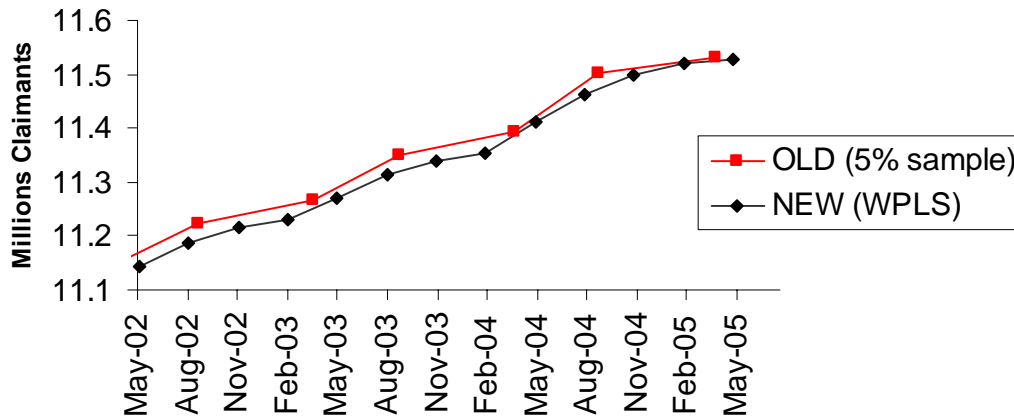


State Pension

Work and Pensions Longitudinal Study (WPLS) data compared to currently published 5% sample data

The WPLS data series for State Pension (SP) is only available from May 2002. Over the May 2002–May 2005 period the WPLS caseload is on average 0.3% lower than that obtained from the 5% sample data (approximately 38,000 cases).

SP: comparison of caseloads, 5% sample and 100% data



The main causes of this difference have been identified. (See the glossary below for an explanation of retrospection) These are:

Retrospection accounts for approximately 15% (12 thousand) of claims which exist on the 100% but not the 5% sample data.

Temporary Removal of cases accounts for 19% (15 thousand) of cases which exist on the 100% but not the 5% sample data. These cases appear in the 100% data at the relevant reference date but are absent from the corresponding 5% sample data. However, they appear in the 5% datasets which precede and follow the reference date.

Different scan dates The 5% data is extracted on a 6 monthly basis in March and September; however the 100% data is extracted quarterly in February, May, August and November. Because the data has different reference dates this will cause the caseloads to differ.

Removal of terminated cases The 5% sample data (but not 100% data) incorrectly contains some cases which have left the benefit due to death. The 100% data reflects more accurately when a claimant leaves benefit. The 5% sample data estimates claim end date, whereas the actual date of death (if known) is utilised in the 100% data, and supersedes the estimated date. This accounts for approximately 34,000 cases (30% of the extra cases on the 5% sample data).

Different scan specifications The scan specifications (computer code) to extract the 5% and 100% data differ slightly. For example the 5% code includes customers who have partially voluntarily relinquished receipt of their pension, whereas the 100% does not. Although this is a small discrepancy (approximately 150 cases per extract), it is possible that there are further, as yet undiscovered, differences between the specifications.

Glossary

Retrospection arises from the delays which occur in practice between some new claims becoming eligible and their entry onto the computer system. If a case has not been added to the computer system by the time the 5% sample data is extracted (the reference date), it will not be counted (although benefit will be paid). The WPLS, however, incorporates information added to the computer system after the reference date.

The addition of such 'late' information is an important source of difference between the caseloads obtained from 5% sample and WPLS data, and generally results in the WPLS including more cases than the 5% sample data.

Late terminations

Delays can exist between a claim terminating and its removal from the computer system. Such cases will be included in the caseload obtained from the 5% sample data. When information is subsequently received that the claim has ended through claimant death, these cases are removed from the 100% data with respect to a point in time. However, any other form of late termination is not accounted for in the WPLS.

Sampling error

By definition, WPLS data yields the true population value. By contrast statistics published from 5% sample data (obtained by rating up by a factor of 20) are subject to sampling error and are actually estimates of the true population value. By chance, an estimate of the population value, obtained from the 5% sample data, may be slightly lower or slightly higher than the true population value.