The Use of Receipts in the Survey of Household Spending Diary

www.statcan.gc.ca

Tom Haymes
Denis Malo

July 2018
The Survey of Household Spending

- Annual, voluntary survey of approximately 17,500 households
- Redesigned in 2010
- Collects information on household expenditure using both a personal interview and an expenditure diary
The Survey of Household Spending

- Personal Interview
  - CAPI
  - Collects common or major expenses (rent, utilities, furniture, etc.) → about 70% of household consumption
  - Varying recall periods dependent on frequency of expense (12 month, 3 month, 1 month or last payment)
  - Approximately an hour in length
  - 65% interview response rate (2016)
The Survey of Household Spending

- Expenditure Diary
  - 50% subsample of original interview sample
  - Collects frequent expenditure and expenditure that would be difficult to recall during retrospective interview (food, gas, tobacco, alcohol, etc.)
  - Approximately 30% of household consumption
    - Receipts account for 12% of total household consumption
  - Two weeks in length
  - One diary per household
  - 43% overall diary response rate, 65% when including only interview respondents (2016)
Collection, Scanning and Capture of the Receipts

- Respondents can provide a combination of transcriptions and receipts
- Diary booklet and receipts are scanned into two separate files
- Booklet captured using OCR but receipts are currently captured manually from scanned image
  - Varying receipt formats initially prohibited auto capture, probably possible with recent technological advances
- Illegible (faded) receipts are rare → flagged for imputation
Capture of the Receipts

- Prior to capture all receipts are manually reviewed to ensure:
  1) Within diary reference period
  2) Transaction was approved
  3) No duplication
     - Within receipts (store receipt and debit slip)
     - Between receipts and transcriptions
       - Respondents are specifically reminded to submit a transcription or a receipt for each item, not both → still, 30% of receipts correspond to a transcription
Coding of Diary Items

- Coding assigns one of over 650 SHS codes to each item to classify the expenditure
- Automated process in place to match description to data dictionary containing common item descriptions with corresponding SHS code
  - Currently requires an exact match but method could be improved
  - Items not autocoded are coded manually
Coding of Diary Items

- Transcriptions have consistently autocoded better than receipts
  - Steady 9%-10% difference since 2014
- Likely due to exact match requirement
  - Receipts often contain more acronyms and abbreviations in the description
  - Receipts often include a brand name or volume/weight
- Room for improvement for both transcriptions and receipts, but especially receipts
  - Possible to close the gap?
## Respondent Reporting Profiles

### Diary Respondent Type

<table>
<thead>
<tr>
<th>Respondent Type</th>
<th>Diary Count</th>
<th>Diary Percent</th>
<th>Average Expenditure</th>
<th>Average Diary Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receipts and Transcriptions</td>
<td>7185</td>
<td>56.1%</td>
<td>$646.81</td>
<td>88.5</td>
</tr>
<tr>
<td>Transcriptions Only</td>
<td>4324</td>
<td>33.7%</td>
<td>$575.44</td>
<td>61.7</td>
</tr>
<tr>
<td>Receipts Only</td>
<td>1227</td>
<td>9.6%</td>
<td>$469.60</td>
<td>74.6</td>
</tr>
<tr>
<td>Empty Diaries</td>
<td>80</td>
<td>0.6%</td>
<td>$0.00</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12816</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>$601.73</strong></td>
<td><strong>77.6</strong></td>
</tr>
</tbody>
</table>

Source: SHS 2013 – 2015 post imputation

- Respondents that provide both receipts and transcriptions exhibit the most expenditure
Respondent Burden

Diary Reporting Mode

<table>
<thead>
<tr>
<th>Reporting Mode</th>
<th>Items</th>
<th>Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>914,785</td>
<td>$7,709,990</td>
</tr>
<tr>
<td>Transcription</td>
<td>391,889</td>
<td>$4,622,402</td>
</tr>
<tr>
<td>Receipt</td>
<td>522,896</td>
<td>$3,087,588</td>
</tr>
</tbody>
</table>

Source: SHS 2013 – 2015 post imputation

- Opposite splits for items and expenditure suggest respondents prefer to use receipts when there are a large number of smaller priced items
  - i.e. groceries (59% receipt expenditure) vs gasoline (24% receipt expenditure)
Respondent Burden

Response Fatigue by Reporting Mode

Source: SHS 2013 – 2015 post imputation
Respondent Burden

Response Fatigue by Respondent Type

![Graph showing average household expenditure over diary days for different respondent types: Receipts Only, Transcriptions Only, and Receipts and Transcriptions.](image)

Source: SHS 2013 – 2015 post imputation
Respondent Burden

- Response fatigue exists for both response modes and all three respondent types
- Contrary to expected results
  - Why isn’t response fatigue less evident in receipts?
  - Embedded experiment would give more reliable conclusions on the effects of receipts on expenditure reporting, response fatigue and response rates
Data Quality

Overall Imputation Rates by Reporting Mode

<table>
<thead>
<tr>
<th>Reporting Mode</th>
<th>Number of Items</th>
<th>Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Imputation Required</td>
<td>Imputation Required</td>
</tr>
<tr>
<td>Transcription</td>
<td>193,978</td>
<td>$1,620,951</td>
</tr>
<tr>
<td>Receipt</td>
<td>46,765</td>
<td>$399,062</td>
</tr>
<tr>
<td>Total</td>
<td>240,743</td>
<td>$2,020,013</td>
</tr>
</tbody>
</table>

Source: SHS 2013 – 2015 post imputation

- Transcriptions much more likely to give insufficient detail
- More uncertainty involved in transcription imputation too
  - Tend to be imputed from a higher level (i.e. totals)
Conclusion – Advantages of Receipts

- **Respondent:**
  - Reduces burden and increases flexibility

- **Agency:**
  - Significant gains in the level of information obtained from respondents
    - 35% of transcription expenditure required imputation compared to 13% of receipt expenditure (50% vs 9% for item imputation)
  - Possibility of increased participation? → Experiment required
Conclusion – Disadvantages of Receipts

- **Respondent:**
  - None!

- **Agency:**
  - Transfer of burden from respondent to agency
    - Manual capture of receipts → Auto capture possible?
    - Decreased auto coding rates → Better linkage methodology?
    - Increase in duplication
Conclusion

- SHS diary functions best with a combination of receipts and transcriptions
  - Transcription only diary would decrease data quality
  - Receipt only diary would underrepresent expenditure where receipts are uncommon
    - i.e. occasional babysitting, prepared coffee
    - Lost receipts
Thank You

- For more information, please contact:
- Pour plus d’information, veuillez contacter :
  
  Tom Haymes: [tom.haymes@canada.ca](mailto:tom.haymes@canada.ca)
  Denis Malo: [denis.malo@canada.ca](mailto:denis.malo@canada.ca)

- Corresponding document available upon request