



9-1-2006

Cross-Cohort Differences in Health on the Verge of Retirement

Beth J. Soldo

University of Pennsylvania, bsoldo@pop.upenn.edu

Olivia S. Mitchell

The Wharton School, University of Pennsylvania, mitchelo@wharton.upenn.edu

Rania Tfaily

Carleton University, rania_tfaily@carleton.ca

John F. McCabe

University of Pennsylvania, mccabef@pop.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/prc_papers

 Part of the [Economics Commons](#)

Soldo, Beth J.; Mitchell, Olivia S.; Tfaily, Rania; and McCabe, John F., "Cross-Cohort Differences in Health on the Verge of Retirement" (2006). *Wharton Pension Research Council Working Papers*. 366.

https://repository.upenn.edu/prc_papers/366

The published version of this Working Paper may be found in the 2007 publication: *Redefining Retirement: How Will Boomers Fare?*.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/prc_papers/366

For more information, please contact repository@pobox.upenn.edu.

Cross-Cohort Differences in Health on the Verge of Retirement

Abstract

Baby Boomers have left a unique imprint on US culture and society in the last 60 years, and it might be anticipated that they will also put their own stamp on retirement, the last phase of the life cycle. Yet because Boomers have not all fully retired, we cannot yet judge how they will fare as retirees. Instead, we focus on how this group compares with prior groups on the verge of retirement, that is, at ages 51-56. Accordingly, this chapter evaluates the stock of health which Early Boomers bring to retirement and compare these to the circumstances of two prior cohorts at the same point in their life cycles. Using three sets of responses from the Health and Retirement Study, we find some interesting patterns. Overall, the raw evidence indicates that Boomers on the verge of retirement are in poorer health their counterparts 12 years ago. Using a summary health index designed for this study, we find that those born 1948 to 1953 share health risks with the War Baby cohort. This suggests that most of the health decline instead began before the late 1940's. A more complex set of health conclusions emerges from the specific self-reported health measures. Boomers indicate they have relatively more difficulty with a range of everyday physical tasks, but they also report having more pain, more chronic conditions, more drinking and psychiatric problems, than their HRS earlier counterparts. This trend portends poorly for the future health of Boomers as they age and incur increasing costs associated with health care and medications. Using our health index, only those at the 75th percentile or higher are likely to be characterized as having good or better health.

Disciplines

Economics

Comments

The published version of this Working Paper may be found in the 2007 publication: *Redefining Retirement: How Will Boomers Fare?*

Chapter 7

Cross-Cohort Differences in Health on the Verge of Retirement

Beth J. Soldo, Olivia S. Mitchell, Rania Tfaily, and John F. McCabe

The demographic cohort known as the Baby Boom has always had profound impacts on society, first with a tsunami of young children washing through the educational system, and later with a wave of young people inundating job and marriage markets. Now the oldest Boomers are poised to flood into retirement with important implications for public and private pension systems, health care programs, and cross-generational transfers. Yet because this cohort has not yet fully retired, it is difficult to project how well Boomers will fare in retirement. This chapter compares the health of the Early Boomer cohort to that of previous generations, as they were poised on the verge of retirement. Our work will help evaluate whether evidence from past cohorts can be used for projecting Boomers' future health and retirement security.

Our analysis relies on responses to the Health and Retirement Study (HRS) to examine individual determinants of health for respondents born in 1948–53, the so-called Early Baby Boomer (EBB) cohort, with two older cohorts born 1936–41 and 1942–47. Our goal is to determine whether and why stocks of health capital differ across cohorts on the verge of retirement. We estimate fixed-effect models with life-cycle and lifestyle factors, both past and present, as 'inputs' into a model of health capital accumulated by age 51–56, that is, the age of entry into the HRS. We conclude that Boomers do not appear to be entering retirement better positioned than their recent predecessors.

In what follows, we first discuss the motivation for cohort models of preretirement health. Next, we summarize our analysis sample and statistical framework including an Item Response Theory (IRT) model of health. Empirical findings regarding the stock of health are then provided using first a fixed-effect model with no interactions, and then allowing for sex and cohort interactions with a number of background predictors. The last section offers conclusions and draws out policy implications.

Cohort Models and Methods

Demographers consider each birth cohort unique because it represents the singular intersection of historical time and chronological age. People born just before World War II, for example, enjoyed the benefits of penicillin and antibiotics throughout most of their adult lives, as well as economic growth during their 20s, the decade of labor market entry and family formation, and also during their 40s, their peak earnings period. This cohort also is smaller than subsequent ones, giving it the benefit of little competition from peers. By contrast, the Baby Boom generation was substantially larger than all precursor cohorts, which Easterlin and colleagues (1987) have argued exposed it to extraordinary competition over its life cycle.

Our motivation for examining differences in the stock of health of sequential birth cohorts on the verge of retirement is to dissect the 'influences of the past' (Hobcraft 1982) that shaped their life histories to date and will imprint on the remainder of their lives. Norman Ryder (1965) was among the first to recognize the inherent potential of a birth cohort as 'an agent of social change.' More recently, a number of demographers and sociologists have broadened the cohort concept by embedding it in a life cycle based on the principle:

... that the influence of historic events var(ies) depending on the stage of life at which they are experienced. Tracing cohorts through time is one way to examine the influence of [such] historic events on aggregates of individuals' different ages.

(O'Brien 2000: 124)

Analytic leverage can also be gained by comparing birth cohorts at the same age or life-stage at different points in time. This strategy also may suggest the factors differentiating cohort experiences and the outcome of interest. A cross-sectional array of cohorts by period and by age defies easy analysis, however, because the three temporal dimensions are linearly dependent on the value of the other two (Mason et al. 1973). To achieve identification, various mathematical transformations have been proposed, such as imposing an equivalency assumption on any two adjacent age groups, periods, or cohorts. Others have estimated models in which only two of the three dimensions are assumed to affect the outcome or that the effect of one of the temporal domains is assumed to be proportional to a substantive variable.

These approaches and others (Tarone and Chu 1996; Brewster and Padavic 2000) require strong theoretical assumptions that cannot be easily verified empirically. Moreover, mathematical adjustments for the sake of identification fail to specify the mechanisms by which adjacent age groups or cohorts are differentiated. Because age-*a* is nested within cohort-*c* at

time- t , we estimate fixed-effect models that account for variance between, but not within, cohorts.

Data and Sample

The Health and Retirement Study (HRS) is a nationally representative longitudinal survey of Americans over the age of 50. Supported primarily by the National Institute on Aging (NIA), the study tracks health, assets and liabilities, and patterns of well-being in older households over time.¹ Beginning in 1992, a 90-minute core questionnaire has been administered every two years to age-eligible respondents and their spouses or partners. The initial or 'original' HRS cohort was aged 51–61 when first interviewed in 1992 (along with their spouses of any age). Subsequently, two new cohorts have 'aged-into' the survey. For this research, we focus on the three birth cohorts for whom we have comparable HRS data obtained at the same ages (51–56). We define three 6-year birth cohorts and designate them following the conventions of the HRS as follows. The original HRS (born 1926–41) was first interviewed in 1992, the War Babies (WBS) (born 1942–47) was inducted in 1998, and the EBBS (born 1948–53) was first introduced to the survey in 2004. These three cohorts span eighteen years of accelerated change in nearly all economic and demographic aspects of life.

The HRS ages-in new cohorts every six years. This design feature has the advantage of making the survey a representative sample of the noninstitutionalized² population aged 50 and over in waves where a new cohort ages-in. Furthermore, the 1998 and 2004 waves of the HRS also are representative cross-sections of the new cohort. Thus, the new age-eligible respondents, in combination with extant respondents born in the same years, are representative of their respective birth cohorts.³ In what follows, therefore, we define birth cohorts in terms of year of birth rather than year of first interview.

We also note that the HRS poses one statistical issue common to analyses based on data collected using a multistage cluster design. Specifically, error terms in the HRS are correlated at the household level when two spouses or partners coreside and each participates in the HRS.⁴ Such is the case in the following analysis where we pool male and female respondents, some of whom are spouses/partners. We adjust for clustering at the household level by deriving robust standard errors.

Determinants of Health on the Verge of Retirement

Previous studies have suggested that the notion of health is fruitfully conceptualized as a multidimensional state defined by physical (Fonda and

Herzog 2004), affective (Steffick 2000), functional, and cognitive (Ofstedal et al. 2005) domains associated with pathology (Fisher et al. 2005). In the HRS, all health indicators derive from self-reports rather than performance or clinical assessments, with the exception of cognitive measures. Chronic disease reports are predicated on a health care professional ever having told the respondent that he or she had a specific condition, namely, diabetes, hypertension, cancer, heart disease, cerebrovascular disease (e.g. stroke or transient ischemic attacks, TIAs), arthritis, or respiratory diseases (e.g. asthma or emphysema).⁵ Self-reports of chronic diseases usually yield lower prevalence rates than clinical assessments, although differentials by age, sex, and race are typically of the same order of magnitude for the same chronic conditions collected by the National Health and Nutrition Examination Study (NHANES).⁶ Furthermore, self-reports are less reliable than clinical assessments and are affected by recall bias, length of the recall period, and saliency. On the other hand, only respondents can gauge the overall level of their own health, the degree of difficulty they experience in performing common physical tasks, and the severity of their pain. Both survey and clinical interviews also depend on respondents to communicate their accumulated or episodic health risks, such as smoking or drinking.

The full range of health variables included in our empirical analyses is shown in Table 7-1, arrayed by cohort and sex. With several exceptions, we include only variables that were identical in question wording and response set across the three 'intake' interviews.⁷ A two-tailed ANOVA tests whether the sex-specific means of the two more recent cohorts, WB and EBB, are statistically different from the estimated means of the original HRS cohort.

The first panel of Table 7-1 shows the health index computed for each respondent. This scoring index is usually centered on 0, ranges from -4 to +4, and corresponds to a *Z* score. We discuss the derivation of this index in the next section. In terms of the descriptive data in Table 7-1, the index behaves as one would anticipate: that is, the score for men exceeds that for women in all three cohorts.⁸ Members of the two more recent cohorts have statistically lower scores indicating worse health than those in the original HRS cohort. Most of this decline occurred by the time the WB cohort entered the HRS.

The next panel shows the components used to craft the summary index of health. In spite of advances in diagnosis and surgical and pharmacological treatments, members of both the WB and the EBB cohorts are less likely than the cohort born prior to World War II to evaluate their overall health as 'excellent or very good'. The younger cohorts report more difficulty, on average, than the original HRS respondents in performing most of the ten physical tasks listed, especially the more physically demanding ones such as climbing several flights of stairs without resting, lifting or carrying more

TABLE 7-1 Weighted Descriptive Statistics^a for HRS Sampled Birth Cohorts^b:
Health-Related Variables, HRS, 1992, 1998, and 2004

<i>HRS Cohort Birth Years Yr. aged 51–56</i>	<i>Original HRS born 1936–41 1992 (5,354)</i>		<i>War Babies born 1942–47 1998 (5,078)^c</i>		<i>Early Baby Boomers born 1948–53 2004 (5,030)^c</i>	
	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>
<i>Health Status^d</i>						
Health index mean	0.1706	–0.007	0.01***	–0.222***	–0.04***	–0.262***
<i>Health index components</i>						
SRH: excel./very good	57.41	56.92	53.62**	52.19**	49.92***	50.21***
<i>Problem^e with</i>						
walk 1 blk	3.21	4.6	4.28†	6.86***	5.95***	7.27***
several blks	9.14	13.1	11.98**	18.58***	11.72**	18.84***
1 flight of stairs	5.07	8.68	6.28	10.8*	7.17**	14.18***
several flights	16.63	26.74	19.1†	36.63***	19.96**	38.43***
sit for 2 hrs	14.45	15.37	12.65	18.68**	14.62	19.5***
up from chair	10.62	13.74	21.45***	31.91***	22.15***	33.12***
lift 10 lbs.	5.56	15.78	8.5***	18.92**	8.88***	18.11*
kneel or crouch	13.71	21.82	25.95***	33.75***	26.24***	35.11***
push large object	6.81	14.46	12.19***	22.08***	10.9***	21.65***
arms over head	3.27	4.87	10.74***	12.21***	10.15***	11.68***
Pain	16.76	23.86	22.94***	29.88***	29.09***	32.9***
<i>Demographics</i>						
Sex	47.99	52.01	47.01	52.99	47.8	52.2
Race: white	87.04	85.85	87.74	84.88	81.59***	79.96***
Married/partnered	82.12	73.74	78.26**	69.42***	76.61***	67.83***
HS grad	78.8	77.32	86.19***	83.99***	88.63***	89.08***
Mother is HS grad	37.4	30.18	56.27***	49.36***	66.62***	60.09***
Father is HS grad	27.83	24.59	45.82***	46.35***	55.38***	53.87***
Born in US	88.32	88.34	93.94***	91.7***	88.85	88.72
Mother is alive	52.63	51.73	54.15	51.03	52.98	53.69
Father is alive	22.67	22.05	27.87***	27.48***	28.93***	28.52***
SRH as a child: ex./vg	77.87	75.58	82.03**	79.31**	82.67***	79.14**
SES as a child: ex./vg	70.26	72.17	72.37	74.66†	76.4***	76.24**
<i>Health indicators</i>						
CAGE ^f score > 1	21.41	7.18	23.07	10.67***	28.08***	11.63***
More than 2 drinks day ^g	9.74	2.03	9.08	1.12*	9.55	2.08
Ever smoked	73.51	54.43	68.69**	55.07	65.41***	55.39

7 / Cross-Cohort Differences in Health 143

TABLE 7-1 (continued)

HRS Cohort Birth Years Yr. aged 51–56	Original HRS born 1936–41		War Babies born 1942–47		Early Baby Boomers born 1948–53	
	1992 (5,354)		1998 (5,078) ^c		2004 (5,030) ^c	
	Male	Female	Male	Female	Male	Female
Psychiatric problems	7.87	12.03	17.23***	26.88***	21.26***	27.87***
No chronic conditions	46.73	41.85	46.45	41.21	40.07***	35.35***
Chronic disease count (mean if condition counts > 0)	1.51	1.56	1.57*	1.57	1.68***	1.71***

† p < 0.10; * p < 0.05; ** p < 0.01; *** p < 0.001.

^a Differences of means are *t*-tests for equivalent means, with the sex-specific means for the original HRS cohort as the contrast for comparison to the means for both the WB and EBB cohorts.

^b Sample cohorts based on birth year only.

^c War Babies and Early Boomer cohorts contain fewer respondents than the HRS cohort so are weighted to make the cohort sizes equivalent.

^d Item Response Theory was used to compute the components of the health index, shown in the first panel of the table. In this sample, the health index ranges from –2.62 to 1.03.

^e Because of changes in the response set across waves for the Nagi items, we use the RAND recoded variables that allow for comparison across waves. In 1998 and 2004, responses were for matted as simple ‘yes’ or ‘no’ questions, but in 1992, three yes affirmative responses were offered: ‘yes—a little difficult’, ‘yes—somewhat difficult’, and ‘yes—very difficult’. In the RAND files, those responding ‘yes—a little difficult’ are combined with the ‘no problem’ response.

^f Respondents coded as having a potential drinking problem if they responded positively to more than one of the four standard CAGE items: ever felt should cut down on drinking, ever criticized for drinking, felt bad or guilty about drinking, or ever taken a drink first thing in the morning. A score greater than 1 is used clinically to screen for alcoholism (Mayfield et al. 1974).

^g In 1992, alcohol use was assessed using one question that asked respondents to report the number of drinks they had per day, without a time frame. In 1998 and 2004, respondents were first asked how many days per week did they drink alcohol, on average, over the last three months. For respondents who reported that they had anything to drink in the past ninety days, a follow-up question asked about how many drinks they had on these days. For 1998 and 2004, we used the two items to calculate a weekly drink total, divided by seven, to obtain a daily average.

Source: Authors’ calculations.

than ten pounds, or kneeling/crouching. This downwards drift is evident for both men and women, with the exception of one activity in which the reported level of difficulty for men in the EBB cohort is statistically indistinct from that reported by their counterparts in the HRS cohort.

Relatively more men and women also report having more frequent and severe pain than those born prior to 1942. The last panel of Table 7-1 describes health indicators, and it offers a nuanced picture. More recent cohorts are just as likely to have chronic health problems, and in about the same number, as those in the original HRS cohort (cf. Weir, Chapter 5, this volume).

The second panel of Table 7-1 confirms several demographic trends documented in a variety of statistical publications. The EBB cohort, the most recent of the three cohorts we consider, is proportionately less white but better educated as are the mothers and fathers of the most recent cohort. Changes in marital status also are noteworthy. While a clear majority of both men and women were married at the time of the original HRS interview in 1992, the proportions married or partnered dropped in subsequent cohorts. Men are more likely to be married/partnered than are women in all three cohorts. The persistent male mortality disadvantage, as well as higher probability of men remarrying if divorced, accounts for the lower proportion of married/partnered women at the baseline interview in all cohorts. Both men and women in all 3 cohorts have at least a 50:50 chance of having a living mother while they themselves are in their 50s. In contrast the unadjusted probability of having a living father is only about 0.27 but increasing across cohorts. This change reflects both improvements in male survivorship and delayed age at fathering a child but these trends do not offset the persistent differences in life expectancy that favor women and the normative pattern of women marrying men about three years older than themselves.

Regardless of cohort, most respondents have rosy memories of their childhood health and family status. Nonetheless, more recent cohorts are more likely than those in the original HRS cohort to recall their childhood health as excellent or very good. Only respondents in the EBB cohort recall the socioeconomic status of their families as being very good or better when they were aged 10 or under.

In the last panel of Table 7-1, we document differences in health behaviors across the cohorts, by sex. The first variable is a standard indicator of drinking problem, the CAGE index. Respondents are coded as having a problem with drinking if they reported any three out of four CAGE items: ever felt should cut down on drinking, ever criticized for drinking, felt bad or guilty about drinking, or ever taken a drink first thing in the morning. A score greater than 1 is used clinically to screen for alcoholism (Mayfield et al. 1974; Ewing 1984; Bush et al. 1987; Ewing et al. 1998) The proportion of women considered as having a drinking problem is consistently lower by half of the relative proportion of men who are potential alcoholics, but significantly higher than that of women born in the 1930s. The proportion of men in the WB cohort with a drinking problem is indistinguishable from a comparable proportion of men in the original HRS. By the EBB cohort,

the proportion of both men and women screen for drinking problem was statically distinct from the proportion of their counterparts in the original HRS cohort. Cohort differences in drinking, however, are not reflected in the proportion of men and women who acknowledge having more than two drinks per day. This inconsistency may be associated with a change in question wording after 1992.

Smoking is a leading, but preventable cause of death. The proportion of the three cohorts stating that they have ever smoked trends down for both men and women. The lifetime prevalence of smoking has declined substantially over the three cohorts and at about the same rate. In spite of this, EBB women have a one-third lower risk of having ever smoked compared to men in the same cohort.

Whether because of increasing social acceptance or availability of psychotropic medications, the self-reporting of prior psychiatric problems is higher for women in both of the recent cohorts. Only men in the EBB cohort acknowledge psychiatric problems at a higher rate than the original HRS counterparts.

Creating a Summary Health Index

For the most part, Americans on the verge of retirement present a bimodal health picture. A large group of the respondents aged 51–56 reports few chronic conditions, little pain, no restrictions in activity, or cognitive problems. But a small fraction is in very poor health, with multiple chronic conditions, regular and severe pain, or moderate cognitive impairment. The remaining group indicates some problem on one or more health domains that are neither severe nor negligible.

To summarize all these health indicators succinctly into a single index, we use IRT to construct a score for each individual in the analysis sample (McHorney and Cohen 2000; Dor et al. 2003). This method is used to evaluate the measurement of survey questions and to estimate individuals' scores on a derived index. It postulates that an individual's response to a health question is a function of the individual's unobservable, or latent, 'true' health status, and the characteristics of the health items in question. The item characteristics include the slope (or discrimination) and threshold (or difficulty). The slope is a measure of the steepness of the item curve such that a steeper curve indicates a more reliable item, while the threshold describes the location of the item on the trait scale. The threshold of a binary item corresponds to the item inflection point, the trait value at which the respondents have an equal probability of reporting that they have/do not have the health condition in question. The item parameters are independent of each other (Andrich 1988; Embretson and Reise 2000; Baker 2001).

The IRT computes an overall score for each individual based on his or her responses to the four different health components discussed above and shown in Table 7-2. The scoring of individuals is done in two steps. Item characteristics (slope and thresholds) for each health component are estimated, and these estimates are then used in computing an overall score for each individual. The scale of measurement generally has an arbitrary midpoint of 0, a unit measurement of 1, and values that range from -4 to $+4$, corresponding to that of Z scores (Camilli and Shepard 1994). The mathematical relationship between trait level and the characteristics of the item in a two-parameter IRT model is expressed by the following equation:

$$P(X_{is}) = 1 | \theta_s, \beta_i = \exp(a_i(\theta_s - \beta_i)) / [1 + \exp(a_i(\theta_s - \beta_i))]$$

where X_i is the response of respondent s to item i ; θ_s is the trait level of respondent s ; β_i is the difficulty value/threshold of item i ; and a_i is the discrimination value/slope of item i .

In the analysis below, we use the graded-response model (Ostini and Nering 2006), a generalization of the two-parameter IRT model because the health components are categorical rather than binary. Multiple dichotomizations are used to estimate the item parameters, that is, category one versus categories two and above; categories one and two versus categories three and above; categories one, two, and three versus category/categories four and above, and so on. Each health component has one slope and $k - 1$ between category thresholds, where k corresponds to the number of categories of a health component (Embretson and Reise 2000). Figure 7-1 shows the response pattern for the five-level self-reported health (SRH) to illustrate IRT results. Note first of all that each of the category curves has the same slope. Only the extreme categories of an ordered polytomous variable, such as SRH, are monotonically decreasing or increasing, as shown in Figure 7-1. The interim categories, 'fair', 'good', and 'very good', are also shown in Figure 7-1. Consider the curve for category one, 'poor health'. The threshold value for the contrast of 'poor' versus 'fair' or better SRH is -2.22 . This is the point as which the curves for the two categories indicating 'poor' and 'fair' health intercept. It is also the point on the trait scale at which the probability of choosing 'fair' or higher categories is equiprobable, 0.5 and over; threshold₂ is the point at which the probability of choosing 'good' or higher is 0.5 and over, while threshold₃ and threshold₄ correspond to the probability of choosing 'very good' or higher and the probability of choosing 'excellent', respectively (du Toit 2003). In this sense, the graded-response model is an exercise in curvefitting (Ostini and Nering 2006) across multiple domains.

The components of the health index we use are similar to those in the Short Form Health Survey SF-36 and 18 (Ware et al. 1995). Specific

7 / Cross-Cohort Differences in Health 147

TABLE 7-2 Slope and Threshold IRT^a Parameters for the Four Components of Health Index

<i>IRT Parameters</i>	<i>Self-Reported Health (SRH)</i>	<i>Mobility</i>	<i>Agility</i>	<i>Pain</i>
Slope ^b	1.67	2.98	3.17	2.03
	-0.03	-0.06	-0.06	-0.05
Threshold ^c ₁ ^d	-2.2	-2.08	-2.28	-2.36
	-0.04	-0.03	-0.03	-0.05
Threshold ^e ₂	-1.12	-1.68	-1.83	-1.23
	-0.03	-0.02	-0.02	-0.02
Threshold ₃	0.01	-1.19	-1.48	-0.84
	-0.02	-0.02	-0.02	-0.02
Threshold ₄	1.14	-0.58	-1.13	
	-0.03	-0.01	-0.02	
Threshold ₅			-0.75	
			-0.01	
Threshold ₆			-0.23	
			-0.01	

^a The graded-response model, a generalization of the two-parameter IRT model and which allows for ordered categorical items, is used in estimating the slope and threshold parameters (Embretson and Reise 2000).

^b The slope is a measure of item discrimination and steepness of the curve. The slope indicates how rapidly the probability of choosing a particular response category changes as the trait level increases. Items with larger slopes are more discriminating and more reliable. In the graded-response model, different categories of the same item are assumed to have the same slope (Embretson and Reise 2000).

^c Threshold describes the location of items (in cases of binary items) or the item categories (for categorical items) on the trait scale. Larger (and more positive) thresholds indicate more difficult items/item categories that are further located on the trait scale (Embretson and Reise 2000).

^d Threshold₁ corresponds to the trait level at which the respondents have a probability of 0.5 and higher of choosing any but the lowest response category. This is: fair, good, very good, or excellent (rather than poor) in the case of SRH; three, two, one, or no difficulties (rather than four difficulties) in the case physical mobility; five, four, three, two, one, or no difficulties (rather than six difficulties) in the case of agility; moderate, mild, or no pain (rather than severe pain) in the case of pain rating.

^e Threshold₂ corresponds to the trait level at which the respondents have a probability of 0.5 and higher of choosing any but the lowest two response categories. This is: good, very good, or excellent (rather than poor or fair) in the case of SRH; two, one, or no difficulties (rather than three or four difficulties) in the case of physical mobility; four, three, two, one, or no difficulties (rather than five or six difficulties) in the case of agility; mild or no pain (rather than moderate or severe pain) in the case of pain rating.

Source: Authors' calculations.

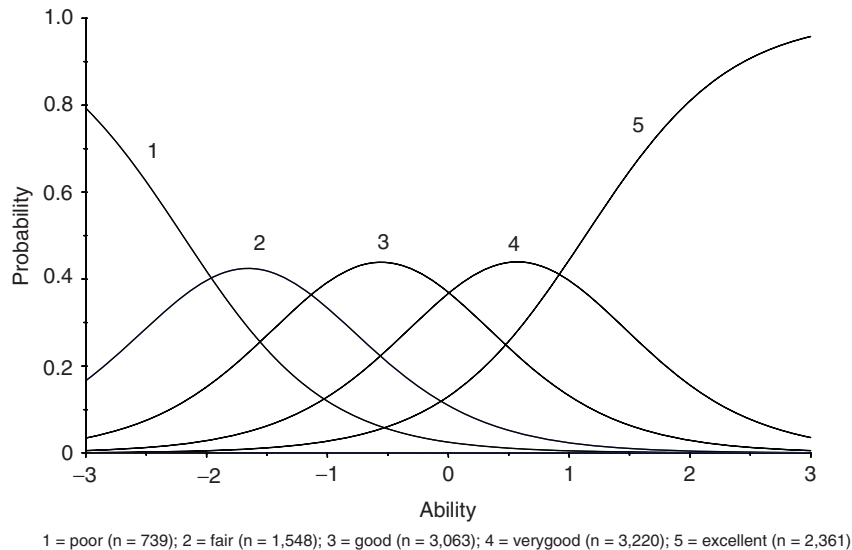


Figure 7-1. Item characteristic curve for self-reported health: HRS respondents Aged 51–56 in 1992, 1998, and 2004. (Source: Authors' calculations.)

components considered are: self-rated health (poor, fair, good, very good, and excellent), perception of physical pain (severe, moderate, mild, and no pain), and difficulty in physical mobility (can't do/has difficulty doing; does not do, and has no difficulty), and difficulty in agility (can't do/has difficulty doing; does not do, and has no difficulty). The physical mobility items include standard items (Nagi 1976) measuring lower body function, including degree of difficulty experienced in: walking several blocks, walking one block, climbing several flights of stairs, climbing one flight of stairs, sitting for about two hours, getting up from a chair after sitting for long periods, stooping, kneeling, or crouching. Upper body agility items include reaching or extending arms above shoulder level, pulling or pushing large objects, and lifting or carrying weights over ten pounds.

Table 7-2 shows the slope and thresholds estimates for all four components of the Summary Health Status index. Each of the components has a relatively large slope parameter, indicating that they are reliable health items. Physical mobility and agility, which have the largest slope estimates of 2.98 and 3.17, respectively, are more related to the unobservable health trait continuum than either SRH or pain. The threshold parameters correspond to the location of various response categories on the health trait scale. The SRH, and to a lesser extent the agility component, however, tap a wider range of trait levels than the other two components. The item

parameters, shown in Table 7-2, are then used to compute an overall health score for each individual. In our pooled sample the health status index ranges from -2.618 to 1.034 , with a median of 0.665 .

Fixed-Effect Models of Health

We model health, measured with the Health Status Index described above, using a fixed-effect OLS specification in which we pool male and female respondents and use all the variables listed in Table 7-1 as right-hand side variables. At ages older than those of respondents included in our analysis, gender differences typically emerge with increasing proportions of men reporting potentially fatal chronic conditions while women report higher levels of disability. The decision to use a pooled sample, rather than separate models for males and females, was based on testing two-way interactions with sex for each of the demographic variables, including the binary variables for WB and EBB, and all of the health indicators. Most of these interaction terms were not statically significant. We retain only those that were significant at $p < 0.10$. These are shown in Table 7-3 in the second set of columns. A positive coefficient for a main effect in Column 1 indicates a direct association with the health index outcome.

Of overall importance is the effect of the WB and EBB cohorts on health, relative to the original HRS cohort. The effects of the binary cohort variables are significant at $p < 0.001$ in both the main effects and the interaction models. In the interaction model, the cohort coefficients are approximately equal in size and negative. Although we use different health outcomes, these findings are consistent with those reported by Weir (Chapter 5, this volume). Adjusting for demographic factors, childhood conditions, and individual health behaviors, the more recent cohorts, on average, have poorer health than the original HRS cohort.

Based on our preliminary analysis, the main effect of sex on the health index is significant and consistently negative, that is, women in the younger cohorts have worse health than men in the same cohort. Compared to white respondents, the predicted health index score is lower for both black and 'other race' respondents, although only the effect of this latter category is significant. Most of these respondents describe themselves as Hispanic, whether US-born or foreign-born.

Either as a main effect or as an interaction with sex, education has a positive effect on health. In both equations contained in Table 7-3, the main effect of education indicates that the rate of return to the Summary Health Index in the gender interaction model is about 0.03 . In the model excluding interactions, each additional year of education only modestly shifts the intercept, but in the interaction model, a high school education,

TABLE 7-3 Fixed Effect Models of Select Demographic and Socioeconomic Variables on IRT-Derived Health Index: HRS Birth Cohorts Interviewed in 1992, 1998, and 2004

<i>Predictors</i>	<i>No Interactions</i>		<i>With Interactions</i>	
	<i>Coeff.</i>	<i>S.E.</i>	<i>Coeff.</i>	<i>S.E.</i>
Cohort (ref. = HRS)				
WB	-0.137***	0.018	-0.211***	0.046
EBB	-0.153***	0.018	-0.214***	0.051
Sex (ref. = Male)				
Female	-0.137***	0.015	-0.396***	0.068
Race (ref. = White)				
Black	-0.016	0.023	-0.011	0.023
Other	-0.105***	0.033	-0.1**	0.033
Marital status (ref. = currently married)				
Union ended	-0.012	0.020	-0.012	0.020
Never married	-0.032	0.039	-0.03	0.039
Education years	0.033***	0.003	0.032***	0.003
Mother's education (ref. = <HS)				
High school	0.03	0.019	0.033†	0.019
> High school	0.045†	0.026	0.045†	0.026
Missing	-0.028	0.035	-0.029	0.034
Father's education (ref. = <HS)				
High school	0.038†	0.020	0.038*	0.020
> High school	0.053*	0.024	0.052*	0.024
Missing	-0.009	0.028	-0.007	0.028
No. of chronic conditions	-0.324***	0.008	-0.298***	0.010
Chronic*WB			-0.063***	0.018
Chronic*EBB			-0.014	0.016
Smoking status (ref. = never smoked)				
Former	-0.022	0.017	-0.05*	0.024
Current	-0.125***	0.020	-0.172***	0.028
Missing	0.037	0.048	0.098	0.095
Former*Female			0.039	0.034
Current*Female			0.075*	0.038
Missing*Female			-0.068	0.108
Drinking problems ^a (ref. = 0)				
1	-0.081***	0.022	-0.08***	0.022
>1	-0.104***	0.023	-0.105***	0.023
Missing	-0.019	0.033	-0.006	0.033
# Alcoholic drinks (ref. = 0)				
0 < # drinks < 1	0.123***	0.017	0.076***	0.023
1 to 2	0.178***	0.025	0.112***	0.031
> 2	0.167***	0.035	0.136***	0.039

7 / Cross-Cohort Differences in Health 151

TABLE 7-3 (continued)

Predictors	No Interactions		With Interactions	
	Coeff.	S.E.	Coeff.	S.E.
<1*Female			0.084**	0.031
1 to 2*Female			0.159***	0.048
> 2*Female			0.116	0.086
Mental health problems? (ref. = none)				
Mental health prob(yes)	-0.344***	0.022	-0.484***	0.043
Mental*Female			0.084*	0.041
Mental*WB			0.134**	0.050
Mental*EBB			0.103*	0.049
Health in childhood (ref. = fair/poor)				
Good	0.155***	0.037	0.102†	0.055
V. good	0.21***	0.035	0.13*	0.052
Excellent	0.335***	0.034	0.221***	0.050
Missing	0.132†	0.076	-0.047	0.088
Good*Female			0.084	0.074
V. good*Female			0.124†	0.070
Excellent*Female			0.193**	0.067
Missing*Female			0.355***	0.083
SES in childhood (ref. = fair/poor)				
Average	0.037*	0.018	0.037*	0.018
Well off	0.04	0.033	0.037	0.033
Missing	0.154*	0.065	0.144*	0.065
Region of residence (ref. = rural south)				
Urban south	0.037	0.028	-0.013	0.033
Rural north	0.04	0.025	-0.024	0.030
Urban north	0.06*	0.024	-0.017	0.029
Not in US	0.118***	0.030	0.051	0.036
Missing	0.055†	0.032	-0.009	0.038
Urban south*WB			0.146*	0.059
Rural north*WB			0.1†	0.054
Urban north*WB			0.135**	0.052
Not in US*WB			0.167*	0.068
Missing*WB			0.297*	0.120
Urban south*EBB			0.008	0.068
Rural north*EBB			0.089	0.059
Urban north*EBB			0.086	0.055
Not in US*EBB			0.047	0.067
Missing*EBB			0.055	0.071
Mother (ref. = deceased)				
Alive	0.06***	0.015	0.058***	0.015
Missing	0.058	0.077	0.059	0.078

(cont.)

TABLE 7-3 (continued)

Predictors	No Interactions		With Interactions	
	Coeff.	S.E.	Coeff.	S.E.
Father (ref. = deceased)				
Alive	0.009	0.017	0.009	0.017
Missing	-0.058	0.056	-0.056	0.057
Constant	-0.352		-0.147	
R^2	0.4		0.41	

Source: Authors' calculations.

^a Respondents coded as having a potential drinking problem if they reported positively to at least two of the four standard CAGE items: ever felt should cut down on drinking, ever criticized for drinking, felt bad or guilty about drinking, or ever taken a drink first thing in the morning. A score greater than 1 is used clinically to screen for alcoholism (Mayfield et al. 1974).

[†] $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

ceteris paribus, shifts the Summary Health Index by 0.237, and a college education by 0.512, the equivalent of scoring in the 40th and 75th percentile on the health index. Similarly, higher levels of parental education, particularly father's education, are significantly and positively associated with better health in midlife. The consistent positive effect of education on health is consistent with the pioneering work by Marmot (2001, 2006) on the SES gradient in all dimensions of health, including mortality.

We now turn to examine how lifestyle and life-cycle factors affect the health index score. Each additional chronic disease reduces the health index by 0.298, one of the strongest effects in Table 7-3.⁹ Relative to the original HRS cohort, those in subsequent cohorts not only have more chronic conditions, but the effect of these is amplified for persons born between 1942 and 1947.

Smoking is recognized as having an enduring negative effect on health. Relative to having never smoked, being a current smoker reduces the health index score by 0.172. Compared to persons with no drinking problems, persons who acknowledge that they had a drinking problem sufficient to elicit a positive response to even one of the four CAGE drinking behaviors reduces the Summary Health Index by 0.08, while those reporting two or more problem behaviors reduce their score by 0.105. Controlling for all other factors in the interaction model, the number of drinks per day has a positive effect on overall health with the greatest gain accruing to those who have two or more drinks per day. Interacting drinks-per-day with being female increase the improvement relative to that for men.¹⁰ Having psychiatric problems has an inverse effect on health relative to those who

do not report such problems. With the advent of new treatment modalities, the health implications for those born after 1942 are attenuated. Because psychiatric problems are often associated with women, the consequences of reporting such problems are attenuated for women but remain negative for men.

Variables that capture early life conditions are included in order to capture what has been described as the 'long reach of childhood' on adult health (Case et al. 2002; Hayward and Gorman 2004). Relative to having poor health as a child, those reporting better early life health also have better health in midlife. There also is a distinct gradient in the childhood effect such that those who consider their health as excellent have a greater return on the health index than those who report even 'very good' health before age 10. There is an added benefit of being a woman in excellent health rather than a comparable man. Note that the coefficient for 'missing' on childhood health is positive and highly significant. Here, as in other variables where we code the effect of having incomplete data, we interpret as an adjustment for statistical noise. Finally note that the effect of childhood socioeconomic status at the same period has minor effects on adult health. It is reasonable to assume that economic status of the family of origin has a positive effect on childhood health, and that the latter variable captures some of the SES effect from childhood.

The last life-cycle variable we include is the region in which the respondent attended elementary school. We do so because at the time when the members of the three cohorts of interest began their elementary education, schools in the Rural South, the reference category, were deemed inferior to those in urban areas in the North. Growing up outside the United States is positively associated with midlife health. The interaction terms included in the second equation provide additional insight. Members of the WBs cohort whose early schooling was in the Urban North have a midlife health advantage compared to members in the original HRS cohort who were first schooled in the Rural South. No effect is discernable for the EBBs regardless where they attended elementary school.

Finally note that having a living, rather than a deceased, mother has a positive effect on midlife health in the two equations shown in Table 7-3. This may indicate a hereditary advantage that accrues to those with long-lived mothers, but more likely it signifies the psychological benefit to have the parent with whom most adult children have the strongest attachment. Most adult children lose their father first and a surviving mother protects adult children from transitioning to 'orphan hood'.

We also analyze the data using multilevel analyses with individual respondents nested within birth years. Table 7-4 provides a summary of the fit of three multilevel models. The first column shows results from a model that includes only the constant. This equation yields the baseline for the

TABLE 7-4 Determinants of Health and Intraclass Correlation Coefficient
(Standard Errors in Parentheses)

	<i>Model 1</i> <i>Constant</i> <i>Only^a</i>	<i>Model 2</i> <i>Micro Variables</i> <i>No Interactions^b</i>	<i>Model 3</i> <i>Micro Variables</i> <i>With Interactions^c</i>
<i>Variance</i>			
Birth year level	0.012	0.004	0.004
	-0.004	-0.001	-0.002
Individual level	0.665	0.409	0.407
	-0.009	-0.006	-0.006
Intraclass correlation coefficient (%)	1.77	0.97	0.97
-2 log-likelihood	26,608.85	21,285.59	21,229.54
Number of observations	10,931	10,931	10,931

^a The model includes a random intercept only.

^b The model includes a random intercept in addition to the microlevel variables shown in Table 7-3.

^c The model includes all the variables in the previous model in addition to interaction terms between sex and each of the following: smoking status, alcoholic drinks, mental problems, and health status during childhood.

Source: Authors' calculations.

variance decomposition. The constant only model attributes almost all of the variance in our data to within birth year differences rather than between birth year differences. In the second column, we summarize the decomposition of variance in the main effects model shown in the first column of Table 7-3. The ICC shown in the third row of Table 7-4 is a measure of the total variance associated with a given model. The ratio of the variance in the main effect model to the baseline variance (0.97/1.77) indicates a 54.8 percent reduction in the initial variance. While the model that includes the interaction terms, column 3 in Table 7-4, does not contribute any additional explained variance, this decomposition is not a test of the significance of specific variables.

To summarize these findings, we derived a health index measure for respondents on the verge of retirement based on four domains: SRH, mobility, agility, and pain. An IRT measurement model is used to predict a Summary Health Status score for each respondent. We find that the health of both recent cohorts, the WBs and the Early Boomers, has deteriorated relative to their HRS counterparts. In addition to own education, the main predictors of good health are years of education, parental schooling, and childhood health. Smoking, heavy drinking, a large number of chronic health problems, a history of psychiatric problems, and attending

elementary schooling in the Rural South have negative consequences for health on the threshold of retirement.

Conclusions and Discussion

Baby Boomers have left a unique imprint on US culture and society in the last sixty years, and it might be anticipated that they will also put their own stamp on retirement, the last phase of the life cycle. Yet because Boomers have not all fully retired, we cannot yet judge how they will fare as retirees. Instead, we focus on how this group compares with prior groups on the verge of retirement, that is, at ages 51–56. Accordingly, this chapter evaluates the stock of health which Early Boomers bring to retirement and compare these to the circumstances of two prior cohorts at the same point in their life cycles.

Using three sets of responses from the Health and Retirement Study, we find some interesting patterns. Overall, the raw evidence indicates that Boomers on the verge of retirement are in poorer health than their counterparts 12 years ago. Using a Summary Health Index designed for this study, we find that those born 1948–53 share health risks with the WB cohort. This suggests that most of the health decline instead began before the late 1940s. A more complex set of health conclusions emerges from the specific SRH measures. Boomers indicate they have relatively more difficulty with a range of everyday physical tasks, but they also report having more pain, more chronic conditions, more drinking and psychiatric problems than their HRS earlier counterparts. This trend portends poorly for the future health of Boomers as they age and incur increasing costs associated with health care and medications. Using our health index, only those at the 75th percentile or higher are likely to be characterized as having good or better health.

We are not the first to signal the eroding health of middle-aged persons in the United States. Using comparative data from sources including the HRS, Banks et al. (2006) obtain similar findings even after controlling for health insurance coverage and other health inputs such as weight, exercise patterns, and other covariates omitted from our models. Moreover, they conclude that adult health in Britain is superior to Americans at ages 50 and older. What should we make of our conclusions in the context of increased public and private health care spending? There are several hypotheses that warrant consideration. Promising ones included: the very act of seeking care for even minor health problems increases awareness of other, seemingly unrelated health problems; the barrage of advertisements for prescription medication increases disease or symptom awareness as much as it encourages care seeking; and changing notions of health in

aging increase intolerance of minor pain, slight loss of stamina, or even minute loss of muscle strength to the extent that younger cohorts are less accepting of physiological changes that are not pathologic. Future research will need to consider unobserved factors that correlate with health, such as cognition, obesity, and use of health care services.

Acknowledgments

The authors are grateful for research support from the Pension Research Council and the Boettner Center for Pensions and Retirement Research at The Wharton School, the Population Aging Research Center (PARC) at the University of Pennsylvania, and NIH/NIA grant no. 1R13AG028231-01.

Notes

¹ For more on the Health and Retirement Study, see <http://hrsonline.isr.umich.edu/>

² At baseline, all members of the entering cohort are community residents. As they age, sampled persons continue to be followed even if they enter a nursing home or other type of facility categorized as an institution. It is unlikely that the entry restriction comprised the representativeness of the sample at baseline because very few of those aged 51–56 are institutionalized in any type of facility.

³ Neither the age-in supplement nor the ‘younger spouse’ component are independently representative of a given birth cohort. The WB and EBB cohorts, however, contain individuals who were previously interviewed as the younger spouses of age-eligible respondents. So, for example, a woman born in 1948 may have been first interviewed in 1998 as the younger spouse of a WB entrant born in 1945. In 1998 her record would have had a zero case weight. Her first cohort interview would have been conducted in 2004 as other members of her birth cohort, the EBB, entered the study. At this time her record would carry its own nonzero person weight.

⁴ That is, the probability of observing one spouse is conditional on the probability of observing the other spouse/partner. This issue does not arise in households where we observe only one respondent at baseline.

⁵ To maintain comparability across cohorts we use the initial or screener question for each disease: ‘Has a doctor ever told you that you had _____?’

⁶ These data cover 1997–2003; see http://www.cdc.gov/nchs/health_data_for_all_ages.htm

⁷ The exceptions are the Nagi (1976) items and questions eliciting number of drinks per days on occasions when respondent drinks. These differences especially are described at the end of Table 7-1.

⁸ Although women have lower mortality risks, women regularly report higher rates of disease and disability.

⁹ The HRS asks if a physician ever told respondents that they ever had hypertension, diabetes, arthritis, lung disease, cancer (other than skin cancer), heart attack, and stroke. The count variable ranges from 0 to 7.

¹⁰ These results are somewhat surprising, although Hurd and McGarry (1995) report similar results using the 1992 wave of HRS.

References

- Andrich, D. (1988). *Rasch Models for Measurement*. Beverly Hills, CA: Sage.
- Baker, F. B. (2001). *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Banks, J., Marmot, M., Oldfield, Z., and Smith, P. (2006). 'Disease and Disadvantage in the United States and in England', *Journal of the American Medical Association*, 295(17): 2037–45.
- Brewster, K. L. and Padavic, I. (2000). 'Change in Gender-Ideology, 1977–1996: The Contributions of Intracohort Change and Population Turnover', *Journal of Marriage and Family*, 62(2): 477–87.
- Bush, B., Shaw, S., Cleary, P., Delbanco, T. L., and Aronson, M. D. (1987). 'Screening for Alcohol Abuse Using the CAGE Questionnaire', *American Journal of Medicine*, 82(2): 231–5.
- Camilli, G. and Shepard L. A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: Sage.
- Case, A., Lubotsky, D., and Paxson, C. (2002). 'Economic Status and Health in Childhood: The Origins of the Gradient', *American Economic Review*, 92(5): 1308–34.
- Dor, A., Sudano, J. J., and Baker, D. W. (2003). 'The Effect of Private Insurance on Measures of Health: Evidence from the Health and Retirement Study', NBER Working Paper 9774.
- du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTIFACT*. Lincolnwood, IL: Scientific Software International.
- Easterlin, R. A. (1987). *Birth and Fortune: The Impact of Numbers on Personal Welfare*, 2nd edn. Chicago, IL: University of Chicago Press.
- Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Ewing, J. A. (1984). 'Detecting Alcoholism. The CAGE Questionnaire', *Journal of the American Medical Association*, 252(14): 1905–7.
- Bradley, K. A., and Burman, M. L. (1998). 'Screening for Alcoholism Using CAGE', *Journal of the American Medical Association*, 280(22): 1904–5.
- Fisher, G. G., Faul, J. D., Weir, D. R., and Wallace, R. B. (2005). 'Documentation of Chronic Disease Measures in the Health and Retirement Study (HRS/AHEAD)', HRS Documentation Report DR-009.
- Fonda, S. and Herzog, A. R. (2004). 'Documentation of Physical Functioning Measures in the Health and Retirement Study and the Asset and Health Dynamics among the Oldest Old Study', HRS Documentation Report DR-008.

- Hayward, M. D. and Gorman, B. K. (2004). 'The Long Arm of Childhood: The Influence of Early-Life Social Conditions on Men's Mortality', *Demography*, 41(1): 87–107.
- Hobcraft, J., Menken, J., and Preston, S. H. (1982). 'Age, Period, and Cohort Effects in Demography: a Review', *Population Index*, 48(1): 4–43.
- Hurd, M. and McGarry, K. (1995). 'Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study', *Journal of Human Resources*, 30(Special Issue on the Health and Retirement Study Supplement): S268–92.
- McHorney, C. A. and Cohen, A. S. (2000). 'Equating Health Status Measures with Item Response Theory: Illustrations with Functional Status Items', *Medical Care*, 38(9): II: 43–59.
- Marmot, M. (2001). 'Economic and Social Determinants of Disease', *Bulletin of the World Health Organization*, 79(10): 988–9.
- (2006). 'Status Syndrome: A Challenge to Medicine', *Journal of the American Medical Association*, 295(11): 1304–7.
- Mason, K. O., Mason, W. M., Winsborough, H. H., and Poole, W. K. (1973). 'Some Methodological Issues in Cohort Analysis of Archival Data', *American Sociological Review*, 38(2): 242–58.
- Mayfield, D., McLeod, G., and Hall, P. (1974). 'The CAGE Questionnaire: Validation of a New Alcoholism Screening Instrument', *American Journal of Psychiatry*, 131(10): 1121–3.
- Nagi, S. Z. (1976). 'An Epidemiology of Disability among Adults in the United States', *Milbank Memorial Fund Quarterly Health and Society*, 54(4): 439–67.
- O'Brien, R. M. (2000). 'Age Period Cohort Characteristic Models', *Social Science Research*, 29(1): 123–39.
- Ofstedal, M. B., Fisher, G. G., and Regula Herzog, A. (2005). 'Documentation of Cognitive Functioning Measures in the Health and Retirement Study', HRS Documentation Report DR-006.
- Ostini, R. and Nering, M. L. (2006). *Polytomous Item Response Theory Models*, Thousand Oaks, CA: Sage.
- Ryder, N. B. (1965). 'The Cohort as a Concept in the Study of Social Change', *American Sociological Review*, 30(6): 843–61.
- Steffick, D. E. (2000). 'Documentation of Affective Functioning Measures in the Health and Retirement Study'.
- Tarone, R. E. and Chu, K. C. (1996). 'Evaluation of Birth Cohort Patterns in Population Disease Rates', *American Journal of Epidemiology*, 143(1): 85–91.
- Ware, J. E. J., Kosinski, M., Bayliss, M. S., McHorney, C. A., Rogers, W. H., and Raczek, A. E. (1995). 'Comparison of Methods for the Scoring and Statistical Analysis of SF-36 Health Profile and Summary Measures: Summary of Results from the Medical Outcomes Study', *Medical Care*, 33(4): AS264–79.