

The Value of the Health and Retirement Study for Health Economics Research

Courtney C. Coile
Wellesley College

Nicole Maestas
Harvard Medical School

June 2016

Acknowledgements:

This paper was prepared at the request of the Health and Retirement Study's Data Monitoring Committee. We are grateful to Katherine Baicker for her helpful suggestions on the paper and to the following individuals for generously sharing their views on the HRS: Jason Abaluck, John Cawley, David Cutler, Angus Deaton, Carlos Dobkin, Jonathan Gruber, Amanda Kowalski, Adriana Lleras-Muney, Lee Lockwood, Ellen Meara, Matthew Notowidigdo, Mark Shepard, Dan Silverman, Kosali Simon, Jonathan Skinner, Benjamin Sommers, Courtney Van Houtven, and Heidi Williams. Comments shared with us by these individuals and used in this paper have in some cases been minimally edited to protect the anonymity of the author or for clarity.

Since the Health and Retirement Study (HRS) was launched in 1992, the study has been used in over two thousand papers and publications on the subjects of health care, health conditions and status, or health insurance, according to the HRS online bibliography.¹ Yet it would be premature to conclude that the HRS has been highly useful for health economics research on the basis of this evidence alone. For one thing, this bibliography includes publications in all disciplines, not just economics. As reported in Table 1, the number of these papers that were published in the fifty most highly-ranked economics journals and in selected well-known health economics journals is far fewer, at 111 papers. While this count relies on a somewhat arbitrary choice of journals and misses some key health economics articles, such as those published in medical journals, this relatively low figure nonetheless raises several important questions. Are health economists using the HRS as much as one might expect? If not, what factors may be contributing to this low usage, and are there changes to the survey that could help to reverse this phenomenon?

To help us grapple with these questions, we sought input from leading health economists by fielding a short survey. We first describe the results of this effort below. We then offer some thoughts regarding the future direction of health economics research and the advantages and disadvantages of the HRS in this context. We conclude by making some suggestions that might help to enhance the value of the HRS for research in health economics.

¹ <http://hrsonline.isr.umich.edu/index.php?p=biblio>. A search conducted on May 10, 2016 using the categories “health care,” “health conditions and status,” and “Medicare/Medicaid/Health Insurance” and a retrieval threshold value of 1 (which will tend to generate a more expansive list of publications) generated a list of 2,149 papers and publications.

Survey of Health Economists

We sent a short survey to a number of leading health economists at U.S. universities. We received responses from 18 of the 29 we contacted, for a 62 percent response rate. The survey questions are listed in Table 2.

We first asked whether survey respondents believe that they and other health economists are relatively familiar with the HRS and HRS-linked data sets. Virtually all respondents said yes (16 of the 18; 1 said no and 1 gave no answer), although 5 of these went on to express some reservations about the extent of their own or others' knowledge of the survey. For example, respondents said they had a "vague sense" or were "somewhat familiar" with the included variables and that there were "many linked data sets... that people are unfamiliar with." There were also several cases where respondents suggested survey improvements that have already been undertaken, such as the addition of biomarkers or the creation of harmonized international data. Table 3 provides a list of all comments that reflect some lack of confidence or knowledge about the contents of the HRS.

Next, we asked respondents whether they found the HRS to be useful for their own work. Two-thirds (12 of 18) answered yes, while the remainder (6) said no. Among those saying no, respondents cited: their tendency to work with administrative data sets (2); their need for a larger sample size (2), younger survey respondents (3) or earlier life information (1); and the changes over time in the survey instrument (1). We return to some of these issues below.

Our next question asked respondents to elaborate on the ways in which they see the HRS as being well-suited for producing high-quality research on key health economics questions. Here respondents were virtually unanimous in praising the HRS for the breadth

of information collected, the longitudinal nature of the data, and the linkages to administrative data sets such as Medicare claims data, biomarkers, and geographic data. One respondent called it the “premier social science panel study, period” for its “uncommonly rich socio-economic data with high quality and detailed information about health, including links to claims data from Medicare... its combined length (t) and size (n) are almost unmatched in social science.” A number of respondents stated that combining these aspects in a single data set is the HRS’ greatest strength:

“[The HRS is the] only national data set with excellent work/wealth/health questions [that is] longitudinal [and] can be linked to state data to take advantage of state level instrumental variables.”

“The integration of the data from all of these domains, including administrative data, gives an unusually complete view of the economic forces behind health-related outcomes and behaviors.”

Several respondents also mentioned the value of having data on both members of a couple and their children. A full set of these responses is available in Table 4.

Respondents were also queried as to the ways in which they see the HRS as being ill-suited for health economics research, with the aim of gaining insight into why the HRS has not been used more for this type of work. These responses are compiled in Table 5. Most respondents referred to one or more of the following factors: 1) trends in research questions away from areas where the HRS has a comparative advantage; 2) rising use of large administrative data sets instead of survey data; 3) reliance on specific empirical strategies that may pose challenges for the HRS; and 4) the HRS’ relatively small sample size. The following comment illustrates how these various factors may be intertwined:

“I think the key barrier to the HRS being used is topical. Most of health economics now focuses on health care system questions -- health insurance, provider choice and behavior, costs. For these, large administrative datasets (e.g., Medicare data) are much better suited. In comparison, HRS has two weaknesses: (1) it is a much smaller sample, and (2) the health care use variables in the public data are self-reported. Issue #2 can be overcome by using the restricted Medicare link, but #1 cannot. A lot of papers use variation that affects a small subset of the population. This is possible when you start with 40 million observations but not when you start with 30,000.”

Several other comments support the notion that these concerns are often interrelated and provide more detail regarding the empirical strategies issue:

“Identification strategies... such as RD [regression discontinuity] and panel methods require much larger samples than the HRS has to provide precise estimates. I typically use administrative data sets with sample sizes orders of magnitude larger than the HRS.”

“Sample sizes [are] inadequate for research using geographic policy variation.”

Overall, sample size was by far the most frequently mentioned concern, with 11 respondents making reference to it. Use of administrative data received 4 mentions.

Another common theme was the difficulty of accessing restricted data, an issue raised by five respondents. One respondent called the process of obtaining restricted data “incredibly complicated” and referred to the HRS as “more security-conscious” than other sensitive data sources. Another respondent spoke of the time cost of obtaining access to the restricted data, while several others pointed to the need to have federal funding or access to a research data center (RDC) as important obstacles, particularly for graduate students: “I think that a lot of my students are deterred from using the HRS because state-level identifiers require RDC access, and there was no RDC [near] our campus.”

A similar number of respondents spoke about the HRS’ focus on the elderly and near-elderly population as a potential drawback. One respondent noted that the survey’s

age limitation “is, of course, by design, but does limit what one can study with the HRS,” while other respondents said “we would have liked to have had [a] broader age range of prime-age adults” and “picking up new cohorts even earlier would be useful.” Finally, two respondents referred to the HRS’ biannual frequency as insufficient in some cases, while two other respondents said the survey was of limited value for studying insurance issues either because most of the elderly are covered (so there are few uninsured) or because there is insufficient administrative data available on health insurance factors.

The final survey question asked how the HRS might be changed to make it more useful for health economics research, including whether there was any existing health-related content in the survey that respondents saw as less useful. Respondents had a number of interesting suggestions. The first set of suggestions concerned increasing awareness of the HRS and access to the public data. Two respondents suggested producing a handout or other materials providing an overview of the HRS and highlighting its strengths relative to other data sets. Two other respondents suggested making it easier for people to use the public data, either by making the RAND version of the data even easier to use or allowing users to create cross-tabs or other summary measures of the public data without investing deeply in learning the HRS.

A second recommendation was that the HRS should make it easier for researchers to access restricted data such as the Medicare claims data, Social Security records, and geographic identifiers. Three respondents raised this issue, including one who suggested allowing exceptions to the policy of requiring federal funding to access the restricted data and another who suggested allowing simultaneous access to geographic data and Social Security data.

Adding administrative data on health insurance to the HRS was another common suggestion, although there were some differences regarding the specifics. One respondent suggested collecting administrative data on health insurance plans, presumably much in the same way as was done for employer-provided pension plans. Another recommended adding “linkage ability to private insurance claims,” especially for members of the under-65 population who are not observed in the linked Medicare claims. A third respondent suggested that there could be “matching of hospital and nursing home characteristics or other summary measures of health care utilization,” presumably by geographic location such as state, metropolitan statistical area (MSA), or hospital referral region (HRR).

Several respondents also made suggestions regarding how to improve the health measures in the survey. One respondent suggested reducing the amount of information collected on self-reports of chronic conditions, on the grounds that these data are unreliable due to the high frequency of individuals who have undiagnosed conditions or misreport their conditions, and also that the numbers of individuals with any given condition are too small to be useful for analysis. This respondent suggests that the HRS instead focus on collecting “more [and] better reports of: pain, depression, prescription drug use, cognitive impairment, and other limitation of roles.” This could be accomplished, the respondent proposes, by collecting data on back pain at every wave, by asking everyone about antidepressant use and prescription pain relief (eliminating skip patterns), by collecting photographic evidence of prescriptions, and by administering a better screen for depression. Two other respondents also weighed in on this topic, recommending collecting more complete health information for those under age 65 – for example, complete prescription drug use data – and richer biomarker data.

A number of respondents also suggested making improvements to the HRS' economic, family, and early life variables. One respondent noted that wealth and pension data are noisy, while another asked for more broad-based consumption measures, possibly derived by linking credit report data to the survey. Other respondents asked for more information on work productivity, on adult children of survey respondents, and on the early labor market experiences of respondents. Finally, two respondents indicated that it would be useful to raise the frequency of the survey, either by doing so for a subset of respondents or by collecting some of the modules more often.

Analysis and Recommendations

Through our survey of prominent health economists, we have identified a number of challenges to the use of the HRS for health economics research as well as some opportunities for strengthening the HRS. We begin with some general thoughts before turning to specific recommendations.

In brief, the challenges identified in our survey are: 1) a lack of familiarity with the HRS; 2) the difficulty of accessing restricted data; 3) trends in health economics research questions and methods; 4) sample size; and 5) specific concerns about some economic and health variables. Which of these areas offers scope for improvement? The first issue, a lack of familiarity with all of what the HRS has to offer, seems like an area where progress can certainly be made, and we include a number of specific suggestions to that end below. The difficulty of accessing restricted data is more challenging, in that protecting the privacy of HRS respondents is obviously critically important, but there may be policies and practices affecting restricted data access that merit reexamination. It may also be possible to create

new composite variables that are calculated using the restricted data and make them part of the public data; we expand on this below.

What should one make of the concern that health economics research is moving in a direction that focuses on questions the HRS was not designed to answer and favors the use of large administrative data sets rather than survey data? First, it is important to acknowledge that a single survey, even one as rich as the HRS, cannot be the ideal data set for every research question. Nonetheless, there may be some strategic opportunities to augment the survey in ways that increase its value. The recent guidance from the National Institutes of Health (NIH) stating that *health* should be the outcome of interest in NIH-supported health economics research represents an opportunity for researchers to rediscover the value of the HRS, with its rich health outcomes data, including self-reported and biomedical measures. The HRS may seem relatively less well-positioned to address other key questions in health economics, such as issues related to costs and competition in health care markets. Yet where those issues intersect with areas of strength for the HRS, there are opportunities – for example, to explore how health care market conditions affect health outcomes. For this to be possible, the HRS would need to be augmented with some contextual data on health care markets, linked to individuals by geography; naturally any such data would need to be provided in the form of composite variables that could not be easily decomposed, to prevent the possibility of re-identification of respondents based on their geographic location. We discuss this suggestion in more detail below.

Taking a broader perspective, the HRS has been tremendously valuable for the estimation of structural models of retirement and saving, work done by researchers in the economics of aging that relies critically on the availability of restricted Social Security

earnings records and employer-provided pension data as well as HRS survey data. Some health economics research also employs a structural approach, though the goal is typically to identify some parameter of interest rather than to specify the entire utility maximization problem. But much of the best work in the field relies on identifying a good source of exogenous variation – for example, arising from state-level policy differences or regional differences in health care market conditions. The HRS maximizes its utility to health economics researchers when they are able to link that variation to all of the other data in the HRS, including the many variables that cannot be found elsewhere – as one of our survey respondents put it, “there are certain types of things [in the HRS] that would never be measured in administrative data.” In sum, it may be that the best way to keep the HRS relevant for health economics research is to facilitate access to restricted data as much as feasible, given privacy concerns, and also to strategically augment the survey with variables derived from the restricted data, which give researchers some ability to exploit exogenous variation in conditions without having direct access to the restricted data.

Sample size was the concern mentioned most often by our survey respondents. Unfortunately, this concern may also be the most difficult to address, absent a substantial increase in funding for the survey. The HRS simply does not offer a comparative advantage in the exploration of phenomena that only affect a small minority of the population or in the estimation of models where hundreds of thousands of respondents would be needed to uncover a statistically significant effect, and this is unlikely to change unless the sample size were to be increased dramatically.

We turn next to our specific suggestions for increasing familiarity with the HRS, enhancing restricted data access, and improving health and economic measures.

1. Increase Familiarity with the HRS

We believe that several specific steps could increase researchers' familiarity with the HRS and its contents, and that investing a small amount of money in this area could yield high returns. Our specific recommendations are:

- Update the HRS website by: adding content that compares the HRS to other data sets and highlights its strengths; providing more information about restricted data sets (e.g., by incorporating them into the concordance or providing a more detailed overview of their content on the data products page); and replacing outdated content (e.g., "Growing Older in America: The Health and Retirement Study").
- Create an email mailing list from HRS registrations and use it to share survey updates and other select information (e.g., could have a "Did You Know?" feature that highlights lesser-known aspects of the survey).
- Organize a special issue of a journal for papers using the HRS for health economics research.
- Offer a small grant program for health economics research projects using the HRS. The NIA might be an ideal partner for this, given its new focus on health as the outcome of interest and the HRS's rich collection of health measures. The grant program could potentially be targeted at graduate students and young scholars, who are making key investments in data sets at this point in their career and may be more likely to keep working with those data sets that they learn now.
- Partner with a high-profile research organization like the Kaiser Family Foundation, Urban Institute, or RAND to produce short, eye-catching content making use of the

HRS, such as graphs showing of the variation in health outcomes by sex and age group, across census regions, by cohort, and by race/ethnicity. These partners would likely have their own distribution networks, allowing them to post this content on their website and share it with interested researchers, journalists, and others via email blasts. These descriptive snapshots may inspire some researchers to use the HRS to investigate them further.

- Leverage the training sessions at the annual RAND Summer Institute to have more impact in the health economics arena. The “how-to” workshops on using the HRS and HRS family of data sets could be videotaped and posted on the HRS website. A session could be added on using the different kinds of health measures in the HRS.
- Step up outreach to the health economics community via the biennial ASHEcon conference. There could be a booth on the HRS (or RAND HRS), as there is at other meetings like the Gerontological Society or Population Association of America; there could also be HRS-themed sessions featuring research using the HRS.

2. Expand Access to Restricted Data

The HRS has invested heavily in collecting many kinds of restricted data, and it is important to consider how to best leverage this substantial investment. We suggest the HRS review its current policies and procedures regarding restricted data access to see if there are changes that could increase access while maintaining the appropriate degree of protection against re-identification of de-identified data. Making it easier for potential users to understand what information is included in the restricted data sets (as recommended above) would be a good first step. An important second step would be to

reduce the amount of time it takes to approve restricted data applications. Lengthy approval times deter researchers from using the HRS. The linked Medicare claims data—of particular interest to health economists—take even longer to obtain because of the larger number of institutions that must sign off. Data access delays are particularly costly for graduate students and may dissuade them from using the HRS altogether. As a result, new generations of scholars will be less familiar with the data set and less likely to use it. Finally, data access barriers impede replicability efforts. There is increasing pressure on social scientists to ensure that published findings can be replicated, and as a result, a growing number of funders (including, recently, NIH) and scientific journals have incorporated transparency criteria into their proposal and manuscript review standards. As investigators across the social sciences adapt to these higher standards and incorporate new tracking tools into their research workflows, they will increasingly be expected to make their analysis data files (not just their statistical programs) available to other scientists.

We recommend several steps to increase access to the restricted data products. Additional staff could be hired to help process applications more quickly. The processes themselves could be streamlined so that applications from previously approved (and thus “known”) investigators are expedited. Simplified re-use mechanisms or institutional umbrella agreements could be used more widely, rather than only with select institutions. Further development of the remote access enclave option would reduce the risks associated with having physical copies of restricted-use data reside at many institutions. The policy forbidding simultaneous access to geographic and other restricted data could be reconsidered for projects conducted inside the enclave. Finally, as open science initiatives

gain further momentum, data providers will be called upon to tailor their policies and practices to *realistic* re-identification risks so that the analysis files of one research team can be safely re-examined by other research teams.

Another approach to improving access to restricted data is to increase the number of variables included in the public release files that are derived from restricted data. This provides some of the benefits of restricted data access to researchers who do not have access to the underlying confidential data. Social Security wealth is a good example of something like this that is already available as a researcher contribution file and is now included in the RAND HRS data release. This measure is likely to be of interest to many researchers, but in most cases, having an estimate calculated by someone else is sufficient and it is not necessary to obtain the detailed Social Security earnings in order to make one's own calculation. In the health area, there are likely to be similar cases where Medicare claims data or biomarker data could be aggregated to a level that eliminates concerns about confidentiality – for example, total Medicare expenditures in a number of broad categories (inpatient, outpatient, prescription drug) or health risk indices compiled from biomarker data. Data about regional health care markets (e.g., the approximately 300 hospital referral regions) – including cost, quality, and utilization measures – would be another extremely valuable addition, if this could be aggregated in a sensible way that maintained some variation in these market measures while preventing re-identification of the respondent's geographic location. On this note, a number of researchers have obtained NIA funding over the years to create contextual data characterizing aspects of the environment in which respondents live and work that bear on health outcomes. These aspects include the built environment (relevant for studies of disability), resources for

medical care, the economic environment, the natural environment, public health risks, local governance, and so forth. We suggest creating incentives to encourage researchers to deposit their contextual data files with HRS so they can be made available to other users (for primary analysis and replication). The contextual data could be linkable to the public data files on the basis of the anonymous individual identifiers, rather than the restricted geographic codes, which would reduce the need for providing the actual geographic identifiers to researchers. Re-identification risk could be controlled through a simplified DUA process. HRS could in turn consider whether any of the contributed contextual variables are candidates for inclusion in the public release (or the RAND HRS) as derived variables.

3. Suggestions for health and economic variables

It is worth exploring possible additions to HRS data collection efforts that might strengthen its value for health economics research, while acknowledging that concerns about the additional resources required would also need to be considered before adopting these suggestions. First, the HRS could collect administrative data on employer-sponsored and other private health insurance plans, much as has been done for private pension plans. This could be compiled in a new restricted data set, although to maximize the value of this data collection, it would be ideal to use the information from plan descriptions to construct common measures likely to be of interest to many researchers and to include these measures on the public version of the data. These might include things like premiums, deductibles, copayments, coinsurance, and out-of-pocket maximums, as well as information

about the type of health insurance plan, breadth of the provider network, coverage for drugs and other services that are sometimes carved out, and so on.

In addition to collecting health plan data, the HRS would benefit from the inclusion of more information on health care utilization before age 65. Administrative records would be ideal, but it is difficult to imagine how HRS would collect these given the decentralized nature of the U.S. healthcare system. One approach would be to create a process by which researchers could apply to link administrative data they have obtained access to with the HRS. Increasingly, researchers have developed relationships with commercial health insurers, with state Medicaid agencies, and with growing number of states that are building all-payer claims databases. If HRS opens the door to researcher-initiated external linkages, researchers would do the legwork of building relationships with and negotiating data use agreements with external data providers. The current effort by researchers to link the HRS to state unemployment insurance wage records in the Census Bureau's Longitudinal Employer-Household Data may offer lessons that could inform such a process.

While HRS should continue tracking self-reported chronic diseases to allow measurement of health histories and the onset of chronic diseases, it should put more emphasis on measurement of functional capacity, by collecting better measures of pain, depression, cognitive impairment, and other functional health limitations. This would parallel evolving frameworks in geriatric research and medical care that prioritize improving function and quality of life over treatment of disease. A benefit of this approach is that it would be possible to track *recovery* of functional capacity in the HRS, and not just the onset of new diseases. Finally, it would also be beneficial to collect more complete self-

reported information on prescription drug use before age 65, as several of our survey respondents recommended.

In terms of economic variables, the availability of such data, particularly in combination with rich health information, is a great strength of the HRS. Although collecting feedback on the economic variables was not a major focus of our efforts, the HRS should continue to look for ways to improve this part of the survey, including providing better measures of consumption, more information on the adult children of HRS respondents, and more information on early life experiences. HRS' recent effort to collect life history data is commendable.

In conclusion, the HRS offers many strengths for health economics research and is greatly valued by health economists. However, there are significant obstacles that currently prevent it from being used more widely. In this paper, we offer some thoughts on what these obstacles are and how they might be overcome through a combination of changes to the HRS and other actions to increase awareness of and access to the survey. We hope that this analysis may help to make the HRS even more relevant for health economics research in the years to come.

Table 1: HRS-Based Health Economics Papers in Top 50 Economics Journals and Selected Health Economics Journals

Journal Rank	Journal Name	Impact Factor	Number of HRS Papers
1	The Quarterly Journal of Economics	77.74	1
2	Journal of Economic Literature	72.09	2
3	Journal of Political Economy	67.50	3
4	Econometrica	65.50	1
8	Journal of Finance	38.93	1
9	Journal of Economic Perspectives	38.68	3
10	Economic Policy	36.27	1
11	American Economic Review	34.80	6
15	Journal of Labor Economics	30.88	1
16	Journal of Econometrics	27.84	6
17	Journal of the European Economic Association	25.25	1
19	Economic Journal	24.55	1
20	Journal of Applied Econometrics	23.75	3
23	The Review of Economics and Statistics	23.07	4

27	Journal of Public Economics	20.63	4
30	American Economic Journal: Applied Economics	19.70	1
32	Journal of Environmental Economics and Management	19.19	2
39	European Economic Review	17.09	1
40	Journal of Human Resources	16.68	12
41	International Economic Review	16.59	3
45	Journal of Risk and Uncertainty	15.10	3
59	Journal of Health Economics	12.89	18
180	Health Economics	4.90	11
355	Forum for Health Economics and Policy	2.21	7
--	Health Affairs		15
<hr/>			
	Top 50 Journals		60
	Selected Health Journals		51
	Top 50 and Selected Journals		111
<hr/>			

Notes:

Table displays counts of papers from a 5/10/16 search of the HRS online bibliography (see paper text for search details) that were published in top 50 economics journals (according to IDEAS/RePEc Simple Impact Factors for Journals, as of 5/10/16) and in selected health economics journals.

Table 2: Survey Questions

1. Do you believe that you and other health economists are relatively familiar with the HRS (especially its health-related content) and with HRS-linked data sets?
 2. Do you find the HRS to be useful for your own work? Why or why not?
 3. More generally, in what ways do you see the HRS as being well-suited for producing high-quality research on key health economics questions?
 4. In what ways do you see the HRS as being ill-suited for such research?
 5. Do you have any thoughts about how the HRS might be changed to make it more useful for health economics research? Is there any existing health-related content in the HRS that you see as not being particularly useful?
-

Table 3: Survey Responses Suggesting Imperfect Knowledge of HRS

“I am aware of its existence and have a vague sense of the variables included.”

“I’d say I’m somewhat familiar with the variables available in the HRS.”

“Yes I do believe I am relatively familiar with the HRS, as are other health economists (to varying degrees). Products like the concordance website are VERY helpful in that regard.”

“I believe health economists are generally familiar with the HRS but might not know its relative strengths over other data sets.”

“Yes but there are many linked data sets and differences in cohorts/surveys that people are unfamiliar with.”

“I was only exposed to it via my work on Social Security/retirement. Otherwise, I wouldn’t know much about it, and definitely not that it was linked to Social Security/Medicare data.”

“Yes they are aware of linked data sets. At least Medicare link and geographical link, likely not SSA link.”

“I don’t know what ‘HRS-linked data sets’ are.”

“How timely is it? I don’t know what the turnaround time is on it. I also don’t know – can you use it to do state-level analyses?”

“Is the addition of biomarkers prohibitively expensive?”

“Linking of parents, children and grandparents would also be fantastic. I think some of this is done but I am not sure.”

“Harmonization with other similar data on other countries would be most welcome.”

**Table 4: Survey Responses on Ways in Which the HRS is Well-Suited
for Health Economics Research**

“In my view, it is the premier social science panel study, period. For those interested in health it is especially useful as it combines uncommonly rich socio-economic data with high quality and detailed information about health, including links to claims data from Medicare. Moreover, it has come to include quite good information about consumption expenditure and about subjective data like expectations and "deep" preferences. Finally its combined length (t) and size (n) are almost unmatched in social science.”

“The main advantage is that it gives a broader view than most administrative datasets – only in survey data does one typically find both detailed data on income and wealth, some health outcomes, and data on insurance and expenditures.”

“The HRS is a great source. The breadth of information and the longitudinal nature of the data are probably its main strengths.”

“It’s useful... because it has longitudinal data on the near-elderly, and good detail on financial health as well as physical/mental health.”

“Yes, it is a rare data set that has lots of interesting info on a longitudinal basis.”

“[The HRS is valuable because of] its relatively large sample size, long panels, and high quality economic data, paired with a diverse set of (admittedly imperfect) health indicators.”

“The long time span is hugely helpful for following patterns of health and disability, as well as household and economic circumstances. I especially appreciate measures of functional status, which have [been] found to be very predictive of behavior for older workers.”

“The HRS is well suited to studying life cycle health issues and the interaction between economic variables and health.”

“More and more people are using administrative data for research, but there are certain types of things that would never be measured in administrative data. So the HRS is extremely valuable as a complement to larger scale datasets.”

“The key thing is that it has amazing individual-level data on health and wellbeing, it is longitudinal, and (with some difficulty) can be linked to death certificates, Medicare claims data, and the MDS.”

“The integration of the data from all of these domains, including administrative data, gives an unusually complete view of the economic forces behind health-related outcomes and behaviors.”

“The combination of survey and administrative data, as well as information on households instead of just individuals, is a very important element of the HRS. The longitudinal nature of the HRS is also an excellent asset (the MEPS follows individuals for only 1.5 years total, the HRS surveys individuals every two years).”

“[The HRS is notable for its] longitudinal data, rich domains, linkages to other data.”

“[The HRS is the] only national data set with excellent work/wealth/health questions [that is] longitudinal [and] can be linked to state data to take advantage of state level instrumental variables.”

“It's a great data set for looking [at] within-individual-over-time changes as they experience onset of chronic illness or have hospital admission. Good sample size; lots of measures of income, wealth, etc. Spousal part is also really nice (can't link spouses in credit reports, for example).”

“It contains information on both members of married households, as well as transfers to and from their children; it contains detailed information on finances; it contains detailed information on health (especially interesting to be are the more recently collected biomeasures); it is longitudinal and contains retrospective information from earlier in life; it can be linked to useful administrative data.”

“[The HRS offers] rich variables on long-term care and informal care and while there are limitations, having information on three generations is super helpful for family/intra-household decision-making work.”

**Table 5: Survey Responses on Ways in Which the HRS is Ill-Suited
for Health Economics Research**

Trending topics/empirical strategies/administrative data/sample size

"I think the key barrier to the HRS being used is topical. Most of health economics now focuses on health care system questions -- health insurance, provider choice and behavior, costs. For these, large administrative datasets (e.g., Medicare data) are much better suited. In comparison, HRS has two weaknesses: (1) it is a much smaller sample, and (2) the health care use variables in the public data are self-reported. Issue #2 can be overcome by using the restricted Medicare link, but #1 cannot. A lot of papers use variation that affects a small subset of the population. This is possible when you start with 40 million observations but not when you start with 30,000."

"I see the HRS's weaknesses in the light of the recent trend toward administrative data sources. In this light, the HRS's main weaknesses are its relatively small sample size and its greater measurement error compared to the best administrative datasets, at least along some dimensions."

"Since I work a fair amount on health care systems, requiring very large datasets, I can't use the HRS for all the projects."

"In general the profession is moving away from survey data towards administrative data."

"Identification strategies... such as RD [regression discontinuity] and panel methods require much larger samples than the HRS has to provide precise estimates. I typically use administrative data sets with sample sizes orders of magnitude larger than the HRS."

"Sample sizes [are] inadequate for research using geographic policy variation, or genetic markers."

"In practice the sample size means that it is hard to use for some studies."

"Sample size is small, and does not allow for enough heterogeneity tests to get at hypotheses in enough detail."

"Sample size can be an issue."

"Sample size, though this can't be fixed."

"It's not so much ill-suited, it's just the natural limitations of sample size."

“[The HRS has a] great consumption module with detailed questions but [sample size is] very small.”

Access to restricted data

“Accessing restricted data can be incredibly complicated. The process of applying to use Medicare-linked data seems designed to put hoops in the way. The inability to combine SSA and geographic identifiers limits their value. The HRS seems more security-conscious than other restricted data sources.”

“I think that a lot of my students are deterred from using the HRS because state-level identifiers require RDC [research data center] access, and there was no RDC [near] our campus.”

“Have to have federal funding to link to restricted data. Huge time investment to compile large panels...that said RAND contribution [is] massively valuable.”

“It can be costly in time to get permission to link the HRS to restricted administrative data, and to utilize it after approval.”

“[A limitation can be] difficulty in accessing restricted access data including geographic identifiers, bio-markers etc.”

Focus on elderly/near elderly population

“Age limitations can be a problem, sample size sometimes... Picking up new cohorts even earlier would be useful. ”

“A limitation is that it doesn’t contain information on younger ages/cohorts. That is, of course, by design, but does limit what one can study with the HRS.”

“We would have liked to have had [a] broader age range of prime-age adults.”

“Only covers near elderly and elderly.”

“All of these areas [such as health reform and cost-sharing in employer plans] have implications for the elderly and near-elderly, but they cannot be studied with data on individuals aged 50+ alone.”

Frequency/insurance data

“Two year lag [is] a limitation for some persons.”

“For some purposes, its low frequency data collection is not ideal. It is hard, for example, to observe how individuals respond to health (spending) shocks.”

“On some dimensions it is also not as strong on administrative data. For example, it can be difficult to know with sufficient certainty what sorts of insurance people have and what it covers.”

“Given that so many elderly individuals are insured, the HRS does not tell us much about expansions in health insurance coverage.”

**Table 6: Suggestions from Survey Respondents on Improving the HRS
for Health Economics Research**

Increasing awareness of data and public data access

“You might think about making a short (e.g. 2-page) summary of what is available in the HRS that could be used as a class handout, along the lines of this:
www.povertyactionlab.org/sites/default/files/documents/AdminDataCatalog.pdf.

“Is there a resource that shows briefly its pros and cons relative to other data sets? For example, I wonder if it would be useful to have infographics that compare it on dimensions that matter for some often-studied topics to other data sets. Highlighting strengths: HRS is best placed of the usual suspect list of data sets in detail and reliability of answers for [some] list of outcomes (such as dementia or other highly fundable topics), sample size (comparable to MEPS when it comes to the 50+), studying consumption and savings behavior (relative to the Consumer Expenditure Survey), studying elderly healthcare (relative to the MCBS), etc.”

“Like most rich panel data sets, the costs of entry [for using the HRS] are high, because longitudinal data are complex. RAND (or if appropriate a different contractor) should be compelled to make things even simpler, and they must answer their phones/e-mails when researchers write to them with questions.”

“If there were some way to access the public data more easily and produce cross-tabs of interest without diving too far into the data.”

Increasing access to restricted data

“Things that could be fixed include: 1) making it easier to apply for Medicare-linked data and 2) allowing simultaneous access to geographic data and other restricted data.”

“Loosening of data permissions for some of the low risk restricted data, e.g. state identifiers, [and] exceptions to having ‘federal funding’ for some of the higher risk restricted data, e.g. claims.”

“Somehow make it easier to get permission to link the HRS to restricted-use administrative data (e.g. Social Security earnings, Medicare claims, disability program interactions).”

Adding administrative health insurance information

“It would be useful to collect administrative data on health insurance plans – for some questions you need the exact cost sharing, or you want to study how things change when respondents hit Medicare eligibility but lack a good baseline.”

“Linkage ability to private insurance claims to help with comorbidity adjustment.”

“Perhaps more matching of (e.g.) hospital and nursing home characteristics or other summary measures of health care utilization.”

Improving health measures

“[Provide] richer data on biomeasures (e.g. genes/SNPs, measured weight/height/fatness, etc.).”

“[It could be useful to] add medical information before age 65 – for example, you don't know what drugs respondents are taking for any mental health conditions they already had at their first interview.”

On self-reports of chronic conditions: “Over time, self-reports of chronic disease diagnoses have revealed themselves to be particularly poor measures of anything approximating health. It's a good time to revisit what, if anything these would be useful for. HRS could use that space to put in something more useful. I have yet to see analyses where they offer meaningful information about the things the researchers actually think they are measuring. Some health systems diagnose everyone, others do not. Some individuals have undiagnosed conditions, so their behavior maps to conditions that are unreported. And among those correctly diagnosed, older adults get confused, so reports are way off, even for things you think would be simple, like AMI or stroke or cancer [see work by Laura Yasatis]. Specific diagnoses convey little, except stroke, which cannot be studied on its own given small numbers of stroke survivors in the data. Once a researcher decides to focus on a subset of individuals (heart attack survivors, for example) numbers become small quickly. [In one paper, researchers] used the conditions to categorize people into chronic condition groups and no chronic conditions, and that was helpful, but specific diagnoses were not really useful. I cannot think of a good example of papers where the self-reported chronic diagnoses yielded important, well-measured insights beyond what one could get with the questions asking whether respondents were limited by pain, depression, cardiovascular disease (heart attacks, strokes for example), lung disease, cognitive impairment, or cancer. You might simply want to know any versus no major chronic illness, but use limitations to guide additional detail on these chronic symptoms/conditions. You would need to figure out what to do about hypertension (which affects most HRS respondents) and diabetes.”

On better health measures: “Knowing that nuanced analyses of specific diagnoses are not possible, HRS should consider dropping some self-reports of diagnoses, in favor of more/better reports of: pain, depression, prescription drug use, cognitive impairment, and

other limitations of roles. It would be important to publish for HRS users why these self-reported diagnoses of conditions are such bad measures. Thinking about good candidates to discard, my guess is that reports of pain will be way more informative than reports of arthritis, for example. Similarly, cognitive tests will be more informative than self-reports of cognitive impairment, as will a revisited better screen for depression (combined with complete prescription drug info) compared with psychiatric diagnoses. [The HRS should] consider dropping chronic condition reports that are not predictive and replacing with alternative measures – biologic measures like grip strength, for example, or blood pressure, given that these are easier than ever to collect. If those are too costly, sticking with a simple measure of hospitalizations and ER visits, combined with reports of limiting conditions (i.e., self-reports of pain, depression, heart conditions that limit activity, cancer that limits activity, lung disease like COPD etc. that limit activity) would likely yield as much or more information combined with the measures you already have of functional limitations, ADLs, IADLs. Reports of back pain could/would/should be helpful, but should return to every wave. The every other wave format is useless.”

On mental health measures and prescription drug use: “Depression measure (the shortened CES-D) is not predictive of much. I don’t know if that’s because it is not as helpful in this age group or if the short version is less helpful, but it does not behave in the HRS the way depression measures like the full CES-D behave in other data sets. This could and should be revisited, since 20% of Medicare beneficiaries are taking SSRIs. Someone thinks they are depressed so it would be important to understand what symptoms are important. Also, medication questions around antidepressants [and] prescription pain relief should be asked of EVERYONE. The skip patterns make this truly rich health information nearly useless. We know that some patients who don’t need things get them and some patients who do need things do not, so forming an appropriate numerator and denominator is nearly impossible. Better yet, can’t respondents just line up prescription bottles, labels forward and snap a photo? Then you would have EVERY prescription in their cabinets, dates, doses, branded vs. generic, etc.”

Improving economic, family, and early life information

“It seems like the best strategy is to improve the HRS in its strong point: the link between economic variables and health. The health variables are already pretty good. But many of the economic variables -- particularly wealth and pensions -- are extremely noisy. Although it's counter-intuitive, I think improving the economic variables is the best way to promote more health economics work on the HRS.”

“[It would be useful to have] broad-based consumption measures alongside out-of-pocket medical costs and household income measures. One "off-the-wall" idea would be to merge HRS data to consumer credit reports (and use data in those credit reports to construct consumption proxies). Or just expand size of CAMS module.”

“[The HRS would benefit from the inclusion of] measures of work productivity not just labor force participation.”

“Ramping up effort to get better data on the adult children of R’s [respondents].”

“Information about individuals while growing up and their early labor market experiences is very important because so much research today finds that these early experiences shape aging.”

Increasing survey frequency

“Raising the frequency of the study, at least for a subset, would be useful. More generally, using new technologies to make the survey less onerous on respondents could help make it easier to field the study more frequently.”

“Collecting some of the ‘every 4 years’ modules more frequently to help with power (e.g. looking at disease specific questions you have limited power sometimes).”
