Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

Data quality from a mobile app survey to collect expenditure data as part of a large-scale probability household panel survey

Carli Lessof
Annette Jäckle
Mick Couper

ESRA, Lisbon 23, July 18 2017

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public

Qualification

- This presentation was given at ESRA 2017
- All findings are based on early analysis and are subject to change
- Please email cl19g15@soton.ac.uk for updates

Acknowledgements

- ESRC Transformative Research and NCRM support

- Annette Jäckle (PI), Jon Burton, Paul Fisher, Mike Brewer, Alessandra Gaia (ISER)
- Tom Crossley (Economics, Essex)
- Mick Couper (Michigan)
- Cormac O'Dea (IFS)
- Joachim Winter (Ludwig-Maximilian, Munich)

- Kantar Worldpanel and Kantar Public UK
- European Central Bank

- Patrick Sturgis and Dave Martin (Southampton)

Project aims

- Improve survey measurement of household finances
- Develop ways to capture the full household budget

  (income - expenditure = changes in assets and debts)
- Produce new evidence on outstanding puzzles in household finance

  Who is poor?

  Do the rich really save more?

Activities

Components of household finances

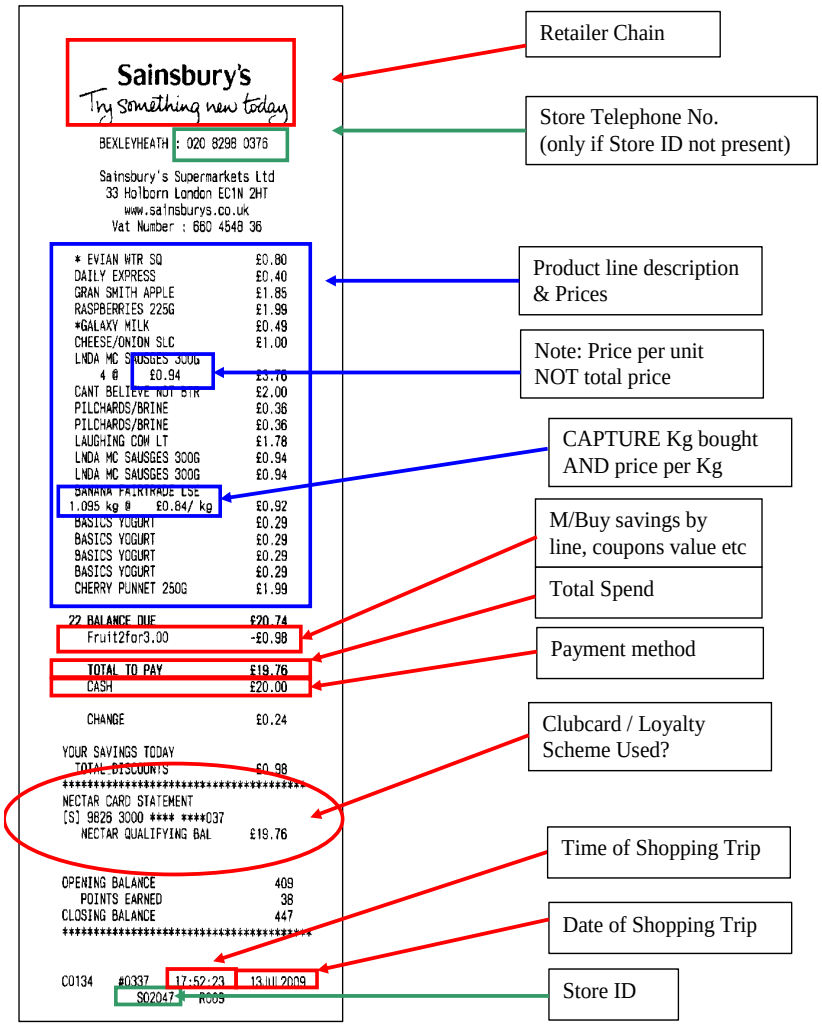
1. Review literature and practice
2. Experiment on Innovation Panel Wave 9 (IP9)
3. Test mobile technologies to reduce reliance on recall

Improve estimation of full household budget

1. Data combination and imputation
2. Budget reconciliation survey

Spending survey

- October 2016-January 2107
- IP sample (issued c.2400)
- Invite between IP9 & IP10
- App to own smartphone/tablet
- Record spending for 1 month
- Scan, direct entry or ‘no-spend’
- £3/£6 to download and 50p daily
- Notification reminders at 5pm
- End of week survey sent Sundays
- End of project survey
- Kantar transcribed receipt data



Evaluation of quality of data

- No direct comparison
- Compliance with the process and consistency over time
- Compare quality indicators against other approaches
- Self-reported accuracy
- Compare spending against benchmarks
- Acknowledge the limitations – and learning

Compliance with the process over time

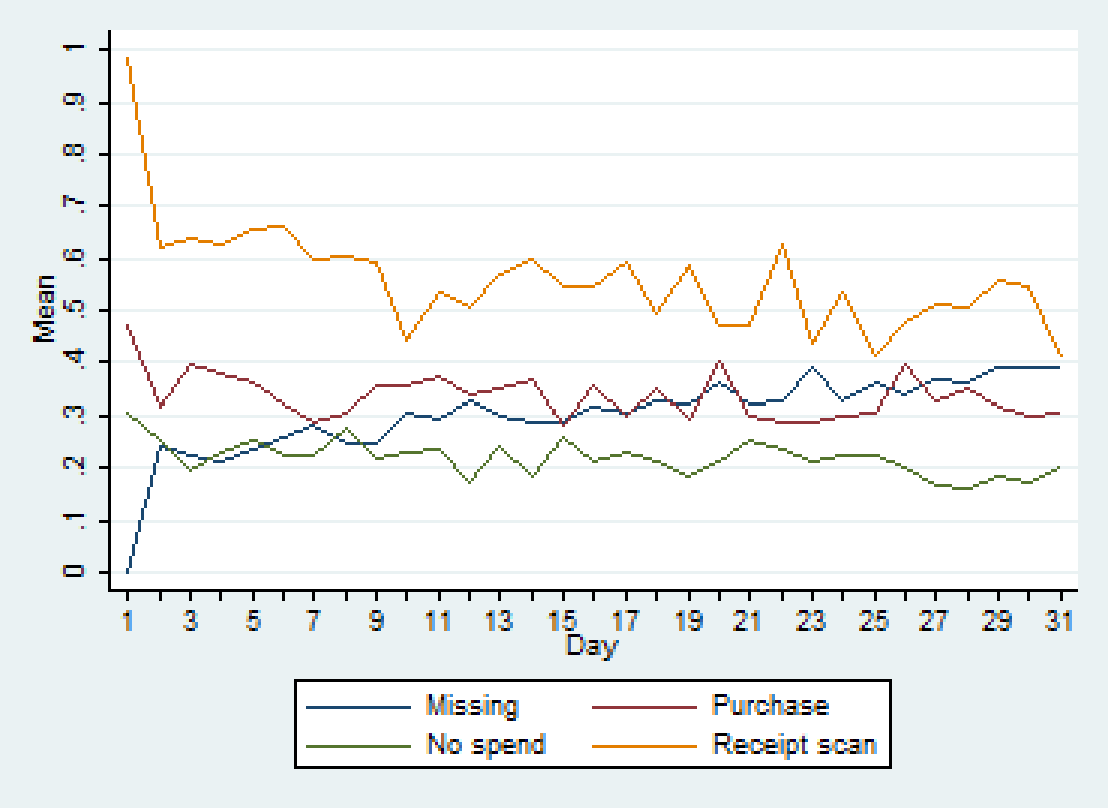
1. How do respondents make app entries?
   a) Do the number of app entries fall?
   b) Is there a shift between scanning and direct entry?
2. What is the pattern of reported spending?
   a) Does the pattern of reported spending change over time?
   b) Is there a shift between receipt scanning and direct entry?
3. How quickly do Rs report spending?
   a) Does it change over time?
4. How much processing error do we observe?
5. How do Rs assess their own accuracy?

1: How do respondents make app entries?

- 270 participants over 31 days
- An average of approximately 35 app entries
  - 50% scanned receipts (c17 entries)
  - 30% directly entered (c10 entries)
  - 20% “no spend” days (c7 entries)
- The 31 days can be characterised, in aggregate
  - 15 days reported purchases (48%)
  - 7 days reported as no spend days (22%)
  - 9 days are missing (30%)
- Generates 5541 receipts and 37,760 items

1a: Do the number of app entries fall?

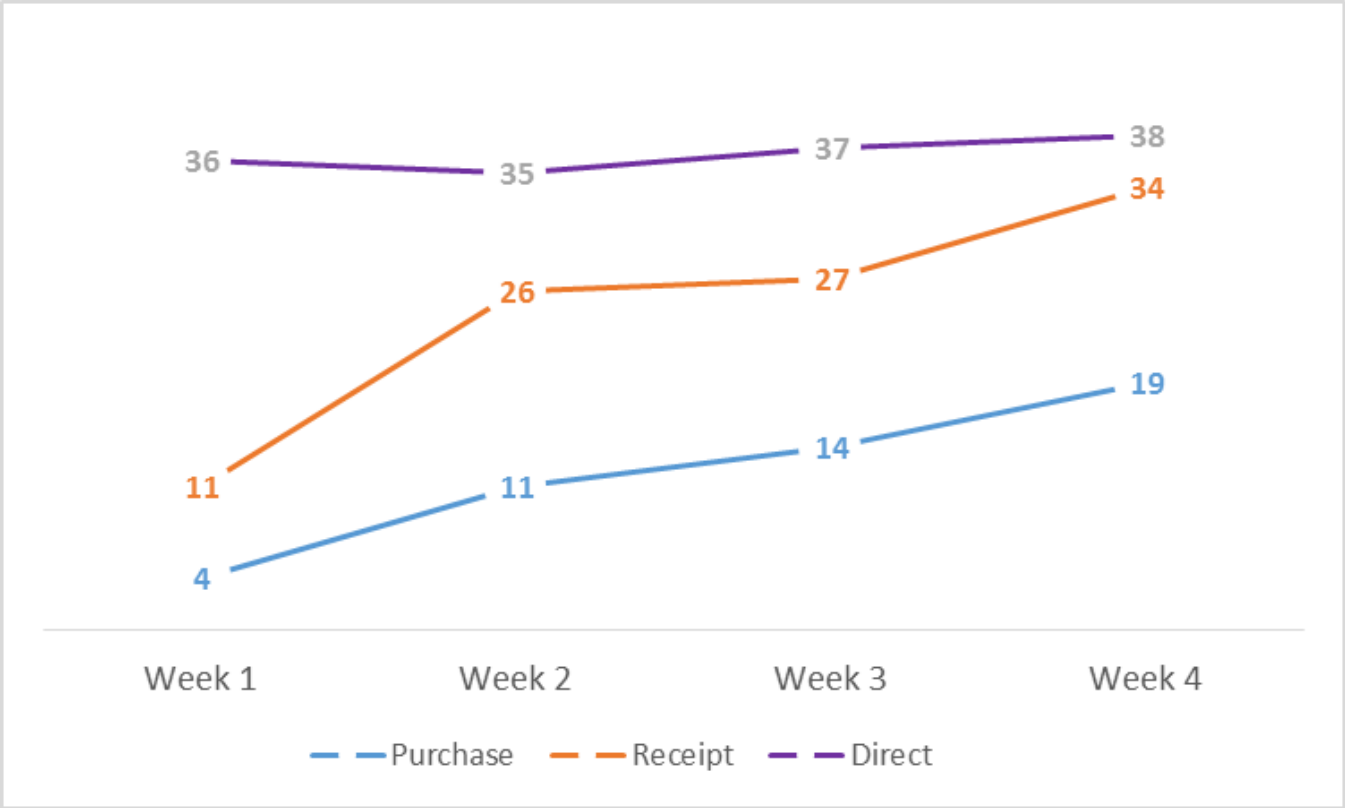
Mean number of receipts, purchases, no spend and missing days



	Day 1-31 (Day 2-31)
All entries	-0.012*** (-0.008***)
Receipts	-0.007*** (-0.005***)
Direct:	-0.002* (-0.001)
No spend	-0.002*** (-0.002**)
Missing	0.007 (0.005)

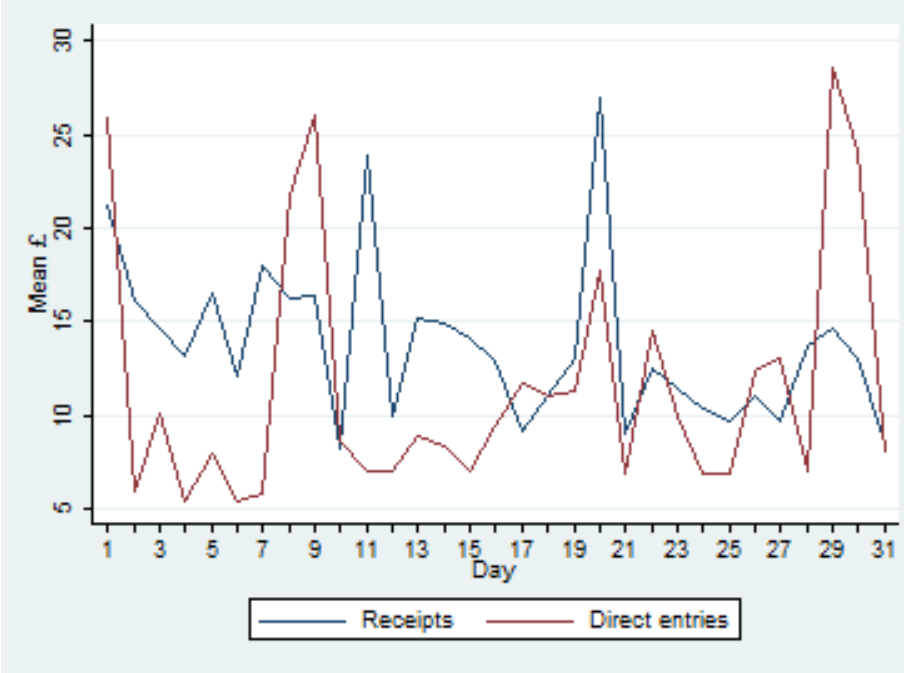
1b: Is there a shift between scans and direct entry?

% participants not reporting a purchase each week

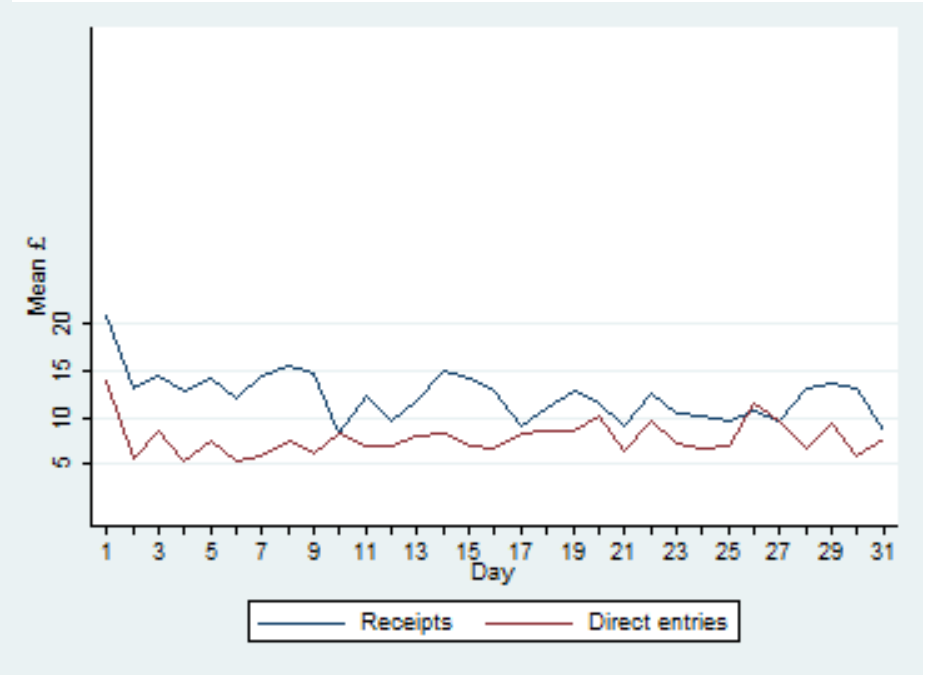


2a: Does the pattern of reported spending change over time?

(a) Means (all)



(b) Mean (values<£250)



(b) Total £ (b=-0.126* SE 0.055) and receipt £ (b=-0.146*** SE=0.041) Day1-31

Excluding day 1 reduces effect: slight fall in receipt £ (b=-0.103*, SE=0.423)

2b: Is there a £ shift between scans and direct entry?

- Model amount reported using receipt scanning
- Small effect of amount of direct spend (b=0.031)*

- Predominantly driven by app activity or inactivity

  missing days (b=-30.80, CI= -28.90, -32,70)***

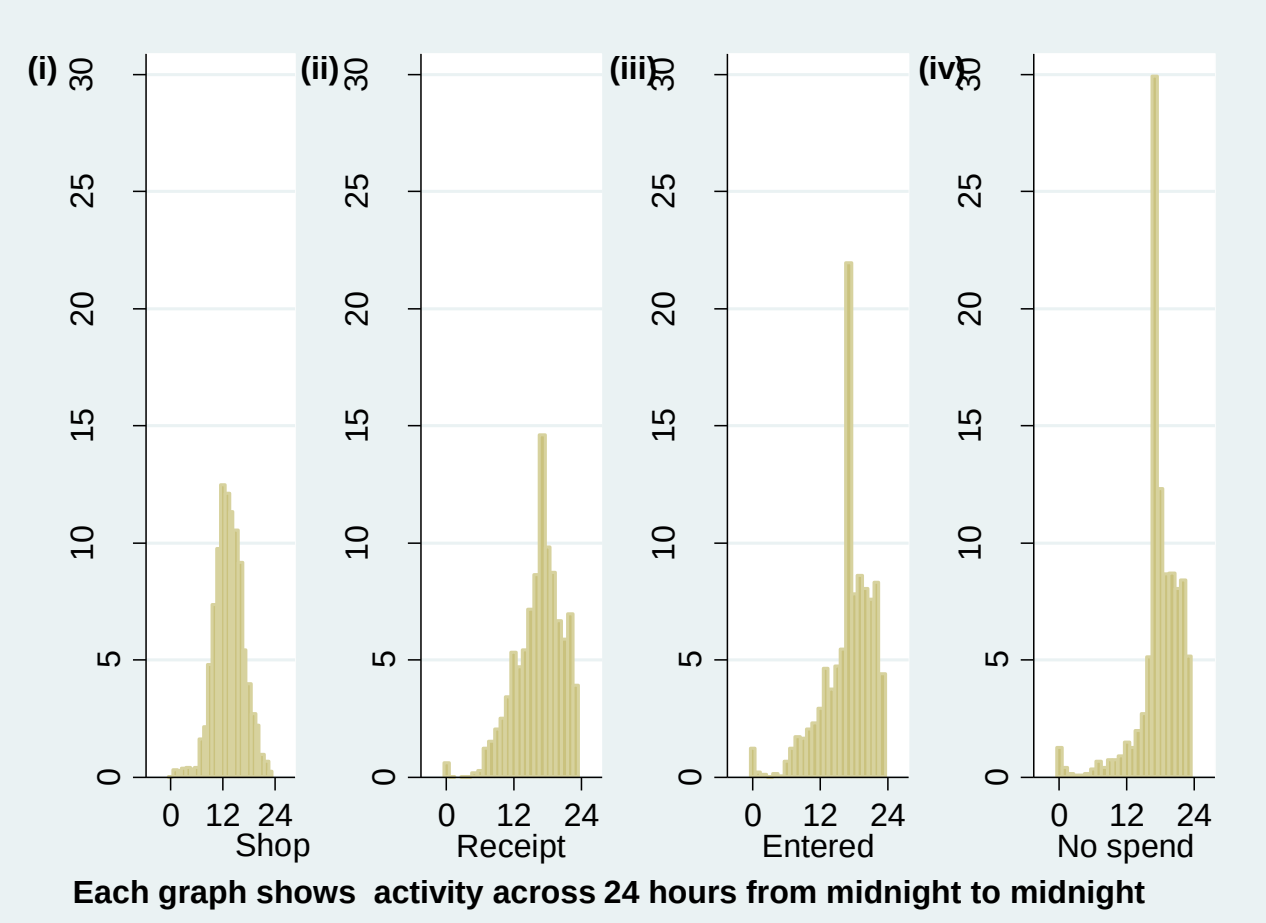
  no spend days (b=-26.97, CI= -25.19,-28.76)***

  direct entry days (b=-11.15, CI= ,-9.84,-12.46)***

- Further work to follow

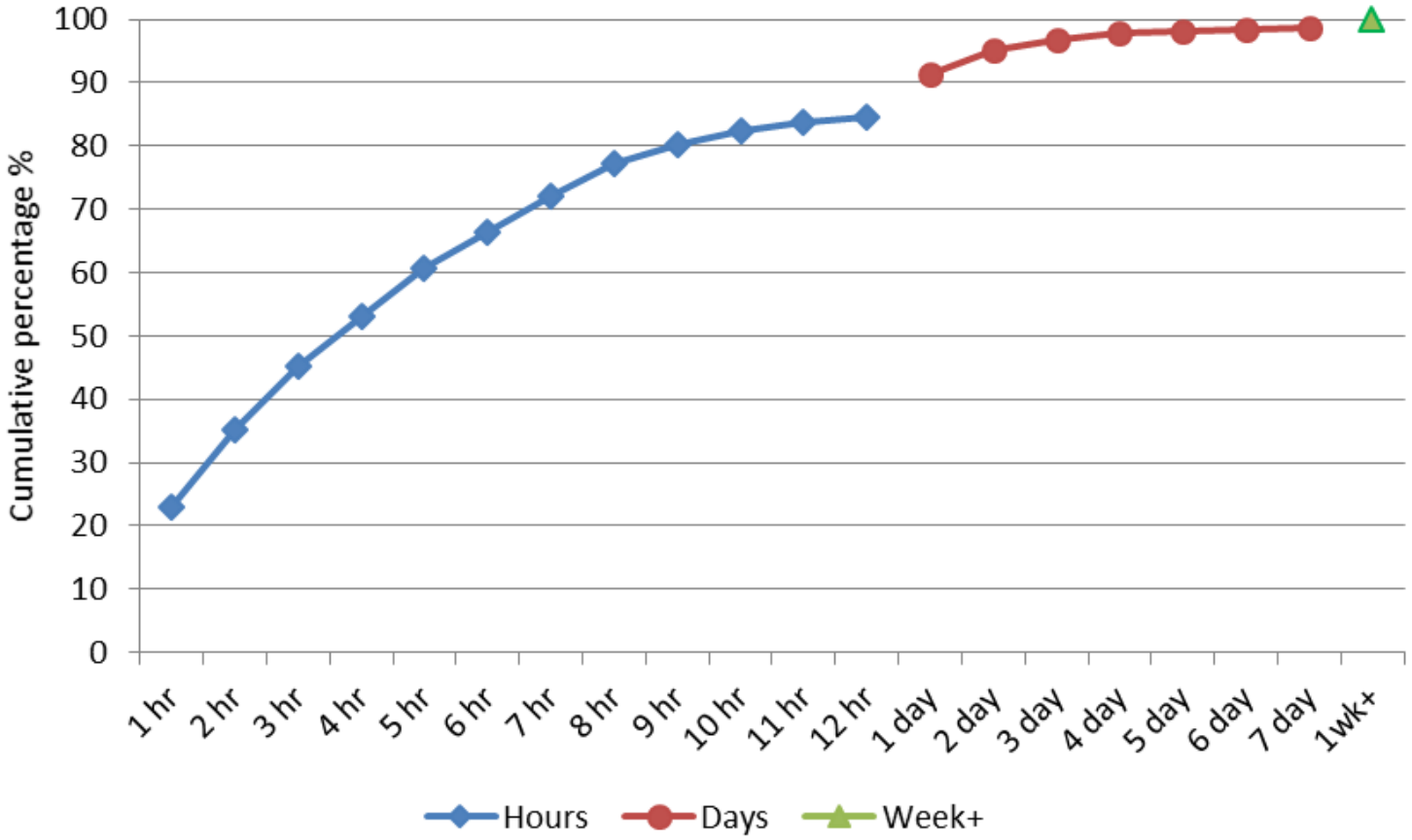
3: How quickly do Rs report spending?

In aggregate, when people shop (i) and enter data in the app (ii- iv)



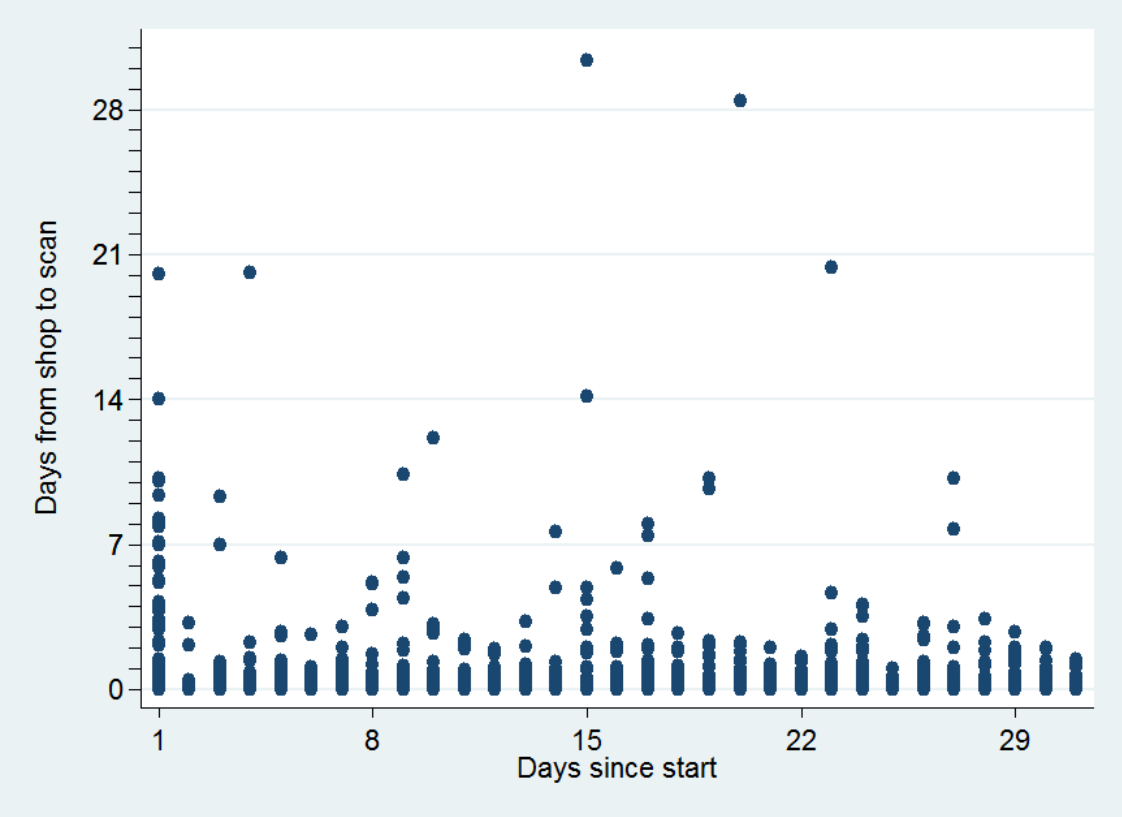
3: How quickly do Rs report spending?

For scanned receipts only, time from shop to scan



3a: Does speed of reporting change over time?

Days between shop and scan, by day of entry



With day 1: respondents speed up (b = -0.158**, SE 0.058)

Without day 1: no change in speed (b= 0.021, SE 0.056)

4: How complete are the data?

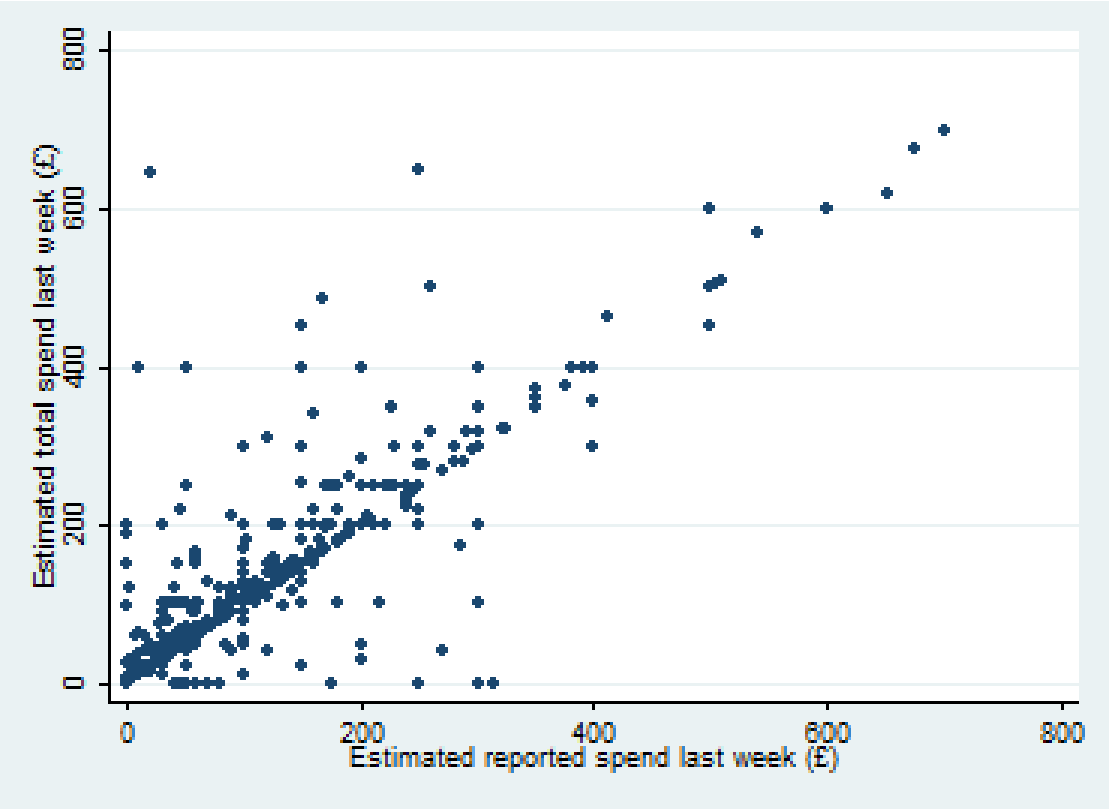
- Of 5541 app mentions of a receipt scan, 387 (7.5%) unavailable
- Of 5228 receipts transcribed, 74 (1.4%) not matched

- Time: 82% date and time, 3% date, 1% time, 14% neither
- Seller: 2.5% name missing but c.one third still code-able
- Categories: largely code-able, c5% ‘amount’ only but some info

- Further detailed work to do before drawing firm conclusions

5: How do Rs assess their own accuracy?

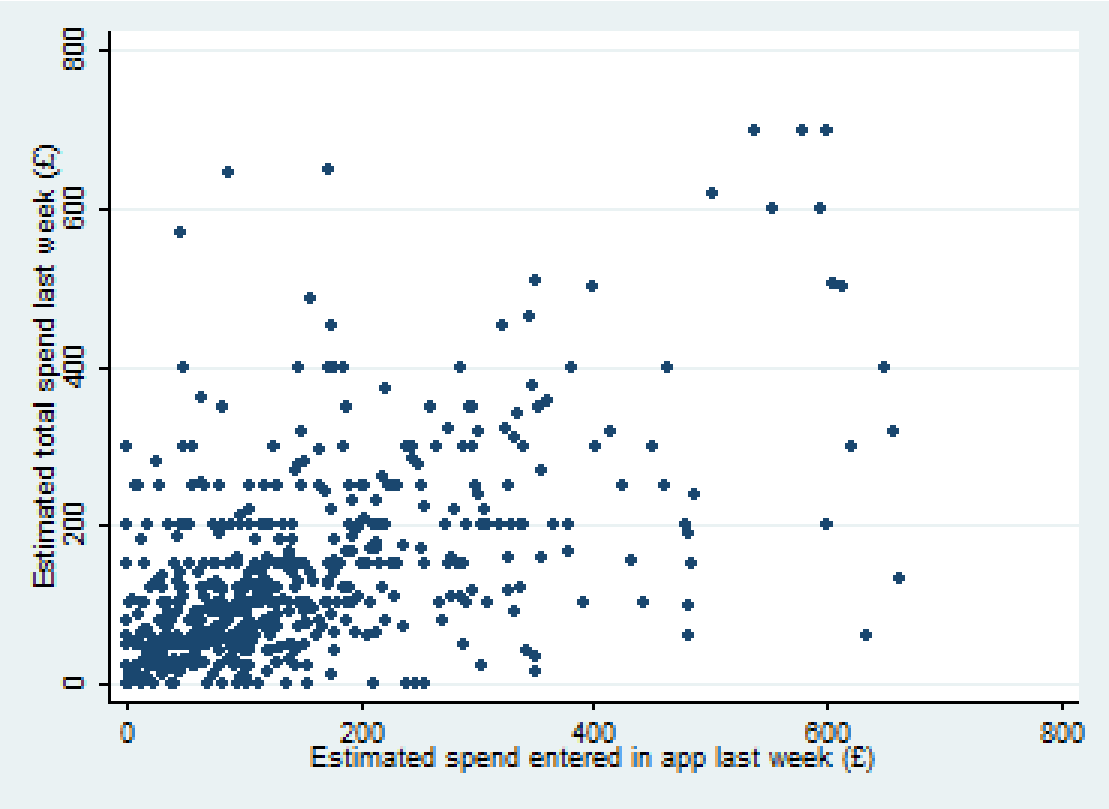
Reported spending v reported app entries



Accurate (52%), under-report (39%, £59), over-report (10%, £63),

5: How do Rs assess their own accuracy?

Reported spending v £ entered in app



Accurate (52%), under-report (39%, £59), over-report (10%, £63), +/- £20 (27%), under-report (28%), over-report (45%)

Next steps

- Sequence analysis to look at clusters of response patterns
- Latent class analysis to identify classes by quality
- Assess plausibility of amounts of spending reported
- Would quality increase or fall if response increased?

Interim conclusions

- Low take up but participants continue over time
- Some positive indicators
- Other warning signs e.g. missing days
- Receipt entry declines and is not recovered fully by direct entries

Understanding Society

THE UK HOUSEHOLD LONGITUDINAL STUDY

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public