Guidelines for measuring statistical quality

London: Office for National Statistics Version 3.1

ISBN 978-1-85774-665-5

Contact points

For enquiries about this publication, contact:
The Quality Centre
Tel: **01633 455631**
Email: **quality.measurement@ons.gov.uk**

For general enquiries, contact the National Statistics Customer Contact Centre on:
Tel: **0845 601 3034**
Minicom: 01633 812399
Email: **info@statistics.gov.uk**
Fax: 01633 652747
Letters: Room D115, Government Buildings, Cardiff Road, Newport, NP10 8XG

You can also find National Statistics on the Internet at: **www.statistics.gov.uk**

A National Statistics Publication

National Statistics are produced to high professional standards as set out in the National Statistics Code of Practice. They undergo regular quality assurance reviews to ensure that they meet customer needs. They are produced free from any political influence.

Preface

Quality is central to National Statistics. As National Statistician it is my responsibility to ensure that we deliver statistics that are of high quality and integrity, are fit for purpose and win the trust and confidence of the public.

A great deal of research has gone into the preparation and updating of the following guidelines to ensure that they promote the high standards of statistical quality essential in the UK and across Europe and that they reflect the quality agenda of other major National Statistics Institutions across the world. The guidance is a tool that will help us to ensure that the quality requirements of the National Statistics Code of Practice are met and that our vision to become 'world class' is realised. It will also provide consistency in information about our statistics that will allow users to judge for themselves the quality and appropriate uses of the data in accordance with their needs.

Using these guidelines when planning and producing statistics and compiling statistical reports and publications, will help ensure their quality. I know this is important to us all.

Karen Dunnell
National Statistician
November 2007

Contents

Section A: Introduction

The Guidelines for Measuring Statistical Quality (Version 3.1) provide a checklist of quality measures and indicators (see A.3 below) for use when measuring and reporting on the quality of statistical outputs. They are not a National Statistics protocol but represent best practice for measuring quality throughout the statistical production process.

This version of the guidelines replaces Version 3.0 (April 2006) and features new measures for information loss after disclosure control methods have been applied.

The purpose of this document is to promote a standardised approach to measuring and reporting on quality across the Government Statistical Service (GSS). Many of the measures and indicators in the guidelines will be familiar to government statisticians – for example, standard errors and response rates. Others will be less familiar as they have been newly developed.

Although the guidelines are primarily aimed at producers of official statistics, they can be used by anyone wanting to report on the quality of statistical outputs.

Future versions of the guidelines will also include specific quality measures for areas that are not comprehensively covered by this version, such as outputs derived from other sources.

This document consists of two parts. Section A presents background information on the guidelines. Section B provides the checklist of quality measures and indicators for consideration when producing statistical outputs.

A.1 Background

The GSS is committed to providing users with information on the methods that have been used to compile its statistical outputs. This commitment is expressed in the National Statistics Code of Practice:

> 'Processes and methods used to produce National Statistics will be sufficiently detailed to allow users to assess fitness for particular purposes.'

In addition, the GSS has declared its intention to provide information on the quality of statistical outputs in the National Statistics Quality Strategy:

> 'The quality measures for National Statistics will be systematically reported, alongside results and will enable [the user] to judge the suitability of their application for their intended uses.'

The guidelines aim to fulfil both of these commitments. They replace the GSS Statistical Quality checklist and the Office for National Statistics (ONS) Quality Measurement and Reporting Framework, which have been integrated to provide a single set of guidelines for quality measurement and reporting.

A.2 Aim and purpose of the guidelines

The overall aim of the guidelines is to outline best practice for measuring and reporting on the statistical quality of GSS outputs. In particular, the emphasis is upon helping users to understand:

- the context in which the data were collected, processed and analysed;
- methods adopted and limitations they impose;
- the reliability of the figures; and
- the way they relate to other available data on the same subject.

The measures and indicators can also be used by producers of official statistics to monitor data quality for the purpose of continuous improvement.

A.3 What is ‘quality’?

The word ‘quality’ has many different meanings, depending on the context in which it is used. The quality of statistical outputs is most usefully defined in terms of how well outputs meet user needs, or whether they are ‘fit for purpose’. This definition is a relative one, allowing for various perspectives on what constitutes quality, depending on the intended uses of the outputs.

Quality measurement for statistical outputs is concerned with providing the user with sufficient information to judge whether or not the data are of sufficient quality for their intended use(s).

In order to enable users to judge for themselves whether outputs meet their needs, it is recommended that output providers report quality in terms of the six quality dimensions of the European Statistical System (ESS), which are shown in Table A.1. The quality measures and indicators in Section B of the guidelines have been developed around these six dimensions. A good summary of quality should contain quality measures and indicators for each of the six ESS quality dimensions.

Quality measure or quality indicator?

Quality measures are defined as those items in the Guidelines that directly measure a particular aspect of quality. For example, the time lag from the reference date to the release of the output is a direct measure. However, in practice many quality measures can be difficult or costly to calculate. Instead we can use quality indicators to give insight in quality.

Quality indicators usually consist of information that is a by-product of the statistical process. They do not measure quality directly but can provide enough information to provide an insight into quality. For example, in the case of accuracy it is almost impossible to measure non-response bias as the characteristics of those who do not respond can be difficult to ascertain. In this instance, response rates are a suitable quality indicator that may be used to give an insight into the possible extent of non-response bias.

The guidelines include both quality measures and quality indicators, which can either supplement or act as substitutes for the desired quality measures.

Table A.1 Dimensions of quality

Definition	Key components
1. RELEVANCE	
The degree to which the statistical product meets user needs for both coverage and content.	Any assessment of relevance needs to consider: • who are the users of the statistics; • what are their needs; and • how well does the output meet these needs?
2. ACCURACY	
The closeness between an estimated result and the (unknown) true value.	Accuracy can be split into sampling error and non-sampling error, where non-sampling error includes: • coverage error; • non-response error; • measurement error; • processing error; and • model assumption error.
3. TIMELINESS AND PUNCTUALITY	
Timeliness refers to the lapse of time between publication and the period to which the data refer. Punctuality refers to the time lag between the actual and planned dates of publication.	An assessment of timeliness and punctuality should consider the following: • production time; • frequency of release; and • punctuality of release.
4. ACCESSIBILITY AND CLARITY	
Accessibility is the ease with which users are able to access the data. It also relates to the format(s) in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the metadata, illustrations and accompanying advice.	Specific areas where accessibility and clarity may be addressed include: • needs of analysts; • assistance to locate information; • clarity; and • dissemination.
5. COMPARABILITY	
The degree to which data can be compared over time and domain.	Comparability should be addressed in terms of comparability over: • time; • spatial domains (e.g. sub-national, national, international); and • domain or sub-population (e.g. industrial sector, household type).
6. COHERENCE	
The degree to which data that are derived from different sources or methods, but which refer to the same phenomenon, are similar.	Coherence should be addressed in terms of coherence between: • data produced at different frequencies; • other statistics in the same socio-economic domain; and • sources and outputs.

A.4 Key Quality Measures

Key Quality Measures (KQMs) are those quality measures and indicators that are considered to be the most important and informative in giving users an overall summary of output quality. In addition, the KQMs can be used to provide management information to monitor performance and any quality improvements in statistical outputs.

The KQMs are shown in Table A.2 and are denoted throughout Section B. There are KQMs for five of the six ESS quality dimensions, but not for the dimension Accessibility and Clarity.

It is recommended that these KQMs are a minimal reporting requirement for all statistical outputs, where they are relevant.

Table A.2 Key Quality Measures

	KEY QUALITY MEASURE	ESS QUALITY DIMENSION	GUIDELINES REFERENCE
1	Where possible, describe how the data relate to the needs of users	Relevance	**B1.3**
2	Provide a statement of the nationally/internationally agreed definitions and standards used	Comparability	**B1.13**
3	Unit response rates by sub-groups, weighted and unweighted	Accuracy	**B3.4 (Household surveys)** **B3.5 (Business surveys)**
4	Key item response rates	Accuracy	**B3.7**
5	Total contribution to key estimates from imputed values	Accuracy	**B4.7**
6	Editing rate (for key items)	Accuracy	**B4.11**
7	Estimated standard error for key estimates	Accuracy	**B5.2 (for key estimates of level)** **B5.3 (for key estimates of change)**
8a	Time lag from the reference date/period to the release of the provisional output	Timeliness and Punctuality	**B8.1**
8b	Time lag from the reference date/period to the release of the final output	Timeliness and Punctuality	**B8.2**
9	Estimated mean absolute revision between provisional and final statistics	Accuracy	**B8.21**
10	Compare estimates with other estimates on the same theme	Coherence	**B8.28**
11	Identify known gaps between key user needs, in terms of coverage and detail, and current data	Relevance	**B8.29**

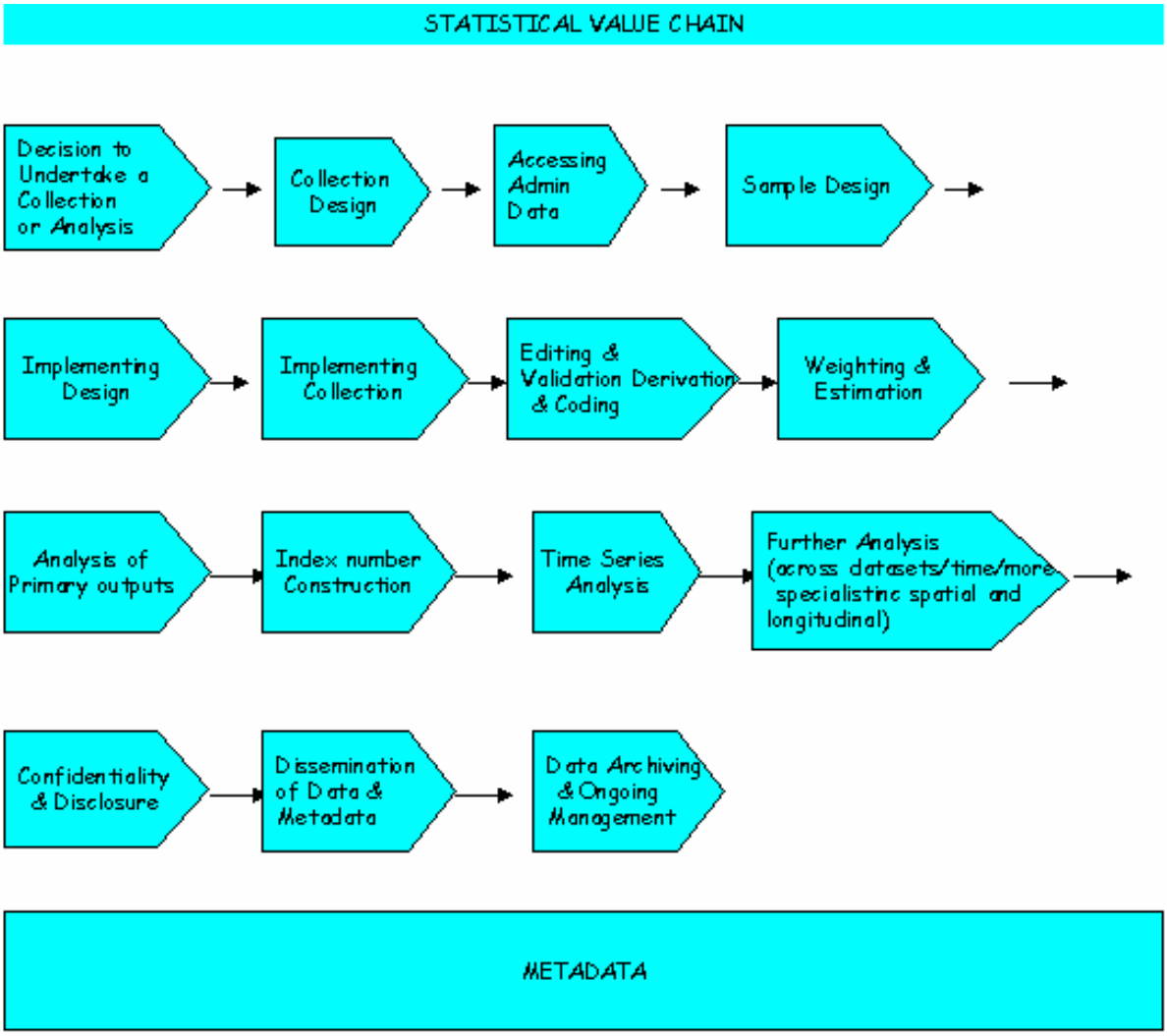
Section B: Quality measurement guidelines

A checklist of items to consider when reporting on the quality of statistical outputs is presented over the following pages. It is not the intention that all quality measures should be addressed for all outputs. Instead, the user is encouraged to select those quality measures and indicators that together provide an indication of the overall strengths, limitations and appropriate uses of a given dataset.

The quality measures and indicators are grouped together into stages of the statistical production process: design, data collection, data processing, weighting and estimation, time series analysis, statistical disclosure control, and dissemination.

ONS has developed a more detailed representation of the statistical production cycle, known as the Statistical Value Chain (SVC). Figure B.1 shows the 15 links in the SVC.

Figure B.1 The ONS Statistical Value Chain



The SVC categorisation is too detailed for classifying the quality measures in these guidelines. However, the seven categories we have chosen for the quality measures roughly correspond to links of the SVC as shown in Table B.1. The section on Administrative data has links to most of the stages of the SVC as it deals with everything to do with Administrative data from the quality of the data at source to the processing done to the data and use of it for the statistical product.

Table B.1 Comparison of the categories in these guidelines with the SVC

Quality measures category	SVC category
Design	Decision to undertake a collection or analysis
	Collection design
	Sample design
	Implementing design
Administrative data	Accessing Administrative data
Data collection	Implementing collection
Data processing	Editing and validation, derivation and coding
Weighting and estimation	Weighting and estimation
Time series	Time series analysis
Statistical disclosure control	Confidentiality and disclosure
Dissemination	Dissemination of data and metadata

In addition to the categorisation by stages of the statistical production process, the quality measures have also been grouped into the ESS quality dimensions: Relevance, Accuracy, Timeliness and Punctuality, Accessibility and Clarity, Comparability, and Coherence which can be accessed from the National Statitics website at http://www.statistics.gov.uk/qualitymeasures.

The tables in this section contain quantitative and qualitative quality measures and indicators, together with:

- descriptions of each measure/indicator and notes on use
- likely frequency of production for the measure (see Table B.2);
- an example of the type of information you may want to record when addressing each measure. For qualitative measures, examples have been taken where possible from recently published documents and articles. For quantitative measures, suggested formulae are shown instead; and
- an indication where the item is one of the ONS Key Quality Measures.

Production frequency is categorised in one of two ways, as outlined in Table B.2.

Table B.2 Production frequency categorisation

Production frequency category	Description
	To be produced for each output.
	To be produced once for all instances of an output: to be revised where changes are required.

It is envisaged that some quality measures and indicators will be produced for each output (for example, standard errors would be calculated with each new estimate). These types of quality measures and indicators would be designated an ‘a’ in the production frequency categorisation. Alternatively, some quality indicators would be produced once for all outputs, only to be rewritten where there are changes. For example, a description of data collection methods for a survey would be applicable to all subsequent instances of a survey, except where there are changes in the data collection methods. In this instance, the quality indicator is assigned a ‘b’ for production frequency.

B1. Design

Ref.	Notes	Example	
B1.1	**Describe and classify key users of output.**		Relevance
b	This information can be obtained from requests to carry out the survey, post-survey feedback, or from information on the users of previous similar surveys. The users are classified according to their use of the survey and the type of agency they are affiliated to (for example institutions, international organisations, researchers and students, businesses).	The key users of the figures for estimated number of applications to higher education institutes are: • governmental statistical agencies; • higher education institutes; and • higher education and student funding bodies.	
B1.2	**Describe needs of key users and uses of output.**		
b	This information can be obtained from requests to carry out the survey, post-survey feedback or from information on the uses of previous similar surveys.	The results will be used by both government and industry. The main user is the Office for National Statistics itself, which uses the data to provide estimates of change in inventories, for use in the compilation of all three estimates of gross domestic product (GDP). The change in inventories, or stock building, is part of final expenditure in the National Accounts. Holding gains on inventories (or stock appreciation) is included within the income measure of GDP. Inventories are also used at the detailed level in the compilation of annual current price Input-Output Supply and Use tables, which determine the level of current price GDP. The Treasury uses the results for forecasting, analytical and briefing work on the economy wide output and on the company sector. (ONS 2003h)	

Ref.	Notes	Example	
B1.3 b	**Where possible, describe how the data relate to the needs of users.** **(Key Quality Measure)**		Relevance
	This indicator captures how well the data support users' needs. This information can be gathered from user satisfaction surveys and feedback.	Users require data on the hours and earnings of full-time and part-time adult employees. The data provide virtually complete coverage of full-time adult employees, but the coverage of part-time adult employees is not comprehensive. Many of those with earnings below the income tax threshold are not covered, which excludes mainly women with part-time jobs and a small proportion of young adults.	
B1.4 b	**Describe key statistical concepts.**		
	This should include descriptions of the statistical measure, the population, variables, units, domains and time reference. This information gives users an understanding of the relevance of the output to their needs, for example whether the output covers their required population or time period. For administrative data sources see: B2.3. For statistical outputs derived wholly or in part from administrative data see: B2.23	The monthly inquiry into retail sales is a sample survey carried out by the Office for National Statistics on 5,000 businesses in Great Britain, including large retailers and a representative panel of smaller businesses. From this survey the Retail Sales Index (RSI) is compiled each month. (ONS 2003g)	
B1.5 b	**For outputs based on other sources, describe the key sources.**		
	This should include the known purpose of the data collection and known merits and shortcomings of the data. The information will help users to assess whether the output is relevant and of sufficient quality for their uses. For statistical outputs derived wholly or in part from administrative data see: B2.13	The second source of data used is the ONS Longitudinal Study (LS), from which estimates of the distribution of ages at first childbearing by education of the cohort born 1954–1958 were derived. The LS has linked the birth registration and Census records since 1971 for a one per cent sample of all women in England and Wales. The very large sample size of the LS presents an opportunity to estimate women's reproductive lives with much lower statistical sampling error than would be possible using a survey data source such as the General Household Survey. (Goldblatt and Chappell 2003)	

Ref.	Notes	Example	
B1.6	**Describe results of user satisfaction assessments.**		Relevance
b	The main results of user satisfaction assessments should be reported, giving priority to the results for the most important groups of users.	A survey was carried out to evaluate whether the publication was relevant to user needs. The survey was sent to all persons and institutions requesting hard copies of the publication, and a link to the survey was provided on the web site for online users. In summary, the main findings were that most of the researchers and academics found the publication to be a useful aid to their research, teaching and for their personal interest, but that the level of detail provided was not sufficient to allow further analysis for some users.	
B1.7	**Describe any actions taken to improve relevance based on customer feedback.**		
b	This records how the results of customer feedback and satisfaction surveys are translated into concrete actions to improve the relevance of outputs, for example, changes to how concepts are operationalised as a result of user feedback on lack of relevance for their needs.	The customer satisfaction survey (2004) indicated that some users were unhappy with the comparability of the results for successive years. Having contacted some of the users, it was discovered that the problem lay with the reference period for the survey which did not consistently contain/exclude the school half term. The situation was further complicated by the fact that different areas of the country have half term at different times. It was proposed that the reference period for this survey is moved so that it would always be in term time (given the current term structure). Response to this approach has been supportive and the change has been introduced.	
B1.8	**Describe any gaps between measured statistical concept and user concept of interest.**		
b	The gap between the user's concept of interest and the measured statistical concept is assessed and described. Gaps between what is measured and what users require may be due to definitional variations or differences in the way that concepts are measured.	The user preferred definition of unemployment includes those out of work who want a job, sought work in the last 4 weeks but are not available to start in the next fortnight. However, the definition of unemployment used in this survey is the International Labour Organization (ILO) definition, i.e. it excludes those who are not available to start work in the next fortnight.	

Ref.	Notes	Example	
B1.9	**Describe plans for meeting needs arising from lack of completeness.**		Relevance
b	This allows users to assess whether plans to ensure that outputs are complete, in terms of coverage and detail, are adequate for their needs of the output.	A range of initiatives has been put in place by the ONS over recent years to improve the quality and availability of service sector statistics, for example the newly developing Index of Services (IoS), the Corporate Services Price Index (CSPI) and improvements to the Monthly Inquiry into the Distribution and Services Sector (MIDSS). However a gap exists in the availability of detailed 'product' statistics for services. (Prestwood 2000)	
B1.10	**Describe the sample design.**		Accuracy
b	This enables users to assess the quality of the output in terms of the accuracy of any population estimates. Population estimates can be derived for outputs based on random probability sample designs. For non-random designs (e.g. purposive sampling designs) derived statistics cannot be generalised to a wider population.	The primary sampling units (PSUs) are stratified by 24 regions and 3 other variables derived from the Census of Population. Stratifying ensures that proportions of the sample falling into each group reflect those of the population. Within each region the postcode sectors are then ranked and grouped into six equal bands using the proportion of heads of household in socio-economic groups 1-5 and 13. Within each of these bands, the PSUs are ranked by the total unemployment rate and formed into 3 further bands, resulting in 18 bands. These are then ranked according to the proportion of households that are owner occupied. This set of stratifiers is chosen to have a maximum effectiveness on the accuracy of two variables: household income and housing costs. (OPCS 1996)	

Ref.	Notes	Example	
B1.11 b	**For a continuous survey, have there been any changes over time in the sample design methodology?**		Comparability
	For example, have there been changes in stratum definitions? If so, what impact does this have on the comparability of estimates across time? For statistical outputs derived wholly or in part from administrative data see B2.22	In the 1986 survey a less clustered sample design was adopted using postal sectors (ward size) as the primary sampling units. Some 672 postal sectors were randomly selected during the year after being arranged in strata comprising standard regions, area type, and two 1981 Census variables (proportion of owner-occupiers and proportion of renters). These postal sectors were not revisited as in the previous design. Therefore the greater spread of the sample should result in greater precision of annual expenditure estimates. (OPCS 1986)	
B1.12 b	**Describe classifications.**		Relevance
	This is an indicator of clarity, in that users are informed of concepts and classifications used in compiling the output. The information should be sufficient to allow replication of data collection and compilation. For administrative data sources see B2.3. For statistical outputs derived wholly or in part from administrative data see: B2.23	With effect from 2003, Standard Occupational Classification (SOC) 2000 has replaced SOC90 as the classification used for the variable 'current or last occupation'. The reason for the change is that SOC2000 has now been adopted as the office standard.	
B1.13 b	**Provide a statement of the nationally/internationally agreed definitions and standards used.** **(Key Quality Measure)**		Comparability
	This indicates geographical comparability where the agreed definitions and standards are used.	The National Accounts are based on the European System of Accounts 1995 (ESA95), itself based on the System of National Accounts 1993 (SNA93) which is being adopted by statistical offices throughout the world. (ONS 2003c)	

Ref.	Notes	Example	
B1.14 b	**Provide a statement of international regulations that apply and any laws which have been repealed or abolished.**		Comparability
	The provision of this information allows users to assess whether international regulations or repealed or abolished laws will affect comparability. For statistical outputs derived wholly or in part from administrative data see B2.21	PRODCOM is a survey of manufactured products governed by an EC Regulation (European Community 1991). The product definitions are standardised across the EC to give comparability between member states' data and the production of European aggregates at product level. The PRODCOM regulation stipulates that the survey should cover at least 90% of national production for each NACE (European Community classification of Economic Activity class). (ONS 2004b)	
B1.15 b	**Describe any deviations from nationally/internationally agreed definitions and standards.**		
	This should include reasons for any deviations. Where there are deviations from national or international definitions and standards, these may make data less comparable with other data that conform to these agreed definitions and standards. This indicator allows users to judge whether the data are comparable to other data from the same or other geographical areas.	The interpretation of Rule 3 was broadened by OPCS in 1984 so that certain conditions which are often terminal, such as bronchopneumonia (ICD 485) or pulmonary embolism (415.1) could be considered a direct sequel of any more specific condition reported. The more specific condition would then be regarded as the underlying cause. (ONS 2001b)	
B1.16 b	**Coverage error.**		Accuracy
	Coverage error is the error that arises from not being able to sample from the whole of the target population. In practice, complete and accurate lists of target populations against which to check frame coverage do not usually exist. Estimates of undercoverage, duplication, ineligibility and misclassification may be provided to give an indication of coverage error. Estimates subject to coverage error should be accompanied by a statement to this effect and a description of the main sources of coverage error.	The estimates produced from this survey are subject to coverage error. The survey draws its sample from the Inter-Departmental Business Register (IDBR). Coverage error arises because not all businesses in scope for this survey are contained on the IDBR. In addition to this, the IDBR contains duplicates of some businesses, and others that are ineligible for this survey.	

Ref.	Notes	Example	
B1.17	**Estimated rate of undercoverage.**		Accuracy
b	Undercoverage occurs when there are units in the target population that are not on the sampling frame. Estimators based on data from incomplete sampling frames are likely to be biased. Undercoverage is difficult to measure, since it requires knowledge of every unit in the target population. However, it is sometimes possible to estimate the rate of undercoverage through special studies.	Since the census aims to cover the entire population, a post-enumeration survey is conducted to check the extent to which this has been achieved. After the 1981 Census a rather more thorough post-enumeration check was made. This discovered that there had been a net under-enumeration of 214,000 people as well as 800,000 absent residents who had not been required to return a form for that address. (ONS 1997)	
B1.18	**Estimated rate of duplicate units.**		
a	Duplicate units often occur on sampling frames that are created from multiple sources. Where duplicates are not identified, this can lead to inaccuracies in survey estimates. Duplicate units will have a higher probability of being selected for the sample and may be selected more than once. This may produce a smaller, less representative sample. Estimating the rate of duplicate records on the sampling frame gives an indication of the extent of this problem.	$\frac{\text{Number of duplicate records on the frame}}{\text{Total number of records on the frame}}$	
B1.19	**Estimated rate of ineligible units.**		
a	Overcoverage occurs on sampling frames containing units that are not part of the target population. If these ineligible units are not detected, they can lead to inaccuracies in survey estimates. Estimating the rate of ineligible units gives an indication of the reduction in accuracy.	$\frac{\text{Number of ineligible units on frame}}{\text{Total number of units on frame}}$	
B1.20	**Estimated rate of misclassified units.**		
a	Surveys often use auxiliary information from the sampling frame to improve the accuracy of estimates. When this auxiliary information is incorrect, the accuracy of the estimates will be reduced. Estimating the rate of misclassified units on the sampling frame gives an indication of this loss of accuracy.	$\frac{\text{Number of eligible units misclassified}}{\text{Total number of eligible units}}$	

Ref.	Notes	Example	
B1.21 b	**Describe methods used to deal with coverage issues.**		Accuracy
	Updating procedures, frequency and dates should be described, in addition to frame cleaning procedures. Also, edit checks, imputations or weighting procedures that are carried out because of coverage error should be described. This information indicates to users the resultant robustness of the study sample as a representative sample of the target population, after these procedures have been carried out to improve coverage. For administrative data sources see: B2.7	Work is currently underway to review the register sources used for the survey to improve coverage. In particular, extra resources are being directed towards improving the frame to provide a better estimate of the total number of companies with foreign links.	
B1.22 b	**Assess the likely impact of coverage error on key estimates.**		
	Coverage error is the error arising from not being able to sample from the whole of the target population. This indicates to users how reliable the key estimates are as estimators of population values, in that coverage error may reduce the representativeness of the study sample. The assessment of the likely impact of coverage error on key estimates should draw on information on coverage error for sub-groups where appropriate, and of estimates of rates of under- and over- coverage, misclassifications and ineligible units.	The target population for this survey was first-time mothers in 2002. The sample was derived from child benefit registers. There was an estimated rate of 9 per cent undercoverage due to some first-time mothers not claiming child benefit. This coverage error is likely to have resulted in a bias in key estimates towards those first-time mothers who are aware of benefit systems and whose first language is English. The resultant unweighted key estimates may therefore overstate awareness of nursery voucher schemes among first-time mothers.	
B1.23 b	**Describe the sampling frame.**		
	This tells users how the sampling frame was constructed, and whether it is current or out of date.	The Inter-Departmental Business Register (IDBR) is a list of UK businesses that is maintained by ONS. It is used for selecting samples for surveys of businesses, to produce analyses of business activity and to provide lists of businesses. It is based on inputs from three administrative sources: traders registered for Value Added Tax (VAT) purposes with HM Customs and Excise (HMCE); employers operating a Pay As You Earn (PAYE) scheme registered with the Inland Revenue (IR); and incorporated businesses registered at Companies House (CH). (ONS 2001a)	

Ref.	Notes	Example	
B1.24 b	**Has the frame been updated to take account of changes in the study population and changes in classifications?**		Accuracy
	Populations are rarely constant and it is important that sampling frames are updated with information on births, deaths and any changes in classification to units in the population. Reporting on updating practices gives an indication to users of the quality of the sampling frame.	For businesses on the register making returns to the quarterly or annual Sales Inquiries, industrial classification is reviewed annually and is derived from an analysis of their commodity sales. For other businesses, the classification is obtained either from VAT sources or from the register proving forms. (OPCS 1993)	
B1.25 b	**Define and compare the target population and the study population.**		
	This comparison should be made for the population as a whole and for significant sub-populations. It gives an indication of coverage error, in that the study population is derived from a frame that may not perfectly enumerate the population. Further information on possible coverage error is gained when the study population and target population are stratified according to key variables, then compared. This indicator can be estimated from follow-up surveys.	The target population for the survey is defined as first-time mothers in 2002. The study population has been defined as those mothers making their first child benefit claim for children born in 2002. It will therefore exclude all first-time mothers who do not claim child benefit for whatever reason, e.g. those who are ineligible, those who do not wish to be known to the benefit system, and those cases where the benefit is claimed by someone other than the mother. It is also likely that in areas of higher unemployment, there will be a greater propensity to claim child benefit than in areas of low unemployment. As a result, under-sampling in certain geographic locations and of some socio-economic groups may occur in the survey.	
B1.26 a	**Sampling fraction.**		
	This can be expressed as a fraction or as a percentage. A survey may have different sampling fractions for sub-groups, e.g. in order to over-sample for scarcer groups. Sampling fractions for each sub-group should be given.	$\frac{\text{Number of units in the sample}}{\text{Number of units in the population}}$	

B2. Administrative Data

Measures B2.1 to B2.12 relate to administrative data sources; measures B2.13 to B2.24 relate to statistical outputs derived wholly or in part from administrative data.

Ref.	Notes	Example	
B2.1	**Describe the main uses of the administrative data.**		Relevance
b	Include all the main statistical processes and/or outputs known to require data from the administrative source.	Traders registered for Value Added Tax (VAT) purposes with HM Revenue and Customs (HMRC) are used to identify new businesses and to provide size and location information for new and existing businesses for the Inter Departmental Business Register (IDBR). (ONS, 2001a)	
B2.2	**Describe the primary purpose of data collection by the administrative source.**		
b	Providing information on the primary purpose of data collection enables users to assess whether the data are relevant to their needs.	The VAT trader system operated by HMRC is designed to register all traders eligible for VAT and to collect the VAT due from, or make repayments to, those traders.	
B2.3	**Describe the concepts, definitions and classifications of the administrative populations and variables.**		
b	Whereas a statistical institution can adjust the concepts, definitions and classifications used in its own surveys to meet user needs, the institution usually has little or no influence over those used by administrative sources. By describing these, the user can decide whether the source meets their needs.	For the Working Family Tax Credit Data, Family Type is defined as either couples (married or unmarried) or lone parents. (ONS, 2001d)	
B2.4	**Describe metadata provided and not provided with the administrative source.**		Accessibility
b	All metadata made available by the data supplier should be described along with that which are missing. A description should include how the missing metadata affect the ability to assess the fitness for purpose of the administrative data. The overall quality of the available metadata should be highlighted, taking into account the completeness of the information. Metadata allow users to make appropriate use of data. Links to appropriate metadata ensure that this information is accessible.	There is currently a Metadata template in place for Neighbourhood Statistics (NeSS) within the Office for National Statistics (ONS). The data suppliers commit to providing "clear and comprehensive" metadata when they sign contracts or Service Level Agreements to ensure that users are able to interpret and make appropriate use of the statistics. Fields therefore contain content which can be understood by a range of users, and clearly highlight any specific issues affecting the quality or potential uses of the data.	

Ref.	Notes	Example	
B2.5 b	**Describe administrative data collection procedures.**		Accessibility
	The description should include the mode of data collection and any known problems, e.g. with questions, scanning, keying or coding. The information should be sufficient to allow replication of data collection procedures.	Administrative data are collected from most businesses using a paper questionnaire. All of the questions have explanatory notes. Coding is done by administrative staff using look-up tables, without the use of expert coders or electronic coding tools.	
B2.6 b	**Describe the format in which the administrative data are available.**		
	Administrative data are often available in different formats e.g. paper documents, flat csv files, magnetic tape. The formats available for users should be described.	Individual births and deaths records are available electronically (either on disk or by email) in Lotus Notes format from the register offices and can be loaded directly into the Office for National Statistics (ONS) system.	
B2.7 b	**Describe the extent of coverage of the administrative data and any known coverage problems.**		Accuracy
	This information is useful for assessing whether the coverage is sufficient. The population that the administrative data covers should be included along with all known coverage problems. There could be overcoverage (where duplicate records are included) or undercoverage (where certain records are missed). Special studies can sometimes be carried out to assess the impact of undercoverage and overcoverage. These should be reported where available. If appropriate and available quantitative measures can be used to highlight the extent of coverage issues: • Coverage error (see B1.16); • Estimated rate of undercoverage (see B1.17); • Estimated rate of duplicate units (see B1.18); • Estimated rate of ineligible units (see B1.19), and • Estimated rate of misclassified units (see B1.20).	Administrative data from HMRC include all businesses that are registered for VAT in the UK. However, any businesses that are below the VAT threshold are included where they register voluntarily.	

Ref.	Notes	Example	
B2.8	**Describe the known sources of error in administrative data.**		Accuracy
b	Metadata provided by the administrative source and/or information from other reliable sources can be used to assess data errors. The magnitude of any errors (where known) that have a significant impact on the administrative data should be made available to users. This will help the user to understand how accurate the administrative data are. If appropriate and available quantitative examples can be used: • Non-response error (see B3.2); • Measurement error (see B3.12); • Processing error (see B4.1); • Scanning and keying error rates (see B4.4); and • Coding error rates (see B4.13).	A question without explanatory notes causes errors on some businesses reporting their business activity. Businesses are expected to include the activity that contributes most financially. However this is not always the case, as the business activity may be highlighted because it takes longer or was the major activity at another point in time. A special study found that in 1% of cases the activity was incorrectly reported.	
B2.9	**Proportion of administrative records (units) with missing values**		
a	Missing values often occur in the administrative data received at source. Users need to be informed of the extent of these missing values. Estimating the proportion of missing vales gives an indication of the quality of the administrative data.	Proportion of units with missing value: $\frac{\text{Number of units with missing value}}{\text{Total number of units}}$	
B2.10	**Proportion of missing values in the administrative data by key items**		
a	The proportion of missing values can differ between the key items in the administrative data. The proportion of missing values for each item will allow users to decide whether there is sufficient data for them to perform analyses. Only units in scope for the particular item should be used in the calculation.	Proportion of missing values by key item: $\frac{\text{Number of units with missing value for item}}{\text{Total number of units for item}}$	
B2.11	**Describe the timescale since the last update of data from the administrative source.**		Timeliness
b	An indication of the timescale since the last update from administrative sources will provide the user with an indication of whether the statistical product is timely enough to meet their needs.	The Inter Departmental Business Register (IDBR) is based on a comprehensive range of data from the administrative sources. This is updated frequently, daily in the case of information on VAT traders from HMRC. (ONS, 2001a)	

Ref.	Notes	Example	
B2.12	**Describe the common identifiers of population units in administrative data.**		Coherence
b	Different administrative sources often have different population unit identifiers. The user can utilise this information to match records from two or more sources. Where there is a common identifier matching is generally more successful.	A common business identity code in Finland was established across administrative data in 2001. The Business information system behind business ID is a legal register administered by Finnish Tax Administration and National Board of Patents and Registration. (Statistics Finland, 2004)	

The following measures relate to statistical outputs derived wholly or in part from administrative data.

Ref.	Notes	Example	
B2.13	**Name each administrative data source used as an input into the statistical product.**		Relevance
b	Name all administrative sources and their providers. This information will assist users in assessing whether the statistical product is relevant for their intended use.	The IDBR is based mainly on three administrative sources: • traders registered with HM Revenue and Customs (HMRC) for Value Added Tax (VAT) purposes: • employers registered with HMRC operating a Pay As You Earn (PAYE) Scheme; and • incorporated businesses registered at Companies House (CH). In addition, the IDBR uses company linkages supplied by Dun and Bradstreet.	
B2.14	**Describe the extent to which the data from the administrative source meet statistical requirements.**		
b	Statistical requirements of the output should be outlined and the extent to which the administrative source meets these requirements stated. Gaps between the administrative data and statistical requirements can have an effect on the relevance to the user. Any gaps and reasons for the lack of completeness should be described, for example if certain areas of the target population are missed or if certain variables that would be useful are not collected. Any methods used to fill the gaps should be stated.	Administrative data provide only limited clinical information. One of the requirements of the statistical product is to provide users with clinical information associated with socio-demographic variables such as age, sex and socio-economic status. This information needs to be supplemented from other sources.	
B2.15	**Describe constraints on the availability of administrative data at the required level of detail.**		
b	Some administrative microdata have restricted availability or may only be available at aggregate level. Describe any restrictions on the level of data available and their effects on the statistical product.	Legal restrictions are in place which prevent the transfer of individual data from the HMRC corporation tax and self-assessment systems to ONS, but it is possible to receive aggregate data. (ONS, 2001a).	

Ref.	Notes	Example	
B2.16	**Describe the data processing known to be required on the administrative data source.**		Accuracy
b	Data processing may sometimes be required to check or improve the quality of the administrative data or create new variables to be used for statistical purposes. The user should be made aware of how and why data processing is used. If appropriate and available additional quantitative measures could be used to highlight data processing issues, e.g: • Total contribution to key estimates from imputed values (see B4.7); and • Editing Rate (see B4.11).	Validation checks are applied to the administrative data to make sure that the values are plausible. If they are not then the record is referred back to the data supplier for clarification.	
B2.17	**Describe the record matching methods and processes used on the administrative data sources.**		
b	Record matching is when different administrative records for the same unit are matched using a common, unique identifier or key variables common to both datasets. There are many different techniques for carrying out this process. A description of the technique (e.g. automatic or clerical matching) should be provided along with a description (qualitative or quantitative) of its effectiveness.	Matching is undertaken using software written by Search Software America. The matching process involves creating 'namekey' codes from the names supplied by the administrative departments, and those already stored on the IDBR. The process has resulted in the matching of 99% of records. (ONS, 2001a)	
B2.18	**Calculate match-rates, false positive match rates and false negative match rates for administrative data sources.**		
a	A false negative match is when two records relating to the same entity are not matched, or the match is missed. A false positive match is when two records are matched although they relate to two different entities. False positives and negatives can only be estimated if double matching is carried out. Double matching is where the matching is done twice, any discrepancies between the two versions can then be investigated.	$\text{Match rate} = \frac{\text{Number of records matched}}{\text{Total number of records}}$ $\text{False negative match rate} = \frac{\text{Number of false negative matches}}{\text{Total number of true and false matching pairs}}$ $\text{False positive match rate} = \frac{\text{Number of false positive matches}}{\text{Total number of true and false matching pairs}}$ (ONS, 2002c)	

Ref.	Notes	Example	
B2.19 b	**Describe the extent to which the administrative data are timely.**		Timeliness
	Provide information on how soon after their collection the statistical institution receives the administrative data. The effects of any lack of timeliness on the statistical product should be described.	In general, there is only a small lag between the Regional Registration Centre (RRC) processing the registration and ONS being sent the information, because new computer records are sent daily. (ONS, 2001a)	
B2.20 a	**Describe any lack of punctuality in the delivery of the administrative data source.**		
	Give details of the time lag between the scheduled and actual delivery dates of the data. Any reasons for the delay should be documented along with their effects on the statistical product.	The administrative data were received one week later than timetabled. The delay was due to a technical problem with the electronic delivery system. This had the effect that not all of the in-house validation checks could be run on the data before publication.	
B2.21 b	**Describe any changes in the legislative environment through which the administrative data are provided and the effects on the statistical product.**		Comparability
	Changes in legislation can cause discontinuities in the administrative data which in turn can affect the comparability over time of the statistical product. Detail any changes of this kind to enable the user to take account of them.	The Marriage Act was amended on 1 January 2001 to introduce a revised system of civil preliminaries to marriage in England and Wales. This system has abolished the provision to give notice, and subsequently marry, by certificate and licence. As a result of this, the licence column under civil marriages and the licence column under marriages with religious ceremonies other than the Church of England and the Church in Wales have been omitted. (ONS, 2002d).	
B2.22 b	**Describe changes over time in the administrative data and their effects on the statistical product.**		
	These can include changes in concepts, definitions, data collection purposes, data collection methods, file structure and format of the administrative data over time, as such changes can cause problems with comparability over time. Changes should be highlighted and an explanation of their effects on the statistical product explained to enable the user to assess comparability over time.	When local government areas are reorganised, the registration districts are also reorganised to the same boundaries as the areas they serve. In addition, local authorities can choose to amalgamate registration districts within their area. The result of these reorganisations is that data for registration districts are not always comparable between years. (ONS, 2002d)	

Ref.	Notes	Example	
B2.23 b	**Describe differences in concepts, definitions and classifications between the administrative source and the statistical output.**		Coherence
	There may be differences in concepts, definitions and classifications between the administrative source and statistical product. Concepts include the population, units, domains, variables and time reference and the definitions of these concepts may vary between the administrative data and the statistical product. Time reference problems occur when the statistical institution requires data from a certain time period but can only obtain them for another. Any effects on the statistical product need to be made clear along with any techniques used to remedy the problem.	Within HMRC, the industrial classification system for PAYE employers is not aligned with SIC (2003). The Office for National Statistics (ONS) uses a conversion table to convert those businesses that only have PAYE information to SIC (2003). Conversion is subject to error which may have an adverse effect on the IDBR quality. Administrative data are only available for turnover from April to March whereas the statistical requirements need information from January to December. Data are used from the previous return as well as the current to estimate the turnover for the current calendar year.	
B2.24 b	**Describe any adjustments made for differences in concepts and definitions between the administrative source and the statistical output.**		
	Adjustments may be required as a result of differences in concepts and definitions between the administrative data and the requirements of the statistical product. A description of why the adjustment needed to be made and how the adjustment was made should be provided to the users so they can assess the coherence of the statistical product with other sources.	Administrative data derived from individual tax records are received from England, Wales, Scotland and Ireland. The data from England, Wales and Scotland are reported in British pounds but the Irish data are reported in Euros. A conversion is done from Euros to British Pounds using the exchange rate on the first of the month.	

B3. Data Collection

Ref.	Notes	Example	
B3.1 a	**What were the target and achieved sample sizes?**		Accuracy
	The target sample size is the sample size specified under the sample design used for the survey. The 'achieved' sample size will typically differ from this because of non-response and non-contacts.	The target sample size was 70,000 enterprises, and the achieved sample size was 56,000 enterprises.	
B3.2 b	**Non-response error.**		
	Non-response error is the error that occurs from failing to obtain some or all of the information from a unit. This error cannot usually be calculated exactly. However, a statement should be provided to indicate to users that the data are subject to non-response error.	Not every business sampled responded to the survey. Because of this, the data are subject to non-response error. Non-response error is the difference between the results attained using the businesses that responded, and the results that would have been attained if every sampled business had responded.	
B3.3 b	**Estimated bias due to unit non-response for key estimates.**		
	Non-response occurs when it is not possible to collect information from a unit. Unit non-response occurs where an entire interview or questionnaire is not completed or is missing for the unit. Non-respondents may differ from respondents, which leads to non-response bias. Non-response bias can be estimated by: comparing the distributions of respondents and non-respondents for the same survey with respect to characteristics known for both groups; by comparing the distribution of respondents with associated distributions from other surveys; or by using information from other sources to gain information about non-respondents.	$(1-R)(\hat{\theta}_r - \hat{\theta}_t)$ where $\hat{\theta}_r$ is the mean estimate for respondents; $\hat{\theta}_t$ is the mean estimate for non-respondents; R is the response rate.	

Ref.	Notes	Example	
B3.4 a	**Unit response rate by sub-groups: household surveys.** **(Key Quality Measure)**		Accuracy
	The response rate is a measure of the proportion of units who respond to a survey. This indicates to users how significant the non-response bias is likely to be. Where response is high, non-response bias is likely to be less of a problem than when there are high rates of non-response. The rates opposite can be expressed as a percentage figure by multiplying by 100. NB: There may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders.	**Overall response rate**, unweighted: $\frac{I + P}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$ **Full response rate**, unweighted: $\frac{I}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$ **Co-operation rate**, unweighted: $\frac{I + P}{(I+P) + R + O + e_c UC}$ **Contact rate**, unweighted: $\frac{(I+P) + R + O + e_c UC}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$ **Refusal rate**, unweighted: $\frac{R}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$ **Eligibility rate**, unweighted: $\frac{(I+P) + (R+NC+O) + e_c UC + e_n UN}{(I+P) + (R+NC+O) + (UC+UN) + NE}$ Where: I = complete interview; P = partial interview; R = refusal; NC = non-contact; O = other non-response; NE = not eligible; e_c = estimated proportion of contacted cases of unknown eligibility that are eligible; UC = unknown eligibility, contacted case; e_n = estimated proportion of non-contacted cases of unknown eligibility that are eligible; UN = unknown eligibility, non-contact. Weighted rates are derived by applying the design weight, π^{-1}, to each variable when calculating the above rates. For example, 'I' becomes the weighted number of complete interviews, 'NC' becomes the weighted number of non-contacts, etc. (Lynn et al 2001)	

Ref.	Notes	Example	
B3.5 a	**Unit response rate by sub-groups: business surveys.** **(Key Quality Measure)**		Accuracy
	The response rate is a measure of the proportion of sampled units who respond to a survey. This indicates to users how significant the non-response bias is likely to be. Where response is high, non-response bias is likely to be less of a problem than when there are high rates of non-response. The rates opposite can be expressed as a percentage figure by multiplying by 100. NB: There may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders.	**Overall response rate**: Unweighted: $\frac{FC + FP + PC + PP}{(FC+FP+PC+PP) + RNU + NR + e(U)}$ Weighted: $\frac{\sum_{i \in R_O} w_i x_i}{\sum_{i \in S} w_i x_i}$ **Full response rate**: Unweighted: $\frac{FC}{(FC+FP+PC+PP) + RNU + NR + e(U)}$ Weighted: $\frac{\sum_{i \in R_F} w_i x_i}{\sum_{i \in S} w_i x_i}$ **Non-response rate**: Unweighted: $\frac{RNU + NR + e(U)}{(FC+FP+PC+PP) + RNU + NR + e(U)}$ Weighted: $\frac{\sum_{i \in Nr} w_i x_i}{\sum_{i \in S} w_i x_i}$ Where: FC = full period return with complete data; FP = full period return with partial data; PC = part period return with complete data; PP = part period return with partial data; RNU = returned but not used; NR = non-response; U = unknown eligibility; I = ineligible (out of scope); e = estimated proportion of cases of unknown eligibility that are eligible. e can be estimated as: $e = \frac{(FC+FP+PC+PP) + RNU + NR}{(FC+FP+PC+PP) + RNU + NR + I}$. w_i = weight for unit i; x_i = value of auxiliary variable for unit i; R_O = set of all responders; S = set of all eligible sampled units; R_F = set of full responders; Nr = set of non-responders (note that $Nr = S - R_O$).	

Ref.	Notes	Example	
B3.6 b	**Estimated bias due to item non-response for key estimates.**		Accuracy
	Non-response occurs when it is not possible to collect information from a sample unit. Item non-response occurs where a value for the item in question is missing or not obtained for the sample unit. Non-respondents may differ from respondents, which leads to non-response bias. Item non-response bias can be estimated from follow up studies.	$(1-R)(\hat{\theta}_r - \hat{\theta}_t)$ where $\hat{\theta}_r$ is the mean estimate for respondents; $\hat{\theta}_t$ is the mean estimate for non-respondents; R is the response rate.	
B3.7 a	**Key item response rates.**	**(Key Quality Measure)**	
	Where item response rates are high the effect of non-response bias is likely to be less than where item response rates are low. However, there may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders.	**Unweighted item response rate:** $\frac{\text{Number of units with a value for item}}{\text{Number of units in scope for the item}}$ **Weighted item response rate:** $\frac{\text{Total weighted quantity for item for responding units}}{\text{Total weighted estimate for the item for all units}}$	
B3.8 b	**Estimated variance due to non-response.**		
	Non-response generally leads to an increase in the variance of survey estimates, since the presence of non-response implies a smaller sample. An estimate of the extra variance (on top of the normal sampling variance) caused by non-response gives useful additional information about the accuracy of estimates.	An approximately unbiased variance estimator for VNR can be obtained by finding an approximately unbiased estimator for $V_q(\hat{\theta}^* \mid s)$ *[where q is the non-response mechanism; $\hat{\theta}^*$ is the estimated parameter, adjusted for non-response weighting adjustment or imputation; and s is the realised sample].* Since in practice the non-response mechanism is unknown, a non-response model is needed to approximate it. Therefore, the variance is valid only if the postulated non-response model is a good substitute for the true unknown non-response model. (Beaumont et al 2004)	

Ref.	Notes	Example	
B3.9 a	**Describe differences between responders and non-responders.**		Accuracy
	This indicates to users how significant the non-response bias is likely to be. Where response is high, non-response bias is likely to be less of a problem than when there are high rates of non-response. NB: There may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders.	From follow-up studies, it was found that, on average, individuals were less likely to answer questions related to their annual salary if they earned below £10,000 per annum. The reason for this may be due to a perception that their earnings were less than average.	
B3.10 a	**Rate of complete proxy responses.**		
	Proxy response is a response made on behalf of the sampled unit by someone other than the unit. This is the rate of complete interviews by proxy. It is an indicator of accuracy as information given by a proxy may be less accurate than information given by the desired respondent.	$\frac{\text{Number of units with complete proxy response}}{\text{Total number of eligible units}}$	
B3.11 a	**Rate of partial proxy responses.**		
	This is the rate of complete interviews given partly by the desired respondent(s) and partly by proxy. It is an indicator of accuracy as information given by a proxy may be less accurate than information given by the desired respondent.	$\frac{\text{Number of units with partial proxy response}}{\text{Total number of eligible units}}$	
B3.12 b	**Measurement error.**		
	Measurement error is the error that occurs from failing to collect the true data values from respondents. Sources of measurement error are: the survey instrument; mode of data collection; respondent's information system; respondent; and interviewer. Measurement error cannot usually be calculated. However, outputs should contain a definition of measurement error and a description of the main sources of the error.	Measurement error is the difference between measured values and true values. For the census, the main sources of measurement error are: • questionnaire design; • interviewer errors; and • respondent error.	

Ref.	Notes	Example	
B3.13 b	**Estimate of measurement error.**		Accuracy
	Measurement error is the difference between measured values and true values. It consists of bias (systematic error introduced where the measuring instrument is inaccurate – this error remains constant across survey replications) and variance (random fluctuations between measurements which, with repeat samples, would cancel each other out). Sources of measurement error are: the survey instrument; mode of data collection; respondent's information system; respondent; and interviewer. Measurement error can be detected during editing by comparing responses to different questions for the same value, for example age and date of birth. Alternatively, other records may be consulted to detect measurement error, for example administrative records.	An estimate of the extent of measurement error that occurred during the data collection phase was derived from re-interviewing a sub-sample of survey respondents between four and six weeks after the original interview. A subset of questions from the original survey was administered a second time to the selected respondents in face-to-face Computer Aided Personal Interviews (CAPI interviews). Responses from the re-interview were then compared to respondents' answers in their original interview for the same set of questions. Where any two responses to a question were different, respondents were asked to state which was the correct answer. They were also asked if the different responses reflected any changed circumstances between the original interview and the re-interview. The results suggest that approximately 8 per cent of responses obtained during the interview were answered differently during re-interview, and that half of these differences were due to changed circumstances during the period between interviews. For those respondents whose answers differed for other reasons, it is assumed that these differences were due to measurement error, either because of the interviewer, the respondent, or ambiguities in question wording.	

Ref.	Notes	Example	
B3.14 b	**Describe processes employed to reduce measurement error.**		Accuracy
	Describing processes to reduce measurement error indicates to users the accuracy and reliability of the measures. These processes may include questionnaire development, pilot studies, cognitive testing, interviewer training, etc.	In order to minimise any measurement error associated with question wording, the following procedures were carried out: • For the module on Leisure Expenditure, questions were developed following a series of focus groups which drew together a variety of interested parties, including experts in the field and a purposive sample from the 2002 FES survey. • The questions were then piloted with a randomly selected sample of households drawn from the postcode address file (PAF) using face-to-face Computer Aided Personal Interviews (CAPI). • Following this, a series of cognitive interviews were conducted in a post-hoc assessment of respondents' understanding of the survey questions. • The questions were then revised to reduce possible measurement error associated with ambiguous or misleading wording.	
B3.15 a	**Proportion of respondents with difficulty answering individual questions.**		
	Where respondents indicate that they have difficulty answering a question, this suggests that their responses to the question may be prone to measurement error. Information on difficulty in answering questions can be obtained in a variety of ways, e.g. from write-in answers on self-completion questionnaires, or from cognitive testing based on a purposive sample of the respondents.	$\frac{\text{Number of respondents with difficulty answering a question}}{\text{Total number of respondents who are asked that question}}$	

Ref.	Notes	Example	
B3.16 b	**Are there any items collected in the survey for which the data are suspect?**		Accuracy
	It is possible that data may be influenced, for example, by the personal or embarrassing nature of the information sought. If so, an assessment of the direction and amount of bias for these items should be made.	Because of the sensitive nature of questions on pupils' smoking, it was decided to obtain a more accurate biochemical marker for exposure to tobacco smoke, as well as asking about smoking on the self-completion questionnaire. Pupils were asked to provide a saliva specimen which was later analysed for the presence of cotinine, which is a metabolite of nicotine. The analysis suggested that more pupils under-reported the extent of their smoking than over-reported it.	
B3.17 b	**Assess differences due to different modes of collection.**		
	A mode effect is a measurement bias that is attributable to the mode of data collection. Mode effects can be investigated using experimental designs where sample units are randomly assigned into two or more groups. Each group is surveyed using a different data collection mode. All other survey design features are controlled. Differences in the response distributions for the different groups can be compared and assessed.	The basic univariate statistics, distributions and analysis of variance show no major differences between the face-to-face and telephone modes. (Hlebec et al 2002)	
B3.18 b	**Estimated interviewer variance.**		
	Interviewer variance is the variability between the results obtained by different interviewers. It can be difficult to estimate interviewer variance. In practice, special studies need to be set up to estimate this indicator of measurement error	The interviewer variance study was designed to measure how interviewer effects might have affected the precision of the estimates. The results showed that interviewer effects for most items were minimal and thus had a very limited effect on the standard error of the estimates. Interviewer variance was highest for open-ended questions. (Tsapogas 1996)	
B3.19 a	**Number of attempts to interview (for contacts and non-contacts).**		
	This is an indicator of processing error. For example, if fewer attempts are made to interview non-contacts than contacts, a systematic bias is introduced into the survey process.	Interviewers made up to four separate calls at different times of day for each household, until an answer was obtained. For households where contact was made, the average number of calls was 2.3. For non-contacts, the average number of calls was higher at 3.8 separate calls per household.	

B4. Data processing

Ref.	Notes	Example	
B4.1	**Processing error.**		Accuracy
b	Processing error is the error that occurs when processing data. It includes errors in data capture, coding, editing and tabulation of the data as well as in the assignment of survey weights. It is not usually possible to calculate processing error exactly. However, outputs should be accompanied by a definition of processing error and a description of the main sources of the error.	There are two types of processing error: • Systems errors: errors in the specification or implementation of systems needed to carry out surveys and process • Data handling errors: errors in the processing of survey data (ONS 1997)	
B4.2	**Describe processing systems and quality control.**		
b	This informs users of the mechanisms in place to minimise processing error by ensuring accurate data capture and processing.	Computer programs are run which will carry out a final check for benefit entitlement and will output any cases that look unreasonable. For example, any cases of people in receipt of One Parent Benefit when they are married will be listed. All cases recorded as a result of this exercise have been individually checked and edited when necessary. (OPCS 1996)	
B4.3	**Estimate of processing error.**		
b	Processing error occurs when processing data. It is made up of bias and variance. Processing bias is systematic error introduced by processing systems, e.g. an error in programming coding software that leads to the wrong code being consistently applied to a particular class. Processing variance consists of random errors introduced by processing systems, which, across replications, would cancel each other out, e.g. random keying errors in entering data. Processing error can be introduced via a number of mechanisms, including keying, coding, editing, weighting and tabulating. However, detecting processing bias and variance is not always possible.	In 1990, the Bureau estimated that 45 percent of the undercount was actually processing error, not undercount. (USCMB 2001)	

Ref.	Notes	Example	
B4.4 a	**Scanning and keying error rates (where the required value cannot be rectified).**		Accuracy
	Scanning error occurs where scanning software misinterprets a character (substitution error) or where scanning software cannot interpret a character and therefore rejects it (rejection error). Rejected characters can be collected manually and re-entered into the system. However substituted characters may go undetected. Keying error occurs where a character is incorrectly keyed into computer software. Accurate calculation of scanning and keying error rates depends on the detection of all incorrect characters. As this is unlikely to be possible, the detected scanning and keying error provides an indicator of processing error, not an absolute measure.	**Estimated scanning error rate**: $\frac{\text{Number of scanning errors detected}}{\text{Total number of scanned entries}}$ **Estimated keying error rate**: $\frac{\text{Number of keying errors detected}}{\text{Total number of keyed entries}}$ It is recommended that keying and scanning errors be calculated by field, as this is useful information for quality reporting. They can also be calculated by character, and this is useful for management information purposes.	
B4.5 a	**Key item imputation rates.**		
	Item non-response is when a record has partial non-response. Certain questions will not have been completed but other information on the respondent will be available. Imputation can be used to compensate for non-response bias, in that the known characteristics of non-responders can be used to predict values for missing items (e.g. type of dwelling may be known for certain non-responders and information from responders in similar types of dwelling may be imputed for the missing items). The imputation rate can provide an indicator of possible non-response bias, in that the imputation strategy may not perfectly compensate for all possible differences between responders and non-responders.	$\frac{\text{Number of units where the item is imputed}}{\text{Number of units in scope for the item}}$	

Ref.	Notes	Example	
B4.6	**Unit imputation rate.**		Accuracy
a	Unit non-response is when there is complete non-response for a person, business or household. Weighting is the usual method of compensation. However, imputation can be applied if auxiliary data are available or if the dataset is very large. Unit imputation can help to reduce non-response bias if auxiliary data are available.	$\frac{\text{Number of units imputed}}{\text{Total number of units}}$	
B4.7	**Total contribution to key estimates from imputed values.** **(Key Quality Measure)**		
a	The extent to which the values of key estimates are informed by data that have been imputed. This may give an indication of non-sampling error. This indicator applies only for means and totals. A large contribution to estimates from imputed values is likely to lead to a loss of accuracy in the estimates.	$\frac{\text{Total weighted quantity for imputed values}}{\text{Total weighted quantity for all final values}}$	
B4.8	**Assess the likely impact of non-response/imputation on final estimates.**		
b	Non-response error may reduce how representative the study sample is. An assessment of the likely impact of non-response/imputation on final estimates allows users to gauge how reliable the key estimates are as estimators of population values. This assessment may draw upon the indicator: 'total contribution to key estimates from imputed values' (see B4.7).	The effect of the imputation was to dampen down the observed increase by only 0.04%. This is trivial, given the sampling variability inherent in the LFS. (ONS 1997)	

Ref.	Notes	Example	
B4.9	**Proportion of responses requiring adjustment due to data not being available as required.**		Accuracy
a	Where data from respondents' information systems do not match survey data requirements (e.g. for respondents using non-standard reference periods), responses are adjusted as required. In this process, measurement error may occur. The higher the proportion of responses requiring adjustment, the more likely the presence of measurement error. The proportion is defined as the number of responses requiring adjustment divided by the total number of responses. This is relevant where respondents consult information systems when completing a survey, and may be more applicable to business than to household surveys.	$\frac{\text{Number of responses requiring adjustment}}{\text{Total number of responses}}$	
B4.10	**Edit failure rate.**		
a	The number of units rejected by edit checks, divided by total number of units. This indicates possible measurement error (although it may also indicate data capture error).	$\frac{\text{Number of units rejected by edit checks}}{\text{Total number of units}}$	
B4.11	**Editing rate (for key items).**	**(Key Quality Measure)**	
a	Editing rates may be higher due to measurement error (e.g. poor question wording) or because of processing error (e.g. data capture error). In addition, editing may introduce processing error if the editing method is not a good strategy to compensate for values that require editing. 'Item' here refers to an individual question.	$\frac{\text{Number of units changed by editing for the item}}{\text{Total number of units in scope for the item}}$	
B4.12	**Total contribution to key estimates from edited values.**		
a	The extent to which the values of key estimates are informed by data that have been edited. This may give an indication of the effect of measurement error on key estimates. This indicator applies only for means and totals.	$\frac{\text{Total weighted quantity for edited values}}{\text{Total weighted quantity for all final values}}$	

Ref.	Notes	Example	
B4.13	**Coding error rates.**		
a	Coding error is the error that occurs due to the assignment of an incorrect code to data. Coding error rate is the number of coding errors divided by the number of coded entries. It is a source of processing error. There are various ways to assess coding errors, depending upon the coding methodology employed, e.g. random samples of coded data may be double checked, and validation checks may be carried out to screen for impossible codes. Coding error is found in all coding systems, whether manual or automated. Not all coding errors may be detected so this is more likely to be an estimated rate.	$\frac{\text{Number of coding errors detected}}{\text{Total number of coded entries}}$	Accuracy

B5. Weighting and estimation

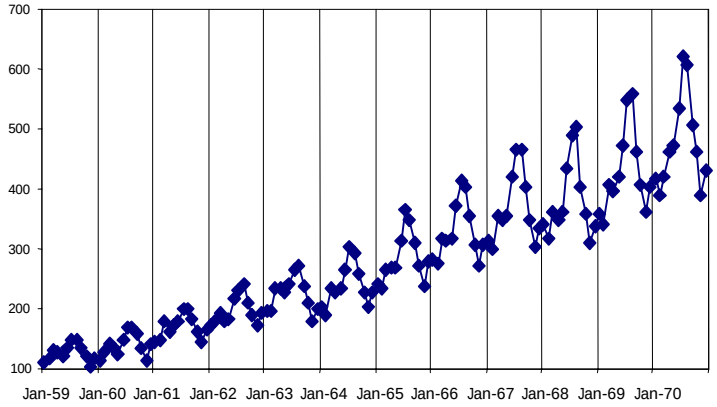
Ref.	Notes	Example	
B5.1 b	**Sampling error.**		Accuracy
	Sampling error is the difference between a population value and an estimate based on a sample. It is not usually possible to calculate sampling error. In practice, the standard error is often used as an indicator of sampling error. Outputs derived from sample surveys should contain a statement that the estimates are subject to sampling error and an explanation of what this means.	Sampling errors in the LFS arise from the fact that the sample chosen is only one of a very large number of samples which might have been chosen from the population. It follows that a quarterly estimate of, say, the number of people in employment is only one of a large number of samples that might have been made. (Risdon 2003)	
B5.2 a	**Estimated standard error for key estimates of level.** **(Key Quality Measure)**		
	The standard error is an indication of the accuracy of an estimate calculated as the positive square root of the variance of the sampling distribution of a statistic. The standard error gives users an indication of how close the sample estimator is to the population value: the larger the standard error, the less precise the estimator. Estimates of level include estimates of means and totals. The coefficient of variation is a measurement of the relative variability of an estimate, sometimes called the relative standard error (RSE).	The method of estimating standard error depends on the type of estimator being used. The standard error estimate is the positive square root of the variance estimate. Estimated coefficient of variation is defined as: $\frac{\text{Standard error estimate}}{\text{Level estimate}} \times 100$	

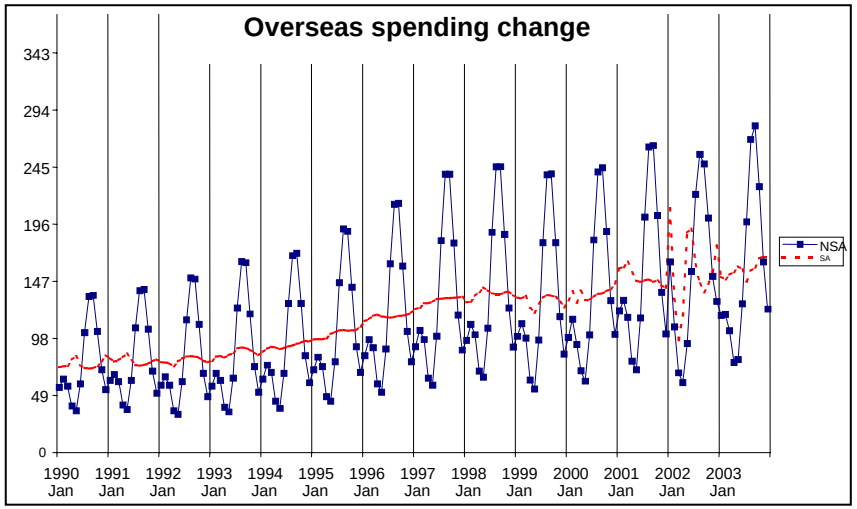
Ref.	Notes	Example	
B5.3 a	**Estimated standard error for key estimates of change.** **(Key Quality Measure)**		Accuracy
	Where the absolute or relative change in estimates between two time points is of interest to users, it is desirable to calculate standard errors for this change.	For absolute changes: $\hat{Var}(\hat{Y}_2 - \hat{Y}_1) = \hat{Var}(\hat{Y}_1) + \hat{Var}(\hat{Y}_2) - 2\hat{Cov}(\hat{Y}_1, \hat{Y}_2)$ For relative changes: $\hat{Var}\left(\frac{\hat{Y}_2}{\hat{Y}_1}\right) \approx \left(\frac{\hat{Y}_2}{\hat{Y}_1}\right)^2 \left[\frac{\hat{Var}(\hat{Y}_1)}{\hat{Y}_1^2} + \frac{\hat{Var}(\hat{Y}_2)}{\hat{Y}_2^2} - \frac{2\hat{Cov}(\hat{Y}_1, \hat{Y}_2)}{\hat{Y}_1 \hat{Y}_2}\right]$	
B5.4 a	**Reference/link to documents containing detailed standard error estimates.**		
	The reference or link should lead to detailed standard error estimates, also providing confidence intervals or coefficients of variation, where possible.	Standard errors are contained in the technical report. This can be found on the National Statistics website: www.statistics.gov.uk	
B5.5 b	**Describe variance estimation method.**		
	Describing the variance estimation method gives an indication of the likely accuracy of variance estimates. The description should include factors taken into consideration (e.g. misclassifications, non-response, etc).	The standard errors of the UK LFS estimates shown in Annex X are produced using a Taylor series approach by treating the Interviewer Area as a stratum and the household as a primary sampling unit (PSU). Currently only a very approximate allowance for the population weighting method is made. It is possible that the standard errors of most estimates would be reduced if population weighting was taken into account – ONS has investigated alternative methods of calculating standard errors to produce valid sampling errors for post-stratified estimates. (ONS 1997)	

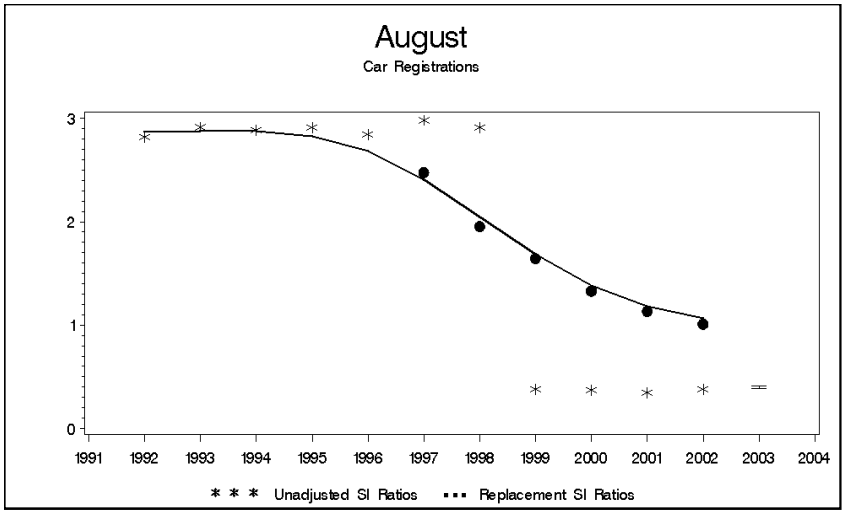
Ref.	Notes	Example	
B5.6	**Design effects for key estimates.**		Accuracy
a	The design effect indicates how well, in terms of variance, the sampling method used by the survey fares in comparison to simple random sampling. The design effect is often used to determine survey sample size when using cluster sampling.	Design effect = $\frac{\text{Var(estimate)}}{\text{Var}_{SRS}\text{ (estimate)}}$ where: Var(estimate) is the variance of the estimate under the sampling method used by the survey. Var_{SRS} (estimate) is the variance that would be obtained if the survey used simple random sampling.	
B5.7	**Effective Sample Size**		
a	The Effective Sample Size (abbreviated to *neff)* indicates the impact of complex sampling methods on standard errors. It gives the sample size, for a particular estimate, that would have been achieved using simple random sampling. The Effective Sample Size is equivalent in terms of precision to that achieved using more complex sampling methods. It is measured as the ratio of the achieved sample to the design effect for each estimate, and can be different across different variables or subgroups.	$neff = \frac{n_{achieved}}{Deff}$ Where: *neff* is the Effective Sample Size $n_{achieved}$ is the Achieved Sample Size *Deff* is the Design Effect	
B5.8	**Kish's Effective Sample Size**		
a	Kish's Effective Sample Size (abbreviated to $neff_{Kish}$) gives the approximate size of an equal probability sample which would be equivalent in precision to the unequal probability sample used. It indicates the approximate impact of weighting on standard errors.	$neff_{kish} = \frac{\left[\sum_{i=1}^{n} w_i\right]^2}{\left[\sum_{i=1}^{n} w_i^2\right]}$ where: w_i is the weight for unit i and is inversely proportional to the unit's selection probability.	
B5.9	**What method of sample weighting was used to calculate estimates?**		
b	Data from sample surveys can be used to estimate unknown population values using sample weighting. The method of sample weighting chosen has a bearing on the accuracy of estimates. Greater accuracy can often be achieved by using an estimator that takes into account an assumed relationship between the variables under study and auxiliary information.	The R&D expenditure total is estimated separately for each 'cell' using ratio estimation with company employment as the auxiliary variable. (ONS 2003f)	

Ref.	Notes	Example	
B5.10	**Model assumption error.**		Accuracy
b	Model assumption error is the error that occurs when assumed models do not exactly represent the population or concept being modelled. It is not usually possible to measure model assumption error exactly. However, estimates produced using models should be accompanied by a description of the model assumptions made and an assessment of the likely effect of making these assumptions on the quality of estimates.	The estimates in this report were calculated by making assumptions about the true underlying population. The model assumed for the population is the most appropriate based on the information available. However, it is impossible to completely reflect the true nature of the population through modelling. Because of this, there will be a difference between the modelled and true populations. This difference introduces error into the survey estimates.	
B5.11	**Description of assumptions underlying models used in each output.**		
b	In order to assess whether the models used are appropriate, it is important to record the assumptions underlying each model.	The standard errors for this index were calculated using a parametric bootstrap. There is an assumption that the data follow a multivariate normal distribution. The calculation of the variance-covariance matrix required to produce these estimates was simplified by making the assumption that the sample selection processes for each input to this index are independent of each other.	
B5.12	**Evaluation of whether model assumptions do/are likely to hold.**		
b	Studies to evaluate how well model assumptions hold give an indication of model assumption error.	The variance of the IoP is fairly insensitive to the assumptions made about the variance of the EPD. This continues to be the case at 4–digit level. Thus the assumption made about the variance of the EPD when deriving formula (4) should be suitable. (Kokic 1996)	
B5.13	**Estimated bias introduced by models.**		
b	Using models to improve the precision of estimates generally introduces some bias, since it is impossible to reflect completely the underlying survey populations through modelling. It is often possible to estimate the bias introduced by models.	There are different formulae to estimate bias depending on the model employed.	

B6. Time series

Ref.	Notes	Example	
B6.1	**Original data visual check.**		Accuracy
a	Graphed data can be used in a visual check for the presence of seasonality, decomposition type model (multiplicative or additive), extreme values, trend breaks and seasonal breaks.	The graph below shows the Airline Passengers original series. From the graph it is possible to see that the series has repeated peaks and troughs which occur at the same time each year. This is a sign of seasonality. It is also possible to see that the trend is affecting the impact of the seasonality. In fact the size of the seasonal peaks and troughs are not independent of the level of the trend suggesting that a multiplicative decomposition model is appropriate for the seasonal adjustment. This series does not show any particular discontinuities (outliers, level shifts or seasonal breaks). **Airline Passengers – Non seasonally adjusted**  	

Ref.	Notes	Example	
B6.2	**Compare the original and seasonally adjusted data.**		Accuracy
a	By graphically comparing the original and seasonally adjusted series, it can be seen whether the quality of the seasonal adjustment is affected by any extreme values, trend breaks or seasonal breaks and whether there is any residual seasonality in the seasonally adjusted series.	By comparing the original and the seasonally adjusted series in the graph below it can be seen that the seasonal adjustment performs well until January 2002 where the presence of an outlier distorts the pattern of the series. 	

Ref.	Notes	Example	
B6.3	**Graph of the seasonal-irregular ratios.**		Comparability
a	It is possible to identify a seasonal break by a visual inspection of the seasonal-irregular ratios. Any changes in the seasonal pattern indicate the presence of a seasonal break.	In the example below, there has been a sudden drop in the level of the seasonal-irregular component (called SI ratios or detrended ratios) for August between 1998 and 1999. This is caused by a seasonal break in the car registration series which was due to the change in the car number plate registration legislation. Permanent prior adjustments should be estimated to correct for this break. If no action is taken to correct for this break, some of the seasonal variation will remain in the irregular component resulting in residual seasonality in the seasonally adjusted series. The result would be a higher level of volatility in the seasonally adjusted series and a greater likelihood of revisions. 	

Ref.	Notes	Example	
B6.4	**Analysis of variance.**		Accuracy
a	The analysis of variance (ANOVA) compares the variation in the trend component with the variation in the seasonally adjusted series. The variation of the seasonally adjusted series consists of variations of the trend and the irregular components. ANOVA indicates how much of the change of the seasonally adjusted series is attributable to changes in primarily the trend component. The statistic can take values between 0 and 1 and it can be interpreted as a percentage. For example, if ANOVA=0.716, this means that 71.6 per cent of the movement in the seasonally adjusted series can be explained by the movement in the trend component and the remainder is attributable to the irregular component. This indicator can also be used to measure the quality of the estimated trend. The tables mentioned in the example opposite (D12, D11, D11A and A1) are automatically generated by X12ARIMA.	$\text{ANOVA}=\frac{\sum_{t=2}^{n}\left(D12_t - D12_{t-1}\right)^2}{\sum_{t=2}^{n}\left(D11_t - D11_{t-1}\right)^2}$ $D12_t$ = data point value for time t in table D12 (final trend cycle) of the analytical output. $D11_t$ = data point value for time t in table D11 (final seasonally adjusted data) of the output. Notes: 1. If constraining is used and the D11A table is produced, D11A point values are used in place of D11 in the above equation. 2. If the statistic is used as a quality indicator for trend, table A1 (final seasonally adjusted series from which the final trend estimate is calculated) should be used instead of table D11. The $D12_t$ values should be taken from the final trend output.	

Ref.	Notes	Example	
B6.5	**Months (or quarters) for cyclical dominance.**		Comparability
a	The months for cyclical dominance (MCD) or quarters for cyclical dominance (QCD) are measures of volatility of a monthly or quarterly series respectively. This statistic measures the number of periods (months or quarters) that need to be spanned for the average absolute percentage change in the trend component of the series to be greater than the average absolute percentage change in the irregular component. For example, an MCD of 3 implies that the change in the trend component is greater than the change in the irregular component for a span at least three months' long. The MCD (or QCD) can be used to decide the best measure of short-term change in the seasonally adjusted series – if the MCD=3, the three-months-on-three-months growth rate will be a better estimate than the month-on-month growth rate. The lower the MCD (or QCD), the less volatile the seasonally adjusted series is and the more appropriate the month-on-month growth rate is as a measure of change. The MCD (or QCD) value is automatically calculated by X12ARIMA and is reported in table F2E of the analytical output. For monthly data the MCD takes values between 1 and 12, for quarterly data the QCD takes values between 1 and 4.	MCD (or QCD) = d for which $\frac{I_d}{C_d} < 1$, where I_d is the final irregular component at lag d and C_d is the final trend component at lag d.	

Ref.	Notes	Example	
B6.6	**The M7 statistic.**		Accuracy
a	This indicates whether the original series has a seasonal pattern or not. Low values indicate clear and stable seasonality has been identified by X12ARIMA. The closer to 0, the better the seasonal adjustment. Values around 1 imply that the series is marginally seasonal. An M7 value close to 3 indicates that the series is non-seasonal. The M7 statistic is calculated by X12ARIMA and can be obtained from the F3 table of the analytical output.	$$M7 = \sqrt{\frac{1}{2}\left(\frac{7}{F_s} + \frac{3F_m}{F_s}\right)}$$ Where: $$F_M = \frac{S_B^2/(N-1)}{S_R^2/(N-1)(k-1)} \qquad F_S = \frac{S_A^2/(k-1)}{S_R^2/(n-k)}$$ F_M = F-statistic capturing moving seasonality. S_B^2 is the inter-year sum of squares, S_R^2 is the residual sum of squares, N is the number of observations. F_S = F-statistic capturing stable seasonality. S_A^2 is the variance due to the seasonality factor and S_R^2 is the residual variance. k = number of periods within a year, k = 4 for quarterly series, k = 12 for monthly series. M7 compares on the basis of F-test statistics the relative contribution of stable (statistic F_M) and moving (statistic F_S) seasonality.	

Ref.	Notes	Example	
B6.7	**Contingency table Q.**		Comparability
a	The contingency table Q (CTQ) shows how frequently the gradient of the trend and the seasonally adjusted series over a one period span have the same sign. CTQ can take values between 0 and 1. A CTQ value of 1 indicates that historically the trend component has always moved in the same direction as the seasonally adjusted series. A CTQ value of 0.5 suggests that the movement in the seasonally adjusted series is likely to be independent from the movement in the trend component; this can indicate that the series has a flat trend or that the series is very volatile. A CTQ value between 0 and 0.5 is unlikely but would indicate that there is a problem with the seasonal adjustment. If constraining is used, point values from table D11A are used in place of D11 in the above equation. Tables D11A and D11 are automatically generated by X12ARIMA.	$Q = \frac{U_{11} + U_{22}}{U_{11} + U_{21} + U_{12} + U_{22}}$ where **Delta C>0** **Delta C <= 0** **Delta SA >0** U_{11} U_{12} **Delta SA <= 0** U_{21} U_{22} $\Delta SA = D11_t - D11_{t-1}$ (Delta SA is the change in the SA data) $\Delta C = D12_t - D12_{t-1}$ (Delta C is the change in the trend)	
B6.8	**STAR (stability of trend and adjusted series rating) – for multiplicative models only.**		Accuracy
a	This indicates the average absolute percentage change of the irregular component of the series. The STAR statistic is applicable to multiplicative decompositions only. The expected revision of the most recent estimate when a new data point is added is approximately half the value of the STAR value, e.g. a STAR value of 7.8 suggests that the revision is expected to be around 3.9 per cent. Table D13 in the example opposite is automatically generated by X12ARIMA.	$STAR = \frac{1}{N-1}\sum_{t=2}^{N}\left\lvert\frac{(D13_t - D13_{t-1})}{D13_{t-1}}\right\rvert$ $D13_t$ = data point value for time t in table D13 (final irregular component) of the output. N = number of observations in table D13.	

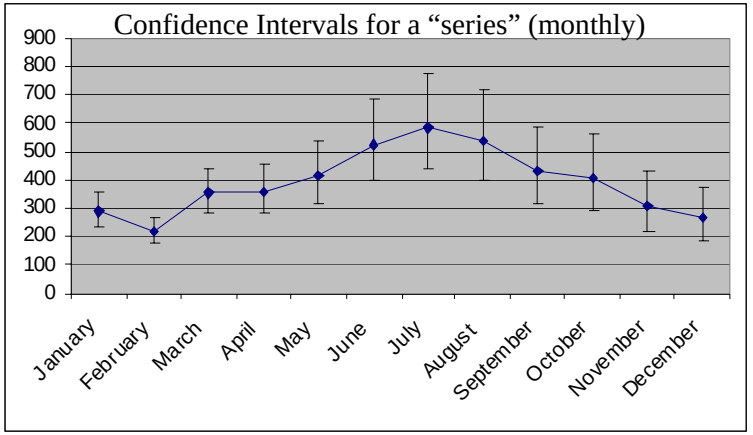
Ref.	Notes	Example	
B6.9	**Comparison of annual totals before and after seasonal adjustment.**		Accuracy
a	This is a measure of the quality of the seasonal adjustment and of the distortion to the seasonally adjusted series brought about by constraining the seasonally adjusted annual totals to the annual totals of the original series. It is particularly useful to judge if it is appropriate for the seasonally adjusted series to be constrained to the annual totals of the original series. The tables in the example opposite (tables E4 and D11) are automatically generated by X12ARIMA.	Formula (multiplicative models): $\frac{1}{n}\sum_{i=1}^{n}\left\|(E4^{D11}_{total(t)} - 100)\right\|$ $E4^{D11}_{total(t)}$ = unmodified ratio of annual totals for time *t* in table E4 E4 is the output table that calculates the ratios of the annual totals of the original series to the annual totals of the seasonally adjusted series for all the n years in the series. Formula (additive models): $\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{E4^{D11}_{t}}{D11_{total(t)}}\right\|$ $E4^{D11}_{total(t)}$ = unmodified difference of annual totals for time *t* in table E4. For additive models E4 is the output table that calculates the difference between original annual total and seasonally adjusted annual totals for all the n years in the series.	

Ref.	Notes	Example	
B6.10	**Normality test.**		Accuracy
a	The assumption of normality is used to calculate the confidence intervals from the standard errors. Consequently, if this assumption is rejected the estimated confidence intervals will be distorted even if the standard forecast errors are reliable. A significant value of one of these statistics (Geary's and Kurtosis) indicates that the standardised residuals do not follow a standard normal distribution, hence the reported confidence intervals might not represent the actual ones. (X12ARIMA tests for significance at the 1 per cent level.)	$\alpha = \frac{\frac{1}{n}\sum_{i=1}^{n}\left\|X_i - \bar{X}\right\|}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}$ $b_2 = \frac{m_4}{m_2^2} = \frac{n\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^4}{\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)^2}$ The above two statistics (Geary's and Sample Kurtosis respectively) test the regARIMA model residuals for deviations from normality. M_4 = fourth moment. M_2^2 = square of the second moment.	

Ref.	Notes	Example
B6.11	**P-values**	
a	This is found in the 'Diagnostic checking Sample Autocorrelations of the Residuals' output table generated by X12ARIMA. The p-values show how good the fitted model is. They measure the probability of the Autocorrelation Function occurring under the hypothesis that the model has accounted for all serial correlation in the series up to the lag. P-values greater than 0.05, up to lag 12 for quarterly data and lag 24 for monthly data, indicate that there is no significant residual autocorrelation and that, therefore, the model is adequate for forecasting.	**DIAGNOSTIC CHECKING** **Sample of Autocorrelation Residuals** Model fitted is: ARIMA (0 1 1)(0 1 1) (Quarterly data so 12 lags will be extracted) ACF = AutoCorrelation Function. SE = standard error. Q = Liung-box Q statistic. DF = degrees of freedom. P = p-value of the Q statistic. The model will not be adequate if any of the p-values of the 12 lags is less than 0.05 and will be adequate if all the p-values are greater than 0.05. Because we are estimating two parameters in this model (one autoregressive and one moving average parameter), no p-values will be calculated for the first two lags as two DFs will be automatically lost hence we will start at lag 3. Here all the p-values are greater than 0.05, e.g. at lag 12 Q=12.44, DF=10 and p-value = 0.256, therefore this model is adequate for forecasting.
B6.12	**Percentage forecast standard error.**	
a	The percentage forecast standard error is an indication of the efficiency of an estimator. It is calculated as the ratio of the square root of the mean square error of the estimator to the predicted value. The larger the forecast standard error, the wider the confidence interval will be.	$\frac{\text{forecast standard error}}{\text{forecast}} \text{x} 100$

Lag	1	2	3	4	5	6	7	8	9	10	11	12
ACF	-0.06	0.12	-0.1	-0.31	-0.25	-0.13	0.21	-0.09	0.21	-0.05	0.09	-0.15
SE	0.19	0.19	0.2	0.21	0.22	0.23	0.24	0.24	0.24	0.24	0.24	0.24
Q	0.09	0.56	0.86	4.04	6.2	6.8	8.51	8.86	10.8	10.89	11.31	12.44
DF	0	0	1	2	3	4	5	6	7	8	9	10
P	0	0	0.355	0.133	0.102	0.147	0.13	0.181	0.148	0.208	0.255	0.256

Accuracy

Ref.	Notes	Example	
B6.13	**Graph of the confidence intervals.**		Accuracy
a	Time to be plotted along the x-axis, and forecast estimates and confidence intervals along the y-axis.		
B6.14	**Percentage difference of constrained to unconstrained values.**		
a	This indicates the percentage difference between the constrained and the unconstrained value. This will need to be calculated for each data point being constrained. The average, minimum and maximum percentage differences (among all data points of all the series) of the unconstrained should be calculated to give an overview of the effect of constraining.	$\left(\frac{\text{constrained data point}}{\text{unconstrained data point}} - 1\right) \text{x } 100$	
B6.15	**Provide an estimate of discontinuity due to change.**		Comparability
a	Discontinuity is a change in data due to changes in statistical methodology or real-world events. Discontinuities may impact upon comparability of data: the greater the discontinuity, the more risk to comparability of data before and after the change. Some discontinuities are quantifiable, while others are not.	Following the new accounting legislation, brought into effect from June 2004, the comparability of new figures with back series has been reduced because of the removal of fixed assets from the Profit and Loss (P&L) Sheet. It is possible to gauge the level of discontinuity by comparing the final 2004 P&L figure with what it would have been if fixed assets were to be included. On average, the discontinuity is in the region of –1 per cent.	

Ref.	Notes	Example	
B6.16	**All discontinuities are flagged and explained.**		Comparability
a	The comparability of data over time may be compromised by discontinuities. Identification of a discontinuity is a judgement based on evidence that may be statistical, contextual or both.	From 1997 onwards, note that for 63.3 (activities of travel agencies and tour operators; tourist assistance activities not elsewhere classified) total turnover, total purchases and approximate gross value added at basic prices – now include monies received and paid in respect of tour operators' invoices. From 2000 the coverage of the ABI financial inquiry has been extended to include Divisions 02 and 05, Forestry and Fishing. (ONS 2003a)	
B6.17	**Describe and assess the impact on data of any real-world events or changes in methods over time.**		
a	This is a description of changes in concepts, definitions, classifications and methods and their causes, whether due to statistical considerations (for example, a change in measurement instrument or in the way that it is applied) or real-world events. The description should include an assessment of the impact of discontinuities on data comparability over time.	These changes include industrial action taken by registration officers, which affected the quality of information about deaths from injury and poisoning. This action meant that details normally supplied by Coroners were not available and the statistics were significantly affected. Information normally requested to aid coding of underlying cause of death was also unavailable. Figures on injury and poisoning for 1981, with the exception of suicides, must thus be treated with caution. (ONS 2001b)	
B6.18	**Describe backseries available.**		
b	Description of which data from previous periods are available to users.	The backseries contains data on the following: public sector finance, central government revenue and expenditure, money supply and credit, banks and building societies, interest and exchange rates, financial accounts, capital issues, balance sheets and the balance of payments. Monthly data are available for this series going back to April 2002.	

B7. Statistical disclosure control

Ref.	Notes	Example	
B7.1 b	**Describe any confidentiality agreements or assurances given to protect participants' confidentiality.**		Accessibility
	This can be a broad textual description, or for fuller information it may display the actual agreement or assurance given to participants. The description may refer to standard agreements, such as the Statistics of Trade Act 1947, which govern the confidentiality of individual data.	As they are at the level of individual businesses, access to these data is restricted. The confidentiality of the data is legally enforced by the Statistics of Trade Act 1947. Confidentiality is a very important concern for the ONS and the respondents. As data custodian the ONS has to be able to give a direct assurance of confidentiality to maintain the trust and confidence of business respondents. (Barnes et al 2001)	
B7.2 b	**Describe how data collection security was ensured.**		
	This should be a broad description of how data collection security was ensured, for example through secure data capture systems or secure data transfer.	Every member of the field team (including contractors) had to sign an undertaking to guarantee not to divulge information obtained during the course of the Census to anyone outside the Census organisation. All field staff were directed to treat completed forms and associated documents with utmost care and ensure that secure accommodation was used for the storage of completed forms. They were also instructed to ensure that access was restricted to staff who had signed the 2001 Census confidentiality undertaking. Requirements were further reinforced by the threat of penalties in the event of there being any divulgence or use of Census information gained during the course of employment on Census duties. (ONS 2002a)	

Ref.	Notes	Example	
B7.3 b	**Describe how processing security was ensured.**		Accessibility
	This should be a broad description of how processing security was ensured, for example through secure office systems or secure access.	A digital image of each 2001 Census form is captured for processing. A microfilm version will be retained in strict confidentiality to be released after 100 years as a public record of great interest to family historians. The paper forms will be pulped and recycled. (ONS 2002b)	
B7.4 b	**What user needs have been considered when assessing which statistical disclosure control methods to apply?**		Relevance
	Producers of statistics should consider user needs when considering what form of statistical disclosure control they should employ. User needs that should be considered may include levels of classification, geographies, etc.	As well as producing tables of turnover by industry at SIC two-digit level, there was also a requirement to produce bespoke tables when requested. To protect against the identification of an individual business, the statistical disclosure control method applied a threshold rule at Enterprise Group level and a modified dominance rule (p% rule) at Reporting Unit level. Cells failing either of these rules were suppressed, and to avoid disclosure by differencing, additional cells were selected for secondary suppressions manually. Only a small number of cells were suppressed so the tables are still of great use and confidentiality is maintained.	
B7.5 b	**How were records anonymised, i.e. removal of direct identifiers?**		Accessibility
	Examples of direct identifiers include name, address, National Insurance number, etc. The removal of identifiers does not in itself ensure data confidentiality: additional measures to protect the data may still be required.	Names, telephone numbers and addresses were recorded for back-checking purposes but this information was kept in a separate file. Where names and addresses were recorded on paper, these were separated from the responses when questionnaires were returned to the office. (ONS 2003j)	

Ref.	Notes	Example	
B7.6	**Describe in broad terms the statistical disclosure control method(s) applied.**		Accessibility
b	This textual description should be broad enough to give an indication of the methods employed, without allowing the user to unpick the methods and derive any disclosive data. For tabular data, reference should be made to both pre- and post-tabulation statistical disclosure control methods employed.	MICRODATA: The Census Bureau also protects confidentiality by considering other issues. For example, sampling information is scrambled (identity of PSUs or control numbers); and if some data items isolate readily identifiable sub-populations, then blanking, imputation, recoding, or noise inoculation are used. (Hawala 2003) TABULAR DATA: Small counts for all tables issued for England and Wales are adjusted. In addition there is swapping of records in the output database, and broad limitations are placed on details in tables to be produced for small populations. There are also minimum thresholds of numbers of persons and households for the release of sets of output. These are 40 households and 100 persons for Census Area Statistics, and 400 households and 1,000 people for Standard Tables. (ONS 2002b)	
B7.7	**Describe the impact of statistical disclosure control methods on the quality of the output.**		Accuracy
b	This should be a broad textual statement of impact on quality (e.g. on the bias and variance of estimates), without allowing the user to unpick the original disclosive data.	MICRODATA: Local suppressions are determined automatically and optimally. In practice, one does not want too many local suppressions as this would introduce bias in the statistical results obtained from the protected micro dataset. (De Waal et al 1996) TABULAR DATA: The techniques of record swapping and small cell adjustment added uncertainty to tabular counts in order to protect the confidentiality of individuals. The methodology was designed to be unbiased. Any changes to the data from disclosure control methods tended to cancel out in the sense of having a mean value that tended to zero. (ONS 2005)	

Ref.	Notes	Example	
B7.8	**Describe the impact of any changes in statistical disclosure methods over time.**		Comparability
b	Users should be informed if there are any changes in statistical disclosure methods that result in incomparable data over time (including historic comparisons).	MICRODATA: Historically, confidentiality was maintained on the annual database by restricting the range of variables made available. However, software designed to link records for the same people has improved considerably meaning that the risk of identification has also increased. Although the risk for most respondents remains negligible, the National Statistician has taken the decision to no longer make available databases with local area identifiers. Thus there are no publicly accessible LADB databases for 2000/01 and 2001/02. (ONS 1999) TABULAR DATA: The magnitude tables released have been anonymised by application of both a threshold level and a dominance rule. The parameters for both these rules were changed in 2000 leading to the suppression of more cells. The frequency tables released were protected by random rounding up to 2003. Since then controlled rounding has been applied, which could damage the data more but retain additivity.	
B7.9	**How well does the output meet user needs after applying statistical disclosure control?**		Relevance
a	This should indicate which user needs have been met, and highlight any areas where user needs have not been met (e.g. on levels of classifications available or geographies) because of statistical disclosure methodologies.	User feedback from previous instances of this survey revealed that there was a demand for small area data based on both administrative geographies and postal geographies. Therefore, small area data have now been provided for both these geographies for the current survey instance. In terms of geography, this output now meets declared user needs. However, the survey outputs are often used in historic comparisons. In order to enable users to make comparisons over time, it is planned to release earlier outputs in both geographies as well.	

Ref.	Notes	Example	
B7.10	**Does statistical disclosure control affect significant relationships in the data?**		Accuracy
b	Any affect of statistical disclosure control on significant relationships in the data should be broadly stated to avoid the risk of users unpicking the data and restoring disclosive data.	Potentially disclosive information was sometimes protected by swapping a sample of records with similar records in other geographical areas. The procedure was designed so that the integrity of swapped data was not significantly different from that of unswapped data. The percentage of records swapped remains confidential. (ONS 2005)	
B7.11	**Have harmonised statistical disclosure control methods been applied?**		Coherence
b	This enables users to assess whether there are differences between datasets in their disclosure control treatment.	There are substantial co-ordination problems to ensure that adequate and reasonably uniform standards of data quality and confidentiality protection are applied. A further complication arises because in some cases knowledge of the parameters used in the disclosure control method sometimes helps an intruder to undo the disclosure protection. For this reason, each agency will apply slightly different confidentiality procedures from the others, using parameters to characterise each variant. (Armitage and Brown 2003)	
B7.12	**When were statistical disclosure control methods last reviewed? Describe the results.**		Relevance
b	This gives an indication of the effectiveness of disclosure control methodologies, in showing how new developments (such as new technologies, the availability of additional information or new publication formats) have been accounted for in updated methodologies. The description should include reference to any changes implemented as a result of the review.	These expanded race/ethnicity distinctions, combined with the increased ability to electronically match different data files, prompted the DRB to review in detail the potential disclosure risk related to the Special EEO Tabulation format. As a result, the DRB issued two rules in 2000 affecting the Census 2000 Special EEO Tabulation. 1. Any occupational category containing fewer than 10,000 people employed nationwide cannot be shown separately, and must be combined with related occupational categories to create aggregates containing 10,000 or more people. 2. Detailed occupational categories, when cross-tabulated by the 12 race/ethnicity categories, cannot be shown for any geographic area containing fewer than 50,000 people. (USCB 2004)	

Ref.	Notes	Example	
B7.13 a	**Have revised outputs been assessed for potential disclosure risks?**		Accessibility
	This assures users that any threats to confidentiality in revised datasets or outputs have been assessed and addressed.	The revised datasets have been subject to the same assessment as the original dataset, which was released on 27/07/2002, and all identified potential disclosure risks have been addressed with the application of statistical disclosure control methods.	
B7.14 b	**Describe any restrictions on access to the dataset.**		
	This should inform users of any restrictions on access, and give broad details on the secure environments in place to protect data (including whether secure environments meet British Standard 7799 and, if not, a broad description of the secure environment that is in place). For statistical outputs derived wholly or in part from administrative data see B2.15	Access to the main 2001 Census results is free and use is unrestricted. This is a major change from previous censuses, and reflects wider policies on access to government information. (ONS 2005)	
B7.15 b	**Describe any restrictions on the use of the dataset.**		
	This informs users of any restrictions governing their use of a dataset, for example restrictions on releasing the data to third parties, disclaimers that must accompany outputs from the data, contractual agreements limiting use, etc. For statistical outputs derived wholly or in part from administrative data see B2.15	The supply of all Census material in electronic form will be subject to end user licences which grant unrestricted use of the material but with conditions that require: • the acknowledgement of Crown copyright and source • that the user shall not attempt to derive information about identifiable individuals nor purport to have done so • that the user is aware of third party rights which apply to some material and any need for licences (ONS 2004a)	
B7.16 b	**Notation for Information Loss Measures:** B7.17, B7.18, B7.19, B7.20, B7.21, B7.22**.**		Accuracy
	Information Loss measures are used to measure the statistical changes between an original table and a protected table allowing suppliers of data to compare disclosure control methods. Information loss measures can also be considered from a user perspective to assess quality of tabular outputs after disclosure control has been applied. The general notation for information loss measures is described in the example.	Let *T* be a collection of tables, and *m* the number of tables in the collection. Denote the number of cells in table $t \in T$ as l_t and let n_{tk} be the cell frequency for cell $k \in t$ When used, the superscript *orig* indicates the original table and the superscript *pert* indicates the protected table. When no superscript is applied the formula applies to both tables, original and protected, separately.	

Ref.	Notes	Example	
B7.17	**Distance Metrics**		Accuracy
a	Distance metrics are used to measure the statistical distance between an original table and a protected table after disclosure control methods have been applied. The metrics are calculated for each table with the final measure being the average across the tables. A variety of metrics can be used: (1) Hellingers' Distance Hellingers' distance emphasizes changes in small cells and is particularly appropriate when assessing sparse tables. (2) Absolute Average Distance AAD is based on the absolute difference between the original and protected cells. This measure is easily interpretable as the average change per cell. (3) Relative Absolute Distance RAD is based on the relative absolute difference between the original and protected cell relative to the original cell value. This measure is particularly suitable for assessing tables with a large variability in cell values.	(see information loss measures [ref. B7.16] example for general notation) (1) $HD_t = \sqrt{\sum_{k \in t} \frac{1}{2} (\sqrt{n_{tk}^{pert}} - \sqrt{n_{tk}^{orig}})^2}$ next calculate: $\overline{HD} = \frac{1}{m} \sum_{t \in T} HD_t = \frac{1}{m} \sum_{t \in T} \sqrt{\sum_{k \in t} \frac{1}{2} (\sqrt{n_{tk}^{pert}} - \sqrt{n_{tk}^{orig}})^2}$ (2) $AAD = \frac{1}{m} \sum_{t \in T} \frac{\sum_{k \in t} \left| n_{tk}^{pert} - n_{tk}^{orig} \right|}{l_t}$ (3) $RAD = \frac{1}{m} \sum_{t \in T} \sum_{k \in t} \frac{\left| n_{tk}^{pert} - n_{tk}^{orig} \right|}{n_{tk}^{orig}}$	
B7.18	**Standard Error for Distance Metric**		
a	When computing the average distance metric, the standard error is often used to take into account the dispersion of values of the metric between tables. An example is shown for the Hellingers' distance.	(see information loss measures [ref. B7.16] example for general notation) SE (for Hellingers' Distance) = $\sqrt{\frac{1}{m-1} \sum_{t \in T} (HD_t - \overline{HD})^2}$	

Ref.	Notes	Example	
B7.19	**Impact on Totals and Subtotals**		Accuracy
a	A user may be interested in changes to the grand total of a variable or changes to the subtotals by the variable categories (e.g. changes to totals of males and totals of females for the gender variable). The impact is expressed either as a relative absolute distance for a total or average absolute distance per cell.	(see information loss measures [ref. B7.16] example for general notation) Let *C(t)* be a defined collection of cells in table *t*, let $l_{C(t)}$ denote the number of cells in the collection *C(t)* and let $n_{C(t)} = \sum_{k \in C(t)} n_{tk}$ Average Absolute distance per Cell: $AAT = \frac{1}{m} \sum_{t \in T} \frac{\left| n_{C(t)}^{pert} - n_{C(t)}^{orig} \right|}{l_{C(t)}}$ Relative Absolute distance for a Total $RAT = \frac{1}{m} \sum_{t \in T} \frac{\left| n_{C(t)}^{pert} - n_{C(t)}^{orig} \right|}{n_{c(t)}^{orig}}$	
B7.20	**Impact on Measures of Association**		
a	A statistical analysis that is frequently carried out on contingency tables are tests for independence between categorical variables that span the table. The test for independence for a two-way table is based on a Pearson Chi-Squared Statistic. In simple terms, this measures the strength of the relationship between the variables in the table. Note that two-way tables can include more than one variable across the rows and columns by setting up a cross-classification of variables. The measure indicates the percent relative difference on the measure of association (Cramer's V) between the original and perturbed tables.	(see information loss measures [ref. B7.16] example for general notation) The Pearson Chi squared statistic for table *t* is: $\chi_t^2 = \sum_{k \in t} \frac{(n_{tk} - e_{tk})^2}{e_{tk}}$ Where $e_{tk} = \frac{n_{t(i\cdot)} \times n_{t(\cdot j)}}{n_t}$ is the expected count for cell *k* in row *i,* denoted $t(i\cdot)$ and column *j,* denoted $t(\cdot j)$ of table *t.* And $n_{t(i\cdot)} = \sum_{k \in t(i\cdot)} n_{tk}$, $n_{t(\cdot j)} = \sum_{k \in t(\cdot j)} n_{tk}$, $n_t = \sum_{k \in t} n_{tk}$ Cramer's V: $CV_t = \sqrt{\frac{\chi_t^2 / n}{\min(r_t - 1, c_t - 1)}}$ Where r_t = number of rows c_t = number of columns, for table *t* $RCV_t = 100 \times \frac{CV_t^{pert} - CV_t^{orig}}{CV_t^{orig}}$	

Ref.	Notes	Example	
B7.21	**Impact on Dispersion of Cell Counts**		Accuracy
a	This information loss measure compares the dispersion of the cell counts across the tables before and after the disclosure control method has been applied. Note that if the table has been protected via suppression, then the suppressed cells would have to be given imputed values in order to calculate this measure. For example suppressed cells can be given imputed values based on a row average of unsuppressed cells.	(see information loss measures [ref. B7.16] example for general notation) For each table *t*, we calculate: $D_t = \frac{1}{l_t - 1}\sum_{k \in t}(n_{tk} - \bar{n}_t)^2$ where $\bar{n}_t = \frac{\sum_{k \in t} n_{tk}}{l_t}$ Next calculate $DR_t = \frac{D_t^{pert}}{D_t^{orig}}$ To obtain a quality measure across several tables we calculate the average of these ratios: $D = \frac{1}{m}\sum_{t \in T} DR_t$	
B7.22	**Impact on Rank Correlation**		
a	One statistical tool for inference is the Spearman's Rank Correlation, which tests the direction and strength of the relationship between two variables. The statistic is based on comparing the ranking (in order of size) of the variable values. Therefore, one important assessment of the impact of disclosure control is to compare rankings of the data before and after perturbation. This information loss measure compares the ordering of tables that all have the same structure, for a given cell in that structure. Normally the tables represent different geographical areas. To calculate the measure, group the tables (for example, into deciles) according to the size of the original counts in a selected cell (for example, unemployment). Repeat this process for the perturbed tables, ordering by size in the selected cell and by the original ranking, to maintain consistency for tied values. The information loss measure is the average number of tables that have moved into a different rank group after disclosure control has been applied.	(see information loss measures [ref. B7.16] example for general notation) For selected cell k, if table *t* is in a different group after protection has been applied let $I_{tk} = 1$ otherwise let $I_{tk} = 0$ $RC_k = \frac{100 \times \sum_{t \in T} I_{tk}}{m}$	

B8. Dissemination

Ref.	Notes	Example	
B8.1 a	**Time lag from the reference date/period to the release of the provisional output. (Key Quality Measure)**		Timeliness
	This indicates whether provisional outputs are timely with respect to users' needs.	The provisional output is published six weeks after the last day of the reference period.	
B8.2 a	**Time lag from the reference date/period to the release of the final output. (Key Quality Measure)**		
	This indicates whether final outputs are timely with respect to users' needs.	The final output is published eight weeks after the last day of the reference period.	
B8.3 a	**Time lag from the end of data collection to the publication of first results.**		
	This indicates whether first results are timely with respect to users' needs. For administrative data see: B2.11.	The first set of results was published seven weeks after the last day of the data collection period.	
B8.4 a	**Time lag between scheduled release date and actual release date.**		
	This is a measure of the punctuality of the output. For statistical outputs derived wholly or in part from administrative data see B2.20	This release is one week later than scheduled.	
B8.5 a	**For customised data requests from existing sources, time lag from receipt of request for data to delivery date.**		
	This is a measure of punctuality for customised data requests from existing sources.	Ward-level tables were delivered 11 working days after receipt of the request.	
B8.6 a	**For new and ad hoc surveys, time lag from the commitment to undertake the survey to the release date.**		
	This indicates whether outputs from new or ad hoc surveys are timely with respect to users' needs.	This survey was commissioned in March 2003. The results are due to be released on 18 April 2003 with a time lag of five weeks between commitment to undertake the survey and release date.	
B8.7 b	**Frequency of production.**		
	This indicates how timely the outputs are, as the frequency of publication indicates whether the outputs are up to date with respect to users' needs.	*Economic Trends* is prepared monthly by National Statistics in collaboration with the Bank of England. (Dickman 2003)	

Ref.	Notes	Example	
B8.8 a	**Timetable for release of data.**		Accessibility
	Providing timetables of future releases enables equal access to outputs for all groups of users.	The First Release of the Retail Sales Index for October 2004 will be published on 18 November 2004.	
B8.9 b	**Describe key user needs for timeliness of data and how these needs have been addressed.**		Timeliness
	This indicates how timely the data are for specified needs, and how timeliness has been secured, e.g. by reducing the time lag to a number of days rather than months for monthly releases.	Users of the 'estimated number of applications to higher education institutes' need to be able to allocate funding for the financial year based on these estimates. Higher education institutes use these estimates in conjunction with their spending review for the previous financial year to predict future departmental budgetary needs. Therefore, these estimates are released in early March so that higher education institutes can use them in their budget planning.	
B8.10 b	**Are outputs accessible to all at the same time, according to a preannounced timetable?**		Accessibility
	Advertising release dates for outputs in advance enables equal access to outputs for all users.	The report will be published on the National Statistics website (www.statistics.gov.uk) on 15 September 2003.	
B8.11 b	**Document availability of quality information.**		
	Users should be informed of what quality information is available to help them judge the strengths and limitations of the output in relation to their needs.	The standard error associated with the computer services and related services aggregates, together with total turnover, has been estimated. At present mechanisms do not exist to measure the non-sampling error although any follow-on surveys would consider these issues. (Prestwood 2000)	

Ref.	Notes	Example	
B8.12	**Describe whether data are accompanied by commentary.**		Accessibility
b	Users should be informed whether data have accompanying commentary to help them interpret the data.	UK Economic Accounts contains two articles incorporating text, charts and tables. (ONS 2003d)	
B8.13	**Are there links to metadata?**		
b	Metadata help users to make appropriate use of the data. Links to available metadata ensure that this information is accessible. For administrative data sources see: B2.4	NDAD produces online catalogues for all material in its archive, covering all aspects of the datasets from a broad view of their administrative and historical background to detailed descriptions of each field in each dataset together with the attributes of such fields. As well as holding the data itself, NDAD also provides essential metadata which describes the relationship between data items, helps the system present data on-screen to users, and allows for export of data to modern software systems. (ONS 2001c)	
B8.14	**Highlight availability of outputs and explain search and navigation tools available within web-based outputs.**		
b	Information should be provided to the user on how to locate and navigate web-based outputs.	Users can download time series, cross-sectional data and metadata from across the Government Statistical Service (GSS) using the site search and index functions from the homepage. Many datasets can be downloaded, in whole or in part, and directory information for all GSS statistical resources can be consulted, including censuses, surveys, periodicals and enquiry services. (ONS 2003c)	
B8.15	**Reference outputs in catalogues and other relevant documents.**		
b	Information on where to locate outputs should be available wherever the output is referenced.	An article describing the development of the current system used in the United Kingdom – the UK farm classification system (Revised 1992) – was published in Appendix 3 of *Farm Incomes in the United Kingdom*: 1991/92 Edition. Further details of the system and an analysis of June census data by farm type for the United Kingdom was published in Chapter 8 of *The Digest of Agricultural Census Statistics*: United Kingdom 1993. (MAFF 1996)	

Ref.	Notes	Example	
B8.16	**Describe procedures to request access to confidentialised public-use micro datasets, including costs.**		Accessibility
b	Procedures to request access to confidentialised public-use micro datasets should be clearly described and any cost implications detailed.	Samples of Anonymised Records (SARs) can be obtained by emailing Census Customer Services on: census.customerservices@ons.gov.uk The cost of Census microdata varies according to individual requirements.	
B8.17	**Describe procedures to request access to non-published data.**		
b	Procedures to request access to unpublished data should be clearly described, and the average time and cost of providing simple tables of non-published data should be detailed.	If you would like to request non-published data, please email info@statistics.gov.uk You should receive an acknowledgement within 2 days and a reply within 10 working days. The cost of providing datasets or tables will vary according to requirements.	
B8.18	**Number of web hits.**		
a	This is an indicator of the accessibility of the statistics on the website. In addition, it may denote the relevance of the output to users' needs.	There was little change of note in the top-line statistics for www.statistics.gov.uk in July 2003. 332,000 people visited the website down from 343,000. (ONS 2003b)	
B8.19	**Detail contact points for further information, including technical information.**		
b	Users should be provided with details of contacts who are able to provide them with further information.	If you have any technical queries, please contact the Quality Centre on 01633 655605. Alternatively, you can email any technical queries to: quality.measurement@ons.gov.uk	

Ref.	Notes	Example	
B8.20	**Describe any differences between domains.**		Comparability
b	Differences between domains may be found in any relevant characteristic of the domain, or in the methods used, including: definitions; target and study populations; sampling unit; sampling methods; data collection methods; processing methods; and data analysis methods. These differences will impact on data comparability between domains.	There were differences between domains in terms of the age composition of the target population. For example, in Oxford there was a higher proportion of people in the 20–29 age range (24.9 per cent compared with an average for England and Wales of 12.6 per cent); while for Eastbourne there was a higher proportion of the population aged 75+ (13.9 per cent compared with an average for England and Wales of 7.6 per cent). The differences in age distributions mean that a direct comparison between domains of health-related variables may be invalid.	

Ref.	Notes	Example	
B8.21 a	**Estimate mean absolute revision between provisional and final statistics.** **(Key Quality Measure)**		Accuracy
	This provides an indication of reliability (which is one aspect of accuracy) in that provisional statistics may be based on less information than is available for final statistics, and may have been processed more quickly. However, the value of provisional statistics lies in their timely provision. Statistics that are not revised are not necessarily more accurate than those that are revised often: it may be simply due to different revisions policies. The number of estimates from which the mean absolute revision is calculated, and the gap between the 'preliminary' and 'later' estimates should be predefined and should be unchanged between publications, in order to be open and clear. For quarterly outputs, this indicator should be based upon at least 20 previous observations for the results to be sufficiently reliable. Where data in a series are comparable over more than five years, this indicator can be based upon more than 20 observations. For monthly estimates, this indicator should be based upon at least thirty previous observations. This quality indicator is not recommended for annual data, due to the uncertainty introduced into the results for this indicator from comparing data over such a long period. The mean absolute revision should be expressed in the same units as the estimate (e.g. percentage points, £m etc). Any known reasons for revisions between provisional and final estimates should be described.	$\frac{1}{n}\sum_{i=1}^{n}\lvert l_i - p_i \rvert$, for $n \geq 20$ for quarterly data, $n \geq 30$ for monthly data. Here $l \in L$ is the set of later estimates; $p \in P$ is the set of preliminary estimates.	

Ref.	Notes	Example	
B8.22	**Estimate the mean revision between provisional and final estimates.**		Accuracy
a	This is an indicator of reliability. It shows whether revisions are being made consistently in one direction - i.e. if early estimates are consistently under or over estimating the later, more information-based figures. The number of estimates from which the mean revision is calculated, and the gap between the 'preliminary' and 'later' estimates should be predefined and should be unchanged between publications, in order to be open and clear. For quarterly outputs, this indicator should be based upon at least 20 previous observations for the results to be sufficiently reliable. Where data in a series are comparable over more than five years, this indicator can be based upon more than 20 observations. For monthly estimates, this indicator should be based upon at least thirty previous observations. This quality indicator is not recommended for annual data, due to the uncertainty introduced into the results for this indicator from comparing data over such a long period. A t-test can be used to assess whether the mean revision is statistically different from zero. An adjusted t-test may need to be used if revisions for different reference periods are correlated. The mean revision should be expressed in the same units as the estimate (e.g. percentage points, £m etc). Any known reasons for revisions between provisional and final estimates should be described, as well as whether the mean revision is significantly different from zero.	$\frac{1}{n}\sum_{i=1}^{n}(l_i - p_i)$, for $n \geq 20$ for quarterly data, $n \geq 30$ for monthly data. Here $l \in L$ is the set of later estimates; $p \in P$ is the set of preliminary estimates.	

Ref.	Notes	Example	
B8.23 a	**Flag data that are subject to revision and data that have already been revised.**		Accuracy
	This indicator alerts users to data that may be, or have already been, revised. This will enable users to assess whether provisional data will be fit for their purposes.	Some figures are provisional and may be revised later, this applies particularly to many of the detailed figures for 2001 and 2002. (ONS 2003c)	
B8.24 a	**Provide timetable for revisions.**		Accessibility
	Information on when revisions will be available enables users to judge whether the release of revisions will be timely enough for their needs and whether revisions are delivered punctually.	The revised migration figures will be made available on the ONS web site (www.statistics.gov.uk) in late summer 2003. (Pereira 2003)	
B8.25 a	**For ad hoc revisions, detail revisions made and provide reasons.**		Accuracy
	Where revisions occur on an ad hoc basis, this may be because earlier estimates have been found to be inaccurate. Users should be clearly informed of the revisions made and why they occurred. Clarifying the reasons for revisions guards against any misinterpretation of why revisions have occurred, while at the same time making processes (including any errors that may have occurred) more transparent to users.	An error in the Labour Market Historical Supplement Table 9 and Table 30b Quarterly Supplement has been corrected. The error pertains to non-seasonally adjusted women 18–24 unemployed over 6 and up to 12 months from September–November 02 onwards. (ONS 2003e)	
B8.26 b	**Reference/link to detailed revisions analyses.**		Accessibility
	Users should be directed to where detailed revisions analyses are available.	The article called ‘Revisions to Quarterly GDP growth’ contains detailed revisions analyses. This is available at: www.statistics.gov.uk	

Ref.	Notes	Example	
B8.27 b	**Have different outputs on the same theme been produced in ways that are coherent?**		Coherence
	This statement advises users on whether outputs on the same theme have been produced in comparable ways, or if there are differences in their production, e.g. in definitions, classifications, sampling methods, data collection methods, processing or analysis.	The information used to compile this report was obtained from many agencies, including government departments, local government and private research organisations. These sources used different methods to collect their data. For example, there are differences between sources in their sampling and data collection methods. Therefore, care should be taken when comparing data from the different sources as they may not be strictly comparable with each other. Documentation of methods and procedures for each source are provided so that users are able to investigate any limitations on comparability between sources and any consequent implications for their own analyses.	
B8.28 a	**Compare estimates with other estimates on the same theme.** **(Key Quality Measure)**		
	This statement advises users whether estimates from other sources on the same theme are coherent (i.e. they 'tell the same story'), even where they are produced in different ways. Any known reasons for lack of coherence should be given.	Estimates of average annual earnings were compared with those from the LFS, which uses the same harmonised questions on income. However, it was discovered that the LFS average earnings were consistently higher than those from this survey. This disparity may have arisen due to undercoverage of higher earners in our sample.	

Ref.	Notes	Example	
B8.29 b	**Identify known gaps between key user needs, in terms of coverage and detail, and current data. (Key Quality Measure)**		Relevance
	Data are complete when they meet user needs in terms of coverage and detail. This indicator allows users to assess, when there are gaps, how relevant the outputs are to their needs.	Some known gaps between the key user needs and the data have been identified, including: • A demand for a question on Income. Testing concluded that for a significant minority of the population it would have been unacceptable to include this question on a Census form that was compulsory. • A campaign mounted in Wales calling for a Welsh tick box as part of the Ethnicity question. ONS introduced a package of measures to improve information about Welsh identity and to promote the option to use the 'write in' option in response to this campaign. Some 3,700 forms were returned with a sticker over the Ethnic group question, showing Welsh as an extra tick-box. (ONS 2005)	

Ref.	Notes	Example	
B8.30	**Present any reasons for lack of relevance.**		Relevance
b	This indicator gives the main reasons why an output is not relevant to user needs. For example, the study population may not be the same as the user's target population, or the operational definition of a concept may not capture the user's required definition of a concept.	The output may not satisfy all users' requirements for ward-level data broken down into sub-groups based on ethnicity. This is because, in many wards, cell numbers were small and potentially disclosive when using the Level 2 Ethnic Grouping (NS interim standard classification of ethnic groups). Therefore, the broader Level I Ethnic Grouping classification was used for ward-level data. In addition, where cell counts were small and potentially disclosive, even when using this broader ethnic classification, data values were suppressed using primary and secondary suppression.	
B8.31	**Describe any standardisation applied to outputs.**		
b	The relevance of an output is increased when adjustments to the data are made, e.g. to weight sub-groups to their population totals, or to separate seasonal effects from underlying trends. While improving the usefulness of data, these adjustments must be described for transparency and to aid users' understanding of potential data uses and limitations.	To compare levels of mortality between different areas it is helpful to remove the effects of varying population structures among these areas. This can be done by using a standard set of sex and age-specific mortality rates and applying these to the population of each area. This will produce an expected number of deaths – that is, the number expected if the standard rates had applied in the area. This expected number is then compared with the number of deaths actually observed in the area. (Pereira 2003)	

References

Armitage P and Brown D (2003) *Neighbourhood Statistics in England and Wales: Disclosure Control Problems and Solutions*. Paper presented at the 2003 UN-ECE/Eurostat work session on statistical data confidentiality in Luxembourg. Available at: http://www.unece.org/stats/documents/2003/04/confidentiality/wp.13.e.pdf [Accessed May 2005]

Barnes M, Haskel J and Ross A (2001) *Understanding productivity: New insights from the ONS business data bank*. Paper prepared for Royal Economic Society Conference, Durham 2001. Available at: http://www.statistics.gov.uk/events/REC01/downloads/Understanding_Productivity.pdf [Accessed May 2005]

Beaumont J F, Haziza D, Mitchell C and Rancourt E (2004) 'New Tools at Statistics Canada to Measure and Evaluate the Impact of Non-response and Imputation'. Available at: http://www.fcsm.gov/03papers/BeaumontHazizaRancourt.pdf [Accessed May 2005]

De Waal A G, Hundepool A J and Willenborg, L C R J (1996) *Argus: Software for statistical disclosure control of microdata*. Paper presented at the 1996 US Census Bureau Annual Research Conference. Available at: http://www.census.gov/prod/2/gen/96arc/iiawille.pdf [Accessed May 2005]

Dickman P (ed) (2003) *Economic Trends* September No. 598. TSO: London.

Goddard E (1990) *Why Children Start Smoking*. Enquiry carried out by Social Survey Division of OPCS on behalf of the Department of Health. HMSO: London.

Goldblatt P and Chappell R (eds) (2003) *Population Trends* summer No. 112. TSO: London.

Hawala S (2003) *Microdata Disclosure Protection Research and Experiences at the US Census Bureau*. Paper presented at the Workshop on Microdata, Stockholm 2003. Available at: http://www.census.gov/srd/sdc/microdataprotection.pdf [Accessed May 2005]

Hlebec V, Kogovsek, T and Polic, M (2002) 'Effects of Data Collection Mode on Measuring Personality Traits'. Denmark: ICS. Available at: http://www.icis.dk/ICIS_papers/E1_4_2.pdf [Accessed May 2005]

Kokic P (1996) 'Measuring the sampling error of the Index of Production'. Internal report (ONS).

Lynn P, Beerten R, Laiho J, Martin J (2001) 'Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys', ISER Working Papers; October 2001-23. Colchester: Institute for Social & Economic Research (ISER); University of Essex.

MAFF (Ministry of Agriculture, Fisheries and Food, Scottish Office) (1996) *Farm Incomes in the United Kingdom 1994/95*. London: HMSO.

ONS (Office for National Statistics) (1997). *Labour Force Survey User Guide Vol 1: LFS Background and Methodology*. London: ONS.

ONS (Office for National Statistics) (1999) *Labour Force Survey User Guide Vol 9, Eurostat and Eurostat Derived Variables*. London: TSO.

ONS (Office for National Statistics) (2001a) *Review of the Inter-Departmental Business Register*, National Statistics Quality Review Series Report No. 2. London: HMSO.

ONS (Office for National Statistics) (2001b) *Mortality Statistics General, 1999, England and Wales*, DH1 No. 32. London: TSO.

ONS (Office for National Statistics) (2001c) *The UK National Digital Archive of Datasets*. London: ONS. Available at: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=2188&More=Y [Accessed May 2005]

ONS (Office for National Statistics) (2001d) *Working Family Tax Credit Metadata Template 2001*. Internal document.

ONS (Office for National Statistics) (2002a) *Census 2001 Data Collection Development Evaluation Report*. London: ONS. Available at: http://www.statistics.gov.uk/census2001/dcdevrep.asp [Accessed May 2005]

ONS (Office for National Statistics) (2002b) *How does the Office for National Statistics keep personal information confidential?* Census background. Available at: http://www.statistics.gov.uk/census2001/cb_4.asp [Accessed May 2005]

ONS (Office for National Statistics) (2002c) Matching Error, One Number Census Steering Committee. Available at: http://www.statistics.gov.uk/census2001/pdfs/sc0202.pdf [Accessed April 2006]

ONS (Office for National Statistics) (2002d) Marriage, Divorce and Adoption Statistics, Series FM2, No. 30. ONS:HMSO. Available at: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=581&Pos=1&ColRank=1&Rank=272 [Accessed April 2006]

ONS (Office for National Statistics) (2003a). Annual Business Inquiry background information. Available at: http://www.statistics.gov.uk/abi/background_info.asp [Accessed May 2005]

ONS (Office for National Statistics) (2003b) *National Statistics Online – Analysis of Online Usage July 2003.* Internal document.

ONS (Office for National Statistics) (2003c) *United Kingdom National Accounts – The Blue Book 2003*. London: TSO.

ONS (Office for National Statistics) (2003d) *United Kingdom Economic Accounts Quarter 2 2003*. London: TSO.

ONS (Office for National Statistics) (2003e) Labour Market Historical and Quarterly Supplements. Available at: http://www.statistics.gov.uk/notices/LMS_FR_HS_12Sep03.asp [Accessed May 2005]

ONS (Office for National Statistics) (2003f) *Research and Development in UK businesses*. Available at: http://www.statistics.gov.uk/downloads/theme_commerce/MA14_2002.pdf [Accessed May 2005]

ONS (Office for National Statistics) (2003g) *Retail Sales September 2003* October. First release. Available at: http://www.statistics.gov.uk/pdfdir/rs1003.pdf [Accessed May 2005]

ONS (Office for National Statistics) (2003h) *Stockbuilding – Business Monitor SQ1 Data for Quarter 1 200*3. London: HMSO.

ONS (Office for National Statistics) (2003j) *The United Kingdom 2000: Time Use Survey Technical Report 2003.* London: HMSO.

ONS (Office for National Statistics) (2004a) *Census 2001 Output Prospectus*. Available at: http://www.statistics.gov.uk/census2001/pdfs/census_prospectus.pdf [Accessed May 2005]

ONS (Office for National Statistics) (2004b) PRODCOM methodological note. Available at: http://www.statistics.gov.uk/downloads/theme_commerce/PRODCOM_Methodological_note.pdf [Accessed May 2005]

ONS (Office for National Statistics) (2005) *Census Quality Report (part A)*. Quantitative information. Basingstoke: Palgrave Macmillan.

OPCS (Office for Population, Censuses and Surveys) (1986) *Family Expenditure Survey 1986 (Revised)*. London: HMSO.

OPCS (Office for Population, Censuses and Surveys) (1993) *Introductory Notes on the Annual Census of Production*, *1993*. London: HMSO.

OPCS (Office for Population, Censuses and Surveys) (1996) *Family Resources Survey, Great Britain 1994-95*. London: TSO.

Pereira R (ed) (2003) *Key Population and Vital Statistics: Local and Health Authority Areas, 2001.* Series VS No. 28. London: TSO.

Prestwood D (2000) *Computer Services Survey (SERVCOM Feasibility Study) – Data for 2000*. Available at: http://www.statistics.gov.uk/downloads/theme_commerce/SERVCOM2000.pdf [Accessed May 2005]

Risdon A (ed) (2003) *The Labour Force Survey*. Available at: http://www.statistics.gov.uk/downloads/theme_labour/REV_LFSQS_0403.pdf [Accessed May 2005]

Statistics Finland (2004) Administrative Data in Business Register. Available at: http://www.insee.fr/en/nom_def_met/colloques/ussa/pdf/laukkanen_doc_en.pdf [Accessed April 2006]

Tsapogas J (1996) *Characteristics of Recent Science and Engineering Graduates: 1993*. National Science Foundation paper. Available at: http://www.nsf.gov/sbe/srs/nsf96309/secta.htm [Accessed May 2005]

USCB (US Census Bureau) (2004) *Census 2000 Special EEO Tabulation*. Available at: http://www.census.gov/hhes/www/eeoindex/privacy.pdf [Accessed May 2005]

USCMB (US Census Monitoring Board) (2001) *A Guide to Statistical Adjustment: How It Really Works*. Report to Congress, May 23. Available at: http://govinfo.library.unt.edu/cmb/cmbc/cmbrept.pdf [Accessed May 2005]