Disclosure Statement
Theft from a shop or stall data

Background

This bespoke data collection exercise was conducted to gather detailed information from magistrates and district judges about how they sentenced the offence of theft from a shop or stall. It was undertaken to support the assessment of the definitive guideline *Theft from shop or stall*[1] and involved administering surveys to a sample of 81 magistrates' courts both pre and post the introduction of the guideline.

The data collected information regarding the date and court location of the sentencing decision, some demographic details of the offenders, the relevant factors taken into account when each offender was sentenced for a principal offence of theft from a shop or stall and details regarding the offence category, subsequent sentence starting point and the offence imposed before and after any reduction for guilty plea. The potential disclosure issues surrounding this data have been carefully considered and the subsequent actions that were taken to ensure that the offenders' confidentiality has been protected are outlined in this document.

Assessment of Risk

In accordance with the Government Statistical Service (GSS) guidance,[2] an assessment of the risk of disclosure was undertaken in order to determine which features of the data may help to protect individuals' identities and which aspects of the data may lead to disclosure risks. This is to help prevent identification of individuals and/or prevent additional knowledge of any previously unknown characteristics for those offenders sentenced for the offence of theft from a shop or stall in the time period of the data collection.

The following information was used to assess the level of risk:

- The sample size of the survey;
- The response rate of the survey;
- The likelihood that the data could be linked/matched to other data sources; and,

---

[1] https://www.sentencingcouncil.org.uk/offences/magistrates-court/item/theft-from-a-shop-or-stall/
[2] https://gss.civilservice.gov.uk/wp-content/uploads/2014/11/Guidance-for-microdata-produced-from-social-surveys.pdf

- The nature and level of detail of the information collected.

Applying Disclosure Control

The following changes were made to the data to reduce the chance of the identification of individuals and thus maintain disclosure control. This included aggregation or banding of variables, where appropriate, in order to reduce unique counts and/or the removal of any variables that could make it easier to identify an individual:

- Removal of court location;
- Original unique case identifiers removed and a new unique defendant identifier created instead by using random numbering;
- Grouping/banding of offender age and removal of date of birth variable;
- Removal of the sentencing date;
- Aggregation of sentence starting point, outcome before guilty plea and final sentence outcome categories where counts were less than 10;
- Cleaning and coding of the 'single most important factor' variable and removal of the original free text answers from the final dataset;
- Removal of other free text fields where the raw data was deemed too disclosive and was not seemed high enough quality to clean and recode; and
- Removal of any harm, culpability, aggravating or mitigating factors with counts fewer than 10.

Comments on this data

- The data collection only sampled around half of all the magistrates' courts open at the time of data collection (81) and additionally achieved a relatively low response rate,[3] thus meaning that the majority of offenders sentenced during the period of the data collection did not have a form filled in for them, and are not captured in the data. Therefore, if someone knew some details about an offender sentenced for theft from a shop or stall during the period of the data collection and found a unique record in the data that seemed to match that individual, they could not be certain that the record in the data relates to that specific person. As such, the Council considers that the data are sufficiently anonymous and feels that any further disclosure control would be detrimental to the usefulness of the data.
- The likelihood has been judged as very low that individuals can be identified in the published data through linking to another dataset, given the lack of other publicly available data sources of the same information and that the sample cannot be replicated.

---

[3] A response rate of 30% in the pre-guideline phase and 26% in the post-guideline.

- The names of the courts and the exact date of the sentence have been removed and as such it is not possible to identify exactly where the survey took place or when any particular form was filled in.
- Where the counts of records were extremely low (less than 10), these variables have been removed from the dataset to reduce the risk of positive identification.[4]
- Banding of specific numeric variables was considered and was conducted for the age variable. However, due to high demand for specific sentence lengths from academic researchers, the decision was made to include the actual custodial and suspended sentence lengths given to offenders, rather than to band them, after balancing the very small risk of identifying individuals with the high usefulness of the data.
- In the raw data, the 'single most important factor' free text field included very specific details about the offender, offence, location and other details that may have risked being disclosive. To mitigate the risk of any offender being identifiable within the data, the data have been cleaned and recoded into a series of less disclosive factors, with the original answers removed from the published data.
- It was also acknowledged that the cases, about which the theft from a shop or stall data relate, would have been heard in open court. Although this may increase the risk of identification due to some of the defendant's details being more accessible, the impact is anticipated to be negligible due to the factors listed above. However, it also means that the impact of any potential disclosure issues is minimal as much of this information is already in the public domain.

If users would like any further details of the disclosure controls applied to the data, please contact the Analysis and Research team at Research@sentencingcouncil.gov.uk.

---

[4] See the metadata file for information on which factors were removed. The number of times that these factors were ticked is provided within the metadata so that users can see overall how many times these features arose in the data.