Progress towards a table builder with in-built disclosure control for 2021 Census

Keith Spicer
UNECE, 22 September 2017
Contents

• Context: 2011 Census

• 2021 Census
  – Targeted Record Swapping
  – Outputs Package (Table Builder)
  – Cell key Method
  – Perturbing Zeros
2011 Census

• Targeted record swapping
  – Targeted to “risky” records

• Table redesign
  – Criteria of % 1s that are “real” and attribute disclosures that are “real”
  – Sparsity gives higher chance of disclosure
  – Sparsity also gives perception of disclosure
2011 Census

Every table had to be checked for disclosure

Timing was affected
......and sometimes........

I wanted this table

They gave me this table
2021 Census

• UK Parliament discussed:
• Aim for 2021 to be the last traditional census in England and Wales
• Look to use administrative and other sources to replace traditional census
• Parallel running of traditional and admin censuses in 2021 census round
2021 Census

User concern from 2011 in three areas:
Flexibility

Accessibility

Timeliness
Targeted Record Swapping I
Targeted Record Swapping II
Targeted Record Swapping III
2021 Census: Outputs Package

Aim to produce an outputs package

- Targeted record swapping
  - Swaps obviously identifiable people / households

- Cell key method
  - Protects all by uncertainty, and differencing

- User-defined tables from a table builder
  - Allows tables quickly
### Cell Key Method

1. Assign each record a random number

<table>
<thead>
<tr>
<th>Record</th>
<th>Rkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_1 \rightarrow )</td>
<td>54</td>
</tr>
<tr>
<td>( r_2 \rightarrow )</td>
<td>4</td>
</tr>
<tr>
<td>( r_3 \rightarrow )</td>
<td>93</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( r_N \rightarrow )</td>
<td>26</td>
</tr>
</tbody>
</table>

2. For each cell, sum rkey and apply a function to get a cell key

<table>
<thead>
<tr>
<th>Age by sex</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-15</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>16-24</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>25-34</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Record Rkey

| \( r_2 \rightarrow \) | 4 |
| \( r_4 \rightarrow \) | 61 |
| \( r_{56} \rightarrow \) | 7 |
| \( r_{72} \rightarrow \) | 90 |

Sum = 162

E.g. take last two digits \( \rightarrow \text{Ckey} = 62 \)

3. Use a look up table to get perturbation value

<table>
<thead>
<tr>
<th>Cell Key</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>61</th>
<th>62</th>
<th>63</th>
<th>...</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Apply pvalue to cell

<table>
<thead>
<tr>
<th>Age by sex</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-15</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>16-24</td>
<td>.</td>
<td>5</td>
</tr>
<tr>
<td>25-34</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Notes for Cell Key Method

- Adapted from “ABS method”
- Method primarily for protecting against differencing
- We are looking at a light touch (record swapping still the primary approach)
- Considering the need to retain 1s and 2s in outputs
- Introduces another layer of uncertainty for intruder
- Consistency in same cell across tables
- Some inconsistencies in breakdowns
Perturbing Zeros I

• Work in progress
• Additional n 1s perturbed to 0s
• Can balance by perturbing n 0s to 1s
  – Counts therefore unbiased
  – Increases protection through uncertainty
  – Means that a 1 does not necessarily represent a record in the microdata
  – Must ensure not changing ‘structural zeros’
Perturbing Zeros II

• Methods for this are being assessed.
• Example:
• Zero cells need a ‘cell key’ but there are no records or record keys
• Each variable has a set of category keys
• Combination (sum) of category keys = cell key
• (any structural zero; cell key set to 0)
• n highest cell keys are perturbed from 0 to 1
Key Points

• Aim to have Table Builder for users
• Targeted Record Swapping
• Cell Key Method
• Try to retain small cell counts
• Benefits of this approach to other collections

• Work continuing.................
• Other areas: microdata products, origin-destination tables, admin census
Questions and Discussion