

# Statistical critique of 203 Home Office press releases from February to September 2008+2009

*Surveys, Design and Statistics Subcommittee of  
Home Office's Scientific Advisory Committee.*

## Background

In June 2008, the Tackling Knives Action Plan (TKAP) was launched to run until March 2009. A News Release and Fact Sheet were published on 11 December to coincide with a TKAP publicity event. This Fact Sheet prematurely published National Statistics data from NHS Information Centre, Leeds against the advice of the Department of Health (DH) and NHS Information Centre's statisticians as the data were not quality assured for publication and were released 3 months earlier than DH intended. Following the intervention of the National Statistician and the UK Statistics Authority, these breaches received major news coverage and new guidance was issued from the Cabinet Office on how statistics should be handled in future.

On 6 January 2009, the UK Statistics Authority launched their Code of Practice with a Monitoring and Assessment Note on the Home Office's 11 December 2008 TKAP Press Release and Fact Sheet. The UK Statistics Authority's assessment according to the Code of Practice for Official Statistics highlighted five areas of concern that the Authority had about the Home Office's TKAP Fact Sheet – specifically:

- Unclear description
- Selective, or otherwise inappropriate, comparisons
- Lack of contextual information
- Inappropriate conclusions being drawn
- Unsubstantiated claims.

Subsequently, on 26 January 2009, the UK Statistics Authority published a further Monitoring and Assessment Note on "Presentation of Statistics in First Releases and Elsewhere". This set out six criteria, formerly used by the Statistics Commission, to be adopted by the Statistics Authority to evaluate the quality of statistical First Releases, namely:

1. clear identification of the statistics that are being released.
2. commentary that is helpful and presents the key messages.
3. commentary that is objective, balanced and professionally sound.
4. commentary that discusses the statistics in the context of their likely uses.
5. readily available metadata about the statistics in the release.
6. statistical presentation that is professional and helpful

*Question posed:* At its meeting on 20 January 2009, the Surveys, Design and Statistics Subcommittee (SDSSC) of the Home Office's independent Scientific Advisory

Committee posed the question: how exceptional were breaches of the type manifested in the 11 December 2008 press release and associated ‘fact sheet’?

If typical, rather than exceptional, it would: a) be less surprising that warnings about using unpublished National Statistics, as above, went unheeded; and b) necessitate an evidence-base on statistical lacunae in press releases so that the most common, or most serious, could be: i) readily identified, and ii) remedied by the Home Office.

Such recognition should be a professional skill across the civil service, not just within the Home Office, and a competence also for parliamentarians, who include Ministers, and for journalists. Remedy, on the other hand, may require professional statistical or other technical input to the drafting of press releases and ‘fact sheets’; or that reference be made back to the research team whose work is being précised.

SDSSC agreed to critique the statistical content of at least 160 press releases by the Home Office for the months of February to September of 2008+2009.

In the interim, by December 2009, the Home Office had introduced guidance (entitled *Standards for the Use of Evidence*) to assist its non-specialists in assessing both the quality of evidence and reputational risk so that an evidence-user might know when s/he should seek specialist advice from staff in the Home Office’s Science and Research Group. The Home Office’s Scientific Advisory Committee endorsed this initiative on *Standards for the Use of Evidence*.

### **Methodology for SDSSC study**

There is an archive of Home Office press releases, see <http://press.homeoffice.gov.uk/>. Only press releases listed on this archive were critiqued, which thereby excluded from SDSSC’s main study statistical news releases if only cited elsewhere, such as at <http://www.homeoffice.gov.uk/rds/stats-release.html>.

Press releases are always dated. Even had dates been obscured, the press releases remained available on the Home Office’s archive, and it was therefore impossible effectively to conceal from SDSSC review-teams the calendar year to which a press release belonged. Dates were not obscured.

*Pilot investigation:* Twenty-six press releases from January 2008+2009 constituted a pilot investigation by SDSSC’s chair with the aims of:

- i) designing an SDSSC survey form onto which the noted statistical lacunae per press release could be summarily encoded, see **Results**;
- ii) estimating what proportion of press releases have statistical lacunae (be they overt or covert) – *the majority, 23/26*;
- iii) documenting from each press release *up to three* “look and learn to recognise” examples; and
- iv) estimating reviewer time required per press release to read and abstract – *allow 35 minutes per press release per independent assessor*.

*In the pilot investigation, 116 lacunae were recorded across 26 press releases. Median and mode were both two, but mean and standard deviation were 4.5, as four press releases accounted for 13, 13, 14, and 16 lacunae respectively. Two of these high-scoring releases were Statistical News Releases. Only three out of the 26 press releases had NIL lacunae. Top-scoring statistical lacunae in the pilot investigation (5 or more mentions) were:*

- [ 3.] **%, absent se or denominator** (13 mentions)
- [24.] **Claimed trend, absent evidence** (12 mentions)
- [ 2.] **% change in counts, absent ‘N1 to N2’** (11 mentions)
- [23.] **Proportionality at issue** (11 mentions)
- [10.] **Missing methodology: survey details** (8 mentions)
- [ 1.] **Count now, absent background or reference count** (6 mentions)
- [ 5.] **Claimed ‘success’ without any quantification** (6 mentions)
- [ 8.] **Missing methodology: re pilot/roll-out** (5 mentions).

*Percentages were frequently quoted without either se or denominator. Trends were described without actual outcomes (counts etc) being documented other than in the main statistical tables to which no page number or table reference was provided. Pilot studies were referenced without any description of their design. Surveys were referenced without adequate description of sampling design, response rate and number of respondents. Statisticians should not fail to label fluctuations that are consistent with random variation as just that. Proportionality was an issue when sums based on available data, such as to translate them to a ‘per day per wand’ or ‘per area-month’ basis, put the data in a revealingly different light . . .*

*SDSSC allocation-design:* Previous experience with statistical critique of published papers in medical or other journals confirms that the abstraction from each text (in this case – Home Office press release) ideally should be done independently by two readers with the text then re-examined by the pair to resolve inconsistencies in their separate summaries.

Unlike the pilot investigation, SDSSC’s main study did not include Statistical News Releases unless they were listed by applying the date-range ‘search’ facility within <http://press.homeoffice.gov.uk/>.

The SDSSC worked in pairs, to which independent members were randomized as follows: Sue Brooker with Brian Francis; Sheila Bird with Mike Hough/Charles Lound (CL joined SDSSC in June 2009). David Blunt, as statistical head of profession, had declared a potential conflict of interest as he might be responsible for signing-off some press releases. Cross-pair calibration was originally planned but would have been too time-consuming and also unnecessary since months (February to September) were anyway randomized to reviewer-pair. A key objective of the SDSSC study was between-year comparison (2008 versus 2009) and monthly releases were critiqued by the same reviewers for 2008+2009.

The random assignment of reviewer-pair to month (in pairs) was as follows. All releases were scored by either SMB or SB with their pair-mate’s scores incorporated for half the months (as shaded):

Month	February	March	April	May	June	July	August	September
Pair	SMB/MH	SB/BF	SMB/MH	SB/BF	SMB/CL	SB/BF	SB/BF	SMB/CL

Serious bias could arise if all press releases for 2008, say, were reviewed in advance of those for the corresponding months in 2009. To limit the potential for such bias, the SDSSC team to review the press releases in February versus March, April versus May, June versus July, August versus September, or October versus November was not only selected at random but was also informed of its task only in second month of the pair. Exceptions were March 2008+2009 and September 2008+2009 because SDSSC required both pairs to have undertaken assessments prior to its April and October meetings.

In practice, pressure of other commitments curtailed the full extent of review. **SMB** and **SB** critiqued all press releases, as assigned. After the first pair of months (February and March) had been scored, MH met with **SMB** and BF with **SB** to discuss how MH and BF, independently, had scored the press releases for February and March of 2008+2009 so that the subsequent assignments by **SMB/SB** might be moderated accordingly. See **Appendix** for results of the **SB/BF** consensus meeting. Independent assignments by BF for the months of March and July 2008+2009 and by CL for the months of June and September 2008+2009 are presented in **Results**. They confirm broad agreement on the press releases which had lacunae, and on the number of lacunae, but less so on assigned classifications.

SDSSC's study progressed in monthly-pairs, from February-March to October-November, until the review-total exceeded 160 press releases for 2008+2009. The month of January was excluded because it had been used for SDSSC's pilot investigation; and December because SDSSC's statistical critique was initiated in response to a much-criticised press release and 'fact sheet' in December 2008.

In principle, the Home Office was forewarned that calendar year comparison between 2008 and 2009 would be made; and of the 29 pre-specified statistical lacunae on which SDSSC would score press releases. In practice, however, neither SDSSC nor the Science Secretariat took special measures to alert the Home Office's press officers. Any notable improvements in the statistical content of press releases for 2009 versus 2008 might therefore be more in reaction to 'knifed by press release' than to specific forewarning.

*Statistical power re calendar year changes:* SDSSC's study was just powerful enough to detect modest differences; and adequately powered to detect a moderate improvement in 2009. For example, if statistical lacunae averaged 4 per press release, as in the January pilot investigation, then – even allowing for over-dispersion relative to Poisson counts (such as: variance of **16** or 8 rather than 4; **9** or 6 rather than 3) - the main study had at least 50% power to detect a modest 25% reduction in statistical lacunae between calendar years: if based on 200 press-releases, and for variances of **16** and **9**. By the same token, based on 160 press-releases, SDSSC's study had 80% power to detect a salutary 33% reduction in statistical lacunae between calendar years such as from mean of 3 (and variance of 6) to 2 (with variance of 4). Lower variances than in the pilot study were possible if Statistical News Releases, some with high numbers of lacunae in the pilot investigation, were listed elsewhere than at <http://press.homeoffice.gov.uk/>.

## Findings

**Table 1** shows the number of Home Office press releases by month in 2008+2009. The pilot months are shaded and in italics. **Table 2** gives the 2008 and 2009 frequency distributions for statistical lacunae per press release as recorded by SMB and SB.

**Table 1** Number of Home Office Press Releases by month and year: 2008+2009.

Month	<i>Jan</i>	Feb	Mar	Apr	May	June	July	Aug	Sept	Feb. to Sept. TOTAL
Reviewer-pair	<i>SMB pilot</i>	<b>SMB</b> MH	<b>SB</b> BF	<b>SMB</b> MH	<b>SB</b> BF	<b>SMB</b> CL	<b>SB</b> BF	<b>SB</b> BF	<b>SMB</b> CL	
2008	<i>9</i>	<b>12</b>	<b>5</b>	<b>3</b>	13	<b>10</b>	<b>15</b>	11	<b>9</b>	<b>78</b>
2009	<i>17</i>	<b>18</b>	16	<b>22</b>	17	<b>11</b>	<b>17</b>	13	<b>11</b>	<b>125</b>

**Table 2** Frequency distribution for statistical lacunae per press release: 2008 and 2009.

Number of statistical lacunae per press release	SMB		SB		2008	2009
	2008	2009	2008	2009		
<b>NIL</b>	5+ 1+ 3+3= <b>12</b>	3+ 7+ 4+ 6= <b>20</b>	2+ 4+ 7+ 6= <b>19</b>	5+10+ 9+ 7= <b>31</b>	<b>31</b> (40%)	<b>51</b> (41%)
<b>1</b>	2+ 1+ 0+1= <b>4</b>	3+ 5+ 3+ 2= <b>13</b>	1+ 2+ 2+ 3= <b>8</b>	3+ 3+ 3+ 3= <b>12</b>	<b>12</b>	<b>25</b>
<b>2</b>	1+ 0+ 5+2= <b>8</b>	2+ 6+ 1+ 2= <b>11</b>	1+ 4+ 3+ 2= <b>10</b>	4+ 1+ 1+ 2= <b>8</b>	<b>18</b>	<b>19</b>
<b>3</b>	1+ 0+ 1+1= <b>3</b>	4+ 3+ 0+ 0= <b>7</b>	0+ 2+ 2+ 0= <b>4</b>	2+ 1+ 3+ 0= <b>6</b>	<b>7</b>	<b>13</b>
<b>4</b>	0+ 1+ 1+1= <b>3</b>	2+ 0+ 1+ 0= <b>3</b>	0+ 1+ 0+ 0= <b>1</b>	1+ 2+ 0+ 1= <b>4</b>	<b>4</b>	<b>7</b>
<b>5</b>	0+ 0+ 0+0= <b>0</b>	2+ 1+ 1+ 0= <b>4</b>	0+ 0+ 0+ 0= <b>0</b>	0+ 0+ 0+ 0= <b>0</b>	<b>0</b>	<b>4</b>
<b>6</b>	2+ 0+ 0+0= <b>2</b>	1+ 0+ 1+ 1= <b>3</b>	0+ 0+ 1+ 0= <b>1</b>	1+ 0+ 1+ 0= <b>2</b>	<b>3</b>	<b>5</b>
<b>7</b>	1+ 0+ 0+1= <b>2</b>	0+ 0+ 0+ 0= <b>0</b>			<b>2</b>	<b>0</b>
<b>8</b>		1+ 0+ 0+ 0= <b>1</b>			<b>0</b>	<b>1</b>
<b>9</b>					<b>0</b>	<b>0</b>
<b>10+ stem</b>			8= <b>1</b>		<b>1</b>	<b>0</b>
<b>Press release TOTALS</b>	12+3+10+9= <b>34 releases</b>	18+22+11+11= <b>62 releases</b>	5+13+15+11= <b>44 releases</b>	16+17+17+13= <b>63 releases</b>	<b>78 releases</b>	<b>125 releases</b>
<b>Lacunae TOTALS</b>	<b>67 Lacunae</b>	<b>114 Lacunae</b>	<b>68 Lacunae</b>	<b>74 Lacunae</b>	<b>135 lacunae</b>	<b>188 lacunae</b>
<b>Mean (sd) per press release</b>	<b>1.97 (2.06)</b>	<b>1.84 (1.94)</b>	<b>1.54 (3.38)</b>	<b>1.17 (1.54)</b>	<b>1.73 (3.51)</b>	<b>1.50 (1.78)</b>

Four features in **Table 2** are worth noting. First, for both calendar years, the mode is NIL lacunae and median one lacuna, with means of 1.7 (se 0.40) and 1.5 (se 0.16) lacunae for 2008 and 2009 respectively. There is notable over-dispersion in the frequency of lacunae per press-release. (Without clustering, we might expect the distribution for number of lacunae per press release have equal mean and variance, where variance is of course  $sd^2$ ). Over-dispersion applies even in 2009 without the hike in variation that was caused by an outlier press-release with a remarkable total of 18 statistical lacunae.

Second, SMB tended to score a higher number of lacunae than did SB. The overall comparison is not formally statistically significant –when comparing means 1.97+1.84 versus means 1.54+1.17 because qualified by a standard error of 0.69, and hence z-score of 1.58 – but the comparison between scorer-rates for 2009, which avoids the above outlier, is (z-score of 2.14,  $p < 0.05$ ).

Thirdly, neither the percentage of press releases with NIL lacunae (40%) nor the mean number of lacunae per press release (1.6) differed notably between 2008 and 2009. The average number of lacunae in the 60% of press releases which had any (that is: one or more) was: 323/121, or 2.7 (se 0.20).

Fourth, only 12/203 (6%) press releases, which are listed in **Table 3**, accumulated a tally of 6 or more statistical lacunae.

**Table 3** Press releases in February to September of 2008+2009 with 6+ lacunae.

<b>Month</b>	<b>Title of Press Release &amp; selected examples of lacunae therein</b>	<b>Statistical lacunae tally</b>
<b>2008</b>		
<b>March</b>	<b>New forces start training programme to further improve performance</b>	<b>18</b>
	On the programme police officers will learn world class tools and techniques . . . for example speed up procedures and improve meeting customer expectations.	
	The QUEST Academy . . . has helped heighten morale, increase performance and raise citizen satisfaction ratings.	
	Officers in Lancashire: 96% of customers who were surveyed following their appointment indicated that they were satisfied with the level of service provided	
	Cheshire Constabulary: 100% of scheduled appointments with the public were serviced at the expected time, with 82% of scheduled appointments within 1 day of an incident (to suit customers)	
<b>Feb.</b>	<b>New Action to tackle drug use to protect families &amp; communities</b>	<b>7</b>
	Compulsory testing on arrest and assessment by a drugs worker . . . contributed to a fall in recorded acquisitive crime of around 20% . . .	
	An innovative new pilot scheme which will help drug users who are on benefits get into treatment, get a job, and live a drug-free life.	
	Parents with dependant children will get better and faster access to specialist drug treatment.	
<b>Sept.</b>	<b>First ID card unveiled</b>	<b>7</b>

Document shows the holder's photograph, name, date of birth, nationality and immigration status. A secure electronic chip will also hold their biometric details, including fingerprints and a digital face image . . . To earn and retain their licence as a sponsor, businesses and education providers must keep records . . . including, in time, a copy of a migrant's identity card. [Doubt about database]		
3,500 cases of identity swap in the UK have been identified and dealt with [How identified?]		
£1 bn saved from the total cost of the scheme [How computed?]		
<b>Feb.</b>	<b>Government gets tough on serious violence</b>	<b>6</b>
Serious violent crime accounts for 1% of all crime.		
Violent crime has fallen by 31% since 1997		
Recent BCS, which highlighted how some crimes such as . . . homicide . . . have reduced . . .		
<b>Feb.</b>	<b>Lowest number of asylum applications in 14 years</b>	<b>6</b>
By the end of last year, BIA was beating its target to have 40% of new asylum cases concluded within 6 months.		
We beat the PM's target and the rate at which foreign national prisoners are being deported has almost doubled.		
<b>July</b>	<b>Dramatic fall in crime in England and Wales</b>	<b>6</b>
<b>2009</b>		
<b>Feb.</b>	<b>The latest statistics on immigration and asylum</b>	<b>8</b>
Although applications . . . rose in the last quarter of 2008, of the 2,540 who made applications only 832 were put forward . . .		
Asylum applications have fallen by 2% - down from 6,870 between Oct. and Dec. 2007 to 6,735 in the same period for 2008. [Consistent with random variation]		
. . . has met key milestone in concluding 60% of new asylum cases within 6 months by the end of 2008.		
<b>Feb.</b>	<b>Operation Staysafe Weekend</b>	<b>6</b>
11 local areas in Nov+Dec 2008, 107 staysafe operations, 244 young people removed to safety, 54 transferred to family or parenting interventions		
27 Friday staysafe operations: > 1,000 young people engaged with (actually 1,612 in 22 Staysafe Friday operations), of whom 120 were returned home & 103 referrals made to family or parenting or alcohol projects.		
Children's permission to report their case-studies (5: no-names)		
<b>March</b>	<b>Ending violence against women and girls – your voice counts</b>	<b>6</b>
. . . for dealing with serial perpetrators of domestic violence [doubt about definitions]		
A new opinion poll by Ipsos Mori shows that a more than a third of respondents know a woman who has been the victim of violence from a man she knows . . .		
Courts may need to control violent perpetrators – particularly serial offenders who move between relationships . . .		
. . . real progress with domestic violence incidents more than halving in the last 12 years		
'One in five of all violent crimes reported are related to domestic abuse, while every year one in six of all murders in the UK are domestic violence-related homicides. The statistics are horrific . . .'		
<b>June</b>	<b>Anti-binge drinking campaign remind people to know their limits</b>	<b>6</b>
Alcohol-related crime & disorder is estimated to cost society between £8B and £13B per year		
Binge drinking is a major factor in accidents, violence and anti-social behaviour		
Last year's adverts – which made two-thirds of 17-24 year olds who saw them say they would reconsider their behaviour - . . .		
<b>July</b>	<b>Home Secretary statement on annual crime statistics</b>	<b>6</b>

<b>Sept.</b>	<b>Tough new powers to help victims break cycle of domestic violence</b>	<b>6</b>
They will be piloted in two police force areas to test the impact on the CJ system. Further details will be published later this autumn		
There are now over 700 IVDA's (Independent Domestic Violence Advisors) and over 200 local areas have MARACs . . . in the last 12 months have worked to protect 29,000 victims of domestic abuse. {But, earlier in release, told: around 750,000 incidents of domestic violence reported to police every year, resulting in 200,000 arrests. ? proportionality as 750,000/700 > 1,000 per IVDA worker pa, or > 4 per working day }		
Conviction rates are improving – 72% of cases charged & brought to court at the end of 2008/09 resulted in a conviction compared with 60% four years ago.		

Notwithstanding some reservations on inter-rater variation, **Appendix Table C** lists the 29 pre-specified statistical lacunae on which press releases were scored by assessors as to frequency of occurrence per press release. For each month, **Table C** shows the only sum of an assessor's assigned frequencies per lacunae for the press releases in a given month. *The same fault may occur more than once in a single sentence, but was counted ONCE ONLY per sentence.*

**Table 4** simplifies more detailed tables in the **Appendix**. Of the 29 specific lacunae, which shared 301 cumulative mentions when scored by SMB/SB, the eight that attracted the highest assigned scores (14+ each; and accounted for 193/301 (64%) lacunae recorded by SMB/SB) are highlighted in **blue**. Those which attracted 7+ mentions when scores by BF/CL (and accounted for 112/151 (74%) of their noted lacunae) are highlighted in **grey**. Three lacunae, which are the top three finally, are in common between the two lists, but it is combined scores of 21+, highlighted in **green**, on which SDSSC bases its summary. Eight of the 29 specific lacunae are thereby identified, which account for 62% (282/452: 95% CI from 58% to 67%) of the assigned scores. The eight fall into three groups - claimed success or trend without any quantification; inadequate reporting of counts or percentage change; and missing methodology – as follows:

[ 5.] **Claimed 'success' without any quantification** (38 mentions)

[24.] **Claimed trend, absent evidence** (21 mentions)

[ 1.] **Count-now, no background or reference count** (27 mentions)

[ 2.] **% change in counts, absent 'N1 to N2'** (86 mentions)

[10.] **Missing methodology: survey details** (46 mentions)

[11.] **Missing methodology: evaluation design** (22 mentions)

[12.] **Missing methodology: re cost estimation** (21 mentions)

[21.] **Doubt re Definitions** (21 mentions).



**Table 4** Statistical lacunae, and their frequency of occurrence by assessor. {blue if lacuna accounts for 14+ of SMB/SB tally across all 29; grey if extra assessor accounts for 7+; green if total 21+}.

Lacunae	Assessor {SMB for 11 Dec. 2008 PR+FS}	SMB/SB	Extra assessor (CL/BF): 4 months	TOTAL
1. Count-now, no background or reference count	{2}	21	6	27
2. % change in counts, absent 'N1 to N2'	{8}	63	23	86
3. %, absent 'se' or denominator	{1}	16	4	20
4. Means compared, no 'se'	{2}	1	0	1
5. Claimed 'success' without any quantification	{4}	16	22	38
6. Claimed 'success' without study design		8	9	17
7. Cited benefits, absent harms		4	7	11
8. Missing methodology: re pilot/roll-out		14	0	14
9. Missing methodology: eg points-system		7	0	7
10. Missing methodology: survey details	{1}	29	17	46
11. Missing methodology: evaluation design ~ before-after, quasi-experiment, RCT etc		6	16	22
12. Missing methodology: re cost estimation		15	6	21
13. Costs: unadjusted for inflation		4	0	4
14. Selective citation: units/centre	{1}	5	1	6
15. Selective citation: time periods	{4}	5	2	7
16. Selective citation: outcomes		11	8	19
17. Selective citation: subgroups		3	2	5
18. Leading descriptive text, eg: "small" %	{1}	7	5	12
19. Target, absent baseline		7	4	11
20. Target versus achievement	{2}	9	3	12
21. Doubt re Definitions	{2}	11	10	21
22. Arithmetic infelicity or inconsistency	{1}	11	4	15
23. Proportionality: at issue	{3}	3	0	3
24. Claimed trend, absent evidence	{2}	19	2	21
25. Absent evidence interpreted as 'evidence of absence'		0	0	0
26. Absent caveats re change-point in series		0	0	0
27. Inferential failure to convey 'consistent with random fluctuation'		4	0	4
28. Inferential confusion: eg annual v. quarterly; indicator A v. B		0	0	0
29. Databases: doubtful claims		2	0	2
<b>Total across 29 specified lacunae</b>	<b>{34}</b>	301	151	452

Finally, the 11 December 2008 press release (PR) and 'fact sheet' (FS) on knife crime were scored by SMB as having 34 lacunae (PR: 10 and FS: 24), as shown in parentheses in **Table 4**. Only one out of 203 press releases scored by SMB/SB, see **Table 2**, had a higher total (18) and no press release came close to the combined score of 34 for PR+FS.

## Discussion

The SDSSC's investigation of statistical lacunae in Home Office press releases for February to September 2008+2009 has established that only one release out of 203 had more errors of omission/commission than the press release of 11 December 2008 (Tackling knives programme – six months on); and none was so extreme as the press release and its "fact sheet" combined.

Forty percent of press releases had no statistical lacunae (82/203) and, at the other extreme, six press releases (3%) had six lacunae, five had 7-9 and one had 10+. In the 60% of press releases with one or more statistical lacunae, the mean number present was 2.7 (se 0.2).

There was evident inter-rater variation in SDSSC’s deciding which on the list of 29 specific lacunae applied, and in the degree of critique. Some simplification of the list was clearly warranted. The 452 assigned scores by four assessors were a useful guide for making mergers or deletions, see **Table 5**. Groups A to C in **Table 5** derive from the top scoring eight of the original lacunae, with Group D (selective citation) retained because selectivity is so potentially under-mining of trust as it was, for example, in the 11 December 2008 press release and accompanying “fact-sheet”.

**Table 5** Merged statistical lacunae (A to D) – after deletions.

Statistical errors of omission	TOTAL
<b>A. Inadequate reporting of counts, percentages, % change, or means: 134</b>	
1. Count-now, no background or reference count	27
2. % change in counts, absent ‘N1 to N2’	86
3. %, absent ‘se’ or denominator	20
4. Means compared, no ‘se’	1
<b>B. Claimed success or trend without quantification or other evidence: 70</b>	
5. Claimed ‘success’ without any quantification	38
24. Claimed trend, absent evidence	21
19. Target, absent baseline	11
<b>C. Missing methodology or definitions: 124</b>	
10. Missing methodology: survey details	46
11. Missing methodology: evaluation design ~ pilot/roll-out, before-after, quasi-experiment, RCT etc	22 + 14
12. Missing methodology: re cost estimation	21
21. Doubt re Definitions	21
<b>D. Selective citation: centres, time-periods, outcomes, subgroups: 37</b>	

The SDSSC’s critique emphasised the need (A) to report basic data in press releases – actual counts, denominators for percentages, and before-after counts not just % change. The public is rightly sceptical about claimed ‘success’ or ‘trends’ (B) when neither quantification nor other evidence backs up the claim. Press releases which fail to differentiate policy aspirations from their achievement lack credibility. Thirdly, methodology matters (C) for public understanding, and for trust, so that survey details need to be outlined (how subjects were selected for the survey, their response-rate, number of respondents), evaluation designs specified (including the numbers of participants), and guidance given on how costs were estimated or forecast. Definitions matter in official statistics and press releases should avoid dubious linguistic licence.

Whereas Ministers habitually rely on their officials to provide the back-up answers ABC when they ask for them, the rest of us do not have this access and so require that the basic ABC information is upfront if the public is to have any chance of making an informed judgement on the matter which a press release brings to attention.

As in planning assumptions, the pooled variance of the number of lacunae per press-release in SDSSC's study ( $6.66 = 2.58 * 2.58$ ) was at least twice the mean. Over-dispersion, or clustering, in number of lacunae per press-release suggests that some drafters are less statistically-adept in conveying what is essential for the public's proper understanding of an evidence-base or some policies have an evidence-base so weak that there is unwillingness for ABC to be transparent.

No statistically significant difference was detected between calendar years of the order targeted: 50% power in respect of 25% reduction in lacunae, and 80% power in respect of a one third reduction. However, getting right the ABC of press-releases is more than capable of delivering at least a one third reduction since ABC includes the top eight of the original 29 lacunae, and they accounted for more than half of the assigned scores (282/452: 95% CI from 58% to 67%).

By dint of other commitments and estimated work-load, some SDSSC members could not complete their full assignment by the agreed dead-lines. However, both MH and BF had undertaken independent ratings for February/March which were used in cross-calibration sessions with their pair-mate who undertook the later assignments in the light of these paired sessions.

## APPENDIX

**Tables A and B** illustrate some difficulty for assessors in agreeing on the assigned lacuna, which suggests the need for simplification of the 29-item coding frame: note the disparity in scores between SMB and CL when rating the same set of 41 press releases, and between SB and BF before their consensus on 21 press releases. Consensus scores tended to add to, rather than deduct from, separate ratings except for random miscodes, re-assignment of codes for ‘other, please specify’ or discounting if a lacuna had occurred within a non-Ministerial quotation.

**Table A.** Variation between **SB** and **BF**, and **SB/BF** consensus scores, on 21 press-releases from March of 2008+2009.

<b>SB</b> Label for Press-Release	<b>SB</b> lacunae codes	<b>BF</b> lacunae codes	<b>SB/BF</b> consensus codes
<b>1. Mar08 R01</b>	1, 1, 1, 2, 2, 6, 6, 6, 6, 16, 16, 16, 16, 17, 17, 30, 30, 30	5, 6, 6, 6, 6, 7, 7, 7, 7, 11, 11, 11, 11, 16, 16, 16, 16, 17, 17, 21, 21, 30, 30, 30	1, 1, 1, 2, 2, 6, 6, 6, 6, 7, 7, 7, 7, 16, 16, 16, 16, 17, 17, 30, 30, 30
2. Mar08 R02	nil	Nil	Nil
3. Mar08 R03	22	22	22
4. Mar08 R04	1, 1	1, 1	1, 1
5. Mar08 R05	nil	Nil	Nil
<b>6. Mar09 R01</b>	2, 6, 22	2, 6, 7, 11, 20	2, 6, 22
7. Mar09 R02	10, 10, 19, 21	3, 3, 10	3, 10, 10, 19, 21
8. Mar09 R03	22	22	22
9. Mar09 R04	21, 24, 30, 30, 30, 30	2, 16	16, 24, 30, 30, 30
10. Mar09 R05	1, 15	3, 15	1, 15
11. Mar09 R06	nil	Nil	Nil
12. Mar09 R07	2, 30	2, 5, 6, 7	2, 5, 6, 30
13. Mar09 R08	7, 30	11, 20	7, 30
14. Mar09 R09	30	Nil	Nil
15. Mar09 R10	2, 16	14, 16	2, 16
16. Mar09 R11	nil	21	Nil
17. Mar09 R12	nil	Nil	Nil
18. Mar09 R13	5	5, 18	5
19. Mar09 R14	nil	Nil	Nil
20. Mar09 R15	nil	Nil	Nil
21. Mar09 R16	2, 16	2, 2, 16, 21	2, 16, 21
22. Mar09 R17	Statistical News Release	21	Not counted
23. Mar09 R18 = dated 1 April 2009 on release	Statistical News Release	21, 30	Not counted
<b>TOTAL Lacunae: for 21 press releases</b>	<b>47</b>	<b>55</b>	<b>53</b>

**Table B.** Variation between **SMB** and CL, and SMB-reviewed consensus scores, on 41 press-releases from June and September of 2008+2009.

<b>SMB Label for Press-Release</b>	<b>SMB lacunae codes</b>	<b>CL lacunae codes</b>	<b>Consensus codes</b>
<b>1. Sept08 R01</b>	nil	5	5
2. Sept08 R02	9	10, 10	9, 9
3. Sept08 R03	nil	10	10
4. Sept08 R04	2, 3	11, 11	2, 3, 11
5. Sept08 R05	8, 8, 11	Nil	8, 8, 11
6. Sept08 R06	1, 8, 8, 18	5, 5, 6	1, 8, 8, 18, 20, 20
7. Sept08 R07	nil	5 (quote)	Nil
8. Sept08 R08	7, 10, 12, 12, 13, 20, 29	1, 5, 5, 10, 12	7, 10, 12, 12, 13, 20, 29
9. Sept08 R09	5, 10	1, 10, 10, 10, 24	1, 5, 10, 10, 24
<b>10. Sept09 R01</b>	7, 12	12	7, 12
11. Sept09 R02	nil	nil	Nil
12. Sept09 R03	nil	nil	Nil
13. Sept09 R04	nil	nil	Nil
14. Sept09 R05	2, 5	5	2, 5
15. Sept09 R06	3	nil	3
16. Sept09 R07	nil	nil	Nil
17. Sept09 R08	nil	nil	Nil
18. Sept09 R09	nil	11	11
19. Sept09 R10	10	11	11
20. Sept09 R11	3, 3, 5, 9, 9, 23	11, 11	3, 3, 5, 9, 9, 11, 23
<b>21. June08 R01</b>	nil	nil	12, 18
22. June08 R02	16, 21	nil	16, 21
23. June08 R03	2, 5, 7, 29	2	2, 5, 7, 12, 29
24. June08 R04	16, 21	nil	16, 21
25. June08 R05	3, 18, 21	30, 30, 30, 30	3, 18, 21, 30
26. June08 R06	19, 20	2, 18, 19, 19, 19, 19	2, 18, 19, 19, 20
27. June08 R07	nil	5, 22	5, 22
28. June08 R08	nil	5	5
29. June08 R09	2, 10	5, 10	2, 5, 10
30. June08 R10	12, 18	12	12, 18
<b>31. June09 R01</b>	nil	nil	Nil
32. June09 R02	nil	5	Nil
33. June09 R03	3, 8, 10, 12, 12, 18	10, 12, 30, 30, 30	3, 8, 10, 12, 12, 18
34. June09 R04	6, 7, 12, 22, 24	5, 5, 12	6, 7, 12, 22, 24
35. June09 R05	12	5	12
36. June09 R06	2, 3	6	2, 3, 6
37. June09 R07	nil	5	Nil
38. June09 R08	2	5, 6	2, 6
39. June09 R09	nil	nil	Nil
40. June09 R10	2, 3, 11, 15	11, 30	2, 3, 5, 11, 15
41. June09 R11	2	11, 11	2
<b>TOTAL Lacunae: for 41 press releases</b>	<b>68</b>	<b>61</b>	<b>93</b>

**Table C (2008)** Statistical lacunae, and their frequency of occurrence (by assessor), in monthly press releases in 2008.

Month in 2008	Feb.	March	April	May	June	July	Aug.	Sept.	Total
<b>Press releases</b>	<b>12</b>	<b>5</b>	<b>3</b>	<b>13</b>	<b>10</b>	<b>15</b>	<b>11</b>	<b>9</b>	<b>78</b>
Assessor {extra assessor}	SMB	SB BF	SMB	SB	SMB CL	SB BF	SB	SMB CL	SMB/SB {extra}
<b>Statistical lacunae: and their frequency of occurrence in monthly press releases.</b>									
1. Count-now, no background or reference count	1	5 2	1			1 1	1	1 2	10 { 5}
2. % change in counts, absent 'N1 to N2'	3	2		8	2 2	12 7	1	1	29 { 9}
3. %, absent 'se' or denominator	1				1	0 1		1	3 { 1}
4. Means compared, no 'se'	1								1
5. Claimed 'success' without any quantification	1	0 1		5	1 3	0 3		1 6	8 {13}
6. Claimed 'success' without study design	1	4 4				1 0		0 1	6 { 5}
7. Cited benefits, absent harms		0 4			1	0 1		1	2 { 5}
8. Missing methodology: re pilot/roll-out	5							3	8
9. Missing methodology: eg points-system	3		1					1	5
10. Missing methodology: survey details				6	1 1	3 2	1	2 7	13 {10}
11. Missing methodology: evaluation design ~ before-after, quasi- experiment, RCT etc		0 4				0 1		1 2	1 { 7}
12. Missing methodology: re cost estimation	1				1 1			2 1	4 { 2}
13. Costs: unadjusted for inflation	1							1	2
14. Selective citation: units/centre			1						1
15. Selective citation: time periods						0 1			0 { 1}

16. Selective citation: outcomes		4 4			2	0 1			6 { 5}
17. Selective citation: subgroups		2 2							2 { 2}
18. Leading descriptive text, eg: "small" %					2 1	0 1		1	3 { 2}
19. Target, absent baseline	4				1 4			1	6 { 4}
20. Target versus achievement	2				1	0 1		1	4 { 1}
21. Doubt re Definitions	1	0 2			3	0 3			4 { 5}
22. Arithmetic infelicity or inconsistency		1 1		1	0 1	0 1			2 { 3}
23. Proportionality: at issue									0
24. Claimed trend, absent evidence	1		2			0 1	3	0 1	6 { 2}
25. Absent evidence interpreted as 'evidence of absence'									0
26. Absent caveats re change-point in series									0
27. Inferential failure to convey 'consistent with random fluctuation'									0
28. Inferential confusion: eg annual versus quarterly; indicator A versus B									0
29. Databases: doubtful claims					1			1	2
30. OTHER, please specify*		3 3			0 4	3 0	1		7 { 7}
<b>LACUNAE TOTAL {extra}</b>	<b>26</b>	<b>21</b> <b>27</b>	<b>5</b>	<b>20</b>	<b>17</b> <b>17</b>	<b>20</b> <b>25</b>	<b>7</b>	<b>19</b> <b>20</b>	<b>67+68</b> <b>{89}</b>
<b>SMB/SB press releases with lacunae {extra}</b>	<b>7</b>	<b>3</b> <b>3</b>	<b>2</b>	<b>9</b>	<b>7</b> <b>7</b>	<b>8</b> <b>8</b>	<b>5</b>	<b>6</b> <b>8</b>	<b>25/44+</b> <b>22/34</b>

**Table C (2009)** Statistical lacunae, and their frequency of occurrence (by assessor), in monthly press releases in 2009.

Month in 2009	Feb.	March	April	May	June	July	Aug.	Sept.	Total
<b>Press releases</b>	<b>18</b>	<b>16</b>	<b>22</b>	<b>17</b>	<b>11</b>	<b>17</b>	<b>13</b>	<b>11</b>	<b>125</b>
Assessor {extra assessor}	SMB	SB BF	SMB	SB	SMB CL	SB BF	SB	SMB CL	SMB/SB {extra}
<b>Statistical lacunae: and their frequency of occurrence in monthly press releases.</b>									
1. Count-now, no background or reference count	7	1 1	1	1		1 0			11 { 1}
2. % change in counts, absent 'N1 to N2'	3	4 4	2	8	4	7 10	5	1	34 {14}
3. %, absent 'se' or denominator	4	0 3	3		3			3	13 { 3}
4. Means compared, no 'se'									
5. Claimed 'success' without any quantification		1 2	1	2	1 6	1 0		2 1	8 { 9}
6. Claimed 'success' without study design		1 2			1 2				2 { 4}
7. Cited benefits, absent harms		1 2						1	2 { 2}
8. Missing methodology: re pilot/roll-out	3				1			2	6
9. Missing methodology: eg points-system	1							1	2
10. Missing methodology: survey details	3	2 1	4	2	1 1	2 5	2		16 { 7}
11. Missing methodology: evaluation design ~ before-after, quasi-experiment, RCT etc	4	0 2			1 3			0 4	5 { 9}
12. Missing methodology: re cost estimation	2		3	1	4 2	0 1		1 1	11 { 4}
13. Costs: unadjusted for inflation			2						2
14. Selective citation: units/centre	3	0 1	1						4 { 1}
15. Selective citation: time periods	2	1 1	1		1				5 { 1}



16. Selective citation: outcomes	2	2 3	1						5 { 3}
17. Selective citation: subgroups			1						1
18. Leading descriptive text, eg: "small" %	1	0 1			1	1 2	1		4 { 3}
19. Target, absent baseline		1 0							1
20. Target versus achievement	4	0 2	1						5 { 2}
21. Doubt re Definitions	1	3 2	3			0 3			7 { 5}
22. Arithmetic infelicity or inconsistency	3	2 1	2	1	1				9 { 1}
23. Proportionality: at issue	1		1					1	3
24. Claimed trend, absent evidence	2	1 0	2	1		6 0	1		13
25. Absent evidence interpreted as 'evidence of absence'									
26. Absent caveats re change-point in series									
27. Inferential failure to convey 'consistent with random fluctuation'	3		1						4
28. Inferential confusion: eg annual versus quarterly; indicator A versus B									
29. Databases: doubtful claims									
30. OTHER, please specify*	2	7 0	1	1	4	2 0	2		15 { 4}
<b>LACUNAE TOTAL {extra}</b>	<b>51</b>	<b>27</b> <b>28</b>	<b>31</b>	<b>16</b>	<b>20</b> <b>18</b>	<b>20</b> <b>21</b>	<b>11</b>	<b>12</b> <b>6</b>	<b>74+114</b> <b>{73}</b>
<b>SMB/SB press releases with lacunae {extra}</b>	<b>15</b>	<b>11</b> <b>11</b>	<b>15</b>	<b>7</b>	<b>7</b> <b>9</b>	<b>8</b> <b>6</b>	<b>6</b>	<b>5</b> <b>5</b>	<b>32/63+</b> <b>42/62</b>