

Appendix A: Technical notes

This appendix provides details on aspects of the study conducted in all states and territories.

Appendix A.1 Sampling strategy

The sampling strategy was developed:

- to ensure a sufficient sample of Indigenous separations by state and/or territory and remoteness level to enable robust estimates of completeness and correction factors to be derived for each remoteness area within jurisdictions; and
- to enable the study to be completed in a cost-efficient and timely manner.

Sample size

The sample size of a study is the main determinant of the quality of data that is able to be produced. Larger sample sizes will produce better quality estimates.

In this study, the AIHW sought a much larger sample size than the previous study, with the view to obtain reliable estimates of the proportion of Indigenous people correctly recorded at remoteness with jurisdiction levels.

On completion of this study and once the proportion of Indigenous people correctly recorded at remoteness within jurisdiction levels is known, it will be possible to design more efficient and smaller samples for future studies.

The sample size for each jurisdictional remoteness area was calculated following three steps:

Step one – sample at total jurisdiction level:

- An indicative sample size required to produce reliable estimates for each jurisdiction is calculated using the formula below and aggregated at state/territory level.

The formula is:

$$Z \geq (1 - s) / (s \cdot y^2 \cdot p), \text{ where:}$$

- Z is the required sample size
- s is the proportion of Indigenous patients correctly identified as Indigenous
- p is the proportion of total patients who were Indigenous
- y is the required relative standard error (RSE) in estimating s .

Note: Jurisdictions with lower proportions of Indigenous patients correctly identified, or with lower proportions of total patients who were Indigenous, will require larger sample sizes.

For example, if we believe that 80% of patients correctly identify as Indigenous but 3% of total patients were Indigenous, the RSE is set to 5%.

then

$$Z \geq \frac{1 - 0.8}{0.8 \times 0.05^2 \times 0.03}$$
$$\therefore Z \geq 3,333$$

Step two – sample at total remoteness area level:

- An indicative sample size required to produce reliable estimates for each remoteness area at national level is calculated using the same formula as above.
- Remoteness areas with lower proportions of Indigenous patients correctly identified, or with lower proportions of total patients who were Indigenous, will require larger sample sizes.

Step three – sample at jurisdictional remoteness area level:

- As the sample size in the previous study was insufficient to produce reliable information about proportions of Indigenous patients correctly identified at this level, the formula used above could not be used to calculate required sample sizes for this study.
- At jurisdiction level, the starting point is the larger of the sample sizes derived in step one above and the sample size used in the previous study.
- Within a jurisdiction, this sample is then distributed across remoteness areas based on the proportion of Indigenous people in each remoteness area.
- Some adjustments are then made to ensure minimum sample sizes at jurisdictional remoteness area level. The following principles are used to determine final sample sizes:
 - The sample needs to be sufficient to achieve a minimum of 25 Indigenous interviews in each remoteness area category within each jurisdiction
 - Time to achieve the sample is within a three month period.

Relative Standard Error (RSE)

To calculate the indicative sample size for each jurisdiction, the relative standard error is set at 5% for all states and territories except in Tasmania and ACT where it is set at 10%.

After the adjustment of sample size, the estimated RSE for each state and territory varies between 1% and 16%. Details are in the Table A1.1.

Table A1.1: Relative Standard Error of state and territories in the sample and the achieved sample

Sample size formula ($Z \geq (1-s)/[s*(x^2)*p]$)	Indicative ample				Achieved sample	
	Weighted Proportion of Indigenous people correctly recorded (s) %	Proportion of total separation that were for Indigenous persons (p) % 2010-11 Data	RSE (x)	Sample size (z)	Sample size	RSE %
New South Wales	84.63	3.9	0.03	6,484	5,109	3.02
Major cities	66.51	3.0	0.08	2,871	2,941	7.55
Inner regional	89.91	4.1	0.05	1,236	1,381	4.45
Outer regional	90.08	6.6	0.05	730	586	5.34
Remote	97.60	14.1	0.04	136	201	2.95
Very remote				65		
Victoria	80.99	1.1	0.06	5,237	4,487	6.90
Major cities	70.75	0.7	0.13	3,270	2,606	15.05
Inner regional	84.48	1.6	0.10	1,230	1,312	9.35
Outer regional	89.50	5.9	0.06	600	569	5.91
Remote						
Queensland	88.27	8.1	0.05	2,741	3,524	2.16
Major cities	71.66	2.9	0.13	865	1,956	8.35
Inner regional	81.57	6.4	0.08	623	974	6.02
Outer regional	94.64	23.7	0.02	908	544	2.10
Remote & very remote	100.00	40.1	0.00	168	50	0.00
Very remote				177		
Western Australia	94.93	8.6	0.05	248	1,482	2.05
Major cities	96.01	5.6	0.04	456	507	3.83
Inner regional	89.54	3.1	0.08	640	712	7.28
Outer regional	100.00	18.2	0.00	122	125	0.00
Remote	89.66	38.2	0.07	60	70	6.57
Very remote	100.00	62.3	0.00	40	68	0.00
South Australia	93.57	5.3	0.04	676	1,263	3.20
Major cities	85.92	3.5	0.08	697	664	8.40
Inner regional	97.34	10.7	0.02	420	319	2.83
Outer regional				155		
Remote	98.71	15.1	0.02	250	280	1.76
Very remote				60		
Tasmania	64.26	2.8	0.16	747	1,772	10.51
Inner regional	64.26	2.8	0.11	1,500	1,772	10.55

Sample size formula ($Z \geq \frac{(1-s)}{[s \cdot (x^2) \cdot p]}$)	Indicative sample				Achieved sample	
	Weighted Proportion of Indigenous people correctly recorded (s) %	Proportion of total separation that were for Indigenous persons (p) % 2010-11 Data	RSE (x)	Sample size (z)	Sample size	RSE %
Australian Capital Territory	57.58	2.3	0.14	1,670	1,857	13.13
Major cities	57.58	2.3	0.14	1,638	1,857	13.13
Northern Territory	97.59	69.8	0.01	371	813	0.66
Outer regional	96.64	56.9	0.02	120	413	1.22
Remote	98.38	80.5	0.01	130	340	0.78
Very remote	97.77	88.6	0.02	58	60	2.07
Australia	93.00	5.8	0.01	18,320	20,299	0.80
Major cities	80.38	2.30	0.03	9,797	10,531	3.18
Inner regional	85.14	5.05	0.02	5,649	6,333	2.34
Outer regional	95.55	21.06	0.01	2,735	2,374	0.96
Remote	97.78	54.97	0.01	744	849	0.70
Very remote	98.76	63.45	0.01	400	212	0.96

At national level, RSEs will generally be lower than those at jurisdiction level.

Appendix A.2 Estimation

Weighting

Indigenous identification characteristics vary by jurisdiction, hospital and remoteness area.

Given that the previous study was based on a small sample of patients, the proportion of surveyed Indigenous patients in a hospital (or remoteness area) compared to the total for the jurisdiction or remoteness area may not be representative of the state or remoteness area overall. As a result, Indigenous patients may be over- or under-represented in the study, potentially leading to biased estimates of correctness.

In order to account for this bias, the AIHW applied weightings to the study results for each hospital and remoteness area within each jurisdiction. These were based on the observed number of Indigenous separations included in the study, compared to the expected number of Indigenous separations. These weightings were applied to the raw estimates of completeness, to produce the final estimates of completeness.

Completeness and correction factors

In this study, estimates of completeness (C) and correction factor (CF) were undertaken at four levels:

1. within-hospital C and CF
2. C and CF for remoteness area (within a state or territory)
3. C and CF by state or territory
4. C and CF by remoteness area (within Australia).

The first level of estimates was an intermediate step to reach the second level of estimates.

Like building blocks, the second level was then applied to the weighting system to form the third and fourth levels of estimates.

1. Within-hospital completeness and correction factor

No within hospital weighting was applied to individual records. This assumes that selection of individuals was random and that each individual had the same chance of selection. This may or may not be true, depending on the method of selection used in each hospital. Given that some patients had multiple separations during the selection period, it is unlikely that all patients had the same chance of selection. However, without knowing the exact method of selection and details of individual patients, it is difficult to apply a more accurate weighting method. Each hospital will have a different selection method which will add further complications.

Alternatively assume that the common selection method is to choose one day and interview all patients in the hospital on that day. Then all patients within the three month period will have a chance of selection which varies depending on the number of days spent in the hospital. For example, if they were in hospital for five days, they have five times the chance of being in hospital on the selected day as someone who spent only one day in hospital.

Assume that the common selection method is to choose more than one (say five) consecutive days and interview all patients on Day One and any new patients on each subsequent day. A patient's chance of selection will depend on the number of days they were in hospital and the number of separations and the gap between separations.

The reason for applying a weighting is that patients with longer hospital stays and with more separations in the period will be over-represented in the sample and therefore their 'Indigenous status correctness' will be over-represented. The information we would need to correctly adjust weights for this would be the number of days spent in hospital for each selected patient (counting each dialysis visit as one day) as well as the total number of patient days in the three month period.

As Indigenous status is decided at admission, the probability of it being correct is unlikely to be impacted by length of stay or the number of visits for dialysis patients. The information needed to correctly adjust for any potential bias is extensive and the adjustment would be quite complex. For this reason no adjustment was made at the individual patient level.

Within-hospital C_i and CF_i were first estimated for each hospital i with Indigenous patients identified in the interview, using the following formulas:

$$C_i = A_i / (A_i + B_i) \text{ and } CF_i = (A_i + B_i) / (A_i + D_i),$$

where:

- A_i was the number of patients identified as Indigenous in both interview and hospital records
- B_i was the number of patients identified as Indigenous in the interview but non-Indigenous or with a 'not stated/inadequately described' Indigenous status in hospital records

- D_i was the number of patients identified as non-Indigenous or with a 'not stated/inadequately described' Indigenous status in the interview, but Indigenous in hospital records.

For some hospitals in Victoria, Queensland, South Australia and Tasmania, the study did not result in any interviews with Indigenous persons. In these cases, neither C_i nor CF_i is an appropriate estimate. These hospitals were excluded from the estimation since no information on their Indigenous identification levels can be sourced from the study.

2. Completeness and correction factor by remoteness area (within a state or territory)

C_s and CF_s for each remoteness area within a jurisdiction were estimated based on the estimated within-hospital C_{is} and CF_{is} in the area, using either W_i or AW_i as the weight.

- W_i , the weight for hospital i in the estimation of the remoteness area CF , was the proportion of separations for Indigenous persons in hospital i , out of the sum of separations for Indigenous persons from participating hospitals in the remoteness area. For this purpose, the number of separations for Indigenous persons was based on separations reported during the period July 2010–June 2011 financial year (benchmark period) as recorded in the AIHW's NHMD. The benchmark period was chosen to assist in avoiding any seasonality issues.
- AW_i , the weight for hospital i in the estimation of remoteness area C , was the proportion of adjusted separations for Indigenous persons in hospital i , out of the sum of adjusted separations for Indigenous persons from participating hospitals in the remoteness area (adjusted by the within-hospital CF).
- For each participating hospital with Indigenous patients identified in the interview, the adjusted number of separations for Indigenous persons was equal to the number of separations for Indigenous persons in the NHMD multiplied by the within hospital CF_i .

The CF for remoteness area r in jurisdiction j was calculated as a weighted average of the relevant within-hospital CFs , based on weight W_i .

$$CF_{(r,j)} = \sum_{i \in (r,j)} [CF_i * W_i]$$

where

$$- W_i = \frac{n_i}{\sum_{i \in (r,j)} n_i} \quad \text{and}$$

- n_i = number of Indigenous separations in participating hospital i in the benchmark period.

The completeness for remoteness area r in jurisdiction j was calculated as a weighted average of within-hospital completeness, based on weight AW_i .

$$C_{(r,j)} = \sum_{i \in (r,j)} [C_i * AW_i]$$

where

$$AW_i = \frac{CF_i * n_i}{\sum_{i \in (r,j)} [CF_i * n_i]}$$

3. Completeness and correction factor by state or territory

The C and CF for each jurisdiction was estimated, based on the remoteness area C s and CF s in the jurisdiction, using either Wj_r or AWj_r as the weight.

Wj_r , the weight of remoteness area r in the estimation of jurisdiction j 's CF , was the proportion of separations for Indigenous persons in remoteness area r , out of all separations for Indigenous persons in jurisdiction j . For this purpose, the number of separations for Indigenous persons was based on separations reported during the benchmark period, as recorded in the AIHW's NHMD.

AWj_r , the weight of remoteness area r in the estimation of jurisdiction j 's completeness, was the proportion of adjusted separations for Indigenous persons in remoteness area r , out of the sum of adjusted separations for Indigenous persons in all remoteness areas in jurisdiction j (adjusted by the remoteness area CF).

For each remoteness area r in jurisdiction j , the adjusted number of separations for Indigenous persons was equal to the number of separations for Indigenous persons in the NHMD multiplied by the remoteness area $CF_{(r,j)}$. The CF for jurisdiction j was calculated as a weighted average of CF s for all remoteness areas in the jurisdiction, based on weight Wj_r .

$$CF_j = \sum_r [CF_{(r,j)} * Wj_r]$$

where

$$- Wj_r = \frac{n_{(r,j)}}{\sum_r n_{(r,j)}} \quad \text{and}$$

- $n_{(r,j)}$ = number of Indigenous separations for remoteness area r of jurisdiction j in the benchmark period.

The completeness for jurisdiction j was calculated as a weighted average of completeness factors for all relevant remoteness areas in the jurisdiction, based on weight AWj_r .

$$C_j = \sum_r [C_{(r,j)} * AWj_r]$$

where

$$AWj_r = \frac{CF_{(r,j)} * n_{(r,j)}}{\sum_r [CF_{(r,j)} * n_{(r,j)}]}$$

4. Completeness and correction factor by remoteness area (within Australia)

The C and CF for each remoteness area (within Australia) was estimated based on the remoteness area (within jurisdiction) C s and CF s, using either Wr_j or AWr_j as the weight.

Wr_j , the weight of jurisdiction j 's contribution in the estimation of remoteness area r 's CF , was the proportion of separations for Indigenous persons in remoteness area r of jurisdiction j , out of all separations for Indigenous persons in remoteness area r across Australia.

AWr_j , the weight of jurisdiction j 's contribution in the estimation of remoteness area r 's completeness, was the proportion of adjusted separations for Indigenous persons in

remoteness area r of jurisdiction j , out of the sum of adjusted separations for Indigenous persons in remoteness area r across Australia.

The CF for the remoteness area r (within Australia) was calculated as a weighted average of CF s for all relevant remoteness areas across Australia, based on weight Wr_j .

$$CF_r = \sum_j [CF_{(r,j)} * Wr_j]$$

where

$$- \quad Wr_j = \frac{n_{(r,j)}}{\sum_j n_{(r,j)}} \quad \text{and}$$

- $n_{(r,j)}$ = number of Indigenous separations at remoteness area r in jurisdiction j in the benchmark period for each jurisdiction.

The completeness for the remoteness area r (within Australia) was calculated as a weighted average of completeness factors for all relevant remoteness areas across Australia, based on weight AWr_j .

$$C_r = \sum_j [C_{(r,j)} * AWr_j]$$

where

$$AWr_j = \frac{CF_{(r,j)} * n_{(r,j)}}{\sum_j [CF_{(r,j)} * n_{(r,j)}]}$$

Confidence intervals

The weighted completeness proportions are reported with 95% confidence intervals, calculated using Wilson's score intervals.

The formulas used were:

$$\text{Lower bound} = \frac{p + \frac{1}{2n} Z_{\alpha/2}^2 - Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n} + \frac{Z_{\alpha/2}^2}{4n^2}}}{1 + \frac{1}{n} Z_{\alpha/2}^2}$$

$$\text{Upper bound} = \frac{p + \frac{1}{2n} Z_{\alpha/2}^2 + Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n} + \frac{Z_{\alpha/2}^2}{4n^2}}}{1 + \frac{1}{n} Z_{\alpha/2}^2}$$

Where:

- p is the weighted correctness proportion
- n is the number of Indigenous persons at interview and
- $Z_{\alpha/2} = 1.96$

Appendix A.3 Possible sources of error or bias

Sample bias

The selection of hospitals can introduce a sample bias. The bias is expected to be relatively small since most non-selected hospitals had between zero and five Indigenous separations during any three month period in 2009-10. Allowing a bias of this type and magnitude was seen to be necessary to enable a cost effective study to be conducted.

Likely within hospital sample bias

An under-representation of children within the sample was likely, as children were required to have a parent or guardian to provide their consent to participate in the study. It is acknowledged that this was not always possible and for this reason it is highly likely that children may be under-represented in the sample.

There is also a strong possibility that some patients who were eligible to participate in the study were missed by the interviewers, for whatever reason, and this may contribute to a sample bias, particularly in regard to those who had very short hospital stays or those who were long stay patients, and therefore may have been over-represented.

Random and systematic error

Random errors occur due to chance variations in the sample and are not considered a source of bias.

Systematic errors are introduced as a result of errors in the sampling method, for example, where the number of Indigenous interviews required to meet the sample over or under estimated the number of Indigenous patients.

Assumptions

The project method was underpinned by the following assumptions that:

1. The patient's Indigenous status reported during the interview was correct.

The accuracy of the answer to the Indigenous status question at interview could vary due to factors including the patient's reaction to the interviewer when asked about his or her Indigenous status and/or interview conditions. Any violations of this assumption could introduce non-systematic (random) sampling errors, necessitating larger confidence intervals for the estimation results.

2. There was no change in admission practices or in the conduct of admission interviews by staff during the data quality study period.

Information obtained from the study was relatively consistent with the usual level of accuracy of Indigenous identification in the hospital.

Conduct of the audit

Timing of the interviews

Due to administrative arrangements, the data quality study was performed during different months of the year for different states and territories. The sampling strategy was not adjusted for seasonality or variation in admission practices over time.

Workforce

There was some variation in the approaches used by the jurisdictions in assigning staff to conduct the interviews. Some jurisdictions used existing hospital staff members to complete the interviews, and some recruited interviewers specifically for the study. As the jurisdictions were supplied with identical training materials, the effect of these differences was assumed to be minimal.