

The Retirement and Health Study - Management of a Large-Scale Longitudinal Study in Singapore

by Jocelyn Ow, Neo Ming Yuan and Koh Phuy Leng
Central Provident Fund Board

Introduction

With one of the most rapidly ageing populations globally, it is imperative to ensure Singapore's social, healthcare, financial, and other systems evolve to cater to the needs of changing demographics.

Timely and pertinent research on retirement planning, health status, income, and living arrangements helps the government and relevant organisations understand the needs and challenges faced by an ageing population, and informs the development of suitable programmes to support the well-being of seniors.

In 2014, Singapore's first large-scale integrated longitudinal study on residents' retirement and healthcare needs, the Retirement and Health Study (RHS), was launched to facilitate evidence-based decision-making.

The RHS is jointly conducted by the Central Provident Fund Board (CPF Board), the Housing and Development Board (HDB), the Ministry of Finance (MOF), the Ministry of Health (MOH) and the Ministry of Manpower (MOM), with the Singapore Department of Statistics (DOS) providing technical advisory.

Modelled after other similar studies overseas (e.g., the Health and Retirement Study in the United States, the English Longitudinal Study of Ageing in the United Kingdom), the multi-disciplinary RHS examines information on various aspects of ageing in Singapore to gain a deeper understanding of the related needs of seniors.

Survey Coverage

The survey covers a sample of Singapore residents aged 45 years and over who are interviewed once every two years.

The sample is selected based on a stratified design, with the sampling frame divided into different groups (or strata) according to age group, sex and race. Systematic sampling by broad dwelling type is then performed within each stratum. Younger cohorts are added periodically to ensure that the RHS sample continues to represent the Singapore resident population aged 45 years and over.

The main unit of analysis is the Main Respondent who goes through a 1.5 hour interview. The Main Respondent's spouse or partner is also invited to participate in a separate 30 minute interview on selected topics from the Main RHS Questionnaire (Table 1).

TABLE I
TOPICS COVERED
IN THE LATEST RHS QUESTIONNAIRES

Topics in Main RHS Questionnaire

Demographics	} Topics in a separate questionnaire for Spouse or Partner
Health Status	
Employment and Retirement	
Assets and Liabilities	
Financial Transfers and Other Income	
Financial Planning Behaviour	
Household Expenditure (Non-healthcare)	
Healthcare Insurance Plans and Schemes	
Healthcare Utilisation and Expenditure	
Health Risk Factors, Social Connectedness and Lifestyle	
Cognitive Status	

To date, four waves of the RHS have been completed, and Wave 5 is in progress. Around 11,700 to 15,100 Main Respondent interviews were completed in each of the past four waves. The longitudinal all-waves response rate for Main Respondent is 60 per cent in Wave 4, with 8,200 Main Respondents who participated in all four waves (Table 2). The response rate for Spouse/ Partner is around 70 per cent from Waves 1 to 4 (Table 3).

The next section provides an overview of the strategies adopted by the Retirement and Health Study Office (RHSEO) in the areas of questionnaire design, panel management and data quality assurance, to achieve a robust longitudinal panel with quality data across waves.

TABLE 2
RESPONSE RATES FOR MAIN RESPONDENTS BY RHS WAVE

RHS Wave	Year Respondents Introduced	Fieldwork Period	No. of Main Respondents	Cross-Sectional Unconditional Response Rate ¹ (%)	Cross-Sectional Conditional Response Rate ² (%)	Longitudinal Response Rate ³ (%)	Longitudinal All-Waves Response Rate ⁴ (%)
1	2014	Jul 2014 – Jun 2015	15,100	61	-	-	-
2		Jun 2016 – Apr 2017	12,900	69	81	80	80
3		Jun 2018 – Mar 2019	11,700	72	83	76	67
4		Sep 2020 – Aug 2021	11,200	83	90	76	60
4 ⁵	2020		3,500	54	-	-	-

1 Cross-sectional unconditional response rate = (No. of respondents who responded in Wave X)/ (No. of eligible sample units released in Wave X), where X = 1, 2, 3 or 4
 2 Cross-sectional conditional response rate = (No. of respondents who responded in Wave Y)/ (No. of respondents who responded in Wave (Y-1) and eligible in Wave Y), where Y = 2, 3 or 4
 3 Longitudinal response rate = (No. of respondents who responded in Waves 1 & Y)/ (No. of respondents who responded in Wave 1 and eligible in Wave Y)
 4 Longitudinal all-waves response rate = (No. of respondents who responded in all waves)/ (No. of respondents who responded in Wave 1 and eligible in Wave Y)
 5 A new panel of individuals aged 45-50 years old was added in Wave 4

TABLE 3
RESPONSE RATES FOR SPOUSES OR PARTNERS OF MAIN RESPONDENTS BY RHS WAVE

RHS Wave	Fieldwork Period	No. of Spouses or Partners of Main Respondents	Response Rate ¹ (%)
1	Jul 2014 – Jun 2015	7,700	69
2	Jun 2016 – Apr 2017	6,500	70
3	Jun 2018 – Mar 2019	6,100	74
4 ¹	Sep 2020 – Aug 2021	7,400	70

1 Response rate = (No. of spouses or partners who responded in Wave X)/ (No. of main respondents who are married or living with a partner in Wave X), where X = 1, 2, 3 or 4

Strategies

There are three broad stages in the collection of RHS data. Firstly, preparatory actions, such as designing the questionnaire and panel, are conducted prior to fieldwork. Secondly, interviews are conducted with respondents during

fieldwork. Lastly, data collected from respondents are checked and processed during and after fieldwork.

To provide researchers with robust data despite the twin constraints of limited time and resources, the RHSO adopts multiple strategies at each stage to:

- i. increase the survey response rate
- ii. enhance process efficiency
- iii. ensure data quality and user-friendly datasets

Pre-Fieldwork: Reviewing Questionnaire Design

Given the longitudinal nature of the study, the questionnaire has to be broadly similar across waves to effectively measure changes in the respondents’ situations and needs. Hence, a well-designed questionnaire is crucial to preserve longitudinal data quality.

With the respective domain agencies, the RHSO conducts a review of the questionnaire before the start of each wave. This ensures that the information collected remains meaningful and relevant for policymaking, keeps the respondent burden manageable, and improves data quality (Table 4).

TABLE 4
OBJECTIVES AND STRATEGIES ADOPTED DURING QUESTIONNAIRE REVIEW

Objective	Strategies Adopted
1 Review relevance of questions to current or new issues	a) Seek inputs from domain agencies on new or emerging issues b) Monitor ageing-related policies for assessment on the application to RHS
2 Reduce respondent burden	a) Re-evaluate value of existing questions, especially for repeat respondents b) Pre-fill responses using past wave survey data to aid respondents' recall c) Tap on administrative data sources to populate survey-equivalent responses
3 Enhance quality of data collected	Identify gaps in question clarity based on quality of data collected in past waves, as well as feedback from respondents and interviewers

Pre-Fieldwork: Maintaining a Robust Panel

Maintaining a robust panel with high response rates is essential to ensure representativeness of the RHS sample. Before initiation of each wave, RHSO reviews the panel management strategies to be implemented, such as assessing if differentiated approaches are needed for different groups.

These strategies seek to retain as many repeat respondents as possible and boost response rates from groups that are relatively harder to reach (e.g., private housing dwellers and respondents from the younger age group).

Fieldwork: Ensuring Operational Efficiency

For effective allocation of fieldwork resources, the RHSO uses administrative information to identify individuals who are out-of-scope due to long-term institutionalisation, changes in residency status and vital status.

Contact information from administrative database and data collected in previous waves are also utilised to aid interviewers in contacting respondents for the survey. To assure respondents on the value and authenticity of the survey, the RHSO sends out invitation letters and emails to respondents. Following which, interviewers contact respondents to schedule appointments via calls and/ or house visits.

To enhance efficiency in fieldwork operations, the RHSO has developed a machine learning model to identify optimal time slots for interviewers to call on respondents without prior appointments. This model uses various information (e.g., respondents' socio-demographic characteristics) to achieve a higher rate of survey completion with fewer house visits.

Fieldwork: Enhancing Participation through Digitalisation

Besides conventional ways of making appointments via the RHS hotline and email, respondents can make appointments via an online appointment form (available in the four official languages viz. English, Chinese, Malay and Tamil). Respondents can indicate their preferred appointment day, time and mode of the interview on the form. A substantial proportion of appointments have been made via this channel since it was introduced in Wave 4.

During the COVID-19 pandemic, the RHS was one of the first government surveys to pilot the use of Video Call Interviews (VCI) as an alternative to face-to-face interviews for respondents who had concerns with in-person meetings.

VCI has enabled the RHSO to reach out to more tech-savvy respondents and those overseas during the survey period. Despite being a novel interview mode at that time, the interviewing process remains effective, with minimal impact on data quality. VCI has also enhanced the resilience of the RHS project to better overcome unforeseen disruptions to fieldwork.

In view of concerns over potential scams, the RHSO has introduced an online link via CheckFirst for respondents to easily verify interviewers' identity. Many respondents have found this useful as they can verify any time instead of calling or emailing the RHSO.

Finally, the RHSO provides updated information for respondents through the RHS webpage, to facilitate their participation and reinforce the authenticity of the study. For Wave 5, an RHS chatbot

named Rosy was deployed to address Frequently Asked Questions from respondents instantly.

Fieldwork and Post-Fieldwork: Engaging Respondents

As a token of appreciation to respondents who committed their time to complete the RHS interview, shopping vouchers are given to them. To encourage continued participation, a loyalty incentive was introduced since Wave 2 for repeat respondents.

The RHSO sends greetings to respondents during festive seasons, as a way of maintaining contact in-between waves. An RHS newsletter is also published regularly. The newsletter provides information on upcoming waves and assurance on data confidentiality, as well as presents useful findings from the RHS data.

Data Management: Ensuring Data Quality

The RHSO implements various measures to ensure that complete, valid and consistent data are collected and compiled so that robust findings can be generated.

To ensure the collected data meet the intent of the questionnaire, interviewer training is conducted before the launch of each wave. Fieldwork audits are conducted regularly, and feedback is provided to interviewers with one-on-one coaching conducted to reinforce learning.

Front-end validation rules to trigger the verification of anomalous responses with respondents are also incorporated as part of the survey program to ensure accuracy of the collected data. Further back-end validation checks and editing are carried out by the RHSO to ensure cross-sectional consistency across related data domains, as well as longitudinal consistency for relatively static data items.

Administrative data are integrated with the survey data for respondents who gave their consent. With over 90 per cent of respondents consenting to linkage since Wave 1, the resultant integrated datasets are more complete and accurate. Key derived variables (e.g., non-work income), supplementary guides and tools (e.g., a search engine which enables discoverability of RHS variables) are also developed to improve data users' experience.

Finally, automation and machine learning techniques are used to improve the RHSO's efficiency in data processing.

A point-and-click interface program was created using Python programming to automate common data processing tasks. In terms of machine learning techniques, the RHSO makes use of natural language processing and a linear support vector machine model to process unstructured job and company-related text responses into the Singapore Standard Occupational and the Singapore Standard Industrial Classifications codes. Another automated valuation model is used to estimate the value of respondents' residential properties and improve the measure of respondents' housing wealth.

Conclusion

Over the past decade, the RHSO has implemented efficient fieldwork processes and robust data management methods, developing the RHS into an authoritative source of longitudinal data for ageing-related policy research in Singapore.

The RHS data have been used by government agencies and academics to undertake studies which have contributed to policy design and thinking in diverse fields such as ageing and retirement, healthcare, private financial wealth, and social relationships. Findings based on the RHS had also been cited in local media and published in international research journals.

The RHSO is committed to promoting greater use of the RHS data to maximise their value, while continuing to safeguard respondents' confidentiality.

