

# Fraud and Error Measurement - Variance and Confidence Intervals

## Distribution of Fraud and Error Values

This guide examines the prevalence of error within our estimates; it will focus on overpayments for simplicity, but the same rules also apply to underpayments.

The majority of the benefit cases that are checked in our fraud and error surveys are correct, and therefore not in error. For the remainder of the cases that have fraud or error (or both), some estimates are calculated. These estimates are then scaled-up to form national estimates for the:

- i) percentage of cases that are incorrect;
- ii) percentage of expenditure overpaid;
- iii) value of the overpayments (the percentage in iii) applied to the actual benefit expenditure.

There is less variability with estimate i) because there is only a binary choice for the response variable, i.e. the case is either correct or incorrect. There is more variability with estimate ii) since the error values are continuous, ranging from pence to hundreds of pounds. Furthermore, for estimate ii), there is a higher volume of small benefit amount errors (in the order of pence) and a lower volume of high benefit amount errors (in the order of hundreds of pounds). This accounts for the high variance in overpayments. The argument is similar for underpayments.

The high variance is particularly problematic in comparing change over years, which the national statistics releases identify as not statistically significant. There are, however, two reasons why it is difficult to compare change over time:

- high variance in over- and underpayments (as already described above);
- the level of fraud and error for the continuously measured benefits is around 5% of benefit expenditure, which is an estimate of small magnitude.

With regard to the second bullet point, since the estimate is of small magnitude and only small changes are measured, this means that it will be difficult to measure any real, significant change. If the level of fraud was of greater magnitude, then measuring a significant statistical difference would be less difficult. Larger magnitude estimates allow the possibility of larger changes that in turn are more likely to be statistically significant.

## Confidence Intervals

A Confidence Interval is used to indicate the reliability of a statistical estimate. It gives an estimated range of values which is likely to include an unknown population parameter (in this instance – a benefit case), the estimated range being calculated from a given set of sample data. The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter. A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.

Assuming that time, budget and human resources were not constraints, repeated random re-samples of the benefit caseload would be taken, and the central estimates of fraud and error would be calculated for each random re-sample. A distribution of the calculated fraud and error central estimates from the repeated re-samples would be built and from this distribution the 95% confidence intervals could be calculated.

However time, budget and resource are restraints so this methodology is impractical for the production of the fraud and error statistics. Instead we choose the largest sample afforded to the survey, that is, one big sample. This sample is designed to ensure we check a specified number of cases of each benefit in order to minimise the variance and therefore increase the precision. We use this sample data to calculate the central estimates of fraud and error. We then calculate the confidence intervals around this central estimate using a method called 'Bootstrapping'.

### What is bootstrapping?

Bootstrapping allows the utilisation of the one sample dataset that is at our disposal to calculate the confidence intervals. Bootstrapping is a computer-intensive statistical technique, which replaces traditional algebraic derivations with data-based computer simulations. The computer is used to take 4,000 random re-samples with replacement from our initial sample dataset. The computer then calculates the rate of fraud and error for each of the re-samples. A Bootstrap distribution of the fraud and error central estimates is then produced. Parametric Bootstrapping, which means we read off the 2.5th percentile (25th record) and the 97.5th percentile (975th record) to obtain our 95% confidence limits, is then used. We also check the median of this distribution with our actual central estimate to ensure that no bias exists.

The main reasons we use Bootstrapping to calculate our confidence intervals instead of other analytical techniques are because:

- we perform a number of statistical adjustments to account for sample bias (for example, an undercount of younger age cases from the point the sample is selected to the time of performing the entitlement checks) and it would be difficult to account for the variance that each adjustment adds analytically using mathematical formulae;
- our central estimate is derived from the "combined ratio estimator" and calculating the variance of this estimator is also difficult analytically.