HOW CENSUS DATA AFFECTS STATE BUDGETS

by Glenn Pure, Stephen Tregea-Collett and Catherine Hull

Commonwealth Grants Commission

SUMMARY

1 The Commonwealth Grants Commission uses disaggregated Census data to calculate notional costs of service delivery in States due to their population characteristics. These cost calculations are updated each year and used by the Commission to develop relativities. The relativities are used by the Australian Government to determine each State's share of a pool of funds currently amounting to about $45 billion (representing almost half of States' revenues). The funds are used by States to provide services such as education, health and law and order.

2 We have calculated what would happen to State shares of the pool of funds if we did not have access to disaggregated Census data. This would effectively prevent us from doing most of the work we do in calculating cost differences between States, and result in a redistribution of almost $2.5 billion among the States. A redistribution of this size would mean that not all Australians would have the opportunity to access services of the same standard.

3 Given the large amount of money that can be redistributed based on disaggregated Census data, we have closely examined possible problems with Census data. The problems include Indigenous under-enumeration, randomisation of small cell counts in highly disaggregated Census tables and errors in estimated resident population. We also outline planned future directions of our work with Census data.

WHAT THE COMMONWEALTH GRANTS COMMISSION DOES

4 The Commonwealth Grants Commission is an Australian Government statutory authority operating under the *Commonwealth Grants Commission Act*.

5 The Commission's main role is to assess the financial support each State should receive from the Australian Government. The current arrangements for payments to the States reflect the

provisions of the *Intergovernmental Agreement on the Reform of Commonwealth—State Financial Relations* (IGA), signed by the Australian and State Governments in 1999 shortly before the goods and services tax (GST) was introduced. The IGA included the following provisions:

- all revenue from the GST, less the costs of collection, is paid to the States to be spent according to their budget priorities;
- the funds are to be distributed among the States on the basis of horizontal equalisation principles; and
- the Commission is to calculate per capita relativities for this purpose.

6 To achieve its mission, each year the Commission calculates the 'relativities' for distributing the pool of GST revenue and Health Care Grants[1] (HCGs). The relativities are used by the Australian Treasury to calculate the shares of the pool of funds that each State and Territory will receive.

7 The funds distributed are important for State budgets. In 2005-06, the GST pool is estimated to be $44.6 billion[2], In 2004-05, the GST and HCG pool was 71 per cent of the funds provided to the States by the Australian Government, and on average provided 42 per cent of State budget revenues. The actual percentage of revenue for individual States varied from 37 per cent for New South Wales to 75 per cent for the Northern Territory.

8 Thus the distribution of the funds has a major impact on the capacity of State governments to deliver comparable services, including schools, hospitals, police services and housing, and also to keep taxes and charges at comparable levels. In calculating the distribution of the GST and HCG pool, the Commission recognises that it costs more to provide services to the same level in some jurisdictions than in others. The pool of funds is divided to compensate for those differences.

HOW THE COMMISSION USES CENSUS DATA

9 In making its funding recommendations, the Commission aims to achieve horizontal fiscal equalisation. In essence, this means that each State should be given funds so that it can provide services to its population to the same level and standard, on average, as all other States[3]. Both revenue raising capacity and expense requirements are considered but it is only the expense assessments that are relevant to this paper.

---

1 The total pool of funds to be distributed includes HCGs, although the size of each State's HCG is not determined by the Commission. See the Commission's latest publications at www.cgc.gov.au for more information.

2 Commonwealth of Australia Final Budget Outcome 2004-05, Table 31, p59.

3 Where and how a State uses the funds given to it is solely a matter for that State. The Commission aims to provide each State with the same capacity only, based on an average of all States.

10 Some population groups use services more or less than others, and a unit of service for some groups may cost more to provide. As a result, a State's costs will differ based on characteristics of its population. The Commission relies heavily on disaggregated Census data to estimate the relevant service populations in each State.

11 State expenses will also vary due to policy choices made by each State about what services will be provided, to what level and to whom. However, the Commission attempts to remove all such policy influences by determining the average policy of all States then utilising demographic and other characteristics that are beyond the direct control of each State to notionally calculate expenses.

12 The Commission uses Census population data dissected by such characteristics as age, gender, Indigeneity, level of income, level of English fluency and region. Table 1 provides a summary of the State services in which cross-tabulated Census data are used in our assessments, and the variables by which they are disaggregated.

Table 1 Census 2001 cross-tabulations used by the Commission

Services	Variables used in cross-classified table
Hospital inpatient services, family and child services, homeless and general welfare	Usual residents by State (URS), SARIA[(a)] , sex, age, income, English fluency, Indigeneity, country of birth
Government schools education – primary Government schools education – secondary Non-government schools education — primary Non-government schools education — secondary	URS, SARIA, major cities, education institution, income, English fluency, Indigeneity (persons aged 5-18 only)
Vocational education and training	URS, SARIA, age, English fluency, employment status, Indigeneity (persons aged 15-59 only)
Housing and housing user charges Aged and disabled services	URS, SARIA, age, Indigeneity, income, household type (persons aged 60 and over only)

(a) The Stated-based version of the Accessibility/Remoteness Index of Australia developed for Commission purposes.

13 These large cross-tabulations of Census data are used to construct socio-demographic composition (SDC) factors. These factors measure the impact on State costs of differences in the characteristics of State populations. The Commission uses Census data, together with use and cost data disaggregated in the same way as the Census data, to calculate notional service populations for each State and from these the SDC factors. The SDC factors are then combined with average Australian expense data to determine a notional 'assessed expense'. This assessed expense is the amount it would cost a State to deliver the service at Australian average cost and efficiency levels.

Example

The 0 to 5 age group in the population uses service X at twice the rate of the rest of the population. For simplification, it is assumed the cost per unit of service is the same for all age groups. To calculate the SDC factor, Census data disaggregated by age would be used to calculate the proportion of the 0 to 5 age group in each State's population. The same would be done for the Australian population as a whole. The 0 to 5 age proportions would then be multiplied by two and added back to the respective proportions of the population aged over 5. This would give a raw SDC factor value. Those raw SDC

factors for each State would then be normalised by dividing them by the raw SDC factor for Australia. This is how it would look, using dummy data:

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Population (m)	6.0	5.0	4.0	3.0	2.0	0.5	0.3	0.2	21.0
A. 0 to 5 year proportion	0.10	0.11	0.12	0.13	0.09	0.08	0.07	0.15	0.11
B. 6+ years proportion	0.90	0.89	0.88	0.87	0.91	0.92	0.93	0.85	0.89
C. Weight 0 to 5 years by 2 (A x 2)	0.20	0.22	0.24	0.26	0.18	0.16	0.14	0.30	0.22
D. Raw SDC factor (B + C)	1.10	1.11	1.12	1.13	1.09	1.08	1.07	1.15	1.11
E. Normalised SDC factor (D_{State} / D_{Aust})	0.99	1.00	1.01	1.02	0.98	0.97	0.96	1.04	1.00

To calculate the per capita assessed expense for service X for each State, the normalised SDC factor for each State is multiplied by the total actual Australian expense for that service, then divided by the Australian population.

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
A. Population (m)	6.0	5.0	4.0	3.0	2.0	0.5	0.3	0.2	21.0
B. Normalised SDC factor (D_{state} / $D_{australia}$)	0.99	1.00	1.01	1.02	0.98	0.97	0.96	1.04	1.00
C. Total expense for service X ($m)									500
D. Assessed per capita expense (B x C_{Aust} / A_{Aust})	$23.61	$23.83	$24.04	$24.26	$23.40	$23.18	$22.97	$24.69	$23.81
E. Total assessed expense for service X in each State (A x D) $m	141.7	119.1	96.2	72.8	46.8	11.6	6.9	4.9	500.0

14 The above example is a simplification. Many SDC factors are constructed using large cross-tabulations of Census data so that populations can be divided into many different sub-populations, each with a separate use weight and/or cost weight. These are combined to produce an overall weighted population and SDC factor. Another potential approach would be to use broader socio-demographic characteristics (for example age), assess factors for these (based on differences in use and cost weight) and combine them into an overall SDC factor. While data may be available for some socio-demographic characteristics to do this, they usually do not reflect how different cost and use weights interact. For example, if we know that Indigeneity results in a two fold increase in service use and location in a remote area results in a 1.5 fold increase in the same service, what is the usage factor for an Indigenous person living in a remote community? It may be 2 x 1.5, it may be 2 + 1.5, or it may be some other value. Identifying many sub-groups where we have corresponding cost and use weights correctly identifies the interactions and overcomes this problem.

What would the funding impact be without disaggregated Census data?

15 The absence of disaggregated Census data would mean that the Commission could not correctly assess differences between States due to the composition of their populations. In a worst case situation, we would be forced to assess expenses in these areas as 'equal per capita' meaning that each State would be assumed to have the same expense per head of population.

16 We have undertaken an analysis for this paper that involved identifying the SDC factors in all of our assessments, then measuring how much redistribution of the GST and HCG pool these cause compared with an equal per capita assessment. The analysis was performed using the relativities we reported to the Australian Government and States in February 2006[4] and applying these to the $44.6 billion estimate for the 2005-06 pool of funds. Table 2 shows the results. A positive redistribution in the table means the State requires more of the GST and HCG pool due to its population characteristics.

Table 2 Redistribution of GST and HCG pool due to SDC factors

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Total[(a)]
	$m	$m	$m	$m	$m	$m	$m	$m	$m
Population dispersion[(b)]	-270	-324	140	207	-9	-41	-33	332	678
Densely urbanised areas[(c)]	40	50	-32	-12	-4	-25	-1	-16	90
Cross border service provision[(d)]	-38	1	1	0	0	0	36	0	38
All other SDC factors	-615	-1115	690	261	40	142	-173	770	1903
Total SDC related	-884	-1388	799	456	27	76	-171	1086	2443
Total due to all expense factors (SDC and non-SDC)[(e)]	-524	-2102	59	892	-64	219	7	1515	2691

(a) The sum of all positive values (or all negative values) for the States
(b) SDC factors based on SARIA classified remoteness characteristics of the population
(c) SDC factors related to location of population in highly urbanised areas (Sydney and Melbourne)
(d) SDC factors that use Census populations close to State borders, where services are procured across the border.
(e) Because redistribution for a particular factor can be positive or negative for a State, the sum due to all factors for a State may be less than the sum for a subset of these factors.

17 The SDC related factors, which are calculated using disaggregated Census data, cause a redistribution of $2.44 billion among the States, representing just under 91% of the total amount redistributed by all expense factors, whether SDC related or not. The impact is the largest in Victoria and the Northern Territory, with impacts of $1.4 b and $1.1 b respectively. Clearly, if reliable disaggregated Census data were not available, major funding inequities for the States on a large scale could result. This would have an impact on the quantity and quality of services able to be provided in each State.

4 The 2006 relativities are calculated using data averaged over the last five years, as this is required by the terms of reference given to the Commission.

PROBLEMS WITH CENSUS DATA AND DEALING WITH THEM

18 Relativities determined by the Commission are as good as the methods and data used in their calculation. Given the large amount of money whose allocation is affected by Census data, the Commission has put significant effort into identifying and understanding problems that may exist with the data. States have a vested interest and will draw perceived problems to our attention as well, although the Commission judges these on their merits and seeks advice from the ABS and others as necessary in doing so. The main areas of concern we have encountered that may be of broader interest to population researchers include:

- Indigenous under-enumeration: Census data are not adjusted for under-enumeration. State governments have raised concerns about this issue, noting discrepancies with Census estimates of under-enumeration and administrative data sets;
- effects of randomisation on small counts in highly disaggregated tables; and
- non-response values, such as 'not-stated', and how to deal with these.

Indigenous under-enumeration

19 Under-enumeration can occur in any population sub-group. The ABS already corrects for this when calculating estimated resident population, and also for broad population groupings (statistical local area, age, sex). The ABS recommends against such adjustments to finely disaggregated Census data.

20 In general, the Commission follows this rule. The situation with Indigenous persons is problematic though, as shown in Table 3. ABS post-enumeration surveys show that Indigenous undercounts are much more significant and are about 4 times higher[5] than whole of population undercounts.

Table 3 Implied undercount for Indigenous persons

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Census count[(a)]	125 077	25 925	117 336	62 157	24 677	16 398	3 634	53 422	428 626
Indigenous ERP (experimental)	134 888	27 846	125 910	65 931	25 544	17 384	3 909	56 875	458 287
Implied Undercount %	7.8	7.4	7.3	6.1	3.5	6.0	7.6	6.5	6.9

(a) After redistributing Indigenous 'Not stated'.
Source: ABS catalogue 3230.0

21 In the Commission's last major review of relativity calculation methods (concluded in 2004), this issue proved somewhat contentious. A number of States argued that even the ABS experimental estimates were too low and provided qualitative data and administrative data that supported this, such as the Community Housing and Infrastructure Needs Survey

5 The effect is magnified for the Commission because substantially higher cost and use weights usually apply to Indigenous groups in the Commission's assessments, compounding the effect of the undercount.

(CHINS) and Indigenous school population data which both suggested higher under-enumeration. In the case of Tasmania, the Commission had concerns that the Indigenous population had been over-estimated. The Tasmanian issue related to a growing propensity for people to identify themselves as Indigenous in the Census.

22 The Commission spent considerable effort investigating the under-enumeration with the ABS and States, and the discrepancies with administrative data sets. The ABS had already investigated many apparent anomalies (resulting in corrections where there was sound evidence to do so). The Commission came to the conclusion that Census collection in Indigenous communities had been the subject of considerable thought, planning and careful execution by the ABS. In assessing whether the number of Indigenous people in Tasmania were over-estimated in the Census, there was evidence that this was an issue in all States, and particularly in urban areas. However, there was no firm evidence that the problem was greater in Tasmania.

23 On balance, the Commission decided that the ABS data were the best available. We recognised that the administrative data was less reliable. However, given that the ABS had put considerable planning and effort into the Indigenous Census and estimating undercounts, the Commission decided that adjusting the Census numbers for the undercount would give a more accurate estimate than simply ignoring the available information. Furthermore, based on advice from the ABS that this undercount would not be counted anywhere else in the Census, the Commission increased the total population estimates by the extent of the undercount of the Indigenous population.

24 Given that Indigenous populations counts are not only relevant to the Commission but to many other areas of government policy formation, we think it is appropriate for further action to deal with the undercount. Possible actions that we think the ABS should consider include:

- formally adjusting Census counts for Indigenous under-enumeration (instead of the current approach of releasing experimental estimates but not adjusting the 'official' figures); and
- further consideration of ways to improve the post enumeration survey (or conduct other analyses) to provide confidence for more formal recognition of the undercount.

25 Under-enumeration of non-Indigenous people was also investigated. Since the ABS already makes adjustments to estimated resident populations for general under-enumeration, the impact is lower and less significant for the Commission. However, the ABS does not use under-enumeration estimates to adjust highly disaggregated data. Analysis by the Commission showed that this had little impact on GST and HCG distribution, although it might be more significant for other users of Census data.

26 We understand that work on the Indigenous enumeration survey is ongoing within ABS. We welcome further research and improvements to technique that will provide more accurate Indigenous counts, and that provide more detailed explanatory material to proactively address concerns that may be raised about this aspect of the Census.

Effect of randomisation on small cell counts in large cross-tabulations of Census data

27 Before Census data are released to the public, for reasons of confidentiality, the ABS replaces cells with small counts in frequency tables with random values. This is known as the 'randomisation' procedure. In general, the larger the cross-tabulations, the more cells would have small counts and therefore more randomisation. The way the procedure is implemented does not guarantee the same total population or subtotals as non-randomised data. Hence, errors can arise in the totals for various subgroups in the population.

28 The ABS was able to assist the Commission in analysing the effect of such errors on our assessments by replacing randomised data with actual data. Due to the ABS's confidentiality obligations, this work had to be done by the ABS using spreadsheets supplied by the Commission. Only highly aggregated results were provided back to the Commission so that the actual Census data would not be exposed.

29 This analysis showed that randomisation results in a small decline in total population count of about 200, although the contribution from each State was different. The maximum difference for an individual State between randomised and actual data was less than 120.

30 To measure what impact randomisation had on SDC factors and redistribution of the GST and HCG pool, an expense category where Census data had a large impact on the redistribution of GST and HCG pool was chosen. This category was acute inpatient services. Another reason for selecting the acute inpatient services category is that it used Census data disaggregated to the finest level used in Commission assessments. None of our assessments use data disaggregated to a greater extent.

31 To test the effect of randomisation, the ABS replaced randomised Census data with actual Census data in the calculation of the SDC factor and passed the result to the Commission. We then compared the result of using the 'randomised' and 'non-randomised' SDC factors on redistribution.

32 Table 4 shows the difference in GST and HCG distribution due to the SDC factor compared to an equal per capita assessment in this services category. The SDC factor based on randomised Census data redistributes a total of almost $284 million. Using actual Census data instead of randomised Census data changed the redistribution by $0.35 million across Australia. The impacts from randomisation were clearly minor, providing confidence to the Commission that it could continue using such data to calculate reliable relativities.

Table 4 Effect of using actual Census data on redistribution of GST and HCG pool for the acute inpatient services category

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	total redist'd
	$m	$m	$m	$m	$m	$m	$m	$m	$m
Contribution of acute inpatients SDC to distribution of GST revenue from EPC, 2005 Update[(a)]	-92.3	-123.6	78.8	3.2	105.5	38.9	-68.0	57.5	283.9
Changes in GST revenue distribution due to randomisation[(b)]	0.02	-0.01	0.30	0.00	-0.22	0.01	-0.03	-0.08	0.35

(a) Based on randomised Census population data.
(b) By applying the acute inpatients SDC factors using actual (non-randomised) Census population data to the 2005 Update relativities.

Issues arising from 'not stated' entries in Census tables

33 Another issue with highly disaggregated Census tables is that cells sometimes contain 'not stated' entries. Because the Commission attempts to apportion population numbers to these cells, they can cause problems of a similar nature to randomisation. At present, the Commission calculates numbers for these cells to reflect the same proportions as shown elsewhere in the Census (for example, the overall split between Indigenous and non-Indigenous). However, this is not necessarily valid. We would welcome analysis and publication of results by the ABS to provide a better basis for the Commission (and other researchers) to estimate values for the 'not stated' cells. Some work has already been done in this area by the ABS and we would like to see more.

Data mismatches

34 The Commission sometimes sees data mismatches when we have obtained use or cost data for sub-groups in the population (for example, provided through surveys by service providers) but Census data shows that there are no people in that sub-group. For example, this issue was observed in the acute inpatients category when comparing morbidity sub-groups in the population with Census data. Possible causes of this mismatch could include:

- Census disaggregations represent a snapshot in time and will be less accurate with time due to normal population movements and changes between States; and
- Our providers of cost or usage data may collect this on a different basis to the Census data, including to different quality standards.

35 Because the Census and usage data are finely dissected, the numbers involved are small and fortunately not material (the size was on a similar scale to the effect of randomisation in the acute inpatients category).

Other Census data issues

36 After the relativities[6] are determined by the Commission, they are used by the Australian Government to calculate shares of the GST and HCG pool for each State. The relativities are per capita amounts and hence have to be multiplied by State population numbers to determine each State's share. Errors in ERP will therefore have an impact on the distribution of the pool and can affect perceptions of the fairness of the process.

37 One area of concern to the Commission is the definition of estimated resident population (ERP). ERP excludes those who are not resident in Australia for at least 10 months in a 12 month period[7]. Hence the ERP estimate will include some foreign nationals who are here on a long-stay basis (such as backpackers and overseas students). These foreign nationals do not have subsidised access to many State services such as education and health services. Hence, when the ERP numbers contain people not entitled to access State services on the same basis as Australian citizens, the Commission will overestimate the cost of providing services to that population. Where there are differences between States in the proportion of such foreign nationals in the ERP figures, the overestimation will vary from State to State, creating potential inequity.

38 Another source of error in ERP is the ability to predict based on a Census conducted every five years. Error will gradually increase in each subsequent year's ERP estimates and will be highest in the last year, before the next Census.

39 The dollar impact of errors can be sizeable. For every error of 1 000 head of population for a State, on average $2.3 million will be potentially misallocated. However, the per capita amount varies from State to State. The Northern Territory sits at the extreme with every 1 000 movement in population translating to around $10 million.

40 There can be significant variation between States in population estimates as illustrated in Table 5. This table shows the error based on estimates from the 1996 Census compared with the 2001 Census results. The final ERPs showed that the inter-Censal error was very low (the Australian population was over-estimated by 0.05%), and was lowest for the Northern Territory and Queensland. The error for the other States was larger. The effect on redistribution of the GST and HCG pool, if this error had applied to our 2006 relativities, would be of the order of $110 million across Australia.

---

6 State and Australian population estimates are also used in the calculation of relativities.

7 The ABS is currently reviewing the definition used for ERP to exclude people who are not resident for 12 months in a 16 month period.

Table 5 Inter-Censal error, 1996 to 2001

	Final intercensal error	Final intercensal error (%)
New South Wales	26 978	0.41
Victoria	-35 593	-0.74
Queensland	-5 783	-0.16
Western Australia	-11 479	-0.60
South Australia	9 586	0.63
Tasmania	1 837	0.39
ACT	4 329	1.36
Northern Territory	219	0.11
Australia[(a)]	-10 550	-0.05

Notes: Intercensal error is the difference between two sets of ERP figures, those based on the 1996 Census and those based on the 2001 Census.
(a) Includes other Territories.
Source: ABS Additional Briefing Information: Population Estimates Released 20 March 2003, Table A2.

41 Due to this substantial impact of such errors, we welcome further work by the ABS to improve the reliability and fitness for purpose of ERP calculations.

FUTURE USE OF CENSUS DATA BY THE COMMISSION

42 There is little doubt that disaggregated Census data will continue to be crucial to the work of the Commission. However, the way in which that data are used may change in the future. Every five or six years, the Commission undertakes a major review of the methods it uses to calculate relativities. The next review will be completed early in 2010. Our terms of reference for that review require that we simplify our methods, among other things by aggregating smaller assessments into larger categories — but without compromising horizontal fiscal equalisation.

43 The Commission is approaching the next review from the other direction by starting at the highest level of aggregation (all State services/expenses) and asking 'can reliable indicators be found to accurately measure differences between States in their costs of delivering services[8]?'. The way we do this is summarised in the following sequence:

- Fundamentally, what are the cost drivers for delivery of services and do they differ from State to State (for example, due to geographical or demographic factors)?
- How best can we disaggregate services into logical groups to reflect common cost drivers within each group but different drivers across groups, bearing in mind a desire to perform the minimum amount of disaggregation?

[8] As noted earlier, the costs the Commission is interested in are those not directly influenced by individual State policy choices about level or type of service.

- Are reliable data (including but not limited to Census) available to measure these cost drivers (in terms of physical, economic and population characteristics and the cost or usage differences within these)?
- How well do these service groups align to service groups for which actual State cost data is available so we can calculate Australian average expenses for each group (the Commission is likely to use ABS government finance statistics to measure actual cost)?

44 We are in the early stages of addressing these questions. This process will continue over the next couple of years. We would be keen to hear from researchers who could help inform our investigations.