Age Effects and Health Appraisal: A Meta-Analysis

Greg Roberts

University of Texas at Austin.

Objective. The purpose was to clarify the effects of methodological variables in the research on age differences in self-rated health and specifically the effects of: (a) the item type used to solicit the appraisal, (b) the approach to sampling, and (c) the criteria for assembling the age groups for study.

Methods. Studies differing on these dimensions were compared using meta-analytic techniques. Seventeen usable samples were grouped according to the age comparisons—seven compared “young” and “old” samples while ten compared “young-old” and “old-old” groups.

Results. The results failed to support the frequently reported health optimism of old-old samples when compared with young-old. Further, the findings suggest that the item type used in eliciting the rating of health might be instrumental in the respondent’s appraisal process. For instance, use of the age-comparative item tends to favor “old” groups when compared to young, whereas the global health item has an effect in the other direction. Effect size was also related to a study’s sampling procedure and to the method used in assembling groups (i.e., age ranges used to represent young and old).

Discussion. Practical implications and areas of needed research are discussed.

One of the more intriguing findings of the research on self-rated health is the difference in the reported health levels of different-aged individuals. Although older age is generally considered to be a time of decline and failing health, a number of investigators have reported that persons 75 years of age and older tend to rate health more optimistically than individuals ranging in age from 65 to 74 years (Ferraro, 1984; Maddox & Douglas, 1973; Murray, Dunn, & Tarnopolsky, 1982; Stoller, 1984). Subsequent research comparing elderly persons to middle-aged persons suggests the possibility of a similar, counterintuitive trend. For instance, Cockerham, Sharp, and Wilcox (1983) found that persons older than 60 tend to rate their health more positively than younger individuals when asked to compare their health “with others your age” (“Would you say your health is better than, the same, or worse than others your age?”). Similarly, Rakowski and Cryan (1990) reported that similar levels of functional disability have a greater negative effect on the self-ratings of persons aged 55 to 64 than those older than 80.

Contrary findings, however, have been reported by Levkoff, Cleary, and Wetle (1987) in a study of 460 individuals aged 45 to 89. Their sample of elderly individuals rated health significantly less optimistically than the group of middle-aged persons. In attempting to explain the discrepant findings (with the Cockerham et al. study), Levkoff and colleagues (1987) suggest that elderly people report more health optimism than middle-aged individuals only when asked specifically to compare themselves with others their age, as in the Cockerham study (1983). They argue that when a more global item is used to elicit an individual’s self-rating (“In general, how would you rate your present physical health?”), results more like those they reported can be expected. (It is worth noting that Levkoff and colleagues’ [1987] results also failed to support the more commonly reported health optimism of old-old persons when compared with young-old individuals.)

The controversy surrounding the salience of “age” in items used in this line of research, as well as the more general age trend that underlies this issue, is best considered within the framework of reference group theory (Maddox, 1962; Shanas et al., 1968). According to this view, individuals tend to compare themselves to relevant groups when asked to evaluate a personal attribute. In the case of self-evaluated health status, same-aged persons appear to be a prominent reference group (Maddox, 1962), suggesting that older individuals may have an advantage when compared with middle-aged and younger persons. The perception that older groups are generally more disabled and less healthy presumably “lowers the bar” for elderly individuals making health appraisals.

However, age may not be the only dimension along which comparisons are made. Persons with chronic illness often use others with similar illness as a comparison group in the appraisal process (Affleck & Tennen, 1991). The use of temporal comparison, where the individual appraises his or her health relative to a point(s) in the past, has also been observed (Suls, Marco, & Tobin, 1991), as has the use of “cognitively assembled” comparison groups (Tornblom, Stern, Pirak, Pudas, & Tornlund, 1993). Evidence also exists that different-aged persons approach the appraisal process in qualitatively distinct ways (Tornblom et al., 1993). For instance, although young adults appear to rely more on comparisons to similar others, middle-aged individuals may prefer dissimilar others whereas older persons are more likely than either of these two groups to engage in temporal comparison. The point here is that a number of salient dimensions may be available to the person rating his
or her health (Singer, 1981), and age may be most prevalent only when specified (in the item used to elicit the appraisal) as the dimension of choice. Thus, although older persons may fare better than middle-aged individuals when age is the predominant basis for comparing, such may be the case only when age is made especially salient by including it in the item's stem.

This argument generally supports Levkoff and colleagues' (1987) position regarding Cockerham and colleagues' findings (1983). Cockerham's use of the age-comparative item may have predisposed the participants of that study to rely primarily on age comparisons, suggesting an advantage for the group of older persons for the reasons discussed earlier. Levkoff's (1987) subjects who responded to the global health item were presumably free to select their own group(s) for comparison, suggesting a less prominent role for "age" and greater pessimism among this older group. This also implies that Cockerham and associates' findings (1983) may be due as much (or at least significantly) to the use of the age-comparative item as to more substantive differences in the health appraisals of older and middle-aged groups. However, as Levkoff and associates (1987) suggest, this is merely speculation. Because the two studies generally fail to clarify the respective effects of the item type being used and the comparison being made (middle-aged persons and older persons), it is difficult to make a conclusive statement on this point.

Other aspects of the research on self-rated health are plagued by a similar lack of clarity. For instance, although the majority of studies in this area rely on probability sampling, a considerable minority have used data collected in less rigorous ways. The use of nonprobability data does not necessarily discount the results of studies relying on such, but it is far from certain that findings of this sort are comparable to those derived from more representative and/or randomly drawn groups.

Another point of potential confusion concerns the ways the age groups used for comparison are operationalized. Different studies have used different age ranges to represent "young," "middle," "young-old," and "old-old." For instance, whereas Kaplan and Camacho (1983) divided their relatively evenly distributed sample of 6,928 adults in Alameda county, California (16–94 years of age) into four age groups (16–29, 30–44, 45–59, and 60+), a study by Jyhla and associates (1986) of health perceptions in men of different ages used three groups, one aged 31 to 35, a second aged 51 to 55, and a third 71 to 75 years of age. Although these two studies both compare the self-rated health of young persons with that of old individuals, they may actually be comparing two very different age groups. In cases where studies comparing the same age groups (young vs old, young-old vs old-old) report different results, the ways in which the groups are operationalized may be partly responsible.

**Purposes of the Study**

Incongruous findings like these are best understood by considering the likely sources of such variation (Cooper & Hedges, 1994). Lipsey (1994) has proposed three possibilities in this respect: (a) method variables, such as sampling, measurement, and/or design; (b) substantive variables, like treatment or population; and (c) extrinsic variables like funding source or type of publication (extrinsic factors generally play a far less prominent role and won't be considered further). Primary research is concerned generally with the substantive variables within a given research domain. When the set of method variables is invariant across the group of relevant studies, the findings that result are generally interpretable in terms of these more substantive issues. However, when the set of method variables also varies across a body of research, a confound with the more substantive factors is possible. The Cockerham/Levkoff situation described earlier, where item type and age differences are confounded, is an example of such a situation. In such cases, clarification requires an unraveling of the respective effects of these two sets of variables (Lipsey, 1994).

Meta-analysis is a useful approach for such a task (Cooper & Hedges, 1994; Lipsey, 1994). Meta-analysis relies on a primary unit of measure, the effect size (Cohen, 1977), which can be calculated for each study and aggregated across studies. The homogeneity of this value can be tested to determine if differences in effect size exist across a group of studies. In cases where such differences do exist (i.e., significant heterogeneity of findings is present), they can be further analyzed to identify the variables that moderate the variation. Once identified, these moderators can be considered in terms of the categories (substantive and method) described by Lipsey (1994). In this case, the "method" variables were of particular interest.

Several of the more prominent method variables in the self-rated health literature were discussed in the opening paragraphs of this article, and these generally drive the research questions (Question 3, in particular) that follow:

**Question 1.** Does the evidence support the claim that older persons (older than 60 or 65 years of age) are more likely than younger or middle-aged individuals (younger than 60) to make positive appraisals of health? Similarly, do "old-old" individuals (over 75) rate health more optimistically than "young-old" persons (60 or 65 to 75)?

**Question 2.** Are there significant differences in the effect sizes reported in this line of research? If not, further analysis is unnecessary, and the pooled effect size(s) calculated in response to Question 1 is adequate for summarizing the findings in this area. However, if significant heterogeneity is present, questions regarding the source(s) of such differences will be of interest.

**Question 3.** As suggested, the presence of significant variation in effect size may be due partly to differences in the design, sampling technique, or instrumentation (i.e., the method variables) of the studies that comprise the literature in question. Several possibilities were outlined previously in this study. They are briefly recapped here:

(a) Levkoff and colleagues (1987) have suggested that older individuals may be more optimistic than middle-aged persons only when asked specifically to make an age comparison. The age-comparative item specifies a reference group, and comparison with age peers may work to the advantage of older individuals. The question is this: Does the age-comparative item result in more optimism on the parts of old individuals when compared to middle-aged persons? Simi-
larly, does item type make a difference in the relative health appraisals of young-old and old-old groups?

(b). Differences in effect size may also result from differences in the approach taken to sampling, suggesting the question: Do the effect sizes in the meta-analysis differ according to the sampling approach taken by a given researcher?

(c). The actual age ranges used to represent a group (i.e., young, old, or older) often differ across studies seemingly devoted to the same comparison. For instance, two studies both comparing middle-aged persons to old persons may actually be comparing two very different age groups. This meta-analysis addressed this question: Do studies on levels of self-rated health and the effect sizes they report differ based on the age ranges used to represent the different age groups being compared?

METHOD

Collection and Selection of Studies

The first step in locating studies was to search the various indexes and databases relevant to this area of research. PSYCINFO, MEDLINE, Social Science Abstracts, and Sociological Abstracts were researched using various combinations of keywords. These included self-rated health, subjective health, self-appraised health, self-evaluated health, age differences, health appraisal, social comparison theory, and reference group theory. Almost 2,000 studies were suggested by searches of the different indexes (this includes duplicates). Of these, the majority were clearly unrelated (based on a review of abstracts) to the questions addressed by this meta-analysis and were not pursued. This left 37 unduplicated, potentially usable studies. The full texts for these were collected. A second step in finding potentially usable studies was to examine the reference lists of the studies collected in the first step. All studies and articles with remotely promising titles were considered. The reference lists of the studies collected in the second step were similarly examined, and potentially useful studies were located and considered. Together, these steps yielded four additional potentially usable manuscripts. Finally, several journals were searched by hand, including The Journals of Gerontology, The Journal of Health and Social Behavior, Health Psychology, Psychology and Aging, The Gerontologist, and The Journal of Aging and Health. No new studies were found as a result of this effort.

Generally, in order to be included, a study had to: (a) report a mean difference in the self-rated health levels of different-aged groups, (b) report enough information for such a difference to be calculated, (c) report a chi-square value for group differences, or (d) report enough frequency data for a chi-square statistic to be computed. Of the 41 potentially usable studies, 14 contained a total of 17 usable samples providing adequate information for calculating the effect size measures of interest in this study. These are listed in the references of this article (indicated by an asterisk). The Cockerham and colleagues (1983) and the Perlmutter and Nyquist (1990) studies were used in both analyses (young-old and old-old) because each reported data across the life span. Also, the Singer (1974) study reported data on the same sample of persons with Parkinson’s disease using two different item types. A weighted average of the two measures of effect was used generally (though not always) for the meta-analysis.

Organization of Studies

Once located, the studies were organized into two groups. Those dealing with differences between younger and older individuals comprised one group whereas studies comparing young-old individuals to old-old persons were taken in a second. Some flexibility was required in assigning an age range to a classification. Generally, however, groups ranging in age from 18 to 44 were considered "young," groups from 45 to 60 or 65 were classified as "middle," and persons older than the age of 60 or 65 were deemed "old." For the young-old to old-old comparison, 80 years was used as a general cutoff. Thus the young-old group in this analysis was generally 60 or 65 to 80 and the old-old group was over 80 years of age. Table 1 displays the studies falling into these respective groups.

Table 1 also shows the other variables coded for the meta-analysis. In the case of the item type used to elicit the health appraisal, there were two possibilities—the global item and the age-comparative item. Sampling technique was classified according to whether a probability or non-probability sample was used. For differences in age range (i.e., operationalization of groups), the mean of a group, when available, was used as a measure of central tendency. When a mean age was not available or not calculable, the midpoint of the range was used.

Plan for Analysis

To address Question 1, the effect sizes of the studies comparing young to old persons were combined using procedures outlined by Rosenthal (1991). Although an effort was made to distinguish between the middle group and the young group when classifying studies by age range (necessary for research Question 3), for this analysis, these two categories were collapsed into a single group labeled "young." The effect sizes of studies comparing "old" with "other" individuals were also combined. To address Question 2, the homogeneity test of effect sizes for the case of unequal N (Hedges & Olkin, 1985) was used. This approach uses the ratio of effect size to variance summed across studies. The resulting statistic is distributed as chi-square.

The three sections of Question 3 were addressed using Hedge’s ANOVA analog (Hedges & Olkin, 1985). The ANOVA analog (Hedges & Olkin, 1985) is a useful tool in cases where there is significant variability among a set of effect sizes. By partitioning the variance among the various studies into independent components, much like sums of squares are partitioned in ANOVA, it is possible to test for differences between the average effect sizes of theoretically derived subgroups. A separate analysis was conducted for each of the questions of interest (i.e., differences due to item type, differences due to sampling technique, differences due to operationalization of groups).

A final note related to analysis concerns the statistics used to represent the effect size. Several measures have
AGING EFFECTS AND HEALTH APPRAISAL

Table 1. Samples Used for Meta-Analysis

<table>
<thead>
<tr>
<th>N</th>
<th>r</th>
<th>Item Type</th>
<th>Sampling</th>
<th>Age Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>684</td>
<td>.165</td>
<td>comparative</td>
<td>probability</td>
<td>37</td>
</tr>
<tr>
<td>360</td>
<td>−.249</td>
<td>global</td>
<td>probability</td>
<td>42</td>
</tr>
<tr>
<td>6908</td>
<td>−.186</td>
<td>global</td>
<td>probability</td>
<td>35</td>
</tr>
<tr>
<td>460</td>
<td>−.229</td>
<td>global</td>
<td>probability</td>
<td>55</td>
</tr>
<tr>
<td>127</td>
<td>.180</td>
<td>comparative</td>
<td>nonprobability</td>
<td>35</td>
</tr>
<tr>
<td>147</td>
<td>.127</td>
<td>global</td>
<td>nonprobability</td>
<td>35</td>
</tr>
<tr>
<td>138</td>
<td>.229</td>
<td>comparative</td>
<td>nonprobability</td>
<td>35</td>
</tr>
<tr>
<td>1071</td>
<td>−.200</td>
<td>global</td>
<td>probability</td>
<td>55</td>
</tr>
<tr>
<td>9757</td>
<td>−.158</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of “Young-Old” Individuals to “Old-Old” Individuals

<table>
<thead>
<tr>
<th>N</th>
<th>r</th>
<th>Item Type</th>
<th>Sampling</th>
<th>Age Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>−.147</td>
<td>comparative</td>
<td>probability</td>
<td>65</td>
</tr>
<tr>
<td>3216</td>
<td>.029</td>
<td>global</td>
<td>probability</td>
<td>70</td>
</tr>
<tr>
<td>2583</td>
<td>−.027</td>
<td>global</td>
<td>probability</td>
<td>70</td>
</tr>
<tr>
<td>69</td>
<td>−.094</td>
<td>global</td>
<td>nonprobability</td>
<td>80</td>
</tr>
<tr>
<td>263</td>
<td>.112</td>
<td>global</td>
<td>nonprobability</td>
<td>69</td>
</tr>
<tr>
<td>251</td>
<td>.208</td>
<td>global</td>
<td>nonprobability</td>
<td>65</td>
</tr>
<tr>
<td>2046</td>
<td>−.113</td>
<td>comparative</td>
<td>nonprobability</td>
<td>70</td>
</tr>
<tr>
<td>64</td>
<td>−.314</td>
<td>comparative</td>
<td>nonprobability</td>
<td>66</td>
</tr>
<tr>
<td>5197</td>
<td>−.0966</td>
<td>global</td>
<td>probability</td>
<td>70</td>
</tr>
<tr>
<td>537</td>
<td>−.046</td>
<td>global</td>
<td>probability</td>
<td>80</td>
</tr>
<tr>
<td>14,379</td>
<td>−.049</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This is a measure of the younger group’s central tendency (mean, midpoint). Group 1 studies with a value of less than 40 were compared with studies with values greater than 40. Group 2 studies with values less than 70 were compared with studies with values of 70 or greater.

Questions 1 and 2

The effect sizes for the samples used in this meta-analysis are displayed in Table 1. The set of studies comparing young to old individuals is shown first, followed by the set of studies comparing young-old to old-old persons. For this first group, the values ranged from −.249 for the Jylha, Leskinen, Alanen, Leskinen, and Heikkinen (1986) study to .180 for the Permlutter and Nyquist (1990) study (when only the average effect size is considered for Singer’s samples). Three of the coefficients were positive, and four were negative. A negative effect size indicates that older groups report lower levels of subjective health than do younger groups. A positive sign suggests the opposite trend. The apparent variability in findings was underscored by the very large value of the homogeneity chi-square ($\chi^2 = 116.23, p < .001$), which suggests that there is considerable systematic variation, due presumably to the moderators (or a subset) included in research Question 3.

For the second set of studies, the effect sizes ranged from .208 (Maddox, 1962) to −.314 (Permlutter & Nyquist, 1990). Also, the ten studies generally were split on the direction of the effect. Seven coefficients had a negative value, and three were positive. As with the comparisons in the first set, though to a lesser degree, there was considerable variation across studies, beyond what one would expect as a result of random error alone ($\chi^2 = 71.23, p < .001$).

The combined effect sizes for the two groups of studies are presented in Table 1, as well. The pooled $r$ for the comparison of young to old persons was −.158 ($p < .001$). Again, the significant negative value indicates that older groups report lower levels of health than do younger groups. The studies comparing the young-old group to old-old individuals suggest a trend in the same direction, though of less magnitude. The pooled effect size was −.049 ($p < .01$).
Question 3

The results of the ANOVA-analog analyses (Hedges & Olkin, 1985) are presented in Table 2. For the comparison of young persons to old individuals, use of the different item types generally led to a different-sized effect ($Q = 118.30, p < .001, df = 1$). The pooled effect size for the studies using the global item was $D = -.392 (p < .001, df = 3)$, whereas the group of studies using the age-comparative item had a combined effect size of $d = .357 (p < .001, df = 2)$. Thus, use of the global item was associated with greater health optimism on the parts of younger persons, whereas the age-comparative item represented an advantage for the older group.

The row labeled “Within Studies” in Table 2 shows the degree of variation within the groups (the groups of studies based on item type). For the group of four studies using the global item (Singer’s Parkinson sample was included in the comparative-item group), there was not a significant amount of systematic error ($Q = 2.32, p = .51, df = 3$), suggesting that these studies report essentially the same-sized effect. The lack of significant within-studies variation for the second set of effect sizes ($Q = .50, p = .78, df = 2$) suggests a similar homogeneity across the three studies using the age-comparative item. Thus, when the studies comparing young persons with old persons are grouped according to the item type used to elicit the health appraisal, the effect size associated with each of the two groups appears to be relatively stable, of moderate size (Cohen, 1977), and of opposite direction (when compared with the other).

For the second set of studies, comparing young-old with old-old persons, the ANOVA-analog results also suggest differences due to the item type used ($Q = 14.15, p < .001, df = 1$). In this case, while the group of studies using the global item generally had a small-sized, though significant, effect ($d = -.071, p < .001$), the age-comparative item was generally associated with an effect of moderate size ($d = -.243, p < .001$). The negative direction of these effect sizes indicates that old-old persons report lower levels of health than do young-old individuals regardless of the item type used. Finally, though the effect sizes of the age-comparative studies did not significantly differ from one another ($Q = 2.76, p = .25, df = 2$), there was considerable within-studies variability for the group using the global item ($Q = 54.29, p < .001, df = 6$), reflecting, perhaps, the variability of comparison groups presumably associated with the global health item.

Differences in sampling technique were also related to effect size. When studies were grouped according to sample type, a significant difference was evident for those studies comparing young with old individuals ($Q = 31.59, p < .001$). Interestingly, the effect sizes for the two groups of studies were about equal in magnitude, but differed in direction ($d = -.340, p < .001$ for probability sampling and $d = .361, p < .001$ for nonprobability samples). This suggests that when probability samples are used, older individuals rate health less optimistically than do younger persons—nonrandom samples result in the opposite effect. Finally, there was considerable variance left unexplained ($Q = 84.84, p < .001$) for the group using probability sampling. The two studies in the nonprobability group did not differ in effect size ($Q = 0.00, df = 1$). For the group of studies comparing young-old to old-old persons, no differences due to sampling technique were evident ($Q = .96, p = .33$).

The analysis of age range differences was based on a measure of central tendency of the younger of the two groups being compared. Four of the studies in the young to old set had a mean or median of less than 38 (see Table 1). The Jyhla and associates (1986) study had a mean of 42, whereas “young” in the Levkoff and associates (1987) study and in the Singer sample (Singer, 1974) of persons without Parkinson’s was about 55. When these latter three samples (i.e., Jyhla, Levkoff, and Singer) were compared with the others, a significant difference in effect size was evident ($Q = 8.56, p < .01$). The results indicate that comparisons of persons older than 40 with the group of old individuals were associated with a larger-sized effect ($d = -.443, p < .001$) than were comparisons of those younger than 40 with the older group ($d = -.291, p < .001$). This suggests that the group of “old” persons fared less well...
when compared with those older than 40 (i.e., a greater difference in self-rated health), than they did with the group of individuals younger than 40 years of age. The considerable within-studies variance \( Q = 107.03, p < .001, df = 3 \) for this latter set of studies, however, suggests that additional variables may moderate this set of effect sizes.

The second set of studies comparing young-old with old-old persons did not differ in this respect \( Q = .01, p = .95 \). The Cockerham and associates (1983), Maddox (1962), and Permlutter and Nyquist (1990) studies had a median or mean of 66 for the young-old group, and the remaining studies in this set had values of 70 or less, with the exception of the Thorlind and Lundberg (1994) and the Larue and colleagues (1979) studies, which had median values of 80. The analysis compared these latter two studies to the eight studies with values of 70 or less.

**DISCUSSION**

The results of this meta-analysis suggest that younger persons tend to be more optimistic about their health than do older individuals—that is, “young” groups rate health more favorably than the group of “old” persons, and “young-old” individuals are generally more optimistic than “old-old” groups. However, as suggested in this article’s introduction, these findings may be subject to qualification. The considerable heterogeneity evident in both sets of studies suggests that a key factor or factors may distinguish some studies from others. This meta-analysis explored this possibility by considering several of the more prominent method variables (Lipsey, 1994) that characterize this literature.

First, item type appears to play a key role in the set of studies comparing old persons with young persons. Although the global-health item results in relatively less health optimism on the parts of older persons (compared with their younger counterparts), the age-comparative item generally produces an effect in the opposite direction, where the older group reports relatively higher levels of health. Aside from supporting Levkoff and colleagues’ (1987) notion about the instrumentality of item type, this finding suggests that the social comparative process implicit in reference group theory (see Feinburg, Lofthus, & Tanur, 1985) is driven, at least partly, by the type of item used in eliciting the appraisal. That is, age may be the most salient dimension for health comparison only when specifically identified as the criterion of interest. When the selection of a dimension (or group) for comparison is left to the respondent (i.e., when the global item is used), age apparently plays a less prominent role. Respondents either choose another dimension altogether, or, more likely, other factors enter the self-assessment process, thereby “diluting” age’s effect.

In cases where old-old persons are being compared with young-old groups, item type appears to be less instrumental. Both the age-comparative item and the global item are associated with greater health optimism on the parts of young-old individuals. This finding is contrary to the frequently reported advantage thought to characterize the health appraisals of old-old individuals (relative to young-old), and it suggests that the conventional wisdom in this respect may need to be reconsidered. The negative direction of the pooled effect for this group of studies underscores this point.

For users of these types of items, these findings suggest a need for caution—the age-comparative item may be appropriate only when an age comparison is desired. For instance, when one’s interest is in a given cohort’s perception of health relative to other members of the same cohort (Cockerham et al., 1983), the age-comparative item may be a more meaningful indicator than the global health item, especially in cases where a sample of persons varying widely in age is being used. At the same time, the global item may offer a more comprehensive assessment of perceived health. Allowing the respondent to evaluate his or her health along a self-selected dimension “fuses information about the respondent’s physical status with the respondent’s judgment about what that physical status means” (Idler, 1992). In cases where the researcher does not wish to “control” the comparative dimensions, the global item may be the more useful of the two.

A second general point suggested by the results of this meta-analysis concerns differences in sampling technique. Although the approach to sampling is associated with no effect in the old-old to young-olds comparisons, in studies comparing young with old samples, probability sampling appears to favor the younger groups. This finding—that the different approaches to sampling are generally associated with different outcomes—underscores the more general precaution against using nonprobability sample data to estimate population parameters. Where nonprobability samples are concerned, measures of self-rated health may be appropriate only for describing the sample in question.

A third key point indicated by these results concerns the age groups used to represent “young.” A comparison of the midpoints of the different young groups in this meta-analysis suggests young and old may not be the only comparison worth exploring. The similarities and dissimilarities in the self-rated health levels of young and middle-aged groups, for instance, may yield additional insight into the health appraisal process generally, and into the nature of age differences in health appraisal more specifically. Considering the possibility that old groups appear to have more in common with young persons than with the group of middle-aged individuals may be an intriguing first step in this respect.

Other areas of future research are suggested by these findings. For instance, it is worth exploring more fully the effect that item type may have on the apparently robust relationship between self-rated health and mortality (see Mossey, 1995, for a discussion). Idler (1992) has suggested that the differences in this respect may be negligible based on her visual inspection of six studies using the different item types. More rigorous investigation is needed to clarify this issue. It may also be worth exploring the utility of item types other than the global health and the age-comparative items (Rakowski, Fleishman, Mor, & Bryant, 1993). Dimensions other than age are of interest to researchers in this area (e.g., functional status, disability), and the behavior of items specifying these dimensions should be more fully explored. Finally, Johnson and Wolinsky (1993) have suggested that the wider use of multi-equation modeling techniques will lead to a fuller understanding of the structure of self-rated health (see Roberts & Stuifbergen, 1998). Because the global and age-comparative items appear to produce different results, they may also reflect qualitatively
distinct aspects of perceived health. Developing and comparing a pair of latent models with the two item types as outcome variables may offer some insight on this point.

As for limitations, a number of promising studies were not included in this meta-analysis because they did not report adequate information for calculating an effect size. Though unlikely, it is possible that this group differs systematically from the set of studies used in this meta-analysis. If so, the findings reported here may not reliably represent the population of effect sizes in this area. There is also the possibility that the results are biased due to systematic reporting of findings (i.e., only studies with a positive result get submitted and/or published), a point to consider when interpreting these findings.

To conclude, the results of this meta-analysis underscore the importance of methodology in this line of research. The selection of item type, the technique used for sampling, and the approach to grouping appear to have an impact on one’s results. For researchers in this area, this highlights the need for thoughtful decision making in the planning process. For others, these findings suggest a need for careful interpretation of the research into this important construct.

ACKNOWLEDGMENTS

The author is a doctoral candidate at the University of Texas at Austin and a research associate in a University Affiliated Program at the University of Texas at Austin.

Address correspondence to Greg Roberts, University of Texas at Austin, S2B252-D5100, Austin, TX 78712. E-mail: gregroberts@mail.utexas.edu

REFERENCES

* Studies included in meta-analysis


Received June 21, 1997
Accepted July 1, 1998

Decision Editor: David J. Ekerdt, PhD