Office for National Statistics

Communicating uncertainty and change: Guidance for official statistics producers

Glenn Everett
Director,
Measuring National Well-being programme
ONS, UK

Background

- Uncertainty is completely normal in statistics, but for many of the users of official statistics, the word “uncertainty” also means “unreliable”.

- As a result, there is understandable concern that making information about uncertainty easily accessible might reduce users’ confidence in the numbers that we publish.

- But this shouldn’t be the case.

Background (2)

- There is no consistent approach across the GSS in how uncertainty and change are communicated.
- The clear communication of uncertainty and change (and the strengths and limitations in the data in general) are key to the appropriate interpretation and use of statistics.
- The National Statistician asked for this work to be given priority.

Approach

- A GSS Task and Finish Group for Communicating Change and Uncertainty was formed to draft GSS guidance.
- Membership included several departments, led by ONS, supported by Good Practice team.
- A limited review of some GSS and other NSIs' publications in relation to communicating change and uncertainty was undertaken.
- Results found little or no evidence of wide spread good practice in communicating uncertainty and change in the GSS - although there are pockets of good practice in some publications.

Challenges

- Get the balance between ensuring confidence in the estimates while describing elements of uncertainty in the data
- Ensure that an element of the analyst's judgement could be used - the balance between 'must', 'should' and 'could'.
- Not to be overly prescriptive.
- Note the draft policy relates to communication, not the measurement of uncertainty and change.

Sources of uncertainty

- Sampling error:
  - Sample size
  - Bias

- Non-sampling error:
  - including administrative data (e.g. definitions, coverage, timeliness, effects of imputation or editing, or the impact of manipulating the data)

- Uncertainty in data from mixed sources:
  - (e.g. rates calculated from both administrative data and population estimates)

Guidance

- The guidance covers two topics - uncertainty and change. Both are fundamental for official statistics, and are closely related.

- Uncertainty describes the sum total of possible errors that affect the accuracy of a statistic. It includes the impact of sampling error and all other sources of error that exist in a data source.

- Change is the difference between measures of the same phenomenon at two different time points.

Principles

Overall, any publication should provide enough information to:

- allow users to judge whether the estimates are **fit for their purpose**;
- maintain and build users' **confidence in the estimates**.

The publication should provide enough information on uncertainty to indicate:

- the **quality** of the data;
- the **uncertainty** in the data.

The publication should provide enough information on change to:

- indicate the uncertainty in the **estimate** of change;
- show the **direction** and relative **size** of any change;
- provide a **longer term** view of change (e.g. trend).

How to apply the principles

- The guidance is split into things producers **should** do and things that they **could** do.
- It can be applied to statistics from all sources
- The guidance contains a range of examples which demonstrate how the principles have been applied in practice.

The Guide

Aim:

- Provide analysts who publish official statistics information and guidance for a common GSS approach to the clear communication of uncertainty and change in publishing estimates from surveys, administrative sources and multiple sources.
- Includes examples of good practice and case studies
- Plus standard wording to be used as appropriate.

Guidance developed is generic, relevant to all sources.

The Guide

- Divided into sections:
    - Early on in the publication
    - In the narrative sections
    - In tables
    - In background or quality notes

- Allow users to get the essential information up front, with further information for the more expert user.

- Includes examples of good practice taken from across the GSS.

The Guide – early on in the publication

Should

- Use the word “estimates”, where appropriate.
- Include a high-level, Plain English description of likely sources of uncertainty.

Could

- Explain if change is statistically significant, with a Plain English description.

The Guide – in the narrative

Should

- Clearly indicate any changes in definitions, methods, etc .... and how they impact on use of the statistics.

Could

- Provide information on change across a range of time periods with appropriate supporting commentary.

The Guide – in tables

Should

- Publish quantitative measures of uncertainty, where available.

Could

- Use colour coding or asterisks to highlight relative levels of uncertainty in the data.

Reference tables – communicating change and uncertainty

1. Communicating uncertainty and change in reference tables

The Welsh Government's National Survey for Wales uses colour coding in reference tables to visually convey the relative levels of uncertainty in the statistics.

The results presented in the Annex have been colour coded according to the CV for each result as follows.

	Estimate is precise	0 ≤ CV < 5
	Estimate is reasonably precise	5 ≤ CV < 10
	Estimate is considered acceptable	10 ≤ CV < 20
	Estimate is not reliable	CV ≥ 20
-	Value is suppressed due to small cell size	(fewer than 30 responses)

Colour coding based on coefficient of variation

As with any survey, the National Survey is also subject to a range of other sources of error: for example, due to non-response; because respondents may not interpret the questions as intended or answer accurately; and because errors may be introduced as the survey data is processed. These kinds of error are known as non-sampling error, and are discussed further in the Quality Report for the survey.

Table 13: Parents' satisfaction with their child's school (a)

	Primary School			Secondary School		
	%	lower CI	upper CI	%	lower CI	upper CI
Very satisfied	**68**	*65*	*70*	**49**	*46*	*52*
Fairly satisfied	**24**	*22*	*26*	**36**	*33*	*39*
Neither satisfied nor dissatisfied	**3**	*2*	*4*	**6**	*4*	*8*
Fairly dissatisfied	**4**	*3*	*5*	**6**	*4*	*7*
Very dissatisfied	**1**	*1*	*2*	**3**	*2*	*5*

Confidence intervals provided

Source: National Survey for Wales, April 2013 - March 2014

Sample size: Primary school 1,900, Secondary School 1,400

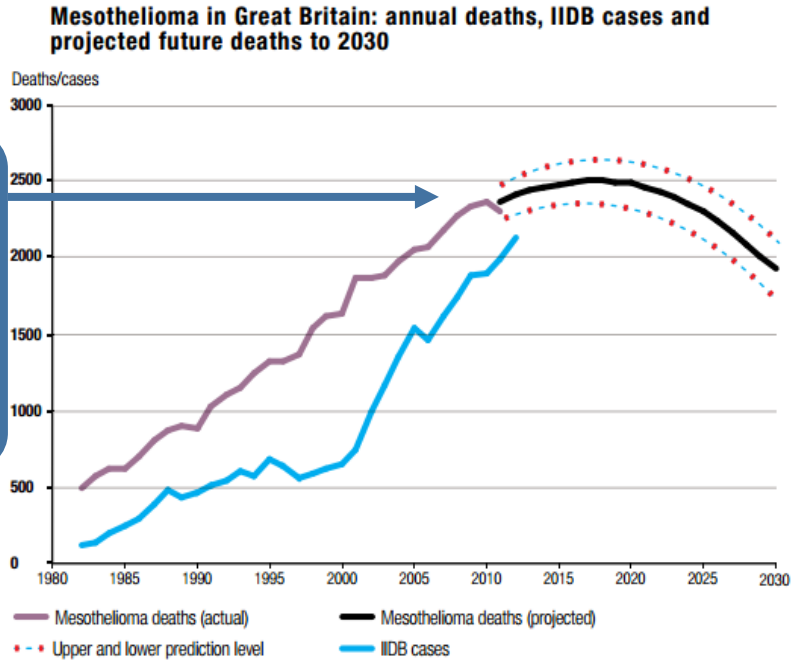
Sample size for these statistics is given

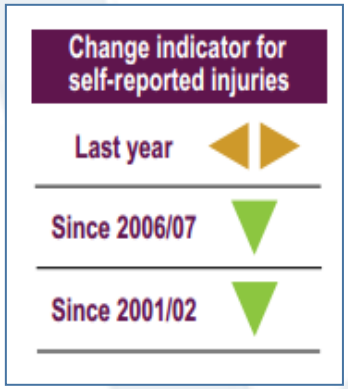
(a) This question was only asked of parents of children who attended a local authority funded primary or secondary school

Visualising change and uncertainty

The Health and Safety Executive's annual statistics report uses a range of techniques for visualising change and uncertainty.

The upper and lower bounds of the prediction are illustrated on the chart





Colour coded arrows illustrate change over a range of time periods

The Guide – in background notes

Should

- Descriptions of data collection and QA processes.
- The place for further detailed descriptions.

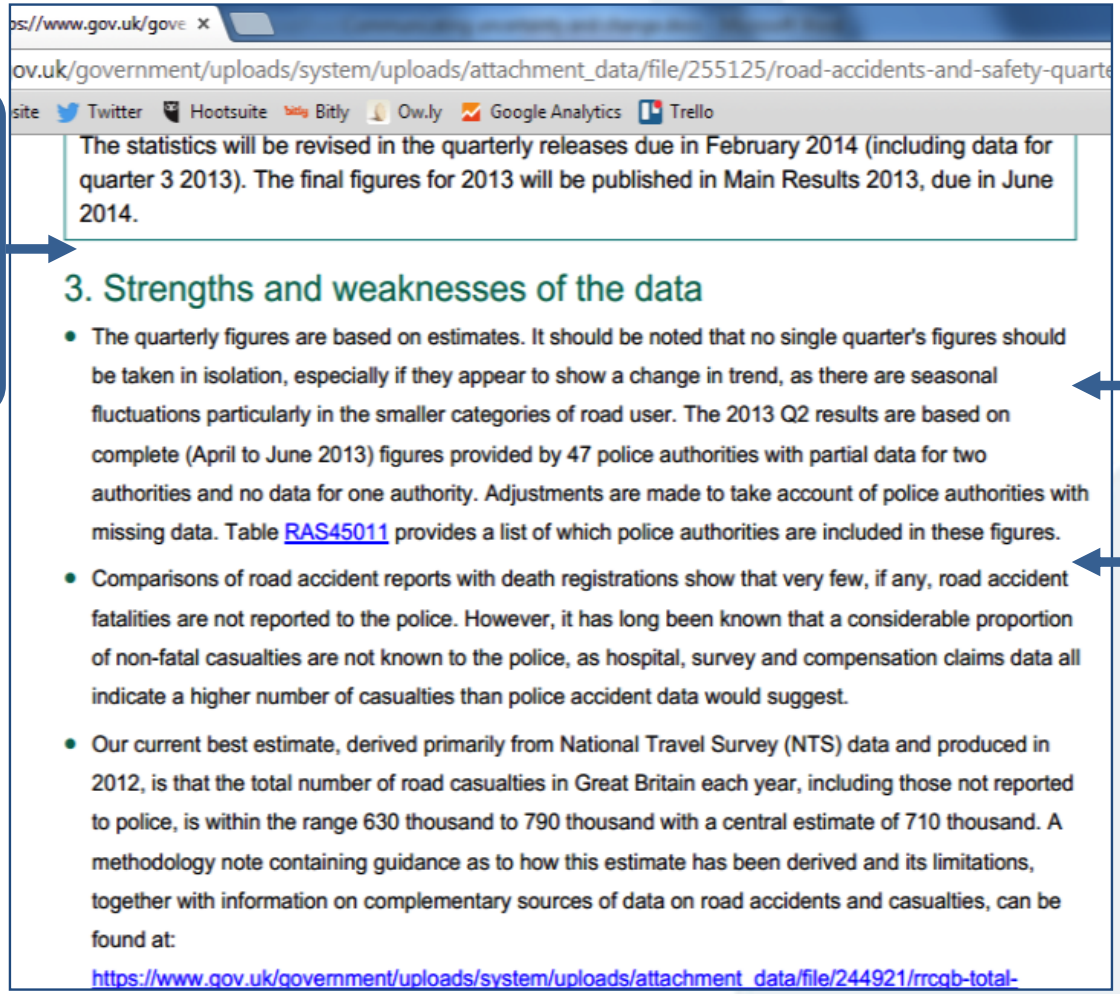
Could

- Include comparisons with other sources
- Consider data visualisations

Background notes – adjustments & other sources

Statistics on road casualties (which are based on administrative data) are clearly presented as estimates

Provides advice on interpreting the statistics and the adjustments that have been made to the statistics



ps://www.gov.uk/gove ×

ov.uk/government/uploads/system/uploads/attachment_data/file/255125/road-accidents-and-safety-quarte

site Twitter Hootsuite Bitly Ow.ly Google Analytics Trello

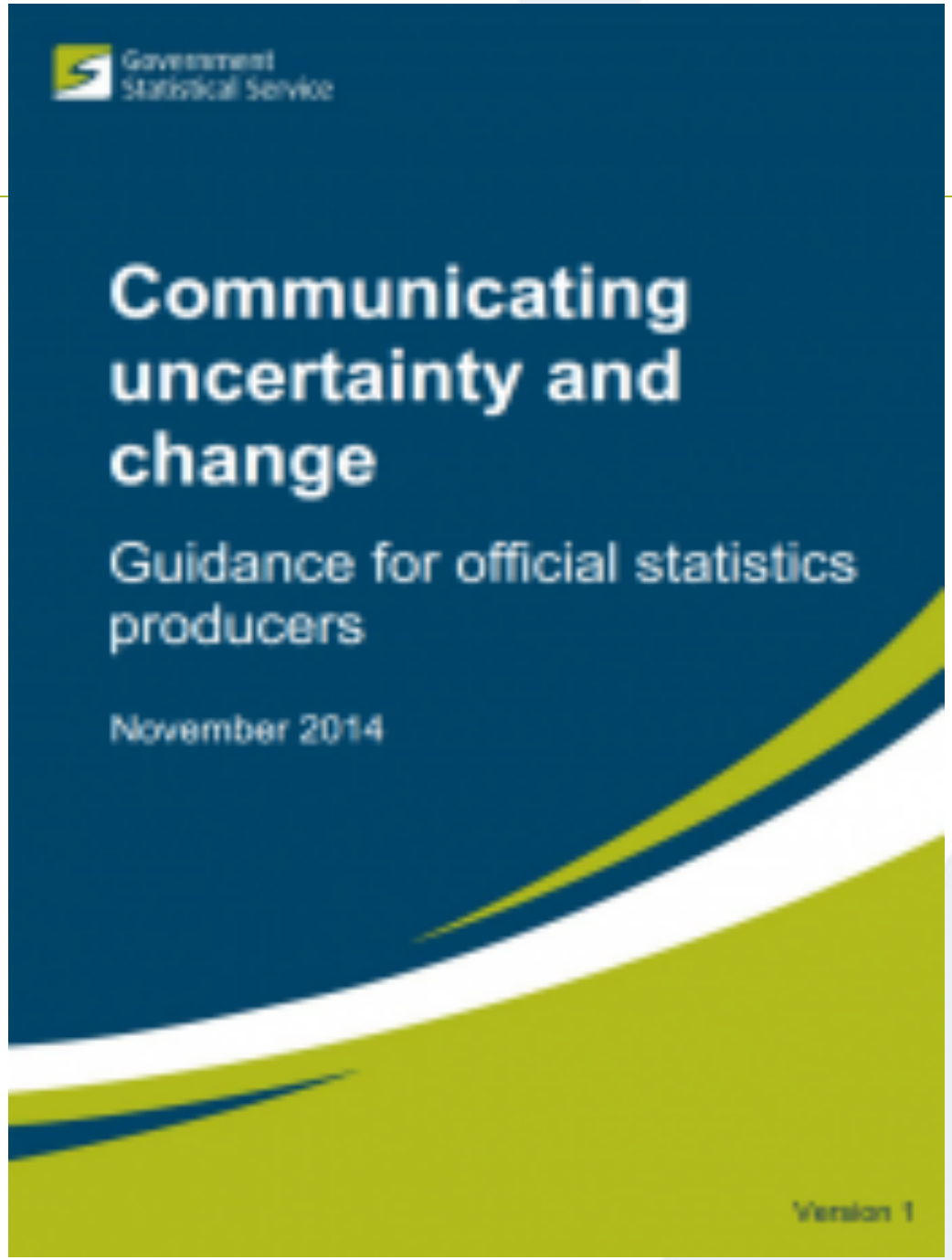
The statistics will be revised in the quarterly releases due in February 2014 (including data for quarter 3 2013). The final figures for 2013 will be published in Main Results 2013, due in June 2014.

3. Strengths and weaknesses of the data

- The quarterly figures are based on estimates. It should be noted that no single quarter's figures should be taken in isolation, especially if they appear to show a change in trend, as there are seasonal fluctuations particularly in the smaller categories of road user. The 2013 Q2 results are based on complete (April to June 2013) figures provided by 47 police authorities with partial data for two authorities and no data for one authority. Adjustments are made to take account of police authorities with missing data. Table RAS45011 provides a list of which police authorities are included in these figures.
- Comparisons of road accident reports with death registrations show that very few, if any, road accident fatalities are not reported to the police. However, it has long been known that a considerable proportion of non-fatal casualties are not known to the police, as hospital, survey and compensation claims data all indicate a higher number of casualties than police accident data would suggest.
- Our current best estimate, derived primarily from National Travel Survey (NTS) data and produced in 2012, is that the total number of road casualties in Great Britain each year, including those not reported to police, is within the range 630 thousand to 790 thousand with a central estimate of 710 thousand. A methodology note containing guidance as to how this estimate has been derived and its limitations, together with information on complementary sources of data on road accidents and casualties, can be found at:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/244921/rrcgb-total-

Provides information on coverage, including comparisons with other sources


Government Statistical Service
Communicating uncertainty and change
Guidance for official statistics producers
November 2014
Version 1

Next steps

- The Guide will evolve as we learn more about what works best in communicating uncertainty and change.

- There are areas that we want to address in future versions, including:
  - Providing a set of standard definitions for common terms, such as statistical significance
  - Communicating uncertainty in rankings and league tables
  - Communicating uncertainty using methods like confidence intervals and coefficients of variation.

Questions

?