HRS/AHEAD
Family-Related Research

Introduction:

It is probably advantageous for the purpose at hand that, while knowledgeable about data files and documentation, I began this review with only a superficial knowledge of HRS. (I once served on an advisory committee, but that was in the earliest days before the study had taken shape for the first wave).

Review of Existing Family-Related Research. I set as my first task the review of research which used the HRS to study family issues. With the help of our librarian, John Carlson, I obtained copies of articles cited in the bibliography on the HRS website, as well as a few additional ones from an internet search. These articles were then examined in terms of which waves were used, whether longitudinal analysis was included, which family variables were employed, what outcome variables were being examined, and, finally, the type of journals in which they were published and dates of publication.

Search for Relevant Variables. This exploration of the website served two purposes. The first was the obvious need to know what was asked as a basis for evaluating the potential of the data and for seeing whether and how these variables had been used. The second purpose was to learn what this search experience might be like for researchers thinking about using these data.

Download of files from 2000 HRS. I had hoped to undertake some elementary longitudinal analysis, but time ran out. It seemed at least necessary to review the process of obtaining and merging data. Our assistant data librarian, Charlie Fiss, carried this out and provided me with feedback.

Request for comments from potential users. The HRS webpage includes a request for comments, but I thought it might be useful to seek advice on the specific issues raised in this charge. To that end, I sent an email message to nearly 500 members of the PAA who had identified “aging” as one of their areas of interest. The response was underwhelming: only 6 replied, and all but one simply provided information on work in progress or recent papers.

Family-Related Research Literature

HRS has received considerable use for the study of family issues, even if far less than its potential. About 95 articles on family issues were identified, involving over 100 authors. (I tried to eliminate some of the redundancy of working papers which were subsequently published). When viewed in the context of the publications and working papers listed on the website, this represents about one-eighth of the product of HRS. (All numbers are approximations for a sense of scale.) My search for articles was less than exhaustive, meeting
presentations were not included even though these often represent work in progress, and I gathered only about 70 of the papers for review.

About half of this family research focuses on intergenerational exchanges: transfers of time, money, and housing are all well represented, with approximately equal attention to transfers in both directions between generations. Transfers of time and money are the most common topics in this set, with about equal attention to each. Articles not concerned with transfers are distributed roughly evenly over employment and retirement, economic well-being, and health and psychological well-being.

The story with respect to relevant measures of independent variables is rather more complicated. A small minority of articles are primarily concerned with marital status as a family condition relevant to the outcomes. More common is attention to the characteristics of both husbands and wives. As important as these topics are, the data are obviously drawn from respondents’ characteristics that would be measured in any event. What then of the more detailed characteristics measured for children, parents, and siblings? The answer is clear for children: almost all of the questions receive extensive use. One possible exception is that I do not believe that there is any use of the question about whether the coresidence is to help the child or the parents.

In contrast to the adult child questions, however, there is almost no use of the characteristics of siblings. Siblings’ marital status age and financial status were used by one or two of the articles I reviewed. Nonetheless, the sibling network is of considerable theoretical importance for understanding flows between a particular adult child and aging parents. I think, however, that one reason these data haven’t been used is that most of the research on transfers to parents was based on the AHEAD interviews with G3. This leaves a gap in the research that deserves attention. The conditions and effects of receipt of assistance by a parent, and the motivations for, and consequences of, adult children’s providing help are different substantive questions. In addition, transfers weighted differently by family size, and the G2 perspective covers better the too-ill and institutionalized elderly than the G3 sample does. Sibling data from the G2 perspective should have the considerable advantage of setting the richness of data for one child’s behavior (the respondent) in the context of the (more limited) information on the rest of the sibling network.

The questions on parent characteristics were used by a number of the papers. As above, one of the reasons that these are less represented is that most of the analyses of transfers to parents were done with AHEAD—hence “parents” characteristics were those of respondents. I was surprised that I saw nothing that examined the questions on parental illness or nursing home stay before death, and even more puzzled by finding nothing on the frequency of contact questions: from either the parent or adult child sequences.

Turning to the dates of publication, despite a focus on the first wave which I will describe next, interest in the data for family research continues to increase. About one-third more papers have appeared in 2001-02 than in 1999-2000 despite the incomplete record for 2002. Approximately, a quarter of the articles have appeared in journals on aging, about another quarter in economics journals, and the remainder are spread over journals in demography, family, and sociology. This last point is relevant to knowledge of HRS among those who study
families.

Two major limitations of this family-related research stand out. First, the vast majority (about three-quarters) of these papers were based only on the first wave of either HRS or AHEAD, with about half from each. Second, of those that used a later wave, or multiple waves, fewer than 10 were longitudinal analyses. I cannot put this in the context of all research with HRS, but it is clear for family research that the longitudinal investment in these data is very seriously underutilized. The interplay between family, health, and retirement is inherently dynamic, so there is an enormous potential for an expanded agenda on this dimension alone.

Factors Limiting the Number of Researchers Using HRS

Substantive issues:

It is essential to begin this discussion with the obvious point that the very design and purpose of HRS are relevant to only a subset of family researchers—surely much more so than for those studying aging. A large proportion of research on family transitions and interactions focuses on the earlier life-course. But even for those who would very much like to use the data, key missing variables or design decisions may preclude their use.

The advantages of HRS should induce some scholars to broaden their work to take advantage of these data, and this has surely happened. However, while I have no information on the issue, I think family researchers outside of aging, including family demographers, are largely unaware of the nature and potential of this study. The name and purpose are probably well recognized, but most researchers probably assume that the data are not relevant to their interests. As noted above, the journals in which HRS articles have been published may be seldom seen by researchers not already committed to this area. For example, two extremely useful introductions to the data (Soldo and Hill, 1994 and Henretta, et al., 1997) were published in the Journal of Human Resources and the Journal of Gerontology, respectively. This may help explain the lack of papers on intergenerational contact, how much couples like to spend together, and how enjoyable they find this time. Those who may be most likely to use these data probably have no idea that these questions were asked in the HRS.

I will address the search for variables in subsequent sections, but if potential users do examine the available variables they may well find that this would be a powerful resource for their research “if only.” The study content has necessarily been streamlined to focus on the analysis of health and retirement issues. It is obvious that only very limited resources can be allocated to new items in order to expand the population of users. Nonetheless, utilization of of the data could be greatly increased if a few key variables were added. I have not thought this through systematically but will cite two examples that come readily to mind.

The first concerns the importance for health and retirement, and the potential for recruiting family researchers, of a question on the perceived quality of marital relationships. The two questions on “free time” are related to a global assessment of the marriage but address specific and limited dimensions, and they are not very common in the family literature. (These “free time” questions are potentially important, so it is worth noting that they are apparently
asked of subsets in 2000 that yield samples that are too small and will preclude longitudinal analysis.) A more global question on “satisfaction with marriage was asked in 1992, but had little variance and understandably did not appear in subsequent waves. There is a standard question about global relationship quality. While asked somewhat differently various studies, while highly skewed, this measure captures enough variance to be an extremely useful. In the NSFH, this question was phrased: “Taking things all together, how would you describe your marriage?”: with responses ranging from Very Unhappy to Very Happy. This item has similar properties to global measures of health or overall psychological well-being, and considerable construct validity: the risk of disruption between NSFH-1 and NSFH-2 was two-thirds lower in the “happiest” third in contrast to the lowest third at wave 1, net of background variables. Divorce is generally not at issue for this population, but the social support provided by marriage (related to health and psychological well-being) is obviously variable. Changes in couple relationships over the course of retirement, serious health changes and increased dependency, or changes in parental health or parental death, are topics that are central to the purposes of HRS. In addition, however, this variable could greatly expand sociological analysis among family researchers on other topics including longitudinal analyses of changes in couple relationships over the later life course. A similar question on parents’ relationship with their children in the sequence on children could have the same payoff.

Relationships with siblings, frequency of contact, and distance from siblings are also very relevant to the purposes of this study, and have a similar potential for expanding sociological research. Cooperation or conflict among siblings can be a critical aspect of the dynamics of providing help to parents. Of course, I have already indicated that the existing sibling data are yet to be exploited, and it must be noted that the current design precludes broader sociological analysis because siblings are asked about only when there is a living parent. The kinds of analysis permitted by the variables collected on children and siblings are limited by the fact that they were chosen as measures of constraints and needs relevant to intergenerational exchanges. This obviously focuses their use on these outcomes, and few of these variables can be treated as outcomes of interest in their own right. The sole exceptions are employment and income, and residential proximity, and the latter is measured crudely except for coresidence, which is rare. Again this is understandable, but an important context in evaluating why HRS hasn’t been used more for family research.

While endogenous, proximity is a critical variable in the provision of physical help. A limitation that does not seem necessary is the way this variable is measured: the “within 10 miles” format sacrifices a great deal of very important variance. Those who live 20 miles from their parents can provide services on a regular basis almost as well as those who live 9 miles away; and those 100 miles or 1000 miles away obviously cannot. My point is that it is no more expensive to ask “how far” in terms of miles, as the NSFH does, and this enables the analyst to create any categories desired, including the 10-mile measure.

**Documentation**

HRS is exceedingly well documented and this documentation is all made available through a sophisticated website with multiple nested links. This would seem to make evaluation of the potential of the data, and access to it, very simple. Unfortunately, that is not necessarily
so. I had to invest a considerable amount in learning which questions were asked, of whom, and at which waves. (I sometimes found it devilishly hard to figure out the subsample for some of the questions.) The more detailed links provided in the codebooks for wave 3 and beyond are very helpful, and the concordance is a wonderful tool. Nonetheless, I think the central issue here is the lack of “big picture,” and this can be readily fixed.

A potential new user coming to the website has easy access to an enormous amount of detail on a broad array of topics and waves—so much so that it may be so overwhelming for many that their search ends here. It is fairly easy to figure out how to use the codebooks to find specific variables, and to use the concordance to find the waves in which these were asked. But suppose a researcher wants to peruse what was asked across a number of domains, and in which waves these questions were asked. This is a formidable task with the current format. I suspect that difficulties of comparing content across waves have contributed to the extreme focus on the first wave, or only on one wave. Of course, file management may be a primary factor, as discussed in the next section.

The difficulty of searching for content could be solved by providing a much less detailed overview at the start. This overview should be indexed in the tabs at the top of the home webpage, and users should be directed to it within the “Intro/Guide” as well. The documentation for the first wave of AHEAD illustrates the kind of approach I have in mind: a simple listing of specific topics addressed within each domain. While prehistoric compared to the HRS website (in part because of resources and in part because it was created in the internet neolithic), the NSFH documentation provides an example this approach. This example from NSFH-2 is drawn from www.ssc.wisc.edu/nsfh/home.htm.

CAREGIVING AND RECEIVING (ME1-ME55) cr1.002, cr2.002
Long-term physical or mental conditions for all household members | Anyone in household need assistance | R given/received assistance in last 12 months to/from anyone not living with them | R given/received help in last 12 months to/from anyone living with them

RELATIONSHIPS WITH PARENTS (MF1-MF120) cr2.002
Mother, father, spouse/partner's mother, spouse/partner's father, step-parents: Living, age, date died, health (not for step), relationship with, marital status, where live, frequency of contact | Parent types alive at T1: Ever live in R's household since T1 | Own parents: Ever live in household of R's sibs since T1; ever patient in nursing home or long-term care facility since T1 | All parent types: help given and received in past month, hospitalization, physical and mental health, income problems

While creating an abbreviated list such as this takes some time, it is a straightforward task. In order to get some grasp on the content, I had to create crude versions of such lists (using Wave 1) for the sections on demographic characteristics, parents, children, siblings, and transfers. Summaries of this sort serve two very critical functions. First, a researcher can quickly judge whether the questions needed for a particular analysis are in the data. If it appears that they may be, then an examination of the codebook will help clarify that point. Second, being able to see the content in this overview fashion will often stimulate new ideas about potential analyses. Thus such a simple overview would help enormously as a minimal solution to the problem of overwhelming detail at first introduction to the data.

However, I think it would be relatively inexpensive and exceedingly helpful to go beyond
the summary suggested above. I realize that the following is only a first approximation that it undoubtedly overlooks many complications, but what I have in mind is a matrix. Each question sequence (as in the above example) asked in any wave would constitute a row, with columns indicating the waves in which these were asked. The simple example below, based on the concordance for children’s marital status, illustrates what I have in mind.

<table>
<thead>
<tr>
<th></th>
<th>92</th>
<th>94</th>
<th>96</th>
<th>98</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Marital change</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage begin/end</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This might even be elaborated by linking the cells to the codebooks, so that clicking on any cell in the matrix would take the user to the appropriate location in the codebook for those variables and wave, and clicking on the variable name at the left would engage the concordance for the variable.
Data Structure

The inherent statistical and file management complexity of longitudinal analysis partially accounts for why there are so few longitudinal analyses. That having been said, the data may be more complex, or appear to be so, than is necessary. Many users who have fresh ideas and the necessary statistical skills may, nonetheless, be unfamiliar with complex data files—and this may be especially so for the broader family research community who could potentially attracted to HRS. While merging the many files is straightforward, we had no difficulty, the appearance of complexity may deter new users, especially from matching across waves once person numbers enter the picture. While having the data broken up this way makes it easier for those with older PCs, it does not provide much advantage for creating data extracts even though unwanted sections can be excluded. Merging all of the files at the respondent level results in about 7000 columns of data. NSFH-1 and NSFH-2 are both about 8000 columns for the main respondent files and we have had no complaints from users on this score. Hence things might seem less daunting to new users if there were only seven files, one for each of the file levels. Indeed it would be worthwhile to append the household file to each respondent file.

Again, this is outside my expertise, and I have no idea how expensive it would be to make all of the waves, and longitudinal files, available through a data extraction facility. It would undoubtedly be costly. Nonetheless, I think even a very substantial expenditure would be well worthwhile if it were to multiply the payoff from this very large investment. As it is, NIA is already supporting these activities, duplicated over and over again, as fresh code is written by numerous researchers to achieve the same end.

The difficulty of matching across waves, real or perceived, is undoubtedly a major factor in the limited number of longitudinal analyses. This includes, of course, both the technical issues of matching files and documentation issues relating to differences in questions (or samples asked those questions) across waves or between HRS and AHEAD that result from the combination of these originally separate surveys. It may be instructive to note that subsequent improvements in documentation and data structure will go unnoticed by those who give up on the basis of experience with the early waves. Further, such experience may help create a folklore that the data are difficult to use, discouraging others from even approaching the website. I know that concerns such as these were a major consideration at the October 10-11 Family Data Workshop, and that efforts are underway to improve the documentation and structure of the data files.

Conclusion

In sum, there has been more use of HRS by family researchers than we might have thought, and papers have addressed the full range of outcome variables central to the study. Nonetheless, the community of researchers taking advantage of this wealth of well-designed study could be expanded, and the potential for longitudinal research is largely unrealized.

While adding a few variables would increase costs, the marginal payoff from the existing investment could be large. Similarly, simplifying the documentation and file format would
increase use of the data for family-related research, as it would for all topics. This is most important in order to facilitate the use of multiple waves and longitudinal analyses.

Further, family researchers may well be largely unaware of the potential of HRS for their interests, and steps can be made to remedy this. I understand that the Family Data Workshop addressed this issue, and is preparing a list of unexplored, or under researched, topics, and that they are also seeking creative ways to distribute this information. I would encourage two extensions. First, a summary overview of a wave of the data (as suggested above) should be included with the examples of research topics. This raises the probability of potential researchers finding a match with their research in progress, and they may well think of additional analyses as well. This mailing should distributed broadly beyond the usual suspects. Second, presentations about HRS should be considered for a wider set of organizations than those that seem most obvious. I expect this would be more effective in the context of general sessions that include other data sets than it would in stand-alone workshops, but both would likely be useful.

Hopefully, longitudinal analysis will be greatly facilitated for all uses of HRS. Nonetheless, at a smaller scale, the one change that might have the greatest payoff for family-related research could well be the creation of matched files across waves on the intergenerational exchange information. This is a central topic, one that received the most attention in the literature I reviewed, and one that has not yet been analyzed in terms of the dynamic interplay with health and retirement.

I will simply conclude this review by noting that I had no plans to ever use the HRS. I will very likely do so.
Bibliography


Byrne, David, Michelle Goeree, and Bridget Hiedemann "Long-Term Care, Formal Home Health Care, and Informal Care," (2002).


Henretta, John C., Emily Grundy, and Susan Harris, "Socioeconomic Differences in Having Living Parents and Children: A U.S.-British Comparison of Middle-Aged Women," *Journal of*


Kutty, Nandinee K., "The Production of Functionality by the Elderly: A Household Production


Silverstein, Merril and Joseph J. Angelelli, "Older Parents' Expectations of Moving Closer to


