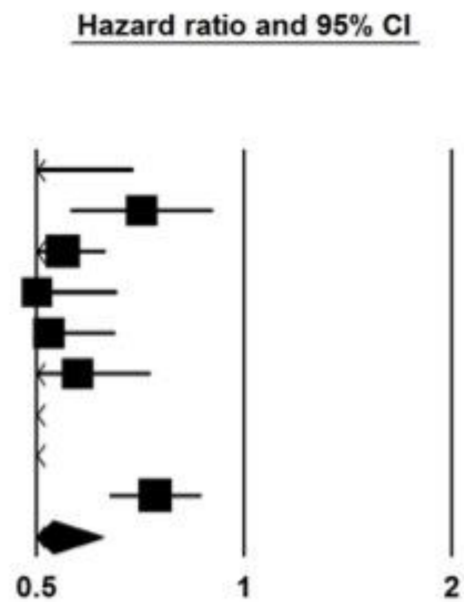

Figure 1. Unadjusted Effect Sizes
Note. For effect sizes less than 1, smaller values indicate larger, beneficial effects of volunteering on mortality risk.

Study name	Hazard ratio	Lower limit	Upper limit	Z-Value	p-Value
Ayalon, 2008	0.770	0.525	1.130	-1.336	0.182
Okun, 2010	0.910	0.750	1.104	-0.957	0.338
Harris, 2005	0.843	0.737	0.965	-2.486	0.013
Rogers, 1996	0.809	0.733	0.892	-4.240	0.000
Gruenewald, 2007	0.723	0.466	1.122	-1.445	0.148
Shmotkin, 2003	0.742	0.624	0.882	-3.379	0.001
Musick, 1986	0.835	0.671	1.040	-1.613	0.107
Oman, 1999	0.767	0.602	0.976	-2.155	0.031
Lee, 2011	0.683	0.569	0.821	-4.058	0.000
Luoh, 2002	0.402	0.294	0.550	-5.696	0.000
Konrath, 2012	0.813	0.679	0.973	-2.260	0.024
	0.763	0.695	0.838	-5.662	0.000

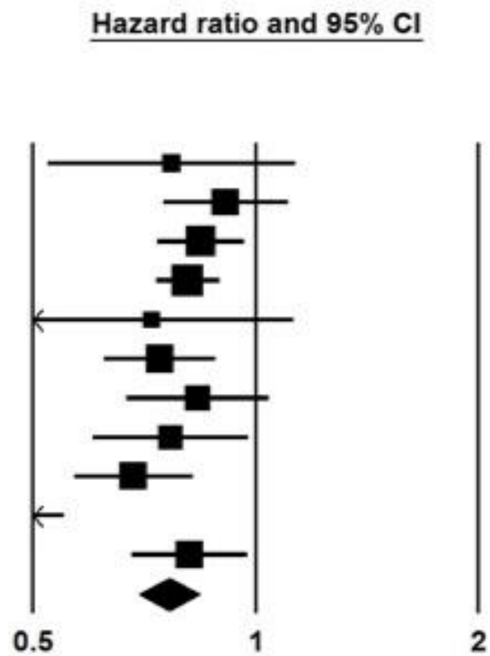

Figure 2. Adjusted Effect Sizes

Note. For effect sizes less than 1, smaller values indicate larger , beneficial effects of volunteering on mortality risk.

Study name	Statistics for each study				
	Difference between hazard ratios	Lower limit	Upper limit	Z-Value	p-Value
Ayalon, 2008	0.275	0.260	0.289	-48.087	0.000
Okun, 2010	0.199	0.190	0.207	-73.696	0.000
Harris, 2005	0.297	0.295	0.299	-341.927	0.000
Rogers, 1996	0.307	0.260	0.363	-13.777	0.000
Musick, 1986	0.312	0.312	0.313	-3678.206	0.000
Oman, 1999	0.193	0.193	0.193	-4066.848	0.000
Lee, 2011	0.276	0.270	0.283	-102.316	0.000
Luoh, 2002	0.093	0.089	0.098	-89.347	0.000
Konrath, 2012	0.069	0.067	0.071	-187.986	0.000
	0.201	0.160	0.251	-14.079	0.000

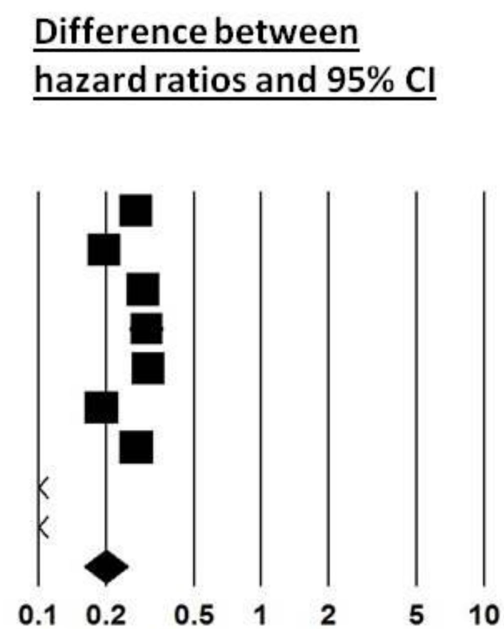

Figure 3. Difference between Adjusted and Unadjusted Effect Sizes
Note. As the difference scores increase, adjusting for covariates reduces the volunteering-mortality association to a greater degree.