

IARIW-World Bank Conference
November 7, 2019

Effects of Data Collection Methods on Estimated Household Consumption and Poverty, and on Survey Costs: Evidence from an Experiment in the Marshall Islands

Michael Sharp and Bertrand Buffiere (SPC)
Kristen Himelein (World Bank)
John Gibson (University of Waikato)

SURVEYS IN THE PACIFIC FOR POVERTY WORK

- Except for 1996 PNGHS that used closed interval recall, all other poverty work for the 21 countries covered by the Secretariat of the Pacific Community is from Household Income and Expenditure Surveys (HIES) that use open form 14-day expenditure diaries
- Very costly to field and process
 - Mean cost of US\$650 per household based on the planned sample size and even higher per completed household with usable data, due to problems of incomplete diaries
 - Max cost was 2009/10 PNG HIES, \$15m total, \$4100 per HH
 - Long gap, of 9 years on average, between surveys due to need to mobilize (donor) resources so limits value for poverty monitoring
 - Surveys sometimes abandoned due to running out of funds (e.g. PNG Urban Household Survey from early 1990s)

IN THEORY IS OPEN QUESTION AS TO WHETHER DIARY OR RECALL SHOULD DO BETTER

For **recall surveys**, data quality can be negatively impacted by:

- **Omission** : Respondents forgetting that certain items were purchased during the recall period.
- **Telescoping** : Respondents “telescoping” purchases into the recall period that in actuality occurred before the start.

For **diary surveys**, data quality can be negative impacted by:

- **Omission** : Respondents forgetting that certain items were purchased.
- **Respondent fatigue**: Respondents becoming disinterested in the diary process and failing to record transactions.
- **Interviewer fatigue**: Interviewers becoming lax about monitoring.
- **Inability to comply**: Respondents not being able to complete the diary due to a structural limitation (such as literacy).

BUT IN PRACTICE, GLOBAL ATTEMPT TO TRY TO MOVE AWAY FROM DIARY-KEEPING SURVEYS

From the 2018 FAO guidelines on “Food data collection in household consumption and expenditure surveys” :

For [HIES], a **seven-day recall period** for food consumption measurement should be adopted. In low- and middle-income countries, **recall surveys are generally preferable to diary surveys**, which should only be deployed with careful and continuous supervision and should not exceed 14 days. While a well-implemented diary is generally held as the gold standard for food expenditure data collection, there is ample evidence that in low-income settings with a prevalence of illiterate respondents, **diaries** are often implemented as a series of short recall interviews, with issues of **respondent and enumerator fatigue** affecting data quality, and with **unsustainable implementation costs**.

IS FATIGUE AN ISSUE IN THE PACIFIC?

Very definitely!
Example: 2009/10 PNG HIES

- Fieldwork staggered start over days of the week and months of year
- Across all households, 37,000 transactions (acquisitions) are listed on day 1 (10 per household)
- By day 14 fallen to 23,000 or 6 per household with no apparent bundling into fewer, larger, transactions
- 3.4% per day fall in number of transactions and 4.4% per day fall in the value of transactions

Households look poorer the longer we observe them.

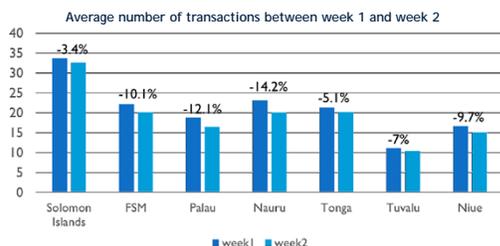
- Headcount poverty rate of 41% if use week 1 diary records, but 47% if use week 2 records
- The fall in compliance will vary with the opportunity cost of time, and so introduces non-random errors in poverty profile.

DIARY FATIGUE – EVEN WITH INTERVIEWERS LIVING IN VILLAGE FOR 3 WEEKS TO MAKE MULTIPLE VISITS

Diary-Keeping Day	Number (LHS) ('000s)	Value (LHS) ('000s)	Average Value (RHS) (Kina)
1	37	25	0.7
2	32	22	0.68
3	30	20	0.67
4	29	19	0.66
5	28	18	0.65
6	27	17	0.64
7	26	16	0.63
8	25	15	0.62
9	24	14	0.61
10	23	13	0.60
11	22	12	0.59
12	21	11	0.58
13	20	10	0.57
14	19	9	0.56

DIARY FATIGUE IN REST OF THE PACIFIC

- While not as dramatic, across other recent HIES in the Pacific there is a median second-week decline of 10% in the number of transactions recorded in the diaries



NOT JUST A COST PROBLEM: HIES ARE NOT WELL DESIGNED FOR POVERTY WORK

- Multiple users of HIES data, including for the CPI and SNA
- Expenditure weights for the CPI has been a dominant consideration (even though Pacific stats offices lack wherewithal to do price surveys outside a few urban areas so a small, urban-focused diary-keeping HIES could be used for the CPI with a more appropriate rural-focused living standards survey for poverty) which leads to:
 - Acquisitions-based record of transactions
 - Consumption of food and other frequently consumed items covered by the diary is obtained as a residual**
 - This entails measuring starting and ending food stocks, which is an idiosyncratic feature of the diary-keeping HIES in the Pacific, not typically found elsewhere

DIRECT VS INDIRECT MEASUREMENT OF HOUSEHOLD CONSUMPTION

- While recall collects consumption directly by asking how much of item x was consumed during the recall period, diary surveys obtain consumption indirectly by measuring inflows and outflows.

$$\begin{array}{r}
 \text{Purchases} \\
 + \text{ Own-production} \\
 + \text{ Gifts received} \\
 - \text{ Sales} \\
 - \text{ Gifts given} \\
 - \text{ Stock increases} \\
 \hline
 = \text{ Consumption.}
 \end{array}$$

- In order to separate expenditure from consumption, it is necessary to collect the **opening and closing stocks** for major food items.

HOW DIFFICULT IS IT TO MEASURE OPENING AND CLOSING FOOD STOCKS?

- Households are **more compliant for measuring stocks on day 1** compared to day 14.
- End stocks seem to be lower than starting stocks even though no reason for net destocking because Pacific agriculture is non-seasonal, and relies on continuous harvest (but with high transactions costs due to travel)

Two examples

- 2009/10 PNG HIES tried to measure stocks of over 100 items.
 - Apparent destocking added 6% to the value of food consumption.
 - This 'extra' food consumption caused the headcount poverty rate to drop by four percentage points.
- 2012/13 HIES for Solomon Islands, calorie totals that included stock changes were 6% higher, on average, with apparent net destocking adding 170 calories per person per day.

THE 2018 RMI SURVEY EXPERIMENT (1)

- Fielded by the RMI stats office (EPPSO), design and analysis by SPC and World Bank, aims to inform choices about future HIES in Pacific
- 14-day diary, highly monitored (with interviewer visits every two days), with transactions recorded with pen and paper (PAPI), and coding and data entry by EPPSO after the field work was completed
 - 14-day diary, less monitored (with interviewer visits after each week), with transactions recorded with pen and paper (PAPI), and coding and data entry by EPPSO after the field work was completed
 - 14-day diary, highly monitored (with interviewer visits every two days), with transactions transcribed into the tablets using CAPI during each interviewer visit
 - 7-day single-visit unbounded recall, using a list of 102 food groups and 20 non-food groups, with data entered into the tablets using CAPI during the interview
 - 7-day two-visit recall, with an initial visit made to the household to indicate the start of the recall period (and to gather other data), using a recall list of 102 food groups and 20 non-food groups, with data entered into the tablets using CAPI during the interview.

THE 2018 RMI SURVEY EXPERIMENT (2)

- Some unknown mix of Arm 1 and Arm 2 reflects the *status quo*
- stats offices may hope that they are implementing a highly monitored diary but under field conditions it often devolves into the less monitored diary
 - Not enough money for transport for frequent visits by the interviewers to check on the diary
 - Not enough supervision resources
- Arm 3 uses same basic design as the *status quo* but uses CAPI
 - More expensive as implemented because costly interviewers transcribe from paper diaries into the CAPI
- Arm 4 reflects the FAO/World Bank guidelines of 7-day recall
- Arm 5 redundant as didn't test bounded recall

THE RMI EXPERIMENT REFLECTS FIELD CONDITIONS – (WARTS AND ALL)

- Researcher-driven survey experiments may be less informative about actual surveys because they are too tightly monitored/controlled
- The RMI experiment was designed and fielded by the agencies who actually carry out HIES in the Pacific, and as such not all went to plan:
 - Was meant to test bounded recall (to deal with telescoping) but wording stayed as “last seven days” while gap between the two visits varied (24% 7 days, 53% 8 days, 9% 9 days, 14% other)
 - Interviewers may have had too much freedom to allocate sample to arms, and in allowing reports of zero consumption (separately from refusals) or extremes of apparent food intake
 - CAPI could have constrained interviewers more
 - Concordances of stocks to COICOP and food recall groups was not in-built at the design stage

12

DIFFERENT SURVEY ARMS ONLY AFFECT PART OF HOUSEHOLD CONSUMPTION

In reality, even diary surveys collect only part of the consumption and expenditure information through diaries. Low frequency transactions, such as the purchase of assets and durables, payment for utilities, home improvements, etc. are collected through recall.

Food purchases, food away from home, high frequency transactions

Diary

Recall

Low frequency transactions, asset purchases, durables, services

Recall

Calculated values: Imputed rental value, “use value” of durables

Calculation

13

LOCATION OF THE EXPERIMENT

- RMI has both remote islands and densely populated atolls, covering range of conditions found in Pacific (except high mountain areas in Melanesia)

RMI due to have official HIES in 2019, so experiment is to directly inform the choice of survey method (as it also did for Kiribati)

14

FIELDWORK ORGANIZATION

- Each interviewer specialized in a particular survey arm within a round, arms fielded simultaneously in the same EAs; either 6 or 18 HHs per interviewer
- Rotated arms between rounds, so got to experience four of the five arms

Interviewer ID	Round 1	Round 2	Round 3	Round 4
	13-Jul to 31-Jul	6-Aug to 26-Aug	27-Aug to 16-Sep	17-Sep to 7-Oct
Team 1: Ebeye	11 Arm 1	Arm 2	Arm 3	Arm 4
	12 Arm 2	Arm 3	Arm 4	Arm 5
	13 Arm 3	Arm 4	Arm 5	Arm 1
	14 Arm 4	Arm 5	Arm 1	Arm 2
	15 Arm 5	Arm 1	Arm 2	Arm 3
Team 2: Majuro	21 Arm 1	Arm 2	Arm 3	Arm 4
	22 Arm 2	Arm 3	Arm 4	Arm 5
	23 Arm 3	Arm 4	Arm 5	Arm 1
	24 Arm 4	Arm 5	Arm 1	Arm 2
	25 Arm 5	Arm 1	Arm 2	Arm 3
Team 3: Outer islands	31 Arm 1	Arm 2	Arm 3	Arm 4
	32 Arm 2	Arm 3	Arm 4	Arm 5
	33 Arm 3	Arm 4	Arm 5	Arm 1
	34 Arm 4	Arm 5	Arm 1	Arm 2
	35 Arm 5	Arm 1	Arm 2	Arm 3

- Surveyed interviewers six months after the experiment to get their impression on the conduct of the different arms

15

SURVEY COSTS (1)

Highly monitored diaries cost over \$3000 per HH for rural EAs

- Expensive transport and per diem
- Over 5-times as expensive as the recall survey
- Less monitored diaries had total costs double that of recall

16

SURVEY COSTS (2)

Very expensive transport (flights/boats to outer atolls) and labour in Pacific

- Also calculated more ‘stylized costs’ due to time requirements for each module that may be more transferrable to other settings
- Includes field costs, printing, coding and data entry
- Highly monitored pen-and-paper diaries cost 4.4 times as much as single visit recall
 - CAPI diaries, as implemented, would be five times as expensive due to slow and expensive transcription from paper diary to CAPI
- Less monitored diaries cost 2.8 times the cost of single visit recall
- Survey of interviewers six months after fieldwork
 - Could handle more than 18 households per round doing single visit recall, struggled with even 6 households in highly monitored diaries
 - Cost ratios could be even wider if recall fieldwork optimized
- Majority of interviewers (and majority of households according to the interviewers) preferred single visit recall

17

WHAT DO OUR RESULTS SHOW?

Overview of the findings:

- Cost of highly monitored diaries is up to 5-times that of single recall
 - 2-3 times as costly for less monitored diaries
- Even for the highly monitored diary approaches, we see a **decline in the number of items reported** in the diary over the course of the 14 days.
- Across the two bounds on *status quo* diaries we get 3.4% daily rate of decline in number of transactions reported (same rate as for PNG HIES)
- Effective completion rate of only 68% for less monitored diaries, while all other arms have 82-88% completion rates
- Non-response affects sample balance so use raking weights to restore
- Based on the raking weights, 4 out of 5 methods (excluding less monitored diary) show similar results for consumption and poverty.
 - Consumption lower, poverty higher, with less monitored diary

18

RESPONSE AND COMPLETION RATES

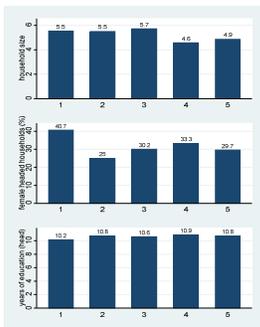
	targeted	replacement rate	completed	no food consumption	analyzable interviews	effective completion rate
highly monitored PAPI diary	72	17.3%	75	16	59	81.9%
low monitored PAPI diary	216	19.9%	181	33	148	68.5%
highly monitored CAPI diary	72	18.8%	64	1	63	87.5%
CAPI 1 visit recall	216	18.6%	199	19	180	83.3%
CAPI 2 visit recall	216	19.3%	197	13	184	85.2%
Total	792		716	82	634	80.1%

19

INCOMPLETIONS AFFECT BALANCE

Despite randomization, after the dropouts/replacements find:

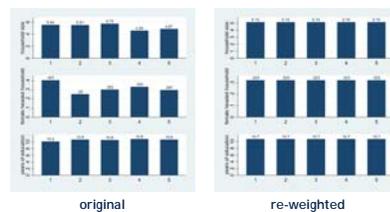
- Household sizes are systematically larger among diary households.
- There is a very low number of female headed households in the low monitored diary arm
- fewer differences in education.



20

RAKING WEIGHTS

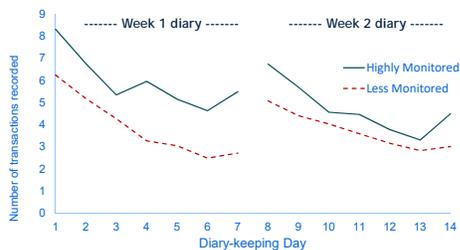
By using re-weighting techniques, we can force certain characteristics to align. In this case we used household size, head literacy, gender of head, years of education for head, broad categories of employment for the head, and the component of the consumption aggregate calculated from survey data (i.e. which does not vary between arms).



21

CONSIDERABLE DIARY FATIGUE WITH THE STATUS QUO DIARY SURVEYS

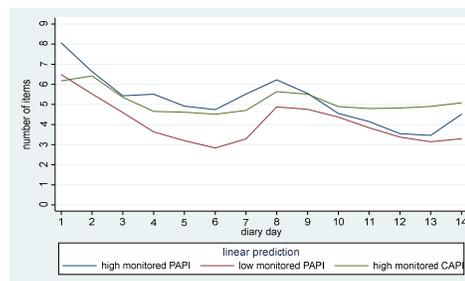
For highly monitored, fall in number of transactions recorded from 8.3 on day 1 to 4.5 on day 14, and fall from 6.2 to 3.0 for less monitored diary. Overall, trend decline in # of transactions of 3.4% per day (same as in PNG) and 35% fewer transactions in the less monitored diary



22

PARTIAL LINEAR MODEL WITH CONTROLS FOR DAY-OF-WEEK AND LOCATION

We see these declines even if we use econometrics (partial linear regression) to remove the effects from the day of the week and the location of purchase.



23

RESULTS FOR FOOD AND NON-FOOD FREQUENT (DESIGN-SENSITIVE ITEMS)

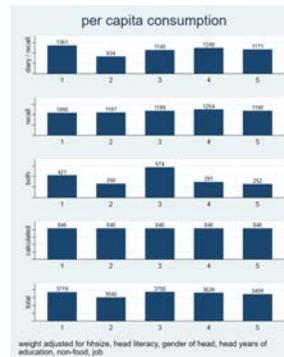
- Less monitored diaries give average consumption just 61% of what the more monitored diary yields
 - And 67% of what single visit recall yields
- Across the two highly monitored diaries, the average value of food and frequent non-food consumption is \$1247
 - Identical to average (\$1248) from single visit recall, despite much greater resource intensity of the diaries

Form a total consumption aggregate from the infrequent (all recall, mixed period), imputed (durables and rents) and a few items that can come from both diaries and recall (following approach SPC use elsewhere)

24

PUTTING ALL COMPONENTS TOGETHER

Less monitored diary gives average per capita consumption of \$3050
 Just 82% of what highly monitored diary (and 84% of single recall) would give
 Across PAPI and CAPI highly monitored diaries, average p.c. consumption of \$3740 is just 2.6% above what single visit recall gives



- Legend
- High monitored PAPI diary
 - Low monitored PAPI diary
 - High monitored CAPI diary
 - CAPI 1 visit recall
 - CAPI 2 visit recall

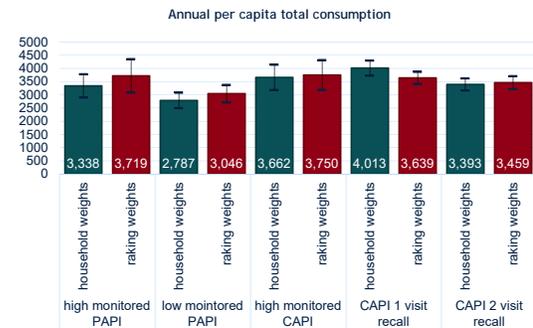
CONSUMPTION FROM DIARY ARMS IS BOOSTED (WRONGLY) FROM STOCKS

Apparent consumption from diary-keeping arms overstated due to the apparent destocking, which really reflects fatigue

- Of the n=270 diary-keeping households with analysable results, 236 have starting food stocks reported but only 211 have ending food stocks reported
 - Of those with both starting and ending stocks measured, 2:1 ratio of having starting stocks seeming to be larger than ending stocks
- Average contribution of the apparent destocking is equivalent to about 4% of total expenditure
 - Removing this component, which is likely due to fatigue and declining compliance, makes poor performance of the low monitored PAPI diary even more apparent (equivalent to almost 400 calories/person/day)
 - C.f. apparent daily per capita calorie contribution of about 130 for the CAPI diary and 220 for highly monitored PAPI diary

26

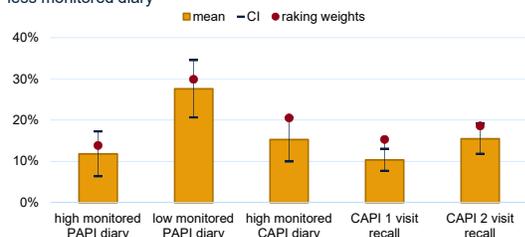
WHAT ARE THE FINAL NUMBERS?



27

WHAT ARE THE IMPLICATIONS FOR NATIONAL POVERTY MEASUREMENT?

if we use a relative poverty line, set as half of the median real per capita consumption, find poverty rates would not differ between recall and highly monitored diary arms, but much higher apparent poverty with the less monitored diary



28

CONCLUSIONS

- Existing HIES used for poverty work in the Pacific are all based on open-form expenditure diaries kept for 14-days
- Costly, unsustainable, infrequent, and give data of uncertain quality
 - Respondents look poorer the longer they are observed, due to the diary fatigue (even with offsetting error of apparent destocking)
- In RMI experiment designs that reflect the status quo are from four to five times as costly (if highly monitored) or two to three times as costly (if less monitored) than a single visit recall survey
- Despite the extra cost of the diary-keeping surveys, they yield data that are, overall, of worse quality. The effective completion rate with less monitored diaries is only two-thirds and apparent consumption is significantly lower (and poverty higher) compared to all other modules. The highly monitored diaries give similar results to using recall, but at much greater cost.

29