Census Statistical Disclosure Control
- The use of record-swapping to protect data confidentiality

Census Division
PAMS Conference 28 November 2013

Balancing the 'rights of privacy'
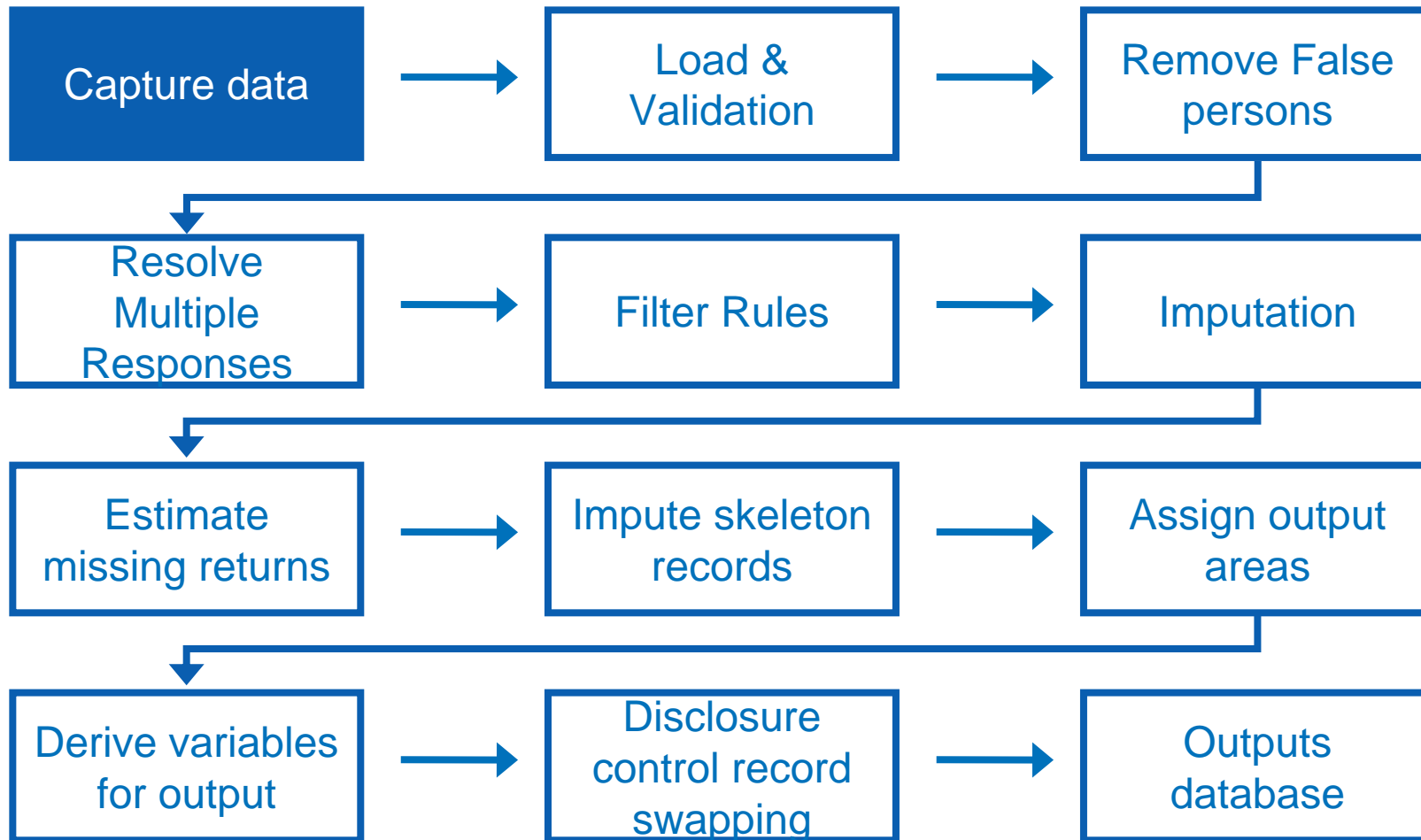versus the 'need to know'

28 November 2013

National Records of Scotland

Scotland's Census 2011
Shaping our future

Statistical Disclosure Control (SDC)

▶ **Statistical Disclosure Control** - aims to **reduce risk** of releasing confidential information directly or by deduction

▶ Can be achieved by **controlling access to data**, for example using safeguarded research settings. **Other methods** can include:

   – **Perturbation** of microdata or **adjustments** to tables of data,

   – **Restriction in levels of detail**

   – **Practical tests** for assessing risks ('intruder testing')

▶ The **choice of method** is context dependent

▶ Trade off between **reducing risk** and **reducing data quality** for users

National Records of Scotland

Scotland's Census 2011
Shaping our future

Census - Background

▶ Census Day - 27 March 2011

▶ Lots of preparatory activity

▶ A 6,000-strong army of 'field staff' hand-delivered 95% of forms (others – generally more remote areas – by post)

▶ Field efforts resulted in a 94% response rate
  - Roughly 80% paper returns, 20% internet

National Records of Scotland

Scotland's Census 2011
Shaping our future

Census - Flow of data

```
Capture data  →  Load & Validation  →  Remove False persons
                                              ↓
Resolve Multiple Responses  →  Filter Rules  →  Imputation
                                                      ↓
Estimate missing returns  →  Impute skeleton records  →  Assign output areas
                                                               ↓
Derive variables for output  →  Disclosure control record swapping  →  Outputs database
```

National Records of Scotland

Scotland's Census 2011
Shaping our future

"The characteristics that make statistical data of high quality also make confidentiality a real concern"

"Statistical Confidentiality: Principles and Practice" – Duncan , Elliot & Salazar-Gonzalez

Scotland's Census 2011
Shaping our future

Statistical Disclosure Control (SDC) - Importance to Census

▶ **Census users** require outputs that contain breakdowns of personal or household characteristics to a **high level of detail** or which relate to **small populations**

▶ **Without any disclosure control protection** being applied, these census outputs would allow individual or households to be identified and therefore **personal information could be deduced**

▶ **This would be in breach of legal and other obligations**, and would likely **affect public trust** in responding to future census collections

National Records of Scotland

Scotland's Census 2011
Shaping our future

*"**Statistical Disclosure Control is a rapidly evolving field**: disclosure control thinking has to keep pace with increase in computing power, developments in matching software and the proliferation of public and private databases"*

*"*Statistical Disclosure Control" - Hundepool, Domingo-Ferrer, Franconi etc

Census 2011 - Development of Statistical Disclosure Control (SDC) Methodology

▶ **The 2006 Registrar General's agreement** set out the **harmonised UK** Census 2011 position on SDC, in that:

▶ **Small cell counts (0,1,2s) would be allowed** in publicly disseminated Census tables provided that

▶ (i) **uncertainty** as to whether the small cell is a true value has been systematically created, and

▶ (ii) that creating this uncertainty **does not significantly damage the data**

▶ Within this framework, an evaluation of SDC methods was carried out and this recommended that **targeted record-swapping** should be the primary disclosure control method

National Records of Scotland

Scotland's Census 2011
Shaping our future

Census 2011 Targeted Record Swapping

▶ Targeted towards records with the **greatest risk of identification**, although all records have a chance of being swapped

▶ For **private** households, entire households are swapped

▶ For **communal** establishments, individuals are swapped between establishments

▶ The level of swapping can vary depending on how much perturbation **coverage adjustment** (imputation) has already added within a particular area

▶ Requires a balance between creating **doubt (perturbation)** and in maintaining **data utility**

National Records of Scotland

Scotland's Census 2011
Shaping our future

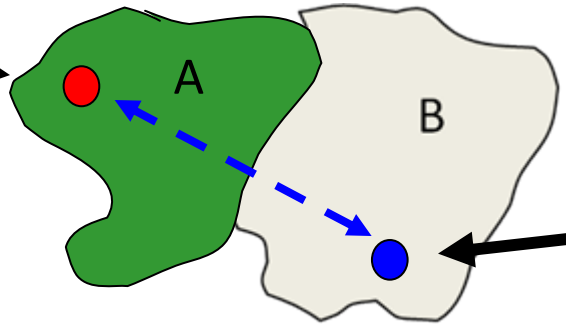# Record-Swapping - illustrative example

Characteristics:

Age: 22,

Sex: Male,

Marital Status: Single

Income Group: High

Region: Area A

A

B

Characteristics

Age: 22,

Sex: Male,

Marital Status: Single

Income Group: Medium

Region: Area B

Unique as only household with 1 person in High Income group in Area A

Treatment:
- Find a different geographical Area
- Identify another household in a different area with virtually all the same characteristics
- Swap the two households

Area B

Matches all variables except Income Group

Swap records

National Records of Scotland

Scotland's Census 2011 Shaping our future

10

UK Census Statistical Disclosure Control (SDC): 2001 v 2011

2001

- Thresholds for Output Area (OA) size
- Table design
- **Random** record-swapping
- **Small cell values were allowed in Scottish** outputs, however
- **England & Wales applied additional small-cell adjustments** (Scotland did not)

2011

- Thresholds for OA size
- Table design
- **Targeted** record-swapping
- **Small cell values are allowed** provided there is sufficient uncertainty in whether these values are real (and that data hasn't been damaged)
- **Additional tests carried out** on output tables to assess disclosure issues

National Records of Scotland

Scotland's Census 2011
Shaping our future

Statistical Disclosure Control (SDC) Census methods used elsewhere

Not the same !!

- ▶ Pre-tabular record swapping – Scotland, England, Wales, Northern Ireland, Belgium, Austria and Israel
- ▶ Post-tabular cell perturbation algorithm - Australia
- ▶ Pre-tabular perturbation algorithm - Germany
- ▶ Combination record swapping and post-tabular rounding - Sweden
- ▶ Rounding of small counts (1 and 2) to 0 or 3 - Norway
- ▶ Tau Argus post-tabulation software - Slovenia

National Records of Scotland

Scotland's Census 2011
Shaping our future

UK Census 2011 – Summary of Statistical Disclosure Control (SDC) Methods
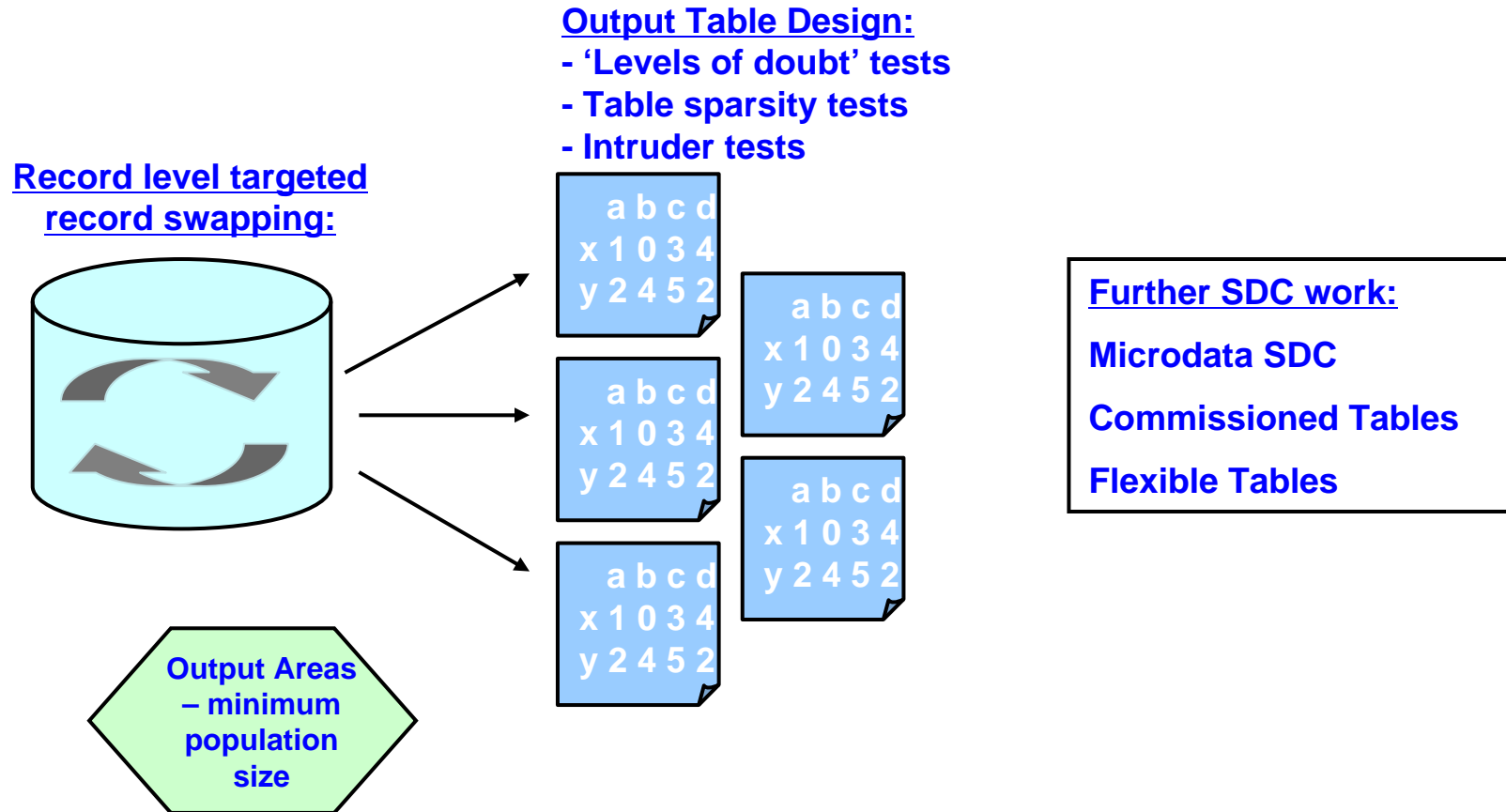
"There is no such thing as a free-lunch when it comes to statistical disclosure control"

- In other words, there will **always be a trade-off** between **reducing risk** and the impact that this has on **data utility**

Illustrative examples of potential census output table redesigns due to Statistical Disclosure Control

▶ **Univariate tables:** Labour Market tables published for 16-74 age group only (and not split into 16-64, 65-74, 75+)

▶ **Local Characteristics tables**:

- collapsed ethnic group categories (e.g. LC2301EW)

- removal of sex, collapsed age categories (e.g. LC3602EW)

▶ **Detailed Characteristics tables:**

- 'caravan or other mobile/temporary structure' combined with 'commercial building' (eg DC4102EW, DC4403EW)

- 'living rent free' combined with 'other private rented' (e.g. DC4103EW, DC4201EW)

- different types of central heating collapsed into 'central heating' (eg DC4402EW, DC4206EW)

National Records of Scotland

Scotland's Census 2011
Shaping our future

Questions or Comments

Census Statistical Disclosure Control

Area 2/3/1, Ladywell House EH12 7TF

email: customer@gro-scotland.gov.uk

0131 314 4299

Scotland's
Census 2011
Shaping our future