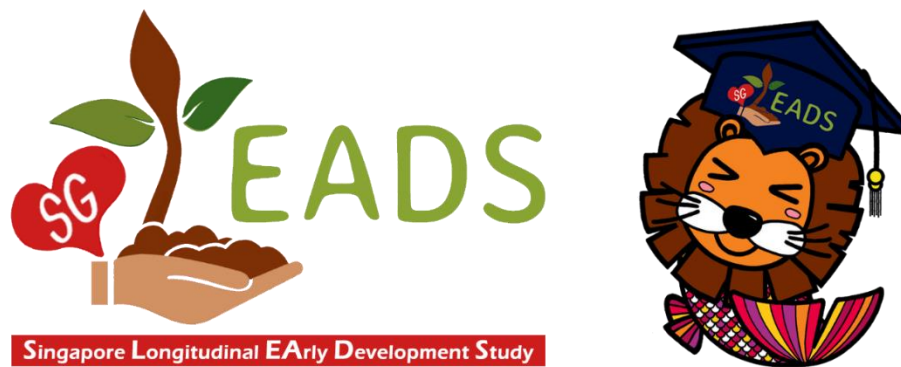# Singapore Longitudinal Early Development Study (SG LEADS)

Panel Survey Wave 2

Technical Report 2

SG LEADS Wave 2 Sampling Weights

Chen Xuejiao & Jean Yeung

Centre for Family and Population Research

National University of Singapore

January 2022

1. Introduction

Singapore Longitudinal Early Development Study (SG LEADS) is a longitudinal study that began in 2018. It provides a nationally representative sample of Singaporean children aged 0 to 6 in 2018 and their families.

In SG LEADS Wave 1 study, 3,476 families with 5,005 children were interviewed. For each household, up to two eligible children were interviewed. The second wave of study was conducted in 2021, and 3,016 Wave 1 households (3017 wave 2 household, including a successful interview of a split-off household) with 4,352 children have been successfully re-interviewed. The study includes several modules: 1) primary caregiver household interview, 2) primary caregiver child interview, and 3) child assessment for children aged 3 and above.

To account for different selection probabilities and response bias, sampling weights were created for SG LEADS wave 1 sample (see technical report of sampling weights for wave 1 study). In wave 2, weights were constructed to account for attrition. The household level weights are used for household-level analysis using data merely from the primary caregiver household interview. The child-level weights are applied for child-level analysis, whenever the data from primary caregiver child interview or child assessment is used.

2. Sample attrition between Wave 1 and Wave 2

In Wave 2, one household reported a death of sample children after Wave 1. Therefore, at the household level, there is 1 ineligible household, with 1 ineligible child at the child level. This returns 3,477 households with 5,006 children in wave 1 dataset. In the second wave, 3,017 households with 4,351 children were successfully re-interviewed. The unweighted response rates in Wave 2 are 86.8% and 86.9% at the household-level and child-level, respectively.

Table 1. Sample Attrition in Wave 2

	Household level	Child level
Total (Wave 1 sample)	3,476	5,005
Wave 2 complete interview	3,016	4,352
Non-response	460	653
Ineligible (sample child is deceased)	1	1
Response rate*	86.8%	87.0%

*Excluding ineligible cases

3. Weight construction procedure

Weights are constructed in sample survey data to adjust for unequal sample selection probabilities, non-response or data that is missing at random. Weights are inversely proportional

to the selection probability of each case, and conditional on the response to the survey questions. In a longitudinal dataset, the joint probability at time t, where the study started at t-1 or earlier, can be expressed as:

$$P(S_t=1) = P(S_{t-1}=1) * P(R_t=1|S_{t-1}=1)$$

In which S_t indicates participation in a study at time t, and R_t refers to response at time t. The probability of being a participant at time t is a product of being a participant at the previous wave (e.g., time t-1) and conditional on the probability of being a response at the current wave. The probability of being a participant at the previous wave (the term P(St-1=1)) is proportional to the weight in previous wave. Therefore, the weight in the current wave is a product of the weight in previous wave and the inverse of probability of response in the current wave (the term $P(R_t=1|S_{t-1}=1)$). Thus, the attrition adjustment factor in wave 2 is 1/ P(Rt=1|St-1=1). The wave 2 weight is a product of wave 1 weight and the wave 2 attrition adjustment factor.

In other Panel studies like Panel Study of Income Dynamic Child Development Supplements (PSID CDS) and The Longitudinal Study of Australian Children (LASC), the probability of a sample person or household being successfully re-interviewed in wave 2 onwards was typically modeled with the linear logistic model. Several wave 1 indicators were taken into consideration in the model specification including: the sample child's age, gender and race, head of household's age, gender, educational level, employment status, whether head of household's spouse lives in the household, Wave 1 income quartile, dwelling types and home ownership, region of residence. Since Wave 1 family income has missing values (n=59 at the household level), a separate multiple imputation for Wave 1 income at the household level was conducted. In the multiple imputation, the head of household's age, age square, gender, race, education, employment status, occupation were used to predict their family income. After the imputation, the logistic model of Wave 1 children's probability of responding in Wave 2 was fitted. The logistic model is presented in Table 2.

As seen, younger children, male-headed household, head of household is a housewife or home maker, head of household's spouse lives in the household, children in rental HDB or owned HDB 1- and 2-Room Flats in north planning area are slightly less likely to response to the Wave 2 survey.

Table 2. Logistic Regression on Wave 1 Children's Probability of Responding in Wave 2 Interview

Variable	Coefficient	robust std. err.	z	P>|z|
Child's age	0.037	0.022	1.650	0.099
Boy	-0.079	0.087	-0.900	0.366
Head of household's age	0.009	0.008	1.150	0.252
Head of household is male	-0.840	0.225	-3.720	0.000
Head of household's race (ref. Chinese)				
Malay	-0.121	0.143	-0.850	0.396

Indian	-0.273	0.177	-1.540	0.124
Others	-0.019	0.309	-0.060	0.951
Head of household's education (ref. secondary and below)				
post-secondary	-0.025	0.150	-0.170	0.867
university and above	0.240	0.190	1.260	0.207
Head of household's employment status (ref. working)				
housewife/homemaker	-0.774	0.296	-2.610	0.009
other-not working	-0.367	0.283	-1.300	0.194
Head of household's spouse lives in the household	0.534	0.252	2.120	0.034
Housing type and homeownership (ref. Owned HDB 4-Room Flats)				
Rental HDB	-0.298	0.204	-1.460	0.144
Owned HDB 1- and 2-Room Flats	-0.587	0.343	-1.710	0.087
Owned HDB 3-Room Flats	-0.020	0.154	-0.130	0.895
Owned HDB 5-Room/Executive Flats	0.243	0.190	1.280	0.201
Owned/rental Condominiums & Landed Properties	-0.203	0.200	-1.020	0.309
Income quartile (ref. Incomeq1_lowest)				
Incomeq2	0.043	0.160	0.270	0.790
Incomeq3	0.150	0.194	0.770	0.440
Incomeq4_highest	-0.143	0.234	-0.610	0.542
Planning region				
East	0.030	0.202	0.150	0.882
North	0.312	0.197	1.580	0.114
North-East	-0.235	0.161	-1.450	0.146
West	0.150	0.171	0.880	0.380
Constant	1.784	0.463	3.850	0.000
N	5,006			
Pseudo R2	2.40%			

Each Wave 1 child's probability of responding to Wave 2 survey (P) was estimated using the model presented in Table 3. The Wave 2 response adjustment factor was constructed for those who have been re-interviewed in Wave 2 by taking the inverse of their response probability (1/P) (refer to Table 3 for the distribution).

Table 3. Distribution of Responding Cases' Response Probability and Nonresponse Adjustment Factor

Percentiles	Probability of response	Wave 2 nonresponse adjustment factor

1%	0.73	1.05
5%	0.79	1.07
10%	0.82	1.09
25%	0.85	1.11
50%	0.88	1.14
75%	0.90	1.18
90%	0.92	1.22
95%	0.93	1.26
99%	0.95	1.36

The last step of weight construction is to censor the extreme weights to reduce their influence on the sample estimation of the population statistics. The weights were top coded and bottom coded at 99th and 1th percentile respectively. The child level weight has been created and stored in the variable child_weight_W2 (normalized weight) and child_raw_weight_W2 (raw weight). According to PSID CDS II user guide, the household level weights are constructed by taking the mean of the W2 child weights of each child observation provided by a caregiver. The household-level weights are stored in HH_weight_W2 (normalized weight) and HH_raw_weight_W2 (raw weight).

Table 4 provides a weighted comparison of some basic demographic, geographic, and socioeconomic variables between Wave 1 sample (weighted by Wave 1 weights) and Wave 2 sample (weighted by Wave 2 weights). AS shown both the household level and child level weighted distribution of Wave 2 sample is close to the Wave 1 sample. It suggests that the Wave 2 attrition adjustment factors used to construct the weights help to compensate for potential attrition bias in the family type and demographic composition of the SG LEADS panel data. We should also note that this comparison does necessarily not rule out the possibility of spurious or more subtle forms of selection bias that may not be associated with the demographic, geographic and socio-economic characteristics of SG LEADS respondents.

Table 4. Weighted Comparison of Selected Variables Between Wave 1 Sample and Wave 2 Sample

	SG LEADS Wave 1 (2018-2019)		SG LEADS Wave 2 (2021)	
	n	weighted %	n	weighted %
Household level				
Planning region	3,476	100.0%	3,017	100.0%
Central	680	19.2%	591	18.7%
East	417	13.2%	363	13.8%
North	571	14.9%	508	15.0%
North-East	864	29.3%	721	28.8%
West	944	23.5%	834	23.9%
Dwelling type	3,476	100.0%	3,017	100.0%
HDB 1- to 2-room flats	388	3.7%	314	4.1%
HDB 3-room flats	957	12.4%	827	12.2%

HDB 4-room flats	1097	36.9%	962	37.0%
HDB 5-Room and Executive Flats	550	28.4%	493	28.4%
Condominiums	428	15.8%	376	15.4%
Landed Properties	56	2.9%	45	3.0%
Education of the household head	3476	100.0%	3,017	100.0%
Secondary and Below	967	23.9%	821	23.9%
Post-Secondary	1143	29.4%	980	29.7%
University	1366	46.7%	1216	46.4%
Race of the household head	3,476	100.0%	3,017	100.0%
Chinese	2187	67.1%	1923	66.5%
Malay	801	14.3%	677	15.2%
Indian	369	13.4%	314	13.1%
Others	119	5.1%	103	5.2%
Child level				
child's gender	5,005	100.0%	4,352	100.0%
boy	2,518	51.2%	2179	51.4%
girl	2,487	48.8%	2173	48.6%
child's race	5,005	100.0%	4,352	100.0%
Chinese	3140	62.7%	2762	67.6%
Malay	1269	25.4%	1083	17.2%
Indian	454	9.1%	382	10.3%
Others	142	2.8%	125	5.0%