Journal of Official Statistics, Vol. 23, No. 3, 2007, pp. 319–344

Benchmarking the Effect of Cell Adjustment on Tabular Outputs: The Shortcomings of Current Approaches

Paul Williamson[1]

Current assessments of the effect of disclosure control lack rigour and are too far removed from the real implications for user analyses. This article uses a case study to illustrate these shortcomings and suggest a remedy. The proposed solution has three main components. First, the use of a set of benchmark counts providing representative population and geographical coverage. Second, consideration of the effect of cell adjustment not only upon counts, but also upon rates derived from them. And third, the use of more transparent statistical summary measures, such as indicative confidence intervals. The benchmark counts and benchmarking software used in the writing of this article have been placed in the public domain, in order to encourage the wider adoption of the proposed evaluation strategy. The general approach outlined, however, could be pursued independently of the use of either.

Key words: Statistical disclosure control; random rounding; small cell adjustment; goodness of fit; fitness for purpose.

1. Measuring the Effect of Cell Adjustment

Within statistical agencies the need to protect respondent confidentiality is a given, almost invariably backed by national statute, and necessitates implementation of a range of statistical disclosure controls. A clear distinction may be made between those controls focusing upon issues of table layout (dimensionality; variable banding; geographical unit; population threshold), and those which involve perturbation of the underlying data (e.g., pretabulation record swapping and noise addition; post-tabulation noise addition, count rounding and cell suppression). In this article the focus is upon the implications of data perturbation, of whatever kind, for user analyses. A general review of disclosure control methods and their implementation is provided by FCSM (1994), whilst readers interested in the most recent developments in tabular protection are referred to, *inter alia*, Cox and Kelly (2003), Giessing (2003), Dandekar (2004), Fienberg and McIntyre (2005) and Salazar-Gonzalez (2005). Methods of assessing the associated risks of disclosure are similarly discussed elsewhere, by authors including Cox (2001), Duncan et al. (2001), Elliot (2001) and Shlomo (2005).

[1] Department of Geography, University of Liverpool, Liverpool, L69 7QU, UK. Email: P.Williamson@liv.ac.uk
Acknowledgments: Results in this article are based upon analysis of benchmark data derived from 1991 England and Wales Census Small Area Statistics and freely available from <http://pcwww.liv.ac.uk/~william/SDC/>. Census data are Crown copyright and are reproduced with the permission of the Controller of HMSO under Click-Use licence C02W0001725.
Many thanks to the anonymous reviewers of an initial draft of this article. I hope they agree that their kind comments have resulted in material improvements in the published version.

The initial stimulus for this article was a user consultation exercise conducted by the England and Wales Office for National Statistics (ONS) in the run-up to the 2001 UK Census. Users were offered a choice of two alternative cell rounding approaches (ONS 2002). The first, random rounding of all cell counts, interior and marginal, to a neighbouring multiple of 3, was modelled on the system implemented by Statistics New Zealand (Statistics New Zealand 2005). The second, based on an alternative approach adopted by the Australian Bureau of Statistics (ABS 1997), and described by ONS as Small Cell Adjustment, involved the random rounding of interior table counts of 1 and 2 to either 0 or 3, with table marginals then being recalculated to fit the sum, post-rounding, of the interior counts. Surprisingly, no formal evaluation of the relative merits of either system exists, at least in print. Instead, to guide their choice users were provided with a small example set of perturbed tables. Different tables, covering different geographical areas, were released for each method, somewhat hindering direct comparison; a step taken in order to safeguard the precise details of disclosure control implementation. For the same reason, broad guidance on the 95% confidence intervals associated with the summation of 5, 10, and 20 modified cells was released, but for random rounding only. As ONS acknowledged at the time, this information was inadequate to support users in making a truly informed choice.

Set against this backdrop, the aim of this article is to outline a new strategy for evaluating the effect of alternative disclosure control methods that is simultaneously far more wide-ranging in scope and far more directly user-relevant. Importantly, the strategy proposed would allow statistical agencies and their user communities to make better informed trade-offs between data utility and data protection, whilst still allowing statistical agencies to avoid releasing disclosive information about the underlying cell adjustment methodologies involved.

The article is structured as follows. The next section briefly surveys the main measures of cell adjustment effect found in the literature. A discussion highlights the ways in which all of them fail to adequately assess the effect of cell adjustment in the context of the real-world situations faced by users. As a result an alternative approach is proposed, that involves applying a series of user-relevant summary measures to a benchmark dataset, containing a range of geographically representative counts and, crucially, rates. Section 3 describes the creation of such a benchmark dataset. This dataset is used in Section 4 as the basis for revisiting the UK cell adjustment debate. The resulting evaluation graphically illustrates the shortcomings of existing approaches, and serves to underline the importance of placing statistical assessments of cell adjustment in a real-world context. Section 5 develops the argument further, explicitly illustrating the place, table and subgroup specific effect of cell adjustment. On the strength of these results a final section summarises the case for a more rigorous evaluation approach based upon use of benchmark data. The article concludes with a brief discussion of two obstacles to the benchmarking strategy proposed: the provision of a benchmark dataset, and the computational challenge posed. Potential solutions are proposed in the form of a public domain benchmark dataset and evaluation software suite.

2. Existing Evaluation Approaches and Their Shortcomings

Existing debates over the choice of methods for evaluating the effects of disclosure control focus principally on the measure of data utility to be used. Proposed measures of note are listed in Table 1. All summarise in some way, at a tabular level, the difference between pre- and post-adjustment cell counts. Their great variety reflects the difficulty of simultaneously capturing in one measure both the absolute and relative effects of cell adjustment. This problem is well-known. A change of 2 in the value of a cell count is small in absolute terms, but might be either trivial or highly significant in relative terms, depending on whether the preadjustment count was 1 or 1,000. Less widely acknowledged, the problems of providing a satisfactory summary measure are exacerbated when the sums of the pre- and post-adjustment counts differ significantly (as can arise, for example, through Small Cell Adjustment). In a review of a number of leading candidates Voas and Williamson (2001) concluded by expressing a preference for the sum of squared Z-scores, a modified version of which, Z_m, can be used when appropriate to make due allowance for the possibility of differing pre-and post-adjustment totals. Other potentially interesting approaches include those outlined by Chambers (2001) and Duncan et al. (2001), although both have significantly higher computational overheads. Of particular significance for this article, however, is that current applications of all of these measures suffer from four major shortcomings.

First and foremost, as inspection of any of the studies listed in Table 1 will reveal, attention is paid only to the direct effect of disclosure control upon cell counts. The "indirect" effect of these modified cell counts upon typical user analyses, including aggregating counts across subgroups and geographical areas, calculating rates, ranking areas and undertaking ecological correlations and regressions, remain unconsidered. Second, none of the measures listed in Table 1 is translatable into a direct and user-friendly indication of the likely effect of cell adjustment upon a typical user analysis. For these purposes measures such as the confidence intervals associated with post-adjustment counts and rates would arguably be more helpful. Third, insufficient attention has been paid to the effects of cell adjustment upon marginal table cell counts. Yet cell adjustment regimes can treat marginal cells in markedly different ways, ranging from independent rounding to accumulation of independently perturbed interior counts. Fourth, existing assessments of cell adjustment methods appear to focus almost invariably upon narrow and potentially biased selections of tables, places and population subgroups. Yet the effect of cell adjustment, particularly for rate-based analyses, is clearly table-, place- and population subgroup-specific.

Gauging the extent to which the identifiable shortcomings in published studies apply equally to evaluations undertaken within statistical agencies is difficult. ONS is hardly alone in providing minimal information to users concerning the effects of disclosure control. Statistics Canada, for example, randomly rounds all census cell counts to a neighbouring multiple of five. Their public acknowledgement of the effect of this measure is limited to the claim that "This technique provides strong protection against direct, residual or negative disclosure, without adding significant error to the census data" (Statistics Canada 2003, p. 24). Similar comments may be made with regard to the information provided by Statistics New Zealand and the Australian Bureau of Statistics. Yet such evidence as does exist in the public domain suggests that within agency evaluations are in fact subject to precisely the same criticisms (cf. Boyd and Vickers 1999; Eurostat 1996; FCSM 1994; ONS undated).

Table 1. Measures of cell-adjustment effect proposed in literature

Measure	Derivation	Cited in
Distribution-free measures		
No. of modified cells*	ΣNC_i where $NC_i = 1$ if $O_i <> E_i$; 0 otherwise	Kirkendall and Sande (1989)
Total value of modified cells*	ΣE_i	Kirkendall and Sande (1989)
Proportion of modified cells*	$(\Sigma NC_i)/n$	Domingo-Ferrer and Torra (2001)
Total absolute error*	$\Sigma\|(O_i - E_i)\|$	Eurostat (1996)
Root mean square error*	$[(\Sigma(O_i - E_i)^2)/\mathrm{n}]^{0.5}$	Voas and Williamson (2001)
Standardised absolute error*	$(\Sigma\|(O_i - E_i)\|)/\Sigma E_i$	Voas and Williamson (2001)
Relative absolute distance	$\Sigma(\|(O_i - E_i)\|/E_i)$	Shlomo (2005)
Average absolute distance	$\Sigma(\|O_i - E_i\|/k)$ where $k =$ no. of nonzero E_i	Shlomo (2005)
Sum of Z_m^{2*}	$\Sigma\{[(Oi/\Sigma E_i) - e_i]/[e_i(1 - e_i)/\Sigma E_i]^{0.5}\}^2$	Voas and Williamson (2001)
Chi-square distributed measures		
Chi-Square*	$\Sigma\{(O_i - E_i)^2/E_i\}$	Fienberg, Makov and Steele (1998)
Freeman-Tukey	$4\Sigma(O_i^{0.5} - E_i^{0.5})^2$	Voas and Williamson (2001)
Cressie-Read	$9/5\Sigma O_i[(O_i/E_i)^{2/3} - 1]$	Voas and Williamson (2001)
Sum of Z^{2*}	$\Sigma\{(o_i - e_i)/[e_i(1 - e_i)/\Sigma O_i]^{0.5}\}^2$	Voas and Williamson (2001)
Wald	$W = (\mathbf{R} - \mathbf{S})^t[\mathrm{diag}(\mathbf{R} + \mathbf{S}) - \mathbf{T} - \mathbf{T}^t]^{-1}(\mathbf{R} - \mathbf{S})$; where **R** and **S** are vectors representing O and E, respectively, for $i < n$, and **T** is the cross-classification of **R** and **S**	Chambers (2001)

Table 1. Continued

Measure	Derivation	Cited in
Entropy based		
Difference in Shannon entropy	$(-\Sigma o_i \log o_i) - (-\Sigma e_i \log e_i)$	Gouweleeuw et al. (1998)
Phi	$\Sigma\{(o_i)\lvert\log(O_i/E_i)\rvert\}$	Fotheringham and Knudsen (1987)
Psi	$\Sigma o_i \log(o_i/s_i) + \Sigma e_i \log(e_i + s_i)$ for $i = 1$ to n where $s_i = 0.5(o_i + e_i)$	Fotheringham and Knudsen (1987)
Other		
Cramer's V*	$[\chi^2/n \min(r-1, c-1)]^{0.5}$ where $r =$ no. of rows and $c =$ no. of columns, in table.	Boyd and Vickers (1999)
Gibson's D*	$0.5\Sigma\lvert(e_i) - (o_i)\rvert$, for $i = 1$ to n	Griffin et al. (1989)
Hellinger distance	$\{0.5\Sigma[(E_i n)^{0.5} - (O_i n)^{0.5}]^2\}^{0.5}$	Gomatam et al. (2003)
Mean squared precision	Not directly derivable from tabulated counts alone	Duncan et al. (2001)

Table key: E_i = expected (preadjustment) cell value; O_i = observed (post-adjustment) cell value; $n =$ no. of cells in table; $\Sigma =$ sum over all i, from 1 to n; $o_i = O_i/\Sigma O_i$; $e_i = E_i/\Sigma E_i$

*Measure calculated by *SDC-i* software suite; see Section 6.3 of article for details.

In the light of the shortcomings identified, a twofold strategy is proposed for placing the evaluation of cell adjustment effects on a more rigorous footing. First, effects should be evaluated over a wide range of cell counts (interior and marginal) and rates, representative in terms of both geographical and population subgroup coverage. Second, not one, but a series of key statistical measures should be considered. In combination these measures should help to identify the main strengths and weaknesses of each cell adjustment approach whilst at the same time offering some practical insight into their potential effects upon end user analyses, both count and rate based. The remainder of this article explores the implementation of such a strategy, using as a case study the choice placed before users by ONS in the run-up to the production of 2001 Census outputs.

3. Creation of a Benchmark Dataset

In order to evaluate the effect of a cell adjustment methodology, access to the original preadjustment data is required. Lacking such access, for the purposes of this article tabular outputs from the 1991 UK Census Small Area Statistics (SAS) have been used as a set of "preadjustment" cell counts. In reality, prior to their release these outputs were themselves subjected to a method of cell adjustment referred to by ONS as Barnardisation. Barnardisation involves the unbiased addition of some random noise (± 0 or 1) to each nonzero interior cell count, followed by recalculation of table marginals on the basis of the modified interior counts. Crucially, unlike for cell adjustment approaches based upon count rounding or suppression, the resulting post-Barnardisation distribution of cell counts may reasonably be assumed to remain representative of the original, although it should be acknowledged that this assumption may not hold in all cases.

From the full range of 1991 UK Census outputs a set of 1,141 cell counts have been identified covering a (subjectively) representative range of tables, population subgroups and geographical areas. The selected benchmark tables and cell counts are outlined in Table 2. To ensure representative geographical coverage, these benchmark counts were extracted for each of the 5,350 1991 Census Enumeration Districts (EDs) in a population-weighted 5% random sample of electoral wards that contained a minimum of 100 residents and 40 households. The minimum size requirement was imposed to ensure greater comparability to Output Areas (OAs), the lowest tier of UK 2001 Census outputs, whilst the geographically stratified sampling strategy was necessary to permit subsequent exploration of the effects of geographical aggregation upon user analyses.

Statistical agencies often point out that spatial units the size of Enumeration Districts are too small to be used reliably for analytical purposes. Rather, they argue, outputs for areas of this size are provided simply to furnish the building blocks for larger user-specified geographical units. Consequently, from the initial 5,350 EDs 1,000 sets of 16 spatially contiguous EDs were randomly selected (with replacement) and aggregated, creating benchmark counts for 1,000 "geographical clusters." A set size of 16 was chosen as this represents the median number of Output Areas in a 2001 Census ward; for at least half of all wards, users seeking to construct larger clusters would minimise the effects of cell adjustment by subtracting the counts of a cluster of less than 16 OAs from the relevant ward total.

Having created a representative set of benchmark cell counts, a variety of cell adjustment methods were used to create a number of directly equivalent post-adjustment

Table 2. Benchmark counts

SAS Table	Variables in table [number of categories in variable]	Table cells		
		Interior	Marginal	All
2	**Age** [21] × **Sex** [2] × **Marital Status** [2]	84	70	154
6a	**Ethnic group** [11] × Sex [2]	22	14	36
6b	Ethnic group [11] × Age [5]	55	17	72
8	**Economic position** [11] × Age [10] × Sex [2]	220	99	319
12	**Limiting illness** [1] × Age [7] × Sex [2]	14	10	24
20	**Tenure** [7] × **Amenities** [4]	28	28	56
21	**Cars in household** [4] × **Household structure** [5]	20	10	30
22	Tenure [4] × **Rooms** [7] × **Residents** [7]	196	124	320
34	Economic position [11] × Marital status [2] × Sex [2]	44	26	70
63	Tenure [7] × **Dwelling type** [4]	28	20	48
71	Resident/Household/Enumeration status [5] × Sampling Fraction [2]	10	2	12
Total	11 Tables / 13 **key** variables	721	420	1,141

datasets. Two variants were created with specific regard to the UK debate introduced in Section 1, one based on randomly rounding all counts to a neighbouring multiple of 3, $RR_{(3)}$, and one based upon small cell adjustment, $SCA_{(3)}$. For the sake of broader comparisons, additional variants were created using three additional cell adjustment methods. $RR_{(5)}$, the random rounding of all counts to a neighbouring multiple of 5, is based on the approach used by Statistics Canada. An alternative method of small cell adjustment, $SCA_{(3+)}$, has also been implemented. In this method interior *and* marginal cell counts of 1 or 2 are independently randomly rounded to either 0 or 3. (This contrasts with $SCA_{(3)}$ in which marginals are based on the sum of adjusted interior cell counts.) $SCA_{(3+)}$, although not officially implemented by any statistical agency, is close to the *de facto* situation in the UK after the 2001 Census, as an independently adjusted set of univariate counts have been published which may be substituted, albeit after some aggregation of categories, for existing table marginals. Finally, a set of adjusted cell counts, $B_{(0.2)}$, were created using Barnardisation. Each original nonzero cell count was modified with a probability of 0.2, with an equal chance of the resulting change in value being either $+1$ or -1. This approach provided the basis for cell adjustment in both the 1981 and 1991 UK Censuses (although the precise probability used to modify nonzero cell counts is a matter of some conjecture – there is some analytical evidence to suggest a lower probability was used, at least for those tables containing many counts). In all five variants each occurrence of a count that appeared in more than one table was independently perturbed.

The probabilities used by Statistics New Zealand when implementing $RR_{(3)}$ are in the public domain (Statistics New Zealand 2005) and follow the algorithm for implementing unbiased random rounding to base 5 set out in FCSM (1994) and more generally for any base by Hatzopoulos et al. (1997). In contrast, the probabilities used by ABS during cell adjustment are unknown. Indeed, were it not for an unguarded comment in a parliamentary submission (ABS 1997), even the fact that only counts of less than four are affected could only be surmised from inspection of published outputs. In line with ABS official practice, ONS have divulged neither the size of small cell counts subject to small cell adjustment, nor the rounding probabilities involved. However, inspection of 2001 Census outputs leaves no reason to doubt that, in line with ABS practice, only interior counts of 1 or 2 are rounded, and that the rounding probabilities adopted follow the pattern set out in Hatzopoulos et al. (1997). Similar deductions underpin the implementation of $B_{(0.2)}$ and $RR_{(5)}$. (See Williamson 2005a for details.)

4. Illustrative Benchmark Results

4.1. Counts

Before presenting a user-relevant empirical analysis of the observed difference between the selected benchmark cell counts and their post-adjustment equivalents, it is worth pausing momentarily to consider the *theoretical* confidence intervals associated with each of the outlined cell adjustment methods. Although no statistical agency appears to have placed these confidence intervals in the public domain, it is nevertheless possible to deduce formulae for their estimation assuming that (i) the details of the implementation of each approach are as outlined in Section 3 and (ii) as the number of potentially perturbed cells, n, increases, using the central limits theorem the distribution of errors approaches a normal distribution with mean 0 and standard error $s_n = (s^2 n)^{0.5}$. On this basis estimates of the relevant 68% (1 standard error) confidence intervals are as follows:

$$\mathrm{RR}_{(5)} \pm 2.0\sqrt{n-1} \tag{1}$$

$$\mathrm{RR}_{(3)} \pm 1.2\sqrt{n-1} \tag{2}$$

$$\mathrm{SCA}_{(3)} \pm 1.0\sqrt{n-1} \tag{3}$$

$$\mathrm{B}_{(0.2)} \pm 0.5\sqrt{n-1} \tag{4}$$

$$\mathrm{B}_{(0.04)} \pm 0.2\sqrt{n-1} \tag{5}$$

Comparison to confidence intervals derived by exact calculation shows that these formulae provide reasonable approximations, accurate to at least one decimal place for a wide range of confidence intervals and values of n (Williamson 2005a).

In the context of the UK choice between $\mathrm{RR}_{(3)}$ and $\mathrm{SCA}_{(3)}$ the story told by these formulae might appear unequivocal. For any given number of potentially perturbed cells the theoretical confidence interval associated with $\mathrm{RR}_{(3)}$ is 20% wider than that for $\mathrm{SCA}_{(3)}$. And this is without taking into account the fact that significantly fewer interior counts will be perturbed by $\mathrm{SCA}_{(3),}$ which leaves all interior counts larger than 3 unmodified. But do these theoretical estimates match the empirical reality? The true relative standings of $\mathrm{RR}_{(3)}$ and $\mathrm{SCA}_{(3)}$, both with respect to each other, and with respect to the other methods considered, depend crucially upon the frequency of preadjustment counts of 1 or 2, which vary from table to table, population to population and place to place. In addition, theoretical confidence intervals for interior cell counts have nothing to say about the relative effect of each cell adjustment method upon table marginals; nor upon user analyses involving rates constructed using post-adjustment cell counts.

Section 3 has already outlined the construction of a set of benchmark cell counts representative in terms of tabular, population and geographical coverage. Through comparison of the pre- and post-adjustment values of these benchmark counts derivation of empirically observed confidence intervals is possible. Figure 1 plots the distribution of these post-adjustment differences across all 1,000 benchmark geographical clusters (each cluster comprising the aggregated counts of 16 spatially contiguous EDs), with both the range and 95% "confidence interval" identified. Dividing the latter by the relevant number of cells in the benchmark dataset (interior; marginal or all) would yield the average (empirical) 95% confidence interval associated with adding together cell counts of that type from 16 geographically neighbouring EDs.

From Figure 1 a reassuring first observation is that, for all cell adjustment methods, the distribution of observed differences is essentially symmetrical around a mean of 0. This would, of course, be expected, and lends broad support to the assertion of statistical neutrality made by statistical agencies concerning their cell adjustment methodologies. In the context of the debate over $\mathrm{RR}_{(3)}$ versus $\mathrm{SCA}_{(3)}$, empirical confirmation is provided that, for interior cell counts, the New Zealand system of random rounding, $\mathrm{RR}_{(3)}$, creates greater post-adjustment disparities than $\mathrm{SCA}_{(3)}$, with the relative ordering of all cell

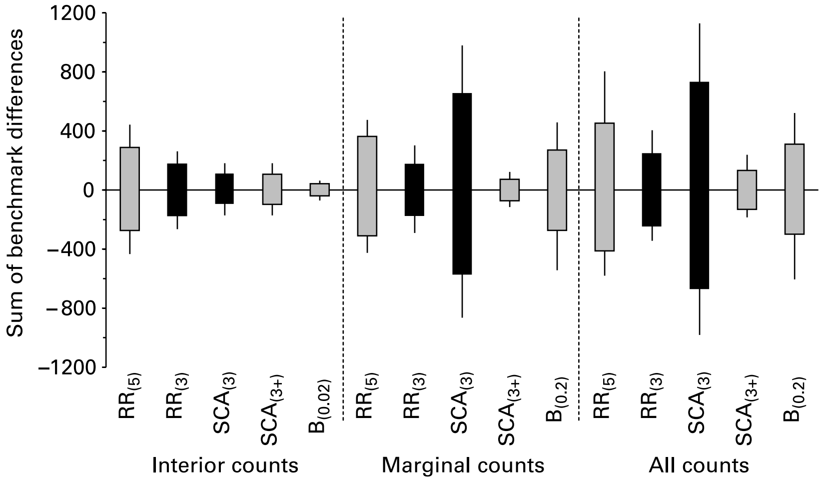

Fig. 1. Range and 95% confidence interval of the total difference between pre- and post-adjustment benchmark counts, as observed across 1,000 geographical clusters, each cluster comprising the aggregated counts of 16 spatially contiguous EDs

adjustment methods following those indicated by their theoretical confidence intervals. Indeed, as a result of $SCA_{(3)}$ leaving interior counts of more than 3 unperturbed, the relative performance gap actually widens from 20% to 50%. It is also worth noting that the SDC method used for the 1991 UK Census, $B_{(0.2)}$, out-performs both $RR_{(3)}$ and $SCA_{(3)}$ for interior counts (because the maximum change in an interior cell value is ± 1), whilst having a significantly larger negative effect upon marginal counts (because $B_{(0.2)}$ perturbs a larger proportion of interior cell counts). But the key message of Figure 1 is that the relative effect of each cell adjustment method varies significantly according to table cell type. For example, the advantage of $SCA_{(3)}$ over $RR_{(3)}$ for interior cell counts appears to be more than outweighed when the far larger negative effect of $SCA_{(3)}$ on table marginals is taken into account. Indeed, errors in marginal cell counts generated by $SCA_{(3)}$ are so large that they continue to drive the ranking of methods even when the effect across all cell types is considered, despite marginal cells comprising only 37% of all table counts.

4.2. Rates

The implication of count-based findings for rate-based analyses is unclear, given that the effect of a particular cell adjustment strategy will depend upon the size and source (interior or marginal) of the numerator and denominator from which the rate is constructed. It may be for precisely these reasons that no statistical agency has issued guidance to users concerning the potential effect of cell adjustment on rate-based analysis. However, following the logic of empirically benchmarking the effect of cell adjustment on counts, it is equally possible to benchmark the effect of cell adjustment on rates. Table 3 lists a set of 17 percentages calculated using numerators and denominators extracted from the benchmark counts listed in Table 2. Between them these percentages attempt to provide a representative cross-section of percentages covering both common and rare population subgroups, and deriving their numerators and denominators from a mixture of single and multiple

Table 3. Benchmark percentages

Benchmark percentage	No. of potentially modified cell counts involved in calculation of				Mean % across sample of 1,000 geographical clusters, each comprising 16 benchmark EDs (95% confidence interval)
	Numerator		Denominator		
	Cell adjustment method				
	$RR_{(5)}$ $RR_{(3)}$ $SCA_{(3+)}$	$SCA_{(3)}$ $B_{(0.2)}$ $B_{(0.04)}$	$RR_{(5)}$ $RR_{(3)}$ $SCA_{(3+)}$	$SCA_{(3)}$ $B_{(0.2)}$ $B_{(0.04)}$	
% Residents					
long-term ill	1	14	1	84	12.5 (7.5–18.3)
unemployed	1	4	1	40	5.9 (2.5–13.8)
employed full-time	1	4	1	40	37.7 (26.6–48.7)
aged 75–84	2	8	1	84	5.7 (2.4–11.4)
aged 85 +	2	8	1	84	1.3 (0.5–3.0)
female	1	42	1	84	51.5 (49.7–54.0)
non-white	2	18	1	20	7.0 (0.3–43.1)
black	3	6	1	20	2.1 (0.0–18.3)
Indian, Pakistani or Bangladeshi	3	6	1	20	3.6 (0.0–29.0)
Bangladeshi	1	2	1	20	0.8 (0.0–8.1)

Table 3. Continued

Benchmark percentage	No. of potentially modified cell counts involved in calculation of				Mean % across sample of 1,000 geographical clusters, each comprising 16 benchmark EDs (95% confidence interval)
	Numerator		Denominator		
	Cell adjustment method				
	$RR_{(5)}$ $RR_{(3)}$ $SCA_{(3+)}$	$SCA_{(3)}$ $B_{(0.2)}$ $B_{(0.04)}$	$RR_{(5)}$ $RR_{(3)}$ $SCA_{(3+)}$	$SCA_{(3)}$ $B_{(0.2)}$ $B_{(0.04)}$	
% Households					
owning/buying a private dwelling	2	8	1	28	68.7 (22.9–91.4)
renting public housing	2	8	1	28	22.3 (2.0–67.3)
with no central heating	2	14	1	28	19.9 (3.8–49.9)
with sole use of dwelling	1	21	1	28	79.7 (19.1–97.5)
sharing use of dwelling	1	7	1	28	0.3 (0.0–2.0)
with no access to a car or van	1	5	1	20	33.3 (12.9–65.5)
with access to 2 or more cars or vans	2	10	1	20	23.1 (5.1–44.0)

benchmark counts. For benchmark purposes these percentages were calculated, pre- and post-adjustment, for each of the 1,000 geographical clusters in the benchmark dataset.

Figures 2 and 3 present two statistical summary measures of the outcome: the average absolute difference in pre- and post-adjustment percentages (averaged across all 17 percentages and all 1,000 geographical clusters); and the average relative difference (i.e., the difference in pre- and post-adjustment percentages expressed as a proportion of the preadjustment percentage). The first columns in each of Figures 2 and 3 (labelled "G" to denote a geographically clustered sampling strategy) show the average effects of cell adjustment upon benchmark percentages. The resulting ordering of cell adjustment methods fails to conform to that observed for either interior or marginal cell counts, and remains almost unchanged whether considering average absolute or average relative differences in pre- and post-adjustment percentages. The one exception is $SCA_{(3+)}$, which produces the smallest average absolute error but is significantly eclipsed by $B_{(0.2)}$ when average relative errors are considered. The explanation for this is that $SCA_{(3+)}$ modifies only counts of 1 or 2, whether found in an interior or marginal cell. Consequently any observed changes to percentages will be very small in absolute terms. On the other hand, any changes in post-adjustment percentages imposed by $SCA_{(3+)}$, involving as they will the change of numerators and denominators of 1 or 2 to values of 0 or 3, are likely to be large in relative terms. In contrast $B_{(0.2)}$ has a slightly larger absolute effect on post-adjustment percentages, because the ± 1 changes applied to 20% of all interior cell counts lead on average to greater cumulative changes in numerators and denominators than $SCA_{(3+)}$, but a much smaller relative effect. This is accounted for by (i) average changes in numerators and denominators tending to cancel out and (ii) a much smaller proportion of small numerators (1 s and 2 s) being adjusted to 0.

Figures 2 and 3, therefore, tell a mixed story. This additional information may make the choice of a preferred method of cell adjustment less straightforward, but at least it shows

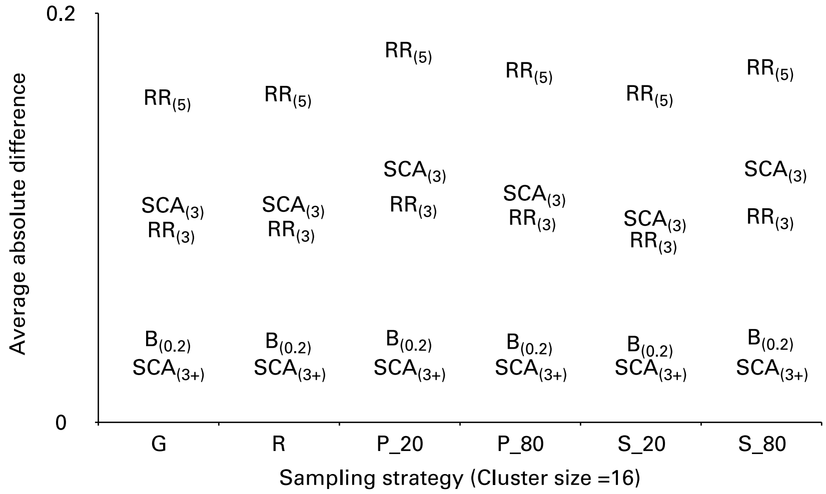

Fig. 2. The average absolute difference between a set of representative pre- and post-adjustment percentages (see text for details)

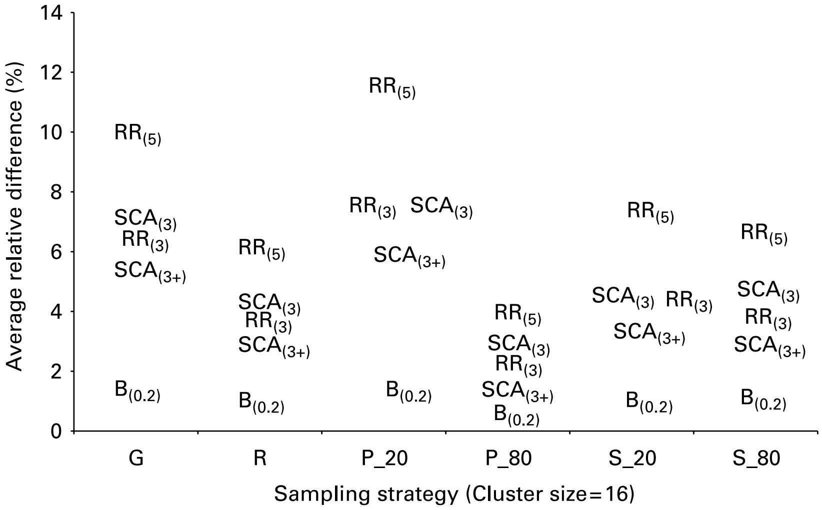

Fig. 3. The average difference between pre- and post-adjustment percentages as a proportion of the preadjustment percentage, for a representative set of benchmark percentages (see text for details)

that paying attention to one aspect of effect alone – such as the effect upon cell counts – can be misleading.

4.3. Ranking

In Section 2 existing measures of cell adjustment effect were criticised for failing to recognise the differential effect of cell adjustment across cell type (marginal or interior) and analysis type (count or rate based). They were also criticised for failing to provide direct measures of the actual effect of cell adjustment upon typical area-based user analyses such as ranking, correlation and regression. To address this final shortcoming, the 1,000 geographical clusters in the benchmark dataset were ranked in turn by each of the 17 benchmark percentages listed in Table 3. Figure 4 depicts the overall percentage of post-adjustment rankings that fell in the same fifth (quintile) and twentieth (vintile) as their pre-adjustment equivalents. Compared to the results concerning counts and percentages previously discussed, which are still arguably rather abstract from a user point of view, these results are of far more direct relevance, at least to some types of analysis, and in many ways rather more salutary.

As Figure 4 shows, even for the relatively unambitious target of ranking each 16 ED cluster within the correct quintile, the cell adjustment methods currently in use by national statistical agencies lead to misclassification rates ranging from 5% for $RR_{(3)}$ (the method adopted by Statistics New Zealand) through to 12% for $RR_{(5)}$ (the method adopted by Statistics Canada). In the UK many government policies are aimed at targeting resources to neighbourhood areas ranked in the top 5%. When ranking areas by correct vintile, no currently implemented cell adjustment method ($RR_{(5)}$, $RR_{(3)}$, $SCA_{(3)}$) achieves an accuracy rate of more than 77%. But to be set against this, areas in the top and bottom vintile are significantly more accurately classified than those in the mid-range (see Table 4). When ranking areas by lowest vintile $RR_{(5)}$, $RR_{(3)}$ and $SCA_{(3)}$ between them

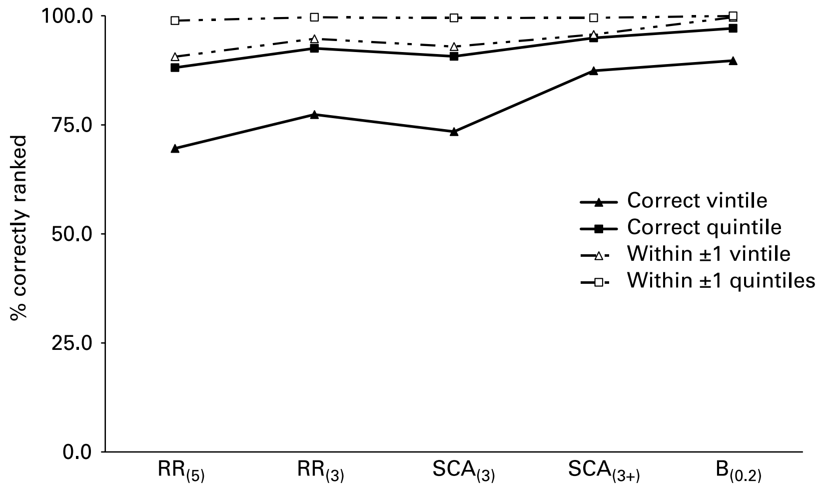

Fig. 4. The average accuracy of the post-adjustment ranking of 1,000 geographical clusters, each comprising the aggregated counts of 16 spatially contiguous EDs post cell adjustment, over each of 17 different representative percentages

yield a minimum benchmark misclassification rate of 6%. Even greater accuracy is achieved for the top vintile, a reflection of the positively skewed distribution of many percentages, which leads to greater differentiation between preadjustment cluster percentages at the top end. In all cases data ranking accuracy was better preserved using $B_{(0.2)}$ than any currently implemented method.

As for ranking, it is of course possible to provide benchmark measures of the effect of cell adjustment on correlation and regression analyses, or indeed upon any other type of user analysis that could be envisaged. However, as Williamson (2007) demonstrates, the rank ordering of adjustment methods essentially follows that reported here for rates whether considering correlation or regression analyses (both bivariate and multivariate). Consequently it can perhaps be reasonably claimed that results of the kind presented in Figures 1, 2, and 3 between them serve to capture the main strengths and weaknesses of each approach – namely their likely cumulative effects upon interior and marginal counts and their average absolute and relative effects upon rates. At the same time Figure 4 (in which the relative ordering of cell adjustment methods follows that in Figure 3) is perhaps best illustrates the general trend observed for most rate-based analyses, providing a measure of effect (ranking accuracy) arguably most directly instructive for user and producer of statistics alike.

4.4. *The Importance of Benchmarking*

Returning to the debate that provided the original stimulus for this article, from the results presented in Figures 1 to 4 it is apparent that the arguments for and against $SCA_{(3)}$ and $RR_{(3)}$ are far more finely balanced than was generally appreciated at the time. By far the most commonly expressed user preference was for adoption of $SCA_{(3)}$, in part on the grounds that fewer interior counts would be perturbed, and in part on the grounds that the resulting marginals would at least be consistent with their associated table interiors.

Table 4. Accuracy of ranking 1,000 geographically clustered aggregates of 16 EDs by vintile, based upon 17 representative percentages.

Vintile	% correctly ranked				
	Cell adjustment method				
	$RR_{(5)}$	$RR_{(3)}$	$SCA_{(3)}$	$SCA_{(3+)}$	$B_{(0.2)}$
(Lowest) 1	93	94	94	97	98
2	73	80	76	88	92
3	70	77	72	86	91
4	64	72	68	83	86
5	58	69	63	80	84
6	56	68	62	81	86
7	57	68	61	81	85
8	54	64	60	76	84
9	53	67	61	78	84
10	58	66	62	81	86
11	60	68	64	84	86
12	61	70	65	85	87
13	64	72	68	86	87
14	65	76	73	88	88
15	69	79	76	89	90
16	72	83	77	91	92
17	77	85	81	94	93
18	85	89	86	96	96
19	91	94	91	98	97
(Highest) 20	97	98	96	99	99
Overall	70	77	73	87	90

The greater adverse effects of $SCA_{(3)}$ on table marginals were generally under-estimated, whilst the possible effects upon rate-based analyses were simply ignored. In fact, as shown in Figures 1 to 4, the implementation of $SCA_{(3)}$ leads to significantly greater uncertainty over the true value of marginal counts and to marginally but measurably less accurate rate-based analyses than $RR_{(3)}$. Which of the above most affects overall data utility ultimately remains a matter of opinion. But the lack of understanding at the time concerning the effect of either strategy upon rate-based analyses clearly represented a serious lacuna in the evaluation process, both inside and outside of ONS, as did lack of access to information concerning the average effect of either approach upon a comparable and representative set of benchmark counts. This points to the importance of adopting a far more rigorous evaluation process when considering any future proposed disclosure control methodology.

Another question that needs to be addressed by any evaluation process is the fitness-for-purpose of the resulting post-adjustment outputs. From this perspective the results presented in Figures 1–4 might be regarded as either reassuring or alarming. Are the effects of cell adjustment demonstrably "negligible" as claimed by, for example, Statistics Canada? The generally low relative effects upon counts and percentages would appear to broadly support this contention. Or are the effects sufficiently damaging to the integrity of the underlying data that they potentially call into question the value of any analyses based upon them? Empirical evidence of the effect of cell adjustment on area rankings perhaps lends some credence to this second viewpoint. But the aim of this article is not to provide the definitive answer to these questions – an answer which in any case will vary from user to user depending upon individual analytical goals. Rather the purpose is to show that any such answer will necessarily lack cogency when, as at present, uninformed by consideration of the overall effect of cell adjustment on a truly representative set of cell counts, or of the consequent "indirect" effects of cell adjustment upon rate-based analyses.

If the case for a more rigorous approach to the evaluation of cell adjustment methods is accepted, two main obstacles remain. One is the objection that overall summary results of the kind presented in Figures 1 to 4 ignore the fact that the effects of cell adjustment are indeed table-, population- and place-specific. This objection is addressed in the following section. The other is that conducting such a thorough and wide-ranging analysis requires too many resources. The final parts of Section 6 outline ways in which such resource demands can be kept to a minimum through the use of "off-the-peg" benchmark datasets and benchmarking software.

5. Table-, Population- and Place-specific Differences

5.1. The Effect of Cluster Size

Before addressing the table-, population- and place-specific nature of cell adjustment effects, the influence of cluster size needs to be briefly considered. It is perhaps self-evident that, all other things being equal, larger absolute effects of cell adjustment can be expected as cluster size increases. However, as the theoretical confidence intervals presented in Section 2 indicate, this rate of increase is nonlinear, the expectation being that the relative effect of cell adjustment will actually decrease with increasing cluster size. Williamson (2007) provides empirical confirmation of these expectations for the behaviour of interior cell counts, and goes on to show that this behaviour applies more generally. Of particular relevance to the discussion that follows, regardless of the measure of effect considered (count-, rate- or rank-based), the relative ordering of cell adjustment methods remains unaffected by cluster size.

5.2. Table-specific Effects

Table 5 helps to illustrate the table-specific nature of cell adjustment. The right hand side of the table presents one simple measure of cell adjustment effect – the proportion of all post-adjustment table cell counts (interior and marginal) that differ in value from their preadjustment counterparts. Once again the measure presented is the average observed across 1,000 benchmark clusters, each cluster comprising 16 geographically neighbouring EDs. The final row of Table 5 presents the average % of cell counts modified across all eleven benchmark tables. Perhaps surprisingly, given the results in Section 4, the overall ranking of the cell adjustment methods by this new measure follows that indicated by their theoretical confidence intervals. However, examination of the table-specific results presented in Table 5 reveals that this is not always the case. In particular the relative rankings of $RR_{(3)}$ and $SCA_{(3)}$ are prone to reversal, as are the relative rankings of $SCA_{(3+)}$ and $B_{(0.2)}$. A second feature of note is that in some tables a far higher proportion of cell

Table 5. Table-specific variations in the proportion of cells modified through cell adjustment, as observed for a set of representative geographical areas

	Cluster size = 1				Cluster size = 16				
	Interior cell counts				All cell counts (interior and marginal)				
	% of cells in table	% containing count of			Cell adjustment method				
					$RR_{(5)}$	$RR_{(3)}$	$SCA_{(3)}$	$SCA_{(3+)}$	$B_{(0.2)}$
SAS Table		0	1 or 2	3 +	% of counts changed by cell adjustment				
2	55	26	17	58	88	85	94	46	63
6a	61	61	15	25	84	78	38	57	42
6b	76	74	10	16	70	65	57	48	32
8	69	65	18	18	76	73	70	53	45
12	58	17	28	55	95	91	84	65	68
20	50	63	13	24	78	74	69	56	43
21	67	29	17	54	91	87	69	44	62
22	61	80	10	10	63	59	59	51	11
34	63	28	21	51	93	89	73	45	65
63	58	67	12	21	80	76	70	55	46
71	83	39	7	54	77	73	33	21	48
All tables	63	60	15	25	76	72	66	51	46

counts are modified, on average, than in others. In all cases the proportion of cells modified is very high – in part the result of summing the modified counts from geographically neighbouring EDs to create larger aggregates.

These findings are explained at least in part by reference to the left-hand side of Table 5. This reveals the distribution of cell count values amongst the EDs from which the benchmark clusters are constructed. Two examples suffice to illustrate the general principles. Of the benchmark tables, SAS Table 12 contains by far the highest proportion of preadjustment cell counts larger than 0. In consequence it experiences one of the highest rates of change in post-adjustment cell counts, regardless of which cell adjustment method is considered. (All methods leave preadjustment counts of 0 unmodified.) In contrast SAS Table 71 contains the lowest ratio of interior cell counts of 1 or 2 to counts of 3 + . As SCA only modifies initial counts of 1 or 2 the result is that a low percentage of the cell counts in SAS Table 71 are modified by $SCA_{(3)}$ and $SCA_{(3+)}$, even though the proportion of cells in this table modified by other cell adjustment approaches lies close to the overall benchmark average.

5.3. Population-specific Effects

Figure 5 illustrates the effect of cell adjustment on four percentages, representing a cross-section of common and rare population subgroups. For each percentage the ranges within which 95% and 100% of the post-adjustment percentages fall are shown. The result is visual confirmation that the effect of cell adjustment varies according to the population of interest. This variation expresses itself in three distinct ways. First, the relative ordering of each cell adjustment method varies slightly across subgroups in terms of effect, with the ordering of those methods shown to be close to each other in Figure 2 swapping places on occasion. For example, the relative rankings of $RR_{(3)}$ and $SCA_{(3)}$ are reversed for % social renting, and are almost indistinguishable for % owner-occupying. The relative rankings of $SCA_{(3+)}$ and $B_{(0.2)}$ with respect to each other vary even more markedly across population subgroups. Second, for some percentages, such as % nonwhite, $SCA_{(3+)}$ has no effect at all upon post-adjustment percentages. The reason for this is that all of the counts involved in the relevant numerators and denominators have values of three or more. Third, although the absolute effect of cell adjustment increases with subgroup rarity, the relative effect decreases. (See the range of differences relative to subgroup means given in Figures 5a–d.) But even for the highest effect approach, $RR_{(5)}$, for the smallest subgroup percentage (% 85 +), the post-adjustment percentage differs by no more than +/− 15% of the original preadjustment value in 95% of the 1,000 benchmark geographical clusters considered.

The reasons for these variations in the effect of cell adjustment between population subgroups are many and varied. In part they relate to those already outlined when considering variations in the % of cells modified across benchmark tables. But they also relate to, amongst other things, the number of post-adjustment counts involved in constructing the numerators and denominators (see Table 3) as well as to their average count size. All things being equal, the more post-adjustment counts a percentage is constructed from, and the rarer the population subgroup involved, the larger the relative effect of cell adjustment is likely to be.

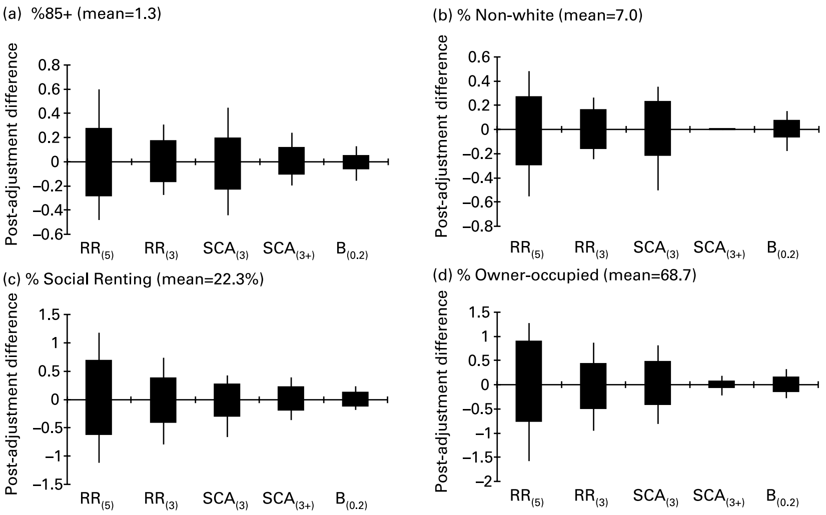

Fig. 5. Range and 95% confidence intervals of post-adjustment difference for four selected percentages, as observed across 1,000 zones, each based on the aggregated counts from 16 spatially contiguous Enumeration Districts

5.4. Place-specific Effects

For consideration of the place-specific effect of cell adjustment, it is necessary to return to consideration of Figures 2 and 3. So far only the results for geographically clustered aggregates (columns labelled G) have been discussed. To illustrate the importance of place, aggregate spatial units were also created using a variety of alternative sampling strategies. First, sets of 16 EDs were assembled through simple random sampling (with replacement) of all 5,350 benchmark EDs (R). Second, stratified samples were taken. After ranking all preadjustment EDs by population density, 1,000 sets of 16 random EDs were drawn from (i) the top fifth of the distribution (P_80) and (ii) the bottom fifth (P_20). By a similar process a set of aggregate benchmark counts were drawn stratified by social class (% persons in UK Registrar General's Social Classes I & II) (S_80 and S_20). The use of alternative sampling strategies is necessary because the stochastic nature of most cell adjustment processes precludes comparison of the outcomes for a selected handful of "representative" geographical aggregates.

As Figures 2 and 3 show, the type of sampling strategy adopted had only a minor effect on the average absolute difference between pre- and post-adjustment percentages, but a significant effect upon average relative difference. For the latter random sampling (R) led to significant under-estimation of the relative effect of all cell adjustment methods. The reason for this is that in a geographically clustered sample (G) a significant number of clusters will be composed entirely of rural EDs. As rural EDs contain lower populations, their average table cell count is lower, leading to larger potential relative effect of cell adjustment upon the numerators and denominators underlying the benchmark rates. Conversely, in a random sample most aggregates of EDs will comprise a representative mix of EDs, urban and rural, resulting in a higher average population density (and hence cell count). For the same reasons the observed effect of cell adjustment upon relative differences in pre- and post-adjustment percentages is highest in low population density areas (P_20) and lowest in high density areas (P_80).

5.5. Implications for Overall Summary Results

Throughout this section the variability of cell adjustment effects has been underlined, emphasising their table-, population- and place-specific nature. Given this variability, are "overall" measures of the kind presented in Section 4 meaningless? On the one hand the answer would appear to be "yes." The precise effects of cell adjustment are highly sensitive to the specific table, population or place being considered. Overall summary measures necessarily cannot highlight the specific counts and rates that, post-adjustment, are most and least reliable. On the other hand, given the wide range of cellular outputs from most censuses, and the far larger possible range of potential rates that could be derived from these counts, attempts to provide such information without reduction into some kind of summary form would almost inevitably prove overwhelming to both data user and provider alike. At the same time the detailed results presented in this section actually provide further confirmation of those summary results already presented in Section 4. The ranking of cell adjustment methods for the majority of tables and rates follows the ranking indicated by the summary measures. And where there is deviation from this ordering, the difference is invariably a simple switch in the ordering of two methods already indicated by the summary measure as being very close in terms of effect. In short, summary measures do appear to provide a robust overall indicator of the strengths, weaknesses and relative standings of the cell adjustment methods considered, provided that they are based upon a wide enough range of representative cell counts to begin with.

6. Conclusion

6.1. The Need for a Benchmark-based Approach

As the political pressures for increased levels of disclosure control continue to mount, the increasing work being undertaken on assessing disclosure risks needs to be matched by more careful scrutiny of the potential effects on the fitness-for-purpose of outputs. The main argument of this article is that current assessments of the effects of disclosure control lack rigour and are too far removed from the reality of effects upon everyday user analyses. None take proper account of the effect of disclosure control upon a genuinely representative set of tables, populations and places; all fail to consider the subsequent effects upon rate-based analyses. At the same time the summary measures used are almost invariably too abstract to offer any real guidance on the resulting fitness-of-purpose of post-adjustment outputs. Using a case study of the choice set before UK users between adoption of an Australian or a New Zealand cell adjustment strategy, this article has attempted to show how these various problems can be addressed. The proposed solution has three main components. First, the creation of a set of benchmark counts to which alternative methods of data perturbation can be applied, allowing direct comparison of effects upon a common set of data with representative population and geographical coverage. Second, consideration of the effect of cell adjustment upon not only counts (interior and marginal) but also rates, allowing variations in effect by data type to be recognised. And third, the use of statistical measures which provide end users with a clear indication of the likely effect of cell adjustment upon their own analyses, including not only the average effects across a set of representative counts/rates, but comparison of absolute and relative effects, indicative confidence intervals/ranges and the post-adjustment accuracy of area rankings.

If the need for a benchmark-based approach of the type outlined above is granted, the resource inadequacies remain to be tackled: lack of appropriate benchmark data and lack of the associated computational overheads. Before drawing to a close this article briefly addresses each in turn.

6.2. *The Provision of a Benchmark Dataset*

Statistical agencies are, of course, best placed to undertake evaluation of their own proposed alternative cell adjustment strategies, as they have full access to the original data to which the strategies will be applied, including, if required, the underlying microdata. For these agencies the extraction of a representative in-house set of pre- and post-adjustment benchmark counts should be a relatively straightforward task. This applies even if the final post-adjustment cell counts are the result of a combination of microdata editing, imputation and record swapping followed by post-tabulation random rounding or suppression. As well as providing a platform for more considered evaluation of alternative disclosure control strategies, the construction of such an in-house benchmark dataset could permit wider public dissemination of any evaluation findings, the empirical derivation of the findings providing protection for the precise details of the method(s) being evaluated. In other cases use of a public domain benchmark might still be more appropriate: for example where the goal is direct comparison with the effect of the overall disclosure control regime implemented by another statistical agency; or where the goal is, as was the case in the UK in 2002, consultation with users over the choice between alternative versions of one element of a broader package of disclosure control measures.

Outside of the statistical agencies the need for a public domain benchmark dataset is even more pressing. The current literature is replete with studies which propose alternative methods of cell suppression, controlled rounding and so on, each evaluated against a different, limited and (as far as can be judged) unrepresentative set of tables. To foster greater comparability it would be helpful if all of these competing methods were tested instead against a standardised benchmark dataset. Ideally this public domain benchmark would include the underlying microdata upon which the final tabulations were based, in order to allow for evaluation of additional micro-level strategies such as record swapping. Unfortunately, all presently available public domain microdata lack sufficient geographical detail to make provision of such a benchmark feasible. Nor is it likely that any statistical agency will feel able to release such a dataset within the foreseeable future. One possible alternative is the creation of a plausible but synthetic set of spatially detailed population microdata (cf. Voas and Williamson 2000). Another is an approach based upon multiple imputation (cf. Abowd and Woodcock 2001; Reiter 2005).

In the meantime ambition must be limited to the provision of a set of public domain benchmark counts. Even here the potential sources are severely restricted. One possibility put forward for consideration is the benchmark dataset of "preadjustment" tabular counts outlined in Section 3, which is freely available to all, subject to use of an acknowledgement similar to that to be found at the beginning of this article. (A copy of the benchmark dataset and a full dataset description is available via http://pcwww.liv.ac.uk/~william/SDC.) These electronically downloadable benchmark counts are available for sets of single EDs, as well as for preaggregated clusters of 16 and 46 spatially contiguous EDs. Each set contains benchmark counts for 1,000 such spatial units. Available separately are benchmark counts for the 5,350 EDs from which these various geographical clusters have been drawn, allowing the creation of additional sets of benchmark counts using alternative cluster sizes and sampling strategies (random or stratified). Users accessing these online benchmark data should bear in mind that they are based upon Barnardised census counts, although the effect of Barnardisation upon the overall distribution of cell counts is believed to have been minimal.

6.3 The Provision of Benchmarking Software

A second potential barrier to the benchmarking and evaluation strategy proposed in this article is the computational overhead involved. Summarising measures of fit across a series of tables, calculating a set of representative pre- and post-adjustment percentages, assessing the effects of adjustment upon area rankings and so on, whilst arguably desirable, all have nontrivial computational overheads if not provided as automated functions within a statistical analysis package. Indeed, the analyses presented in this article would not have been possible without a specially written program suite. To foster wider adoption of a user-centred perspective on the evaluation of statistical disclosure control effects, this software suite, *SDC-i*, together with its accompanying user documentation, has been placed in the public domain (http://pcwww.liv.ac.uk/~william/SDC). As written the suite can be used to compare any two sets of counts (pre- and post-adjustment) and produce a wide range of summary measures, including all of those presented in this article and the majority of those listed in Table 1 (see Table 1 for details). These include the range and 95% confidence intervals of both absolute and relative differences in counts and rates, the assessment of the overall accuracy of rate-based area rankings and summary measures averaged across both single and multiple benchmarks. Simple control parameters allow users to specify the number and layout of their chosen input tables, which can range from simple vectors of interior cell counts through to complex multi-dimensional tables with multiple table marginals and submarginals. Other user-adjustable parameters control the detail of summary output produced, ranging in scope from the total difference observed across all input tables through to measures of difference associated with individual table cell counts. Given a set of user-supplied benchmark counts, other elements of the program suite allow for the creation of geographical, random or stratified clusters of any size, and for the calculation of user-specified percentages. Full details may be found in Williamson (2005b). If the more rigorous approach to disclosure control evaluation proposed in this article finds favour, it is hoped that this software suite might provide at least the springboard for the development of a more professional software tool of the kind well exemplified by ARGUS with regard to the assessment of disclosure control risk (Hundepool et al. 2004; Hundepool et al. 2005).

7. References

Abowd, J.M. and Woodcock, S. (2001). Disclosure Limitation in Longitudinal Linked Data. In Confidentiality, Disclosure and Data Access: Theory and Practical Application for Statistical Agencies, P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.M. Zayatz (eds). Amsterdam: Elsevier, 215–277.

Australian Bureau of Statistics (1997). Information Paper: House of Representative Standing Committee on Legal and Constitutional Affairs Inquiry into the Treatment of Census Forms. Submission from the Australian Bureau of Statistics, 18 July. Canberra: Australian Bureau of Statistics.

Boyd, M. and Vickers, P. (1999). Record Swapping – A Possible Disclosure Control Approach for the 2001 UK Census. Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, Greece. Luxembourg: European Commission.

Chambers, R. (2001). Evaluation Criteria for Statistical Editing and Imputation. National Statistics Methodological Series, 28. London: Office for National Statistics.

Cox, L.H. (2001). Disclosure Risk for Tabular Economic Data. In Confidentiality, Disclosure and Data Access: Theory and Practical Application for Statistical Agencies, P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.M. Zayatz (eds). Amsterdam: Elsevier, 167–184.

Cox, L.H. and Kelly, J.P. (2003). Balancing Data Quality and Confidentiality for Tabular Data. Working Paper No. 4, Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Luxemburg. Luxemburg: European Commission.

Dandekar, R.A. (2004). Maximum Utility-minimum Information Loss Table Server Design for Statistical Disclosure Control of Tabular Data. In Privacy in Statistical Databases, J. Domingo-Ferrer (ed.). Berlin: Springer, 121–135.

Domingo-Ferrer, J. and Torra, V. (2001). A Quantitative Comparison of Disclosure Control Methods for Microdata. In Confidentiality, Disclosure and Data Access: Theory and Practical Application for Statistical Agencies, P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.M. Zayatz (eds). Amsterdam: Elsevier, 111–133.

Duncan, G.T., Fienberg, S.E., Krishnan, R., Padman, R., and Roehrig, S.E. (2001). Disclosure Limitation Methods and Information Loss for Tabular Data. In Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.M. Zayatz (eds). Amsterdam: Elsevier, 135–166.

Elliott, M. (2001). Disclosure Risk Assessment. In Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.M. Zayatz (eds). Amsterdam: Elsevier, 75–90.

Eurostat (1996). Manual on Disclosure Control Methods. Luxemburg: European Commission.

Federal Committee on Statistical Methodology (1994). Report on Statistical Disclosure Limitation Methodology (Statistical Policy Working Paper 22). Washington, D.C.: U.S. Office of Management and Budget.

Fienberg, S.E. and McIntyre, J. (2005). Data Swapping: Variations on a Theme by Dalenius and Reiss. Journal of Official Statistics, 21, 309–323.

Fienberg, S.E., Makov, U.E., and Steele, R.J. (1998). Disclosure Limitation Using Perturbation and Related Methods for Categorical Data. Journal of Official Statistics, 14, 485–502.

Fotheringham, A.S. and Knudsen, D.C. (1987). Goodness of Fit Statistics. Norwich: GeoBooks.

Giessing, S. (2003). Coordination of Cell Suppressions: Strategies for Use of GHMITER. Working Paper 36, Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Luxemburg. Luxemburg: European Commission.

Gomatam, S., Karr, A.F., and Sanil, A. (2003). A Risk-utility Framework for Categorical Data Swapping. Technical Report 132, Durham, NC: National Institute of Statistical Sciences. <http://www.niss.org/technicalreports/tr132.pdf> [date accessed: 29/6/2005].

Gouweleeuw, J., Kooiman, P., Willenborg, L., and De Wolf, P.-P. (1998). The Post Randomisation Method for Protecting Microdata. Qüestiió 22, 145–156.

Griffin, R.A., Navarro, A., and Flores-Baez, L.(1989). Disclosure Avoidance for the 1990 Census. Proceedings of the American Statistical Association, Section on Survey Research Methods, Alexandria, VA, 516–521.

Hatzopoulos, M., Karali, I., Viglas, E., and Murtagh, F. (1997). Specification of Secure Interface Module. Deliverable 3.1, Access to Distributed Databases for Statistical Information and Analysis, Edinburgh: Centre for Educational Sociology, University of Edinburgh. <http://homepages.ed.ac.uk/addsia/deliverables/Del31.doc> [date accessed: 29/6/2005].

Hundepool, A., van de Wetering, A., Ramaswamy, R., Franconi, F., Capobianchi, A., de Wolf, P-P., Domingo-Ferrer, J., Torra, V., Brand, R., and Giessing, S. (2004). μ-ARGUS Version 4.0 Software and User's Manual. Voorburg: Statistics Netherlands.

Hundepool, A., van de Wetering, A., Ramaswamy, R., de Wolf, P-P., Giessing, S., Fischetti, M., Salazar, J-J., Castro, J., and Lowthian, P. (2005). τ-ARGUS Version 3.1 Software and User's Manual. Voorburg: Statistics Netherlands.

Kirkendall, N. and Sande, G. (1998). Comparison of Systems Implementing Automated Cell Suppression for Economic Statistics. Journal of Official Statistics, 14, 513–535.

Office for National Statistics (2002). 2001 Census Disclosure Control in England and Wales. Advisory Group Paper AG(02)03. London: Office for National Statistics.

Office for National Statistics (undated). Census 2001 Data Quality: Variability in Tabular Counts. London: Office for National Statistics. <http://www.statistics.gov.uk/census2001/pdfs/tabular_counts.pdf> [date accessed: 16 June 2005].

Reiter, J.P. (2005). Releasing Multiply Imputed, Synthetic Public Use Microdata: An Illustration and Empirical Study. Journal of the Royal Statistical Society, Series A, 168, 185–205.

Salazar-González, J.J. (2005). Protecting Tables with Cell Perturbation. Working Paper No. 34, Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Geneva. Luxemburg: European Commission.

Shlomo, N. (2005). Assessment of Statistical Disclosure Control Methods for the 2001 UK Census. Working Paper 19, Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Geneva. Luxemburg: European Commission.

Statistics Canada (2003). 2001 Census Handbook. Ottawa: Statistics Canada.

Statistics New Zealand (2005). Standards. In New Zealand: An Urban/rural Profile. Wellington: Statistics New Zealand. <http://www.stats.govt.nz/urban-rural-profiles/standards/> [date accessed: 9/6/2005].

Voas, D. and Williamson, P. (2000). An Evaluation of the Combinatorial Optimisation Approach to the Creation of Synthetic Microdata. International Journal of Population Geography, 6, 349–366.

Voas, D. and Williamson, P. (2001). Evaluating Goodness-of-Fit Measures for Synthetic Microdata. Geographical and Environmental Modelling, 5, 177–200.

Williamson, P. (2005a). Estimating Cell Adjustment Confidence Intervals. Working Paper 2005/1. Liverpool: Population Microdata Unit, Department of Geography, University of Liverpool. Available online: http://pcwww.liv.ac.uk/~william/SDC/Estimating Cell Adjustment Confidence Intervals.pdf.

Williamson, P. (2005b). SDC-i Software Manual v2.0. Working Paper 2005/2, Liverpool: Population Microdata Unit, Department of Geography, University of Liverpool. Available online: http://pcwww.liv.ac.uk/~william/SDC/SDC-i Software Manual v2.0.pdf.

Williamson, P. (2007). The Impact of Cell Adjustment on the Analysis of Aggregate Census Data. Environment and Planning A, 39, 1058–1078.

Received June 2005
Revised September 2006