Retirement in the 1950s: Rebuilding a Longitudinal Research Database

Amy M. Pienta[1], Jared Lyle[2]

Abstract

In 2010, ICPSR began a long process of recovering data from Gordon Streib's Cornell Study of Occupational Retirement (CSOR). Because these unique data fill a gap in our understanding of US retirement history, we determined that an extensive data recovery project was warranted. This paper describes the scope of the data collection and the steps in ICPSR's recovery process. Though the data recovery was ultimately successful, this paper documents the amount of time invested and costs associated with this kind of recovery work. It also highlights the value of these data for future research in understanding gender and retirement in a historic context. In addition to the resulting publicly available data arising from this project, extensive paper medical records are housed at ICPSR for on-site analysis or for a future digitization project. These data would provide unique health information on older women and men traced over a period of time in the 1950s and represents future work for ICPSR to undertake.

Keywords

Data Recovery, Longitudinal Data, Data Archiving, CSOR, Cornell Study of Occupational Retirement

Introduction

As brand new Social Security policies were beginning to influence retirement behavior in the United States, the Cornell Study of Occupational Retirement (CSOR) launched with a plan to follow a large (over 4,000 person), national cohort of retirement-age men and women in 1952. The CSOR remains an important, yet underutilized, piece of retirement history. One of the reasons for renewed interest in these data is that the longitudinal study pre-dates the Longitudinal Retirement History Study (LRHS) of the 1960s, filling in an earlier and critical piece of history. Further, retirement behavior of men and women, married and unmarried, are captured in the data – a limitation present in other historic data around that time period. Another advantage of the data is that health and biomarker data were collected, but not analyzed, for a large portion of the sample. And finally, data associated with the CSOR were never released for public use until this year.

The original research team, led by Gordon Streib, then at Cornell University, focused their efforts at tabulating descriptive results and painting a broad picture of retirement behavior in the 1950s, enabling the possibility of gaining a better, more nuanced understanding of the retirement process at that time. The data also provide an earlier point of comparison for later retirement studies including: the Longitudinal Retirement History Study (1960s and 1970s), The Social Security Beneficiary Study (1980s), and the Health and Retirement Study (1990s+). Health and biomarker data from physicals, lab tests, and interviews supplement the self-report health information in the interviews. Given Social Security policy changes of the 1950s, the study represents an unprecedented data resource for understanding how retirement culture emerged during this landmark time. This paper provides an overview of the substantive contributions past and future of the CSOR and establishes new methods of data recovery relevant to historic data of this type.

Background

Retirement research, regardless of historic time period, has long been viewed as a valuable activity (Eckerdt, 2010). Gordon Streib and his colleagues were among the earliest researchers to provide a comprehensive analysis of retirement behavior in the United States. Their main findings about retirement in the 1950s revealed clustering of retirement at age 65 and 70 with better educated, higher income persons working longer (Streib & Schneider, 1971). They also debunk the "myth" that health declines after retirement by documenting the declining health of those who remain working (Streib & Schneider, 1971). They provide high correlation between self-rating and health and physician ratings (Suchman, Phillips, & Streib, 1958) when subjective measures were first undergoing high scrutiny. They provide little information about how the adjustment to retirement is difficult and note that only poorer adjustment to retirement is found when economic deprivation is reported (Thompson, Streib, & Kosa, 1960).

Retirement research is heavily populated with economic, demographic, sociological, and psychological analyses contributing to a public dialogue and understanding. For today's researchers, the data from the CSOR provide a new depth of understanding to various aging processes, including retirement behavior. The research community can apply a unique set of data to longstanding hypotheses using state-of-the-art statistical methods, further expanding our evidence base. The CSOR is a national, longitudinal study of retirement conducted from 1952 to 1958 and is likely the first large-scale study of retirement behavior in the United States. The data from this study are particularly well suited to advancing the scholarship in the areas described below.

Poor Health and Retirement

Studies abound showing that poor health is associated with retirement (see Bound, Stinebrickner, & Waidmann, 2010 for an example). One of the reasons the relationship between health and retirement has been examined so closely is the historic uptake of retirement-aged men and women claiming disability after this option first appeared as part of receiving Social Security Disability Insurance (SSDI) benefits in 1956. As a result, analysts have sought to explain the health-retirement relationship and to better understand the extent to which older workers leave the labor force due to objectively poor health. Decades of research has shown that persons in poorer health are more likely to leave the labor force in later life (Burr, Massagli, Mutchler, & Pienta, 1995; Hausman & Wise, 1985). And yet, more recent, nuanced analyses have suggested health may be less important to the retirement equation than once thought (Bound, Stinebrickner, & Waidmann, 2010), or be differentially important across population subgroups (Baidwan. Gerberich, Kim, & Ryan, 2016).

While the Health and Retirement Study has provided unparalleled, prospective information about health and retirement (Bound, Stinebrickner, & Waidmann, 2010; Brown & Warner, 2008; Gustman & Steinmeier, 2009), little is known about this relationship during the time period that disability insurance was first offered as a welfare policy. The CSOR offers researchers information about retirement behavior coinciding with the start of SSDI benefits. Also important, the study includes both objective and subjective health indicators. This provides new evidence to evaluate a longstanding critique about the value of self-report health status measures for retirement-aged individuals. Evidence from studies around the world suggests that self-reports are not comparable across individuals or are sometimes reported to rationalize retirement decisions (Ekerdt & Bosse, 1982; Jurges, 2007; Kalwij & Vermeulen, 2008; Muller & Boaz, 1988). The CSOR has the potential to be a valuable data resource for understanding the relationships among subjective and objective health status and retirement behavior in the 1950s--helping to disentangle the causal relationship between health and labor force changes due to its longitudinal design.

Gender and Retirement

Up until the 1970s, little was known about women's retirement behaviors. The Longitudinal Retirement History Study (RHS) from the 1970s offered some of the first national data about women's retirement behaviors (Gustman & Steinmeier, 1994; Henretta & O'Rand, 1983; Pozzebon & Mitchell, 1989). However, the RHS did not interview married women because retirement was viewed as having little meaning for married women of the 1960s (Slevin & Wingrove, 1995). Thus, our understanding of married women's retirement behaviors at the time focused almost exclusively on the relative timing of married partner retirement and the tendency for couple to retire at the same time (Gustman & Steinmeier, 1994; Henretta & O'Rand, 1983; Pozzebon & Mitchell, 1989). The Health and Retirement Study expanded our understanding of women's retirement behaviors in important ways (Pienta, 1999; Forman-Hoffman, Richardson, Yankey, Hillis, Wallace, & Wolinsky, 2008), but with respect to historical information about women, the CSOR is unique in its size and coverage of women – allowing us to extend a lens to an earlier point in US history.

Age-Related Changes in the 1950s

Students of gerontology learn quickly that longitudinal studies are the gold standard for research – allowing one to disentangle aging effects from period and cohort effects to better establish causal relationships. This was known to be true in the 1950s and guided the design of the CSOR (Streib, Thompson, & Suchman, 1958). Despite this fact, there are areas of gerontological study that have been built upon cross-sectional analyses, in part due to data limitations. Thus, it is it is possible to imagine the value to aging research of making longitudinal data available from the CSOR. Because the study lived only on punched cards until recently, the results from analyses in the 1950s and 1960s were largely descriptive. Advances in computing, storage, and analytic capabilities, combined with advances in longitudinal methodologies, have the potential to shed new light on age-related changes in the 1950s – changes in marriage, family status, living arrangements, health status, economic well-being, and life satisfaction can be explored with these historic data.

Health and Biomarker Data from the 1950s

The CSOR includes a wide range of biomarker data including height, weight, pulse, total cholesterol, hemoglobin, systolic and diastolic blood pressures, cardiac function, vision and hearing, and other results of functional assessments conducted by physician, observations and testing. Biomarkers have been used to measure stressful life conditions with implications for health and functional status as a person ages (Singer, Ryff, & Seeman, 2005). These data have become somewhat commonplace today (Ewbank, 2008), but the CSOR collected health and biomarker data in the 1950s – at baseline and in two of the follow-up interviews. What is perhaps more surprising is that much of the health and biomarker data were never digitized (or analyzed to any meaningful extent) and exist only in paper form today. Combined with the interview data, it will be possible for researchers to examine changing social conditions and life events in conjunction with changes in a range of health and biomarker measures.

Data and Methods

Gordon F. Streib, architect of the CSOR, was born in 1918, receiving his M.S.Sc. from New School for Social Research in 1947 and his Ph.D. from Columbia University in 1954. Streib worked at Cornell University in 1949, starting as an instructor and eventually becaming Chairman of the Department of Sociology and Director of the Cornell Study of Occupational Retirement before his departure in 1966. Streib went on to teach at the Universities of Southern California and Florida, retiring in 1990. In the late 1990s, ICPSR obtained the filing cabinets containing the data and records from the CSOR from the Newberry Library of the University of Chicago. ICPSR shipped the cabinets to Ann Arbor, MI where they were stored for years. In 2003, ICPSR invited Gordon Streib to the University of Michigan to

review the contents. Streib provided an oral history of the study and its significance, but given complexities in the punched cards (the cards were double-punched) and available technology to read the cards, the data remained in storage until 2010.

Study Design

Funded by the Lily Endowment, the National Institute of Mental Health, and the US Public Health Service, the CSOR was designed to understand the transition from work to retirement in the 1950s. A baseline interview was conducted with a cohort of workers age 64 in 1952 and were subsequently conducted every 1-2 years for six years. Over the course of the 6-year study, over 50% of the respondents retired (Streib & Schneider, 1971). The survey included a wide range of questions regarding sociodemographic characteristics, family, daily activities, work (type of work and work satisfaction), economic status (income, homeownership, and household size), pensions, age identity, age stereotypes, retirement plans, health, life satisfaction, and adjustment to the retirement transition.

Sample Design

Study participants were born between the years of 1887 and 1889 and were 64 years old when first interviewed in 1952. The sampling strategy was built mainly around selecting companies classified in the US Census across the range of industrial classifications and interviewing all age-eligible employees in the selected companies. Only companies employing 1,000 or more people were selected for the study and all 64 year olds at the company were interviewed. In addition, to better represent white-collar and professional workers, several federal, state, and city/county agencies were included in the sample as were several school systems and a doctors' sample from New York State. Thus, the sample is a hybrid of a stratified design supplemented with additional interviews to get better representation of the professional occupations. In total, 259 organizations were included in the study. 4,032 participants began the study in 1952 with most initial interviews conducted face-to-face. Although the study does not provide a representative sample, it includes a large, diverse sample of the American labor force at the time. The first follow-up interview was conducted in 1954, 12 to 18 months after the first interviews. The second follow-up was conducted in 1955/56 and two additional follow-ups were conducted in 1957 and 1958. Follow-up response rates (for those still living) ranged from 88% to 95%.

Phase 1 – Data Stabilization.

Until 2010, the Cornell Retirement Study was deemed unrecoverable by ICPSR. An early attempt was made to read the punched card data, but because the cards were double-punched (a space saving technique) – the data read from the cards could not be matched with the documentation in the filing cabinets. With funding from the Library of Congress National Digital Information Infrastructure and Preservation Program, ICPSR purchased a refurbished Cardamation punched card reader. All 52 boxes of original punched cards were read and output as raw ASCII text files. Staff used a program developed in tandem with the Roper Center for Public Opinion Archives to convert these raw, complexly structured files into a rectangular, delimited text format. Many of these files were converted to SPSS.

At the same time, project staff inventoried, stabilized, and modernized data and other documentation materials from the CSOR. The results of this work helped to determine the extent to which adequate documentation would be available to recover a usable version of the original data collection. Hard copy materials from seven large filing cabinets were reviewed and inventoried. These included documentation, final reports, publications, correspondence, and completed data collection instruments for physician follow-up surveys and examination(s). Unstable and degrading paper materials (e.g. oily stencils) were replaced with photocopied versions. All materials were moved to 37

archival storage boxes and a full inventory, or finding aid, of study materials was created. The historic materials were grouped into the following: 1) raw data and analysis, 2) medical records, 3) codebooks, 4) blank questionnaires, 5) documentation of study progress, 6) study correspondence, 7) administrative records (e.g., bibliographic material, grant proposals), 8) miscellaneous correspondence, 9) miscellaneous records unrelated to the study, and 10) publications. An in-depth review and inventory of the medical records was conducted to determine the number and content covered.

Phase 2 – Data and documentation matching.

The next phase involved intensive data management and documentation matching. The bulk of the work during this phase was to match the converted punched card data with the large body of documentation. ICPSR conducted an initial inventory of the data to identify respondent populations, as shown in Table 1. The total counts are the total number of respondents interviewed per year. The actual counts are the cases ICPSR was able to match during the initial inventory. Raw rectangular data files were matched by unique IDs associated with each respondent and then merged for each year. In the merged datasets, each observation represents one individual respondent. All variable and value labels were created according to the historic codebook documentation. All missing data were labeled using standardized codes (e.g., -97, -98) to account for system missing data, dropouts, etc.

Table 1: Expected Number of Respondents, by Year, Based on Published Results and Individuals Identified in the Archival Holdings during Pilot Work

	Expected n	Identified n
1952	3,793	3,002
1954	2,857	2,603
1956	2,465	2,465
1957	2,182	1,992
1958	1,969	1,969

Phase 3 - Public-use data file creation

Although none of the respondents are living today, we conducted a full disclosure review to determine if any material should be redacted or removed. The data files were run through ICPSR's data automation system, which creates stable research data products in the following formats: 1) ASCII with SAS, SPSS, and STATA syntax, 2) SAS XPORT (CPORT), 3) SPSS portable files, and 4) STATA .dta files. Data are available through the National Archive of Computerized Data on Aging (NACDA) archive at ICPSR and from the main ICPSR Web site.

Phase 4 – Review and integrate the physician exams and records

Each study participant could have one or more of the following forms in the study files:
- M1-1952 is a Family Medical History that includes questions about illness histories (e.g., diabetes, cancer, heart disease, suicide) for respondents' immediate family members;
- M2-1952 is a Personal Medical History that asks extensive personal questions, ranging from reactions to medication to consistency of bowel movements to use of tobacco, drugs, and alcohol;
- M3-1952 is a Medical History Summary, which is a physician's review and summary of the personal medical history evaluation;

- M4-1952 is a Physical Exam, which assesses respondents' bodily features, from posture to skin and lymph nodes to abdomen to neuromuscular functions;
- M5-1952 is Lab Data, including reports on findings from urine (Albumen, Sugar, RBC, WBC) and blood (Hemoglobin, Serology, Sugar, and Cholesterol) specimens, with some files including special procedures, such as electrocardiograms and x-ray reports;
- M2-1954 is a Follow-up Health Questionnaire asking questions about respondents' health changes since completing the original questionnaires;
- M4-1954 is a Physician's Summary of Follow-Up Examination, which is a physician's review and summary of the follow-up health questionnaire;
- M2-1956 is another Follow-up Health Questionnaire and is the same as M2-1954;
- M4-1956 is another Physician's Summary of Follow-up Examination;
- M2-1958 is another Follow-up Health Questionnaire; and
- M4-1958 is another Physician's Summary of Follow-up Examination.

In the course of our work, an in-depth survey of the physical medical records was conducted to determine the number sampled by content covered (see Table 2). Because some of the variables were captured in the punched cards, ICPSR was able to make those data available with the electronic data files. The lab results and detailed handwritten health data that were not represented in the punched card remain in hard copy, paper form and can be analyzed on-site at ICPSR.

Table 2. Inventory of Completed Health Survey Instruments

	M1 - 1952	M2- 1952	M3- 1952	M4- 1952	M5- 1952	M2- 1954	M4- 1954	M2- 1956	M4- 1956	M2- 1958	M4- 1958	Exams/ Misc.
Total Counts	3,181	3,184	1,110	1,192	1,111	3,085	696	346	463	2,873	325	263

Conclusions

One of the most significant lessons learned by our work at ICPSR was that, even though we did not have a strategy for processing the punched cards when we accepted the data from the Newberry Library, it was eventually possible with changing technologies and new resources to successfully rescue the data. In the course of looking for solutions to read the punched card data, there were many failures by ICPSR staff where the data could not be read and the documentation could not be matched to the data files. However, increasing interest in "at risk" data led ICPSR to partner with the Library of Congress to think through and build new solutions for rescuing older format data including data stored on punched cards. ICPSR maintains this capacity today.

The other significant finding from our work is that the cost of rescuing data from the 1950s was many times more expensive than curating data produced in the last decade. We offer this story as both inspiration to save data that seems forever lost to modern techniques, but also to persuade scientists to plan for and to ensure their research data will be accessible for generations to come. There are many domain repositories like ICPSR (serving the social and behavioral sciences) and NACDA (serving NIA-funded data) as well is institutional repositories on university and college campuses to assist with open access to data. However, there appears to be increasing costs associated with archiving the "older" the data are when they arrive to ICPSR.

This brings us to our final point. Because ICPSR is a domain repository with a commitment to long-term use of data, it was possible to ensure not only "bit-level preservation" of the data, but that the CSOR data would be useable long into the future. Many data repositories and data services that are available to researchers today offer only to take in (ingest) and serve up (disseminate) data in the condition that the researchers donated the data (or published the data). Perhaps this is what the Newberry Library intended when they took in the Streib data. However, without stripping away the myopic storage media (e.g. punched cards, floppy disks, magnetic tape) and data format (e.g. old versions of SPSS), the data become unusable as the media and/or software become obsolete. Best practice in the repository world, and the practice of ICPSR, is to ensure that data storage and format are neutral and/or refreshed over time (in additional to the data being well-documented). This has been the practice of ICPSR for over 50 years.

References

Baidwan, N. K., Gerberich, S., Kim, H., & Ryan, A. (2016). The association between work-related exposures and occupational injuries in the aging United States workforce. *Health Promotion & Chronic Disease Prevention in Canada: Research, Policy & Practice, 36*(11), 265. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5432052/

Bound, J., Stinebrickner, T., & Waidmann, T. (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics, 156*(1), 106-129. https://doi.org/10.1016/j.jeconom.2009.09.010

Brown, T. H. & Warner, D. F. (2008). Divergent pathways? Racial/ethnic differences in older women's labor force withdrawal. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 63*(3), S122-S134. https://doi.org/10.1093/geronb/63.3.S122

Burr, J. A., Massagli, M. P., Mutchler, J. E., & Pienta, A. M. (1996). Labor force transitions among older African American and white men. *Social Forces, 74*(3), 963 982. https://doi.org/10.1093/sf/74.3.963

Ekerdt, D. J. (2010). Frontiers of research on work and retirement. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B*(1), 69-80. https://doi.org/10.1093/geronb/gbp109

Ekerdt, D. J. & Bosse, R. (1983). Change in self-reported health with retirement. *International Journal of Aging and Human Development, 15*(3), 213-223. https://doi.org/10.2190/GXEG-GUMU-T9HG-TE2V

Ewbank, D. C. (2008). Biomarkers in social science research on health and aging: A review of theory and practice. In M. Weinstein, J. W. Vaupel, & K. W. Wachter (Eds.), *Biosocial surveys: Committee on advances in collecting and utilizing biological indicators and genetic information in social science surveys* (pp. 156-174). Washington (DC): National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK62438/

Forman-Hoffman, V. L., Richardson, K. K., Yankey, J. W., Hillis, S. L., Wallace, R. B., & Wolinsky, F. D. (2008). Retirement and weight changes among men and women in the Health and Retirement Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 63*(3), S146-S153. https://doi.org/10.1093/geronb/63.3.S146

Gustman, A. L. & Steinmeier, T. L. (2009). How changes in Social Security affect retirement trends. *Research on Aging, 31*(2), 261-290. Retrieved from http://hdl.handle.net/2027.42/49396

Gustman, A. L. & Steinmeier, T. L. (1994). Retirement in a family context: A structural model for husbands and wives (No. w4629). National Bureau of Economic Research. doi:10.3386/w4629

Hausman, J. A. & Wise, D. A. (1985). Social Security, health status, and retirement. In D. A. Wise (Ed.), *Pensions, labor; and individual choice* (pp. 159-192). Chicago: University of Chicago Press. Retrieved from http://www.nber.org/chapters/c7133.pdf

Henretta, J. C. & O'Rand, A. M. (1983). Joint retirement in the dual worker family. *Social Forces 62*(2), 504-520. https://doi.org/10.1093/sf/62.2.504

Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics, 16*(2), 163-178. doi:10.1002/hec.1134

Juster, F. T. & Suzman, R. (1995). An overview of the Health and Retirement Study. *The Journal of Human Resources, 30*, S7-S56. doi:10.2307/146277

Muller, C. F. & Boaz, R. F. (1988). Health as a reason or a rationalization for being retired? *Research on Aging, 10*(1), 37-55. https://doi.org/10.1177/0164027588101002

Pienta, A. M. (1999). Early childbearing patterns and women's labor force behavior in later life. *Journal of Women and Aging, 11*(1), 69-84. http://dx.doi.org/10.1300/J074v11n01_06

Pozzebon, S. & Mitchell, O. S. (1989). Married women's retirement behavior. *Journal of Population Economics, 2*(1), 39-53. https://doi.org/10.1007/BF00599179

Singer, B., Ryff, C. D., & Seeman, T. E. (2005). Operationalizing allostatic load. In J. Shulkin (Ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation* (pp. 113–149). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781316257081.007

Slevin, K. F. & Wingrove, C. R. (1995). Women in retirement: A review and critique of empirical research since 1976. *Sociological Inquiry, 65*(1), 1-21. doi:10.1111/j.1475-682X.1995.tb00404.x

Streib, G. F., & Schneider, C. J. (1971). *Retirement in American society: Impact and process.* Ithaca, NY: Cornell University Press.

Streib, G. F., Thompson, W. E., & Suchman, E. A. (1958). The Cornell Study of Occupational Retirement. *Journal of Social Issues, 14*(2), 3-17. doi:10.1111/j.1540-4560.1958.tb01402.x

Suchman, E. A., Phillips, B. S., & Streib, G. F. (1958). An analysis of the validity of health questionnaires. *Social Forces 36*, 223-232. doi:10.2307/2573809

Thompson, W. E., Streib, G. F., & Kosa, J. (1960). The effect of retirement on personal adjustment: A panel analysis. *Journal of Gerontology, 15*(2), 165-169. https://doi.org/10.1093/geronj/15.2.165

---

[1] Amy Pienta is Associate Research Scientist of ICPSR at the University of Michigan and can be reached via email: apienta@umich.edu.

[2] Jared Lyle is Director of Curation Services at ICPSR at the University of Michigan and can be reached via email: lyle@umich.edu.