SAAP National Data Collection

Adjustment methods for incomplete coverage

Rosemary Karmel

AIHW

SAAP National Data Collection

Adjustment methods for incomplete coverage

The Australian Institute of Health and Welfare is an independent health and welfare statistics and information agency. The Institute's mission is to inform community discussion and decision making through national leadership in the development and provision of authoritative and timely information on the health and welfare of Australians.

SAAP National Data Collection

Adjustment methods for incomplete coverage

Rosemary Karmel

November 1999

Suggested citation

Australian Institute of Health and Welfare

Board Chair
Professor Janice Reid

Director
Dr Richard Madden

Contents

Preface

The Supported Accommodation Assistance Program (SAAP) National Data Collection is a large-scale and complex data collection involving more than 1,100 SAAP agencies and more than 80,000 clients. Client data are provided to the National Data Collection Agency located at the Australian Institute of Health and Welfare (AIHW). The AIHW is funded by the Commonwealth and the States and Territories to manage the Agency.

The data collected relate to SAAP clients who are homeless or at risk of being homeless. Collecting data from such a population is often difficult as persons in homeless situations (some of whom are escaping from domestic violence, or are long-term homeless, or have a drug-related problem) are unlikely to be motivated to provide personal information, and indeed may not be able to do so. SAAP agencies have shown a great willingness to collect and provide the required information and this has resulted in 95% participation from agencies.

However, many clients have not provided data or have not given their consent for the data to be used. This affects the quality of the data and it is important that the rate of consent is improved. The AIHW, in conjunction with the SAAP agencies, is developing strategies to ensure that SAAP clients are fully aware that the data they provide to the Institute are protected not only by the Commonwealth privacy principles but also by the AIHW Act. It is hoped that such an assurance, the good record achieved in data protection and the gradual use of the data by all, including SAAP agencies, will improve the rate of client consent in the future.

In the meantime, the AIHW has developed methods for adjustment to overcome some of the inadequacies in the data introduced through client non-consent and agency non-participation. It is recommended that adjusted data be used wherever possible. Inaccuracies in estimates of numbers of clients due to identical linkage keys for a small number of clients and changing linkage key information for the same client are not considered in this report.

Rosemary Karmel of the Institute has developed the methods and has written this report. Dr Simon Barry, Department of Statistics and Econometrics, the Australian National University, has provided advice on the statistical methods and commented on various drafts. Dr Ching Choi, Rosangela Merlo and Sheridan Coombes of the Institute provided comments. Their contributions are acknowledged.

Richard Madden
Director
Australian Institute of Health and Welfare

1 Introduction

The Supported Accommodation Assistance Program (SAAP) National Data Collection currently consists of five components: the Client Collection, the Administrative Data Collection, the Unmet Demand Collection, the Casual Client Collection and the special collections. The Client Collection is the main component. It consists of information about clients receiving support under SAAP. Data are collected by service providers during, or immediately following, contact with clients and are then forwarded to the National Data Collection Agency (NDCA) after clients' support periods have ended, and at the end of each reporting period (30 June and 31 December) for ongoing clients. An individual client may receive support on more than one occasion—either from the same SAAP agency or from different SAAP agencies (see AIHW 1997 for details of the Collection, and definitions used).

In the Client Collection, data are collected for each support period that a client has within a financial year. The data include basic sociodemographic information and the services required by and provided to each client (AIHW 1997: 2). The data are used to provide statistics on the characteristics of clients and their support periods. In 1996–97, data was received from nearly 1,100 agencies on 148,873 occasions of support provided to an estimated 100,900 clients (AIHW 1997: xv–xvi, 7).

This paper discusses problems in obtaining estimates of both support periods and clients from the Client Collection caused by service providers not participating in the collection and by clients refusing to provide consent for information to be collected. Ideally, the SAAP Client Collection should provide information on all support periods provided by all agencies funded to provide services under SAAP. At each occasion of support, clients would consent to provide information along with a linkage key (an alpha code) through which repeat use of services could be enumerated. Under these circumstances, there would be complete coverage of SAAP service usage and statistics of interest could be derived simply by counting the number of support periods or clients with particular characteristics. In practice, however, this situation does not occur.

There are two main causes of incomplete coverage of SAAP services in the collection:

- Not all SAAP agencies participate in the collection. In 1996–97, 51 out of 1,119 agencies, or 4.6%, did not participate in the collection and so did not return forms for their clients (AIHW 1997: 7).

- Not all clients consent to provide the information asked for on the collection form. This non-consent could be for any or all of their periods of support. As well, even among support periods for which consent is obtained, incomplete or invalid alpha codes are sometimes recorded. This results in insufficient data to identify repeat use of services, thus making the return unusable. Among forms from participating agencies, in 1996–97 some 36.3% (54,000 forms) did not have either consent or a valid alpha code (AIHW 1997: 7).

The consequences of incomplete coverage are fourfold. First, non-participation by agencies results in estimates of the total number of support periods being too low. Second, non-consent in participating agencies results in a sizeable underestimation of the number of support periods with particular characteristics, for example support periods for women. Third, estimates of the number of clients receiving assistance will also be too low because of

the problem of non-consent.[1] Many clients may only have had support periods during which they did not provide consent and so they would not have been counted as distinct clients in tables showing distributions. Finally, because clients can give consent on some occasions and not others, the distribution of the number of occasions of support per client will be biased downwards. As a consequence, the estimated average number of support periods per client will be too low.

There is no strictly objective method which can be applied to the data so that estimates can be adjusted for the incomplete coverage of the Client Collection. The reason for this is that there is either no or only limited information on support periods during which consent was not obtained. As a result, assumptions about the behaviour of clients and the characteristics of support periods without consent have to be made. For example, assumptions need to be made concerning whether or not clients always give consent (that is, whether consent is given consistently for each and all support periods), and whether support periods without consent are similar in some way to support periods with consent. The assumptions chosen will necessarily affect the method used to adjust estimates and, consequently, the final estimates obtained. That is, different assumptions will lead to different estimates. The validity of the assumptions determines how well adjusted estimates reflect what was actually happening in SAAP agencies.

In order to derive an adjustment scheme, a particular set of underlying assumptions was chosen. Care was taken to ensure that these assumptions were plausible. Methods to adjust estimates were then developed based on this set of assumptions. Because of the greater effects of non-consent and non-participation on estimates of clients than of support periods, different adjustments were required for the two types of estimates. In this paper, the underlying assumptions and the resulting methods derived to obtain estimates are presented. A number of key estimates are given with and without adjustment, and the accuracy of adjusted estimates is discussed. Detailed mathematical specifications of the adjustments, and background analyses carried out to refine the methods, are contained in a number of appendixes.

---

1. In this paper 'support periods without consent' includes both support periods without consent and support periods with consent but without valid alpha codes.

2 The estimation method

The simplest method of adjusting for non-consent, or non-response, is to scale up estimates at the total level. This assumes that the distributional profile of those not consenting is the same as those consenting. The practical implication of this assumption is that we treat the observed sample as a representative sample of the population. If it is known, as in this case, that overall the sample is not representative, then adjusting at the broad level will lead to biases in the estimates. For example, the highest non-consent rate in 1996–97 was for agencies for women escaping domestic violence. As a result, estimates of the number of support periods for women will be too low if the percentage of support periods with informed consent that are for females is applied to the total number of support periods (with and without consent). That is, the representation of females is too low in the 'consenting' support periods. A method of adjusting estimates which allows for such sample biases is therefore required.

To estimate the total number of support periods with certain characteristics it is necessary to make some underlying assumptions concerning the probability of agencies participating and obtaining informed consent. There are several ways this can be done, but the simpler the assumptions the more likely it is that a practical and transparent solution will be found. Two possible approaches are outlined below. In both of these, 'stratum' refers to the level at which adjustment is to be undertaken so that biases in the support periods with consent can be adjusted for, and 'group' refers to the level at which adjustment for agency non-participation is undertaken.

Approach 1: Independent consent for each client

Under this approach it is assumed that:

- If people consent on their first support period in a particular stratum then they will consent on all subsequent periods of support in that stratum. Conversely, if people refuse to give consent on their first support period, it is assumed that they will refuse on all subsequent visits in a particular stratum.
- For all clients in a particular stratum, the probability of giving consent is the same.
- A client has support periods only in one stratum.
- The probability of an agency participating is independent of other agencies and is the same within a particular group. This probability of participating depends only on characteristics of the agency.

Approach 2: Independent consent for each occasion of support

An alternative approach is to assume that:

- The probability of obtaining consent for a support period is the same for each support period in a particular stratum. It is not affected by previous consent or refusal by the client.
- The probability of an agency participating is independent of other agencies and is the same within a particular group. This probability of participating depends only on characteristics of the agency.

2.1 Estimates for support periods

For support periods there are two distinct levels of adjustment—adjusting for agency non-participation and adjusting for client non-consent. Since a range of information is available for all support periods in participating agencies, estimates using these data need only be adjusted for non-participation. However, for estimates using data which requires consent, adjustments for both agency non-participation and client non-consent are required. Now, estimates based on a 95% sample of support periods (all those in participating agencies) will be more accurate than those based on a 61% sample (support periods with consent in participating agencies). Consequently, to take advantage of all available data, two weighting systems are required for support periods: one to adjust for agency non-participation only and one to adjust for both agency non-participation and client non-consent.

When considering support periods, both of the approaches given above lead to the same adjustments or weights. To obtain estimates adjusting only for non-participation, for example when estimating support periods by state, the number of support periods from participating agencies is scaled up using the ratio of total number of agencies to number of participating agencies in an agency grouping. The weight used is then:

$$\frac{\text{Total number of agencies in group}}{\text{Number of participating agencies in group}}$$

To obtain estimates adjusting for both non-participation and non-consent, for example when estimating support periods within age groups, a two-stage weight is used. Details of how these adjustments are derived and used are contained in Appendix A. In essence, the adjustments are a refinement of applying consent rates to counts of support periods with consent, so that the number of support periods with consent are scaled up at the stratum-by-group level using the ratio, for that stratum-by-group, of total number of support periods in all agencies to number of support periods with consent. The weight used is of the form below (see Appendix A, p. 28):

$$\frac{\text{Total number of support periods in participating agencies in stratum}}{\text{Number of support periods with consent in stratum}} \times \frac{\text{Total number of agencies in group}}{\text{Number of participating agencies in group}}$$

The strata on the left-hand side of the weight define the level at which non-consent adjustments are made, while the groups on the right hand side define the level at which non-participation adjustments are made.

The choice of strata is difficult. In particular, questions such as how strata should be defined and how many there should be need to be answered. There is no objective solution to the first question. Analysis suggests that there are different consent rates for different groups. For example, support periods in agencies for women escaping domestic violence have only a 55% consent rate, whereas in agencies for single men the consent rate is 77% (AIHW 1997: 7; see also Appendix B). Not adjusting differentially will preserve any biases in the data with consent. To see this, consider the case where 100 men and 100 women have a single support period each. Assume that 80% of the men give consent but only 50% of the women do, so that only 38.5% of support periods with consent are for women. If no allowance is made for the differential consent rates, using only support periods with consent the estimated number

4

of support periods for men is too high at 123 and that for women is too low at 77 (38.5% of 200) (see Example 1).

Example 1: Effect of not adjusting for differential consent rates

	Actual support periods		Support periods with consent		Estimated total number of support periods
	Support periods	Proportion of support periods	Support periods	Weight	
Men	100	0.5	80	200/130 = 1.54	1.54 × 80 = 123
Women	100	0.5	50	1.54	˙1.54 × 50 = 77
All	200	1.0	130	1.54	200

The identification of strata needs to be based on variables known for all occasions of support in participating agencies. Within a stratum, consenting support periods are assumed to have similar characteristics to those for which consent was not obtained. If consent is less likely to be obtained for support periods with particular attributes, then we want support periods with those attributes with consent to represent those without consent. That is, we want to divide the population into groups of support periods with similar attributes. It is then assumed that, within these groups, support periods with consent represent support periods without consent. Of course, if support periods without consent are always different from periods with consent in some way, this difference cannot be allowed for as consenters cannot, in this case, represent non-consenters.

From the above it can be seen that we want to define a stratification that allows for factors that influence either consent or the characteristics of support periods. There are a reasonable number of variables collected for all support periods in participating agencies. To decide which variables should be used for stratification, an analysis was carried out to identify those variables related to consent. This analysis is described in Appendix B. The final stratification chosen divided support periods into strata based on:

- State/Territory
- primary target group
- length of support period
- whether or not the client was accommodated
- whether or not the agency was a proclaimed place (for New South Wales only)
- whether a general or high volume form was completed for the client.

The first five variables were chosen primarily because of their influence on consent, whereas the last variable was chosen because of its effect on characteristics of support periods.

The above stratification was not the only possible one that could have been used. Because the analysis in Appendix B indicates rather than identifies precisely variables that influence consent, stratifications based on fewer or more variables, or a different combination of variables, could easily have been chosen. Each stratification would result in slightly different estimates.[2] In addition, although stratification is used to adjust for biases in the sample, it cannot be ensured that the sample within a stratum is representative of that stratum.

---

2. A stratification which does not include proclaimed place or high volume agency indicators but which uses 'number of types of specialist assistance' instead gives estimates of 156,478 support periods for 83,330 clients. This compares with estimates of 156,490 support periods for 83,260 clients (from Table E.4) obtained from the stratification used in this paper. There may be greater

Note that if a variable is used to define the stratification, that variable is effectively used as a benchmark. This means that estimates for participating agencies only will closely match the observed numbers for these variables.[3]

Similarly, since agency non-participation (used in the weight to adjust for agency non-participation only and in the right-hand side of the two-stage weight defined above) varies across States and Territories and target groups (AIHW 1997: 7), a differential non-participation adjustment is also desirable. However, as there are only around 1,100 agencies, a fine stratification like that used for non-consent adjustment could not be used. Therefore, non-participation rates were allowed for at the target group level only.

It is possible for the two weighting schemes for support periods described above to give different estimates for the same item. For example, estimates of the number of support periods in each State and Territory may differ slightly when derived using the full non-participation/non-consent weight applied to support periods with consent rather than the non-participation weight applied to all support periods in participating agencies. While the full weight adjusts for biases in support periods with consent, when compared to those without consent, not all differences may have been accounted for. The extent to which biases remain will be reflected in the differences between estimates from the two weighting systems. As agencies become more familiar with the data collection and its purposes, it is expected that differences between support periods with and without consent will diminish, and so too will differences in estimates from the two weighting systems.

Adjusted estimates

As noted previously, the two sets of assumptions (or approaches) set out earlier give the same adjustments for support periods. Adjusted and unadjusted figures are presented in Tables 1 to 4 for a number of classifications. Allowing for non-consent and agency non-participation, it is estimated that SAAP agencies provided around 156,500 periods of support in 1996–97. This compares with 148,900 support periods provided by participating agencies with 94,900 of these having consent.

As expected from the participation and consent rates, the adjustments have a larger effect for tabulations by gender than for the other characteristics considered. In 1996–97, only 40% of support periods with consent were for women. However, when using non-consent and non-participation adjustments, it is estimated that 46% of support periods were for women (Table 3). Such a large change was not observed for any other classification.

---

percentage differences for estimates at a finer level. Due to resource constraints, it has not been possible to investigate this issue further.

3. There is not an exact match because of the measures taken to limit the number of strata (see Appendix B).

Table 1: Support period estimates by State/Territory, adjusted and unadjusted, 1996-97 (%)

State/Territory	Support periods with consent	All support periods from participating agencies	All support periods (adjusted)[a]
New South Wales	41.7	37.0	37.0
Victoria	21.1	23.5	23.5
Queensland	14.1	16.1	16.1
Western Australia	7.6	7.8	7.7
South Australia	7.4	7.3	7.3
Tasmania	3.1	3.3	3.3
Australian Capital Territory	2.3	2.0	2.1
Northern Territory	2.8	3.1	3.2
Australia	100	100	100
Australia (number)	94,885	148,873	156,490

(a) Figures have been weighted to adjust for agency non-participation.

Table 2: Support period estimates by primary target group, adjusted and unadjusted, 1996-97 (%)

Primary target group	Support periods with consent	All support periods from participating agencies	All support periods (adjusted)[a]
Young people	21.4	21.5	21.6
Single men only	29.3	24.8	24.8
Single women only	2.0	2.0	2.0
Families	4.9	4.9	4.8
Women escaping domestic violence	14.1	20.6	20.1
Cross-target/ multiple/general	28.2	26.1	26.8
Total	100	100	100
Total (number)	94,885	148,873	156,490

(a) Figures have been weighted to adjust for agency non-participation.

Table 3: Support period estimates by gender, adjusted and unadjusted, 1996-97 (%)

Gender	Support periods with consent	All support periods from participating agencies with gender data	All support periods (adjusted)[a]
Female	39.9	42.2	46.4
Male	60.1	57.8	53.6
Total	100	100	100
Total (number)	94,320	100,154	155,573
Missing (number)	565	48,719	917

(a) Figures have been weighted to adjust for agency non-participation and client non-consent.

Table 4: Support period estimates by age, adjusted and unadjusted, 1996–97 (%)

Age	Support periods with consent	All support periods from participating agencies with year of birth data	All support periods (adjusted)[a]
Under 15	1.6	1.6	1.7
15–19	19.9	19.5	20.0
20–24	14.0	14.1	14.9
25–44	46.4	47.0	47.2
45–64	15.1	14.9	13.7
65+	3.0	2.9	2.5
Total	100	100	100
Total (number)	94,855	98,138	156,441
Missing (number)	30	50,735	49

(a) Figures have been weighted to adjust for agency non-participation and client non-consent.

2.2 Estimates for clients

Estimating the number of clients is considerably more difficult than estimating the number of support periods. As well, unlike estimates for support periods, client estimates must necessarily be based on support periods with consent. This is because the number of support periods per client can only be determined via the linkage key which is obtained when a client consents. Consequently, there is only a single weighting system for client estimates.

As with estimating the number of support periods, there are several ways that estimating the number of clients can be approached. Either Approach 1 or Approach 2 could be used. However, it is accepted under Approach 2 that clients may have consented for any, all or none of their support periods. While this may be happening in practice, the mathematics involved in estimating the number of distinct clients is very complicated using this approach. For example, without further assumptions it is impossible to estimate the average number of visits per client if, for some clients, the number of visits with consent is less than their actual total number of visits. Therefore, because of its complexity, this approach was not considered. Initially, Approach 1 was used. The resulting weights were then modified significantly to allow for observed client behaviour. Such client behaviour included clients having support periods in more than one stratum, clients giving consent for a proportion of their support periods and clients going to both participating and non-participating agencies.

In order to estimate the number of clients, as opposed to the number of support periods, Approach 1 needs an additional assumption concerning the cross-use of agencies, that is:

• a client either always visits participating agencies or always visits non-participating agencies.

Using the assumptions under Approach 1, estimates of the number of clients within a stratum-by-group subset can then be obtained using the same weight as that obtained for support periods, that is, by multiplying up using the weight below (see Appendix A, p. 28).

$$\frac{\text{Total number of support periods in participating agencies in stratum}}{\text{Number of support periods with consent in stratum}} \times \frac{\text{Total number of agencies in group}}{\text{Number of participating agencies in group}}$$

The total number of clients could then theoretically be found by adding up stratum-by-group estimates. However, there are three reasons this would result in an overestimate of the number of clients:

- Summing the subtotals assumes that a client has support periods only in a single stratum (that is, that there is no stratum overlap).

- The weighting assumes that clients either always give consent or always refuse consent (that is, that there is no mixed consent).

- The weighting assumes that clients either always use participating or always use non-participating agencies (that is, that there is no cross-use of participating and non-participating agencies).

Each of these assumptions is known not to hold in practice. How the weighting system was modified to deal with these problems is discussed below.

Dealing with stratum overlap

That stratum overlap is a problem can be seen from the following example. Consider the simple case where there are 100 clients and two strata: 10 visit both strata whereas 45 visit stratum 1 only and 45 visit stratum 2 only. Since 55 clients visit each stratum, summing the number of clients that visit each stratum gives an estimate of 110 clients rather than 100. If only 5 clients visit both stratum, with 48 visiting only stratum 1 and 47 visiting only stratum 2, the simple estimate is 53 + 52 = 105.

From the above example it can be seen that the effect of stratum overlap on estimates depends on its prevalence. In 1996–97, some 5% of clients with linkage keys visited agencies in more than one State or Territory and 7% visited agencies for different target groups. As the number of strata increases so, too, does the extent of stratum overlap. Using the 46 strata developed for non-consent adjustment (Appendix B), 16% of identified clients visited more than one stratum.

Two steps were taken to avoid double counting of clients due to stratum overlap:

- The number of strata was kept relatively small (46 strata for 94,885 support periods, see Appendix B).

- Clients who still had support periods in more than one stratum were put in a stratum of their own. In this way, multistratum clients with consent represented multistratum clients without consent.

Putting all clients who visited more than one stratum into a single stratum results in a stratum with a relatively large number of clients (16% of clients, see Table A.3). It may be possible to improve the efficiency of the weighting system by dividing this stratum into a number of others. However, such stratification would need to be based on variables intrinsic to the client and collected for all support periods. Currently no such data items are available. In the future it may be possible to use gender to define such strata since gender is to be collected for all support periods from 1998–99. Further work would be required to refine the weighting system in this area.

Dealing with mixed consent

Ignoring mixed consent assumes that all clients consistently give or refuse consent. Support periods without consent are therefore assumed to relate to clients other than those who have given consent at any time. If this is not so, the average number of visits per client estimated

9

Table 4: Support period estimates by age, adjusted and unadjusted, 1996–97 (%)

Age	Support periods with consent	All support periods from participating agencies with year of birth data	All support periods (adjusted)[a]
Under 15	1.6	1.6	1.7
15–19	19.9	19.5	20.0
20–24	14.0	14.1	14.9
25–44	46.4	47.0	47.2
45–64	15.1	14.9	13.7
65+	3.0	2.9	2.5
Total	100	100	100
Total (number)	94,855	98,138	156,441
Missing (number)	30	50,735	49

(a)   Figures have been weighted to adjust for agency non-participation and client non-consent.

2.2 Estimates for clients

Estimating the number of clients is considerably more difficult than estimating the number of support periods. As well, unlike estimates for support periods, client estimates must necessarily be based on support periods with consent. This is because the number of support periods per client can only be determined via the linkage key which is obtained when a client consents. Consequently, there is only a single weighting system for client estimates.

As with estimating the number of support periods, there are several ways that estimating the number of clients can be approached. Either Approach 1 or Approach 2 could be used. However, it is accepted under Approach 2 that clients may have consented for any, all or none of their support periods. While this may be happening in practice, the mathematics involved in estimating the number of distinct clients is very complicated using this approach. For example, without further assumptions it is impossible to estimate the average number of visits per client if, for some clients, the number of visits with consent is less than their actual total number of visits. Therefore, because of its complexity, this approach was not considered. Initially, Approach 1 was used. The resulting weights were then modified significantly to allow for observed client behaviour. Such client behaviour included clients having support periods in more than one stratum, clients giving consent for a proportion of their support periods and clients going to both participating and non-participating agencies.

In order to estimate the number of clients, as opposed to the number of support periods, Approach 1 needs an additional assumption concerning the cross-use of agencies, that is:

- a client either always visits participating agencies or always visits non-participating agencies.

Using the assumptions under Approach 1, estimates of the number of clients within a stratum-by-group subset can then be obtained using the same weight as that obtained for support periods, that is, by multiplying up using the weight below (see Appendix A, p. 28).

$$\frac{\text{Total number of support periods in participating agencies in stratum}}{\text{Number of support periods with consent in stratum}} \times \frac{\text{Total number of agencies in group}}{\text{Number of participating agencies in group}}$$

The total number of clients could then theoretically be found by adding up stratum-by-group estimates. However, there are three reasons this would result in an overestimate of the number of clients:

- Summing the subtotals assumes that a client has support periods only in a single stratum (that is, that there is no stratum overlap).
- The weighting assumes that clients either always give consent or always refuse consent (that is, that there is no mixed consent).
- The weighting assumes that clients either always use participating or always use non-participating agencies (that is, that there is no cross-use of participating and non-participating agencies).

Each of these assumptions is known not to hold in practice. How the weighting system was modified to deal with these problems is discussed below.

Dealing with stratum overlap

That stratum overlap is a problem can be seen from the following example. Consider the simple case where there are 100 clients and two strata: 10 visit both strata whereas 45 visit stratum 1 only and 45 visit stratum 2 only. Since 55 clients visit each stratum, summing the number of clients that visit each stratum gives an estimate of 110 clients rather than 100. If only 5 clients visit both stratum, with 48 visiting only stratum 1 and 47 visiting only stratum 2, the simple estimate is 53 + 52 = 105.

From the above example it can be seen that the effect of stratum overlap on estimates depends on its prevalence. In 1996–97, some 5% of clients with linkage keys visited agencies in more than one State or Territory and 7% visited agencies for different target groups. As the number of strata increases so, too, does the extent of stratum overlap. Using the 46 strata developed for non-consent adjustment (Appendix B), 16% of identified clients visited more than one stratum.

Two steps were taken to avoid double counting of clients due to stratum overlap:

- The number of strata was kept relatively small (46 strata for 94,885 support periods, see Appendix B).
- Clients who still had support periods in more than one stratum were put in a stratum of their own. In this way, multistratum clients with consent represented multistratum clients without consent.

Putting all clients who visited more than one stratum into a single stratum results in a stratum with a relatively large number of clients (16% of clients, see Table A.3). It may be possible to improve the efficiency of the weighting system by dividing this stratum into a number of others. However, such stratification would need to be based on variables intrinsic to the client and collected for all support periods. Currently no such data items are available. In the future it may be possible to use gender to define such strata since gender is to be collected for all support periods from 1998–99. Further work would be required to refine the weighting system in this area.

Dealing with mixed consent

Ignoring mixed consent assumes that all clients consistently give or refuse consent. Support periods without consent are therefore assumed to relate to clients other than those who have given consent at any time. If this is not so, the average number of visits per client estimated

using only support periods with consent will underestimate the average number of visits per client. As a consequence, the estimates overestimate the number of clients.

The effect of inconsistent consent/refusal can be illustrated by considering Example 2 below. For simplicity it is assumed that all agencies are participating.

In Example 2 there are:

- 150,000 support periods over all agencies

- 95,000 support periods with consent (consent rate = 95/150 = 63.3%)

- 60,000 clients relating to the 95,000 support periods with consent

with

- 30% of the 55,000 support periods without consent relating to clients who give consent for at least one support period.

Example 2: The effect of ignoring mixed consent

Total support periods	150,000	
Support periods with consent	95,000	
Support periods without consent	55,000	
Clients with some consent	60,000	
Average support periods with consent per client with consent	= 95,000/60,000	= 1.58
Estimated total number of clients ignoring mixed consent	= total support periods/1.58	= 150,000/1.58 = 94,900.
Per cent support periods without consent that relate to clients who consent at some stage	30%	
All support periods for clients with some consent	= 95,000 + (0.3 x 55,000)	= 111,500
Average support periods per client with some consent	= 111,500/60,000	= 1.86
Estimated total number of clients allowing for mixed consent	= total support periods/1.86	= 150,000/1.86 = 80,600
Alternative scenario: Per cent support periods without consent that relate to clients who consent at some stage	10%	
All support periods for clients with some consent	= 95,000 + (0.1 x 55,000)	= 100,500
Average support periods per client with some consent	= 100,500/60,000	= 1.675
Estimated total number of clients allowing for mixed consent	= total support periods/1.675	= 150,000/1.675 = 89,600

If it is assumed that clients either always give consent or always refuse, then the estimated average number of support periods per client is 1.58 giving an estimated total number of clients of 94,900. However, we know that the 60,000 consenting clients actually had 111,500 periods of support. Thus the average number of support periods per client was really 1.86. This suggests that the total number of clients was 80,600 and not 102,700. If, however, only 10% of support periods without consent had related to clients who gave consent at some stage, the estimated number of clients becomes 89,600.

During 1996–97 valid alpha codes were provided for nearly 12,500 support periods even though consent was not given. Using these cases it was possible to get an indication of the extent of mixed consent and consequently the number of support periods without consent that related to clients who at some stage had given consent. Overall, nearly 37% of these support periods related to clients who consented at some stage during the year.

Using information from the above sample of support periods without consent, mixed consent was allowed for in the weighting scheme. Briefly, the number of support periods without consent that related to clients with some consent was estimated. As in the above

example, these support periods were then treated as support periods with consent when working out the average number of support periods per client. The number of clients relating to all support periods was then derived. Details of the modifications required to the weights are contained in Appendix A, with Appendix C providing a description of background analysis. As with the basic non-consent adjustment, modifications to allow for mixed consent were carried out for sub-populations to allow for varying degrees of mixed consent.

As shown in the example, the existence of mixed consent means that the actual average number of support periods per client is higher than observed. In addition, the spread of number of visits per client is different from that observed using only support periods with consent. Analysis of the 12,500 support periods with valid alpha codes but without consent was also used to adjust the distribution of the number of support periods per client. Based on these data, imputation techniques were used to adjust the number of support periods for a proportion of clients (see Appendix D for details). Using these values it is possible to obtain distributions of numbers of support periods per client which have been adjusted for mixed consent.

Dealing with cross-use of participating and non-participating agencies

Cross-use of participating and non-participating agencies has a similar effect on estimates as mixed consent. That is, ignoring it leads to inflated estimates of the number of clients, and deflated estimates of the average number of support periods per client. This is illustrated in Example 3. For simplicity it is assumed that all clients in participating agencies provide consent.

In Example 3 it is assumed that there are

- 150,000 support periods over all agencies
- 143,000 support periods in participating agencies (95.3%)
- 90,000 clients relating to the 143,000 support periods in participating agencies

and that

- 30% of the 7,000 support periods in non-participating agencies related to clients who had at least one support period in participating agencies.

If it is assumed that clients always go either to participating agencies or to non-participating agencies, then the estimated average number of support periods per client is 1.59, giving an estimated total number of clients of 94,340. However, we know that the 90,000 clients in participating agencies actually had 145,100 periods of support. Thus the average number of support periods per client was really 1.61, suggesting that the total number of clients was 93,200.

There is no direct evidence of the extent of cross-use of participating and non-participating agencies. However, simulations using data from participating agencies, in which 4.6%[4] of agencies were randomly nominated as 'non-participating', suggested that around one-third of support periods in non-participating agencies related to clients in participating agencies in 1996–97 (see Appendix A, p. 33).

---

4. This was the overall non-participation rate for agencies in 1996–97.

Assuming that the above proportion holds in general, weights were modified to allow for cross-use of agencies. An adjustment was applied at the national level since there were no data on which to base finer adjustments (see Appendix A).

Example 3: The effect of ignoring cross-use of participating and non-participating agencies

Total support periods	150,000		
Support periods in participating agencies	143,000		
Support periods in non-participating agencies	7,000		
Clients in participating agencies	90,000		
Average support periods in participating agencies per client	= 143,000/90,000	= 1.59	
Estimated total number of clients ignoring agency cross-use	= total support periods/1.59	= 150,000/1.59	= 94,340
Per cent support periods in non-participating agencies that relate to clients who go to participating agencies at some stage	30%		
All support periods for clients who go to participating agencies at some stage	= 143,000 + (0.3 × 7,000)	= 145,100	
Average support periods per client who goes to participating agencies at some stage	= 145,100/90,000	= 1.61	
Estimated total number of clients allowing for cross-use	= total support periods/1.61	= 150,000/1.61	= 93,200

Given that 95.6% of agencies participated in 1996–97, the adjustment factor for agency non-participation and cross-use is small compared with the adjustments for non-consent, stratum overlap and mixed consent. Remember that if certain types of agencies are less likely to participate than others, for example small agencies, these differences have not been allowed for.

Final client weights

With all of the above modifications made, the weights used to derive estimates of numbers of clients are of the following form (see Appendix A, p. 37):

$$\frac{\text{Total number of support periods in stratum in participating agencies}}{\text{Estimated number of support periods which relate to clients who consent at some stage, in stratum in participating agencies}} \times \frac{\text{Estimated total number of support periods in all agencies}}{\text{Estimated total number of support periods relating to all clients of participating agencies}}$$

The first factor adjusts for both non-consent and mixed consent, and the second adjusts for agency non-participation and cross-use of participating and non-participating agencies by clients.

Using these weights, the total number of clients can *now* be found by adding up stratum subtotals relevant to a particular estimate — the problems of stratum overlap, mixed consent and agency cross-use have all been allowed for.

The weights can be applied to observed stratum-level totals for the characteristics of interest, with the weighted subtotals then being added together to give broad level estimates. Alternatively, each client with consent can be assigned the relevant stratum weight. Totals are then found by summing the weights for each client with the characteristics of interest. (A similar approach can be used for support period estimates.)

Adjusted estimates

Adjusted and unadjusted estimates of numbers of clients are given in Tables 5 to 8. It should be remembered that there are very few data items collected which are intrinsic to the client—only gender, year of birth and ethnicity are expected to remain the same for each support period for a particular client. For other characteristics, the support period for a client that is to be used to define that characteristic needs to be identified. For example, we may be interested in the State or Territory of the first support period or the last; we may like to know the type of family group clients presented in on their first visit or last; or we may want to know whether a client ever had accompanying children.

Using the client weights, it is estimated that some 83,200 people were provided with support through SAAP agencies. This compares with 60,900 clients identified through support periods with consent,[5] and an estimate of 95,400 clients obtained by applying the average number (1.56) of support periods with consent for clients with consent to the number of support periods in participating agencies (see Table 5).[6]

As with support period data, the main effect of weighting on client distributions was by gender. In fact, weighting the estimates resulted in a reversal, with the adjusted estimates indicating that there were more women than men receiving assistance (Table 7). No similar reversal occurred at the support period level because, on average, men have more support periods than women, making up for their smaller numbers. The other main difference between weighted and unweighted estimates was for women using domestic violence agencies, with the estimated percentage of clients using this type of agency on their first visit in 1996–97 going from 18% to 21% (Table 6).

Note that estimates of the average number of support periods for men in New South Wales are particularly affected by the use of proclaimed places. In New South Wales only, police can place inebriated persons in proclaimed places where they are kept (by law) for at least 8 hours until they are sober. There are some 29 proclaimed place agencies. In 1996–97 in some of these agencies each placement was counted as a support period. As a result some SAAP clients had a very large number of support periods (up to 244 in a year). These large numbers of support periods for relatively few clients influence the estimates of average support periods per client. This problem in recording support periods was addressed by the NDCA for the 1997–98 collection.

An item of particular interest for clients is the number of support periods they have. The data adjustment procedure results in increased estimates of the number of support periods per client. This is a direct result of imputing the number of additional support periods without consent provided to consenting clients. The unadjusted estimate of the average

---

5. In the 1996–97 annual report the number of clients with consent and a valid linkage key is given as 64,291 (AIHW 1997: 129). This figure was obtained by counting the number of clients that visited each State/Territory and then summing to get an Australian total. As a consequence, clients that had support periods in more than one State/Territory were counted more than once.

6. In the 1996–97 annual report the number of support periods per client is given as 1.48 (AIHW 1997: 127). This figure was derived by dividing the number of support periods with consent and valid linkage keys by the number clients these support periods related to. However, clients that had support periods in more than one State/Territory were counted more than once in the denominator (see footnote 5). Dividing the number of support periods in participating agencies (148,873) by 1.48 then gave an estimate of 100,900 clients in participating agencies. If all clients are counted only once, the unweighted estimate of number of clients in participating agencies becomes 148,873/1.56 = 95,400 (see Table 5).

number of support periods per client is 1.56. Imputation for mixed consent brings this estimate up to 1.87 support periods per person. (Note that 19% of clients who consented at some stage in 1996–97 were estimated to have additional support periods without consent.) Adjusted and unadjusted distributions of clients by the number of support periods they had are given in Figure 1 and Table 9. By imputing for mixed consent, the proportion of clients estimated to have only one support period drops from 79% to 65%. As a result, the proportions of clients with two and three support periods nearly double.

Table 5: Client estimates by State/Territory, adjusted and unadjusted, 1996–97

| State/Territory[a] | Clients with consent only | | Adjusted[b] | |
	Clients	Average number of observed support periods per client	Clients	Average number of support periods per client
	%	Number	%	Number
New South Wales	32.3	2.01	30.5	2.43
Victoria	26.3	1.26	27.5	1.50
Queensland	16.1	1.39	16.7	1.71
Western Australia	8.9	1.34	8.8	1.62
South Australia	7.9	1.45	7.7	1.76
Tasmania	3.5	1.38	3.6	1.67
Australian Capital Territory	2.5	1.38	2.3	1.71
Northern Territory	2.7	1.59	2.8	1.96
Australia	100	**1.56**	**100**	1.87
Australia (number)	60,878		83,235	

(a) State/Territory of agency of first support period in the financial year.

(b) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.

Table 6: Client estimates by primary target group, adjusted and unadjusted, 1996–97

| Primary target group[a] | Clients with consent only | | Adjusted[b] | |
	Clients	Average number of observed support periods per client	Clients	Average number of support periods per client
	%	Number	%	Number
Young people	23.9	1.42	23.6	1.72
Single men only	25.0	1.90	22.5	2.36
Single women only	2.5	1.31	2.4	1.58
Families	6.6	1.16	6.4	1.32
Women escaping domestic violence	17.6	1.25	20.7	1.64
Cross-target/multiple/ general	24.3	1.70	24.3	1.95
Total	100	**1.56**	100	**1.87**
Total (number)	60,878		83,235	

(a) Primary target group of agency of first support period in the financial year.

(b) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.

Table 7: Client estimates by gender, adjusted and unadjusted, 1996–97

Gender	Clients with consent only		Adjusted[a]	
	Clients	Average number of observed support periods per client	Clients	Average number of support periods per client
	%	Number	%	Number
Female	48.8	1.28	51.8	1.58
Male	51.2	1.83	48.2	2.20
Total	100	1.56	100	1.87
Total (number)	60,498		82,723	
Missing (number)	380		512	

(a) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.

Table 8: Client estimates by age, adjusted and unadjusted, 1996–97

Age[a]	Clients with consent only		Adjusted[b]	
	Clients	Average number of observed support periods per client	Clients	Average number of support periods per client
	%	Number	%	Number
Under 15	2.1	1.32	2.1	1.66
15–19	21.9	1.42	21.7	1.71
20–24	16.0	1.37	16.1	1.68
25–44	45.9	1.58	46.3	1.90
45–64	12.1	1.92	11.8	2.28
65+	2.1	2.21	1.9	2.62
Total	100	1.56	100	1.87
Total (number)	60,848		83,196	
Missing (number)	30		39	

(a) Age at first support period in the financial year.
(b) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.

Table 9: Distribution of support periods per client, adjusted and unadjusted, 1996–97 (%)

Number of support periods per client	Observed support periods per client	Adjusted number of support periods per client[a]
1	79.2	64.6
2	11.8	19.0
3	4.0	8.0
4	1.9	3.6
5	0.9	1.7
6	0.6	0.9
7+	1.5	2.1
Total	100	100
Average (number)	1.56	1.87

(a) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.
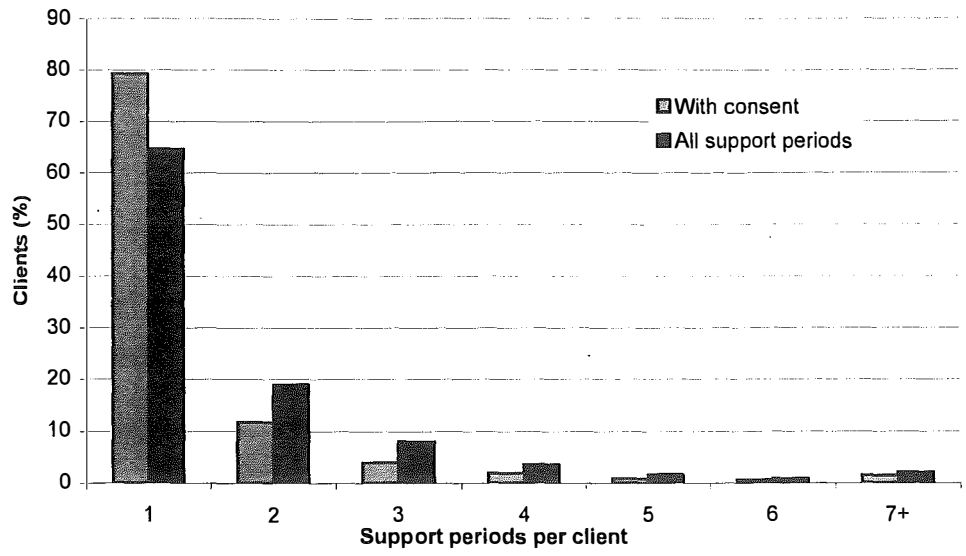
Figure 1: Distribution of support periods per client, with consent only and all support periods, 1996–97 (%)

16

3 Accuracy of estimates

It is difficult to ascertain the accuracy of the adjusted estimates presented in this paper. This is because those support periods and agencies which gave data were not a controlled sample of all support periods and agencies. There are three main sources of inaccuracy in the estimates: sampling of support periods, imputing mixed consent for clients, and bias in both the sample and imputation techniques. These are discussed below.

3.1 Sampling effects

Estimates derived from data obtained from samples are subject to sampling variability. That is, different samples are likely to give different results simply because they have different people in them. Standard errors are often used to measure the accuracy of estimates based on samples. Assuming that consent was obtained from a *random sample* of support periods, there are about two chances in three that the sample estimate will differ by less than one standard error from the figure that would have been obtained if consent had been obtained for all support periods, and about 19 chances in 20 that the difference will be less than two standard errors. The standard error expressed as a percentage of the estimate is called the relative standard error (ABS 1997: 58).

Although the support periods with consent *were not* a random sample of all support periods in that the selection of support periods for the sample was not controlled by the Collection Agency, indicative estimates of the sampling error of estimated numbers of support periods or clients can be obtained by assuming that the data *were* from a simple random sample of support periods and agencies (see Appendix E for detailed discussion). However, estimates of relative standard errors resulting from assuming that random sampling was used are indicative only of the reliability of weighted estimates. They do not measure any of the biases that may be present, but are useful for identifying estimates which are likely to have relatively large errors.

Table 10 presents indicative relative standard errors for broad level estimates. These relative standard errors are small and show that the estimates are usable in terms of sampling error. Given that the number of support periods in participating agencies provides a lower bound on the estimated number of support periods, these indicative relative standard errors probably overstate the level of error in the estimates. However, they suggest that estimates within cross-target agencies and agencies for single women are less reliable than those for other target groups. Similarly, the two Territories and New South Wales would seem to have less reliable estimates than the other States. Given the size of New South Wales, this last result is surprising. However, it is caused by the large range in the number of support periods provided by agencies in New South Wales (see Table E.1).

In general, the accuracy of estimates increases as the homogeneity of the population increases and, in particular, as the number of clients or support periods on which an estimate is based increases. The latter point means that estimates for finer subgroups are likely to be less reliable than estimates at a broader level. Consequently, estimates at a more detailed level will tend to be less accurate than those in Table 10.

Table 10: Indicative relative standard errors by State/Territory and by primary target group, 1996–97

	Estimated number of support periods[a]	Estimated number of clients[b]	Indicative relative standard error
State/Territory[c]	000s	000s	% of estimate
New South Wales	57.9	25.4	5.0
Victoria	36.8	22.9	1.8
Queensland	25.1	13.9	3.4
Western Australia	12.0	7.4	2.0
South Australia	11.4	6.4	3.1
Tasmania	5.2	3.0	3.1
Australian Capital Territory	3.2	1.9	4.6
Northern Territory	4.8	2.3	6.4
Primary target group[c]			
Young people	33.7	19.7	1.6
Single men only	38.8	18.7	3.9
Single women only	3.2	2.0	4.9
Families	7.4	5.4	1.0
Women escaping domestic violence	31.5	17.2	1.5
Cross-target/multiple/general	41.9	20.2	6.7
Australia	**156.5**	**83.2**	**1.8**

(a) Figures have been weighted to adjust for agency non-participation.
(b) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.
(c) For client estimates, this is as on first support period in the financial year.
Note: Any additional uncertainty caused by inaccuracies in the probabilities of consent used to obtain the weights are not included in these estimates.
Source: Appendix E.

3.2 Imputation effects

The imputation used in the weighting system affects two types of estimates: estimates of numbers of clients, and estimates of support periods per client. The effect of randomly assigning mixed consent to clients and then imputing their number of additional support periods can be gauged by repeating the imputation a number of times. The variation in estimates in these two areas was examined using this approach (see Appendix E).

Results indicate that the variation in estimates of numbers of clients due to imputation is quite small, with estimates of the total number of clients varying by, at most, 563 out of 83,261 over 100 estimates (see Table E.4). For broad level estimates, the smallest and largest estimates are generally within 1% of each other. This level of variation is smaller than that estimated to be present due to sampling effects.

There are larger effects on the estimates of the average number of support periods per client. However, for a particular group of clients, the smallest and largest estimates of the average number of support periods per client are generally within 5% of each other (see Table E.5). As with all estimates, the variability of a particular estimate depends on the subpopulation it relates to.

3.3 Bias

There are several possible sources of bias for the estimates. One obvious source is the use of a non-probability sample to obtain estimates. Estimates are based on support periods with consent, equating probabilities of selection with consent rates. However, if there are differences between support periods with and without consent that have not been allowed for in the weighting system, there will be biases in the estimates. In particular, if consent is *never* obtained for certain types of support periods, it is impossible to adjust for them through weights. Consider an extreme example: if consent is *never* obtained for support periods for women, then estimates for women can be obtained only by using support periods with consent for men to represent them. Since a large proportion of women use agencies for women escaping domestic violence, and most men use agencies for single men, using men to represent women will result in biased estimates. Without conducting follow-up surveys of clients who don't consent, it is impossible to gauge the extent of such biases in adjusted estimates.

Another source of bias is the mixed consent adjustments. These are based on a sample of 12,500 non-consenting support periods. If this sample is not representative of support periods without consent, then the adjustments for mixed consent will result in biases. Analysis of this sample indicated that it contained a good spread of support periods, but there may be biases not identified. In addition, the estimates of mixed consent rates are themselves subject to a sampling effect. The actual extent of mixed consent may therefore be different from that observed in the 'sample'. For example, if 40% rather than the estimated 37% of support periods without consent related to consenting clients, then the estimated number of clients should be about 2% lower (around 81,900 rather than 83,200). Similarly, if only 35% of support periods without consent related to consenting clients, the estimated number of clients should be about 84,100, or 1% higher.[7]

Finally, estimates of the number of support periods per client rely on imputation of additional support periods. The distributions used for this imputation are likely to be biased, both because an exact match to what is happening is unlikely and because the number of additional support periods for clients with mixed consent is limited to eight.[8]

That this last bias is present is illustrated by the difference between two estimates of the average number of support periods per client: the value derived by using client weights and that obtained by dividing the estimated number of support periods by the estimated number of clients. The differences between the two types of estimates are caused by the complex methods used to estimate the number of clients and to impute the number of non-consent support periods by consenting clients. In particular, imputation of the number of additional support periods without consent depends only on the number of support periods with consent for a client. Due to data limitations no allowance has been made for other factors which may influence the number of support periods a client has without consent. Therefore, the extent of the difference depends on how relevant the distributions for additional support periods were for certain groups of clients.

---

7. Using a 37% mixed consent rate as the base, at the national level if this is increased by b% (b = (actual mixed consent rate/0.37 − 1) × 100), then the estimated number of clients should be about $1/(1 + b \times 0.2/100)$ times the original estimate.

8. Available data did not support estimates of distributions of additional support periods beyond this level.

Although at the national level there are only very small differences in the two estimates—1.874 using weighted client data[9] compared with 1.880 from the ratio of estimates of total support periods and clients—differences can be seen for other estimates. For example, when looking at the gender split, the number of support periods for female clients divided by the number of female clients gives an estimate of 1.68 support periods per woman compared with a weighted client level estimate of 1.58. The corresponding estimates for men are 2.09 and 2.20, respectively. This suggests that it may be preferable to use different distributions to impute additional support periods for males and females. In 1996–97 there were insufficient data to impute additional support periods by gender; however, it may be possible for the 1998–99 collection as gender will be collected for all support periods.

The above discussion suggests that biases in estimates of number of clients may be as significant as any uncertainty in estimates due to the sampling effect. This source of inaccuracy could be almost completely avoided if it were possible to get alpha codes for all periods of support in participating agencies, rather than just those for which consent is obtained. If this were the case, the number of clients visiting participating agencies would be known, as would the number of support periods each client had in participating agencies. Adjustments would then have to be made only for support periods in non-participating agencies.

Unfortunately, due to resource constraints, it is not possible at this stage to explore further the accuracy of these estimates. However, because imputation has been used at the client level, in particular for support periods without consent by consenting clients, it is recommended that all estimates relating strictly to support periods should be derived using weighted support period and not weighted client data. Client-level weights should be used only when estimating numbers of clients or distributions across clients.

---

9. Weighted client data estimates in this paragraph are those from the single imputation of mixed consent presented in Section 2.2. Mean estimates based on up to 100 imputations are given in Appendix E.2.

4 Conclusion

Published estimates of SAAP usage from the Client Collection have been based on support periods with valid data for the characteristic of interest. No adjustments have been made for support periods without consent or for support periods in non-participating agencies. It is recognised by the National Data Collection Agency that the resulting estimates have significant biases. Therefore a method of adjusting for incomplete coverage of the SAAP Client Collection has been developed.

The methods derived to adjust support period and client estimates as described in this paper rely on a number of underlying assumptions concerning the random nature of obtaining consent, and the independence of consent between clients. The adjustments for non-consent, mixed consent and agency non-participation all have an underlying assumption that, within strata, support periods with consent are a random sample of all support periods and that, within groups, participating agencies are a random sample of all agencies. These assumptions allow us to infer that support periods with consent represent all support periods in SAAP agencies. If there is some systematic way in which non-consent support periods (or support periods in non-participating agencies) differ from those with consent, there will be biases in the adjusted estimates. However, at this stage it is unknown whether or not it is valid to assume that support periods with consent represent all support periods. Such validation would require detailed follow-up studies of support periods without consent.

When estimating numbers of clients, modifications to the adjustments resulting from the initial assumptions were made to allow for observed client behaviour. These modifications have been based on available data. However, in some cases the data on which to base the modifications were limited. In addition, some of these data related to a subset of support periods whose representativeness of other support periods is unknown and which could, therefore, result in some bias.

Through stratification, the methods used to develop the weights allow for some differences in the profiles of support periods with and without consent. However, there may be differences which have not been captured by the stratification, and which therefore are not allowed for in the weighting system. This, too, may lead to some biases in the estimates.

Although the data available on which to base adjustments have a number of limitations, it is quite clear that adjusted estimates better reflect SAAP usage than unadjusted numbers. Using weights it is estimated that, in 1996–97, some 83,300 clients had 156,500 support periods, averaging around 1.87 support periods per client. Adjusting estimates for support period non-consent, agency non-participation and mixed consent by clients affects not only the size of estimates but also distributions across categories. Distributions by gender are particularly affected by the adjustments.

At this stage there are not sufficient resources to examine in depth the accuracy of estimates. However, there are several known sources of error. These include the random nature, or otherwise, of the sample of support periods with consent, unknown biases in the sample that have not been allowed for and imputation related to mixed consent by clients. Inaccuracy in estimates for clients due to imputation could be greatly reduced if alpha codes could be obtained for all support periods in participating agencies. Because of biases in the imputation procedures, support period estimates should be based on support period weights rather than weighted client-level data.

Appendix A: Derivation of support period and client weights

In this appendix weights to adjust for both agency non-participation and client non-consent are derived. Weights to adjust support period estimates for agency non-participation only can be obtained using arguments similar to those used in Section A.1 below.

The derivation of adjustment factors, or weights, is divided into a number of steps. First, a basic derivation of weights, which ignores the problems of agency non-participation, mixed client consent and stratum overlap by clients, is presented. Second, the methods used to modify the weights for these three problems are discussed. Last, the final modified weights are presented.

A.1 Basic derivation of weights

In order to derive weights, a number of underlying assumptions are required. In the following discussion, four underlying assumptions are used (referred to as 'Approach 1' in the main part of this paper):

Approach 1: Independent consent for each client

Within a particular subset, or stratum, of support periods, assume that:

- If people consent on their first support period in a particular stratum then they will consent on all subsequent periods of support in that stratum. Conversely, if people refuse to give consent on their first support period, it is assumed that they will refuse on all subsequent visits in a particular stratum.
- For all clients in a particular stratum, the probability of giving consent is the same.
- A client has support periods only in one stratum.
- The probability of an agency participating is independent of other agencies and is the same within a particular group. This probability of participating depends only on characteristics of the agency.

In this first step towards finding appropriate weights, it is assumed that all agencies are participating in the collection.

Using the above assumptions, weights can be derived if it is assumed that clients who provide consent are a random sample (within strata) of all clients who receive assistance. The probability of consenting is placed on the person within a stratum, and if a person consents on the first visit it is assumed that the person will consent on all subsequent visits. Consequently, under this approach it is assumed that if a person is 'selected' then all visits are 'selected' so that the same weight can be used for both support periods and clients (compare with a household survey where all persons in a selected dwelling are selected, and the dwelling and person weights are therefore the same).

The theoretical derivation of the weights is given below. The following notation is used.

Variables:

p = probability of a client giving consent

w = weight

c = number of consenting clients in participating agencies

C = total number of clients in participating agencies

S = number of support periods for all clients in participating agencies

s = number of support periods for consenting clients only, in participating agencies

Identifiers (or subscripts):

h = number of visits by a client, $h = 1...H$

j = stratum used for non-consent/response adjustment, $j = 1...J$

For example, c_j is the number of consenting clients in stratum j.

The strata represented by j are groups of support periods within which there is a constant probability of consent. These groups and probabilities are not known. Therefore, some analysis is required to try to determine an appropriate stratification to estimate probabilities of consent. Such a stratification should be based on factors that influence consent and characteristics of support periods. The method used in this study to identify these factors and to define a stratification that could be used in practice is described Appendix B.

The following derivation applies to each stratum separately. Therefore, for clarity the stratum subscript j is suppressed.

Using the above notation:

s = Number of occasions of support by consenting clients in stratum

$$= \sum_{h=1...H} hc_h$$

and

S = Number of occasions of support for all clients in stratum

$$= \sum_{h=1...H} hC_h$$

Given that the probability of giving consent for any client in the stratum is the same and equal to p, then c has a binomial distribution $Bin(C, p)$ and c_h has a binomial distribution $Bin(C_h, p)$. Consequently, $E(c) = pC$ and $E(c_h) = pC_h$. We then have:

Expected number of occasions of support with consent in stratum

$$= E(s)$$
$$= E\left(\sum_{h=1...H} hc_h \right)$$
$$= \sum_{h=1..H} hE(c_h)$$
$$= \sum_{h=1..H} hpC_h$$
$$= pS$$

23

Appendix A: Derivation of support period and client weights

In this appendix weights to adjust for both agency non-participation and client non-consent are derived. Weights to adjust support period estimates for agency non-participation only can be obtained using arguments similar to those used in Section A.1 below.

The derivation of adjustment factors, or weights, is divided into a number of steps. First, a basic derivation of weights, which ignores the problems of agency non-participation, mixed client consent and stratum overlap by clients, is presented. Second, the methods used to modify the weights for these three problems are discussed. Last, the final modified weights are presented.

A.1 Basic derivation of weights

In order to derive weights, a number of underlying assumptions are required. In the following discussion, four underlying assumptions are used (referred to as 'Approach 1' in the main part of this paper):

Approach 1: Independent consent for each client

Within a particular subset, or stratum, of support periods, assume that:

- If people consent on their first support period in a particular stratum then they will consent on all subsequent periods of support in that stratum. Conversely, if people refuse to give consent on their first support period, it is assumed that they will refuse on all subsequent visits in a particular stratum.
- For all clients in a particular stratum, the probability of giving consent is the same.
- A client has support periods only in one stratum.
- The probability of an agency participating is independent of other agencies and is the same within a particular group. This probability of participating depends only on characteristics of the agency.

In this first step towards finding appropriate weights, it is assumed that all agencies are participating in the collection.

Using the above assumptions, weights can be derived if it is assumed that clients who provide consent are a random sample (within strata) of all clients who receive assistance. The probability of consenting is placed on the person within a stratum, and if a person consents on the first visit it is assumed that the person will consent on all subsequent visits. Consequently, under this approach it is assumed that if a person is 'selected' then all visits are 'selected' so that the same weight can be used for both support periods and clients (compare with a household survey where all persons in a selected dwelling are selected, and the dwelling and person weights are therefore the same).

The theoretical derivation of the weights is given below. The following notation is used.

Variables:

p = probability of a client giving consent

w = weight

c = number of consenting clients in participating agencies

C = total number of clients in participating agencies

S = number of support periods for all clients in participating agencies

s = number of support periods for consenting clients only, in participating agencies

Identifiers (or subscripts):

h = number of visits by a client, $h = 1...H$

j = stratum used for non-consent/response adjustment, $j = 1...J$

For example, c_j is the number of consenting clients in stratum j.

The strata represented by j are groups of support periods within which there is a constant probability of consent. These groups and probabilities are not known. Therefore, some analysis is required to try to determine an appropriate stratification to estimate probabilities of consent. Such a stratification should be based on factors that influence consent and characteristics of support periods. The method used in this study to identify these factors and to define a stratification that could be used in practice is described Appendix B.

The following derivation applies to each stratum separately. Therefore, for clarity the stratum subscript j is suppressed.

Using the above notation:

s = Number of occasions of support by consenting clients in stratum

$$= \sum_{h=1...H} hc_h$$

and

S = Number of occasions of support for all clients in stratum

$$= \sum_{h=1...H} hC_h$$

Given that the probability of giving consent for any client in the stratum is the same and equal to p, then c has a binomial distribution $Bin(C, p)$ and c_h has a binomial distribution $Bin(C_h, p)$. Consequently, $E(c) = pC$ and $E(c_h) = pC_h$. We then have:

Expected number of occasions of support with consent in stratum

$$= E(s)$$

$$= E\left(\sum_{h=1...H} hc_h \right)$$

$$= \sum_{h=1..H} hE(c_h)$$

$$= \sum_{h=1..H} hpC_h$$

$$= pS$$

An unbiased estimate for p is given by $p' = s/S$.

To estimate the total number of clients we note that:

Expected number of consenting clients

$$= E(c)$$

$$= pC \quad \text{since } c \text{ is binomially distributed.}$$

Now c is the observed number of consenting clients in the stratum, so that we can estimate the total number of clients using substitution:

$$c = p' \, C$$

Giving an estimate C' for C:

$$C' = c \, / \, p'$$

$$= c \, / \, (s/S)$$

$$= c \, (S/s)$$

$$= \text{(Number of consenting clients in stratum)} \times \text{(ratio of } total \text{ to } consenting \text{ support periods in stratum)}$$

Alternatively the estimate can be written as:

$$C' = c \, / \, p'$$

$$= S/(s \, / \, c)$$

$$= \frac{\text{Total number of support periods in stratum}}{\text{Observed average number of support periods per client in stratum}}$$

Putting the stratum subscript back in, for stratum j the above gives

$$w_j = 1/ \, p'_j = (S_j/s_j)$$

as the client weight. Since the probability of having 'consent' at a given occasion of support is also p_j, similar arguments can be used to estimate the total number of occasions of support with given characteristics. Thus each client, or occasion of support, represents $1/p_j$ other clients or occasions.

(Note that, using similar arguments, it can be shown that the same weights are appropriate for *support periods* under Approach 2 mentioned in the main text of this paper. However, under that approach, client-level weights are very difficult to derive, and require additional assumptions.)

At the stratum level the weights can be written as:

$$\frac{\text{Total number of support periods in all agencies in stratum } j}{\text{Number of support periods with consent in stratum } j}$$

The weights can be applied to observed stratum-level totals for the characteristics of interest, with the weighted stratum totals then being added together to give broad-level estimates. Alternatively, each client or support period with consent in stratum j can be assigned the weight w_j. Totals are then found by adding the weights for each client or support period with the characteristics of interest. Because different modifications are required at different levels to adjust the weights for agency non-participation, mixed consent and stratum

overlap, in the following discussion weights are presented at the client or support period level rather than the stratum level.

A.2 Modifying weights to allow for agency non-participation

Although required to, not all SAAP agencies participate in the collection. Whereas participating agencies return information on all support periods, irrespective of whether consent is obtained or not, non-participating agencies provide no information on any of their support periods. As a result, the number of support periods observed in the collection is less than the total number of support periods provided by SAAP agencies.

In 1996–97, 4.6% of agencies did not participate in the collection. For these agencies nothing is known about their clients—it is not even known how many support periods the agencies provided. To provide complete estimates of both the number of support periods and number of clients assisted under SAAP, it is necessary to adjust for this non-participation. This can be done by adjusting the support period and client weights derived from participating agencies.

When adjusting for agency non-participation it is necessary to make certain assumptions concerning the activities of non-participating agencies. It has to be assumed that:

- a client either always goes to a participating agency, or never goes to a participating agency;
- whether or not a client chooses a participating agency is a random event;
- within particular groups of agencies, whether or not an agency participates or not is also a random event;
- the probability of an agency participating is independent of other agencies and is constant within a particular group. This probability of participating depends only on characteristics of the agency.

Under these assumptions, the support periods and clients from participating agencies can be regarded as random samples from all agencies. It is therefore expected that support periods in non-participating agencies have the same profile as those in participating agencies. That is, support periods in participating agencies can be said to represent support periods in non-participating agencies.

Agency non-participation varies across State and Territory and target group (AIHW 1997: 7). Therefore, a differential non-participation adjustment is desirable. However, there are only around 1,100 agencies so a fine stratification could not be used. Consequently, non-participation rates were allowed for at the target group level only. Participation rates ranged from 93% for cross-target agencies to 99% for agencies aimed at families.

Using the above assumptions, weights for support periods and clients can be derived in the same way as before. The following notation is used.

Variables:

p = probability of a client giving consent, given that the agency is participating

θ = probability of an agency participating

w = weight

A = number of agencies

25

c = number of consenting clients in participating agencies

C = total number of clients in participating agencies

N = total number of clients in all agencies

s = number of support periods for consenting clients only in participating agencies

S = number of support periods for all clients in participating agencies

T = number of support periods for all clients in all agencies

Identifiers (or subscripts):

h = number of periods of support for a client, $h = 1...H$

t = agency grouping for non-participation adjustment (defined in terms of primary target group), $t = 1...L$

j = support period stratum for non-consent adjustment, $j = 1....J$

r = group used for combined non-consent and non-participation adjustment—r is defined in terms of j and t, denoted $r = j{:}t$

q = participating agency

For example, c_r is the number of consenting clients in group r.

Each support period can be classified according to its non-consent adjustment stratum j and agency grouping for non-participation t. This disaggregation is illustrated in Table A.1 below. The probability of a support period obtaining consent in stratum j is p_j, and the probability of the support period being in a participating agency is the probability of an agency participating, or θ_t.

Table A.1: Disaggregation of support periods with consent by stratum and agency grouping

	Non-consent adjustment stratum							Probability of participation
	1	2	.	.	j	.	J	Total	
Agency grouping									
1	$S_{1:1}$	$S_{2:1}$			$S_{j:1}$		$S_{J:1}$	$S_{\bullet 1}$	θ_1
2	$S_{1:2}$	$S_{2:2}$			$S_{j:2}$		$S_{J:2}$	$S_{\bullet 2}$	θ_2
.									
.									
t	$S_{1:t}$	$S_{2:t}$			$S_{r=j:t}$		$S_{J:t}$	$S_{\bullet t}$	θ_t
.									
L	$S_{1:L}$	$S_{2:L}$			$S_{j:L}$		$S_{J:L}$	$S_{\bullet L}$	θ_L
Total	S_1	S_2			S_j		S_J	S	
Probability of consent	p_1	p_2			p_j		p_J		

Conditional on the participating agencies, we have as before $s_r \sim \text{Bin}(S_r, p_j)$. This would suggest an estimate of s_r/S_r for p_j. However, this would give a different estimate for p_j for each r where $r = j{:}t$. A better, more accurate estimate is obtained by noting that, conditional on participating agencies, $\displaystyle\sum_{t=1...L} s_r \sim \text{Bin}\left(\sum_{t=1...L} s_r, p_j\right).$

26

Consequently,

$$E(\sum_{t=1...L} s_r) = p_j \sum_{t=1...L} S_r$$

An unbiased estimate for p_j is then:

$$p'_j = \sum_{t=1...L} s_r / \sum_{t=1...L} S_r$$

$$= s_j / S_j$$

To find an estimate for θ_t we consider the expected number of agencies that will participate. Again we use the properties of the binomial distribution since $A_{qt} \sim \text{Bin}(A_t, \theta_t)$. This gives an estimate for θ_t of $\theta'_t = A_{qt}/A_t$.

Now, conditioning on participating agencies we had $s_r \sim \text{Bin}(S_r, p_j)$. But S_r is also a random variable since it depends on which agencies participate. S_r is itself binomially distributed with $S_r \sim \text{Bin}(T_r, \theta_t)$. Thus we have:

$$E(s_r) = E(E(s_r \mid \text{participating agencies}))$$

$$= E(p_j S_r)$$

$$= p_j \theta_t T_r$$

To find an estimate for T_r we substitute in p'_j and θ'_t and the observed number of occasions of support s_r in this expectation equation. That is:

$$s_r = p'_j \, \theta'_t \, T_r \quad \text{where } r = j{:}t$$

giving an estimate T'_r for T_r as follows

$$T'_r = s_r / p'_j \, \theta'_t$$

$$= s_r / \{(s_j / S_j)(A_{qt}/A_t)\}$$

$$= s_r (S_j/s_j)(A_t/A_{qt})$$

$$\underline{\text{Occasions of support}\atop\text{with consent in } r} \times \frac{\text{Observed occasions of support in participating agencies in stratum } j}{\text{Occasions of support with consent in stratum } j} \times \frac{\text{Total number of agencies in grouping } t}{\text{Number of participating agencies in grouping } t}$$

The weight for support period i with consent is therefore:

$$w_i = 1/ p'_j \, \theta'_t$$

$$= (S_j/s_j)(A_t/A_{qt}) \quad \text{where support period } i \text{ is in } r = j{:}t$$

To estimate the total number of clients N_r, and remembering that we have assumed at this stage that clients remain within a single stratum and agency grouping, we note that:

$$c_r \sim \text{Bin}(N_r, p_j\theta_t) \text{ so that } E(c_r) = p_j\theta_t N_r$$

From before, we have estimates for p_j and θ_t. Also, c_r is the observed number of consenting clients in group r, so that we can estimate the total number of clients in group r using substitution. That is,

$$c_r = p'_j \, \theta'_t \, N_r$$

giving an estimate N'_r for N_r

$$N'_r = c_r / \, p'_j \, \theta'_t$$

$$= c_r (S_j / s_j) \, (A_t / A_{qt})$$

$$= \begin{array}{c} \text{Observed number of clients} \\ \text{with consent in participating} \\ \text{agencies in group } r = j{:}t \end{array} \times \frac{\begin{array}{c}\text{Observed occasions of support in} \\ \text{participating agencies in stratum } j\end{array}}{\begin{array}{c}\text{Occasions of support with consent} \\ \text{in stratum } j\end{array}} \times \frac{\begin{array}{c}\text{Total number of agencies in} \\ \text{grouping } t\end{array}}{\begin{array}{c}\text{Number of participating agencies} \\ \text{in grouping } t\end{array}}$$

Alternatively the estimate can be written as:

$$N'_r = c_r / c_j \, \{S_j / (s_j / c_j)\} \, (A_t / A_{qt})$$

$$= \begin{array}{c}\text{Observed proportion} \\ \text{of clients in stratum } j \\ \text{that are in group } r = j{:}t\end{array} \times \left\{ \begin{array}{c}\text{Total number of} \\ \text{support periods in} \\ \text{participating agencies} \\ \text{in stratum } j\end{array} \middle/ \begin{array}{c}\text{Observed average number} \\ \text{of support periods per} \\ \text{client in stratum } j\end{array} \right\} \times \frac{\begin{array}{c}\text{Total number of agencies} \\ \text{in grouping } t\end{array}}{\begin{array}{c}\text{Number of participating} \\ \text{agencies in grouping } t\end{array}}$$

From the above it can be seen that the weight to be added to both client and support period records in group $r = j{:}t$ is:

$$w_i = (S_j / s_j) \, (A_t / A_{qt}) \qquad \text{where support period, or client, } i \text{ is in } r = j{:}t$$

That is:

$$\begin{array}{c}\text{Weight for support} \\ \text{period or client in} \\ \text{group } r = j{:}t\end{array} = \frac{\begin{array}{c}\text{Total number of support periods in} \\ \text{participating agencies in stratum } j\end{array}}{\begin{array}{c}\text{Number of support periods with consent in} \\ \text{participating agencies in stratum } j\end{array}} \times \frac{\begin{array}{c}\text{Total number of agencies} \\ \text{in grouping } t\end{array}}{\begin{array}{c}\text{Number of participating} \\ \text{agencies in grouping } t\end{array}}$$

Similar arguments can be used to estimate the total number of occasions or clients with given characteristics. Thus each client, or occasion, represents $1/ \, p'_j \, \theta'_t$ other clients or occasions. Remember that, when weighting up to get the average number of occasions of support, each occasion represents similar occasions of support by similar clients, and not by the client observed. That is, the number of visits by a particular client is not affected by the weight.

Limitations in client weights

The above derivation of weights has several limitations when estimating the number of clients. These problems arise from the underlying assumptions. Under these, it is assumed that clients fall into distinct groups:

- clients who visit non-participating agencies
- clients who visit participating agencies and who consent
- clients who visit participating agencies and who refuse.

In addition, it is assumed that:

- each client visits only one non-consent adjustment stratum.

As a result, estimates of number of clients based on consenting clients in participating agencies are simply scaled up at the subgroup level using observed 'response rates'.

It is known that such separation does not occur in practice. Investigations show that 65% of clients with two support periods with consent use two agencies and 81% of clients with three support periods with consent use two or more agencies. Since clients are unlikely to know whether or not an agency participates in the collection, cross-use of participating and non-participating agencies seems likely. Also, analysis indicates that, in 1996–97, over one-third of support periods without consent related to clients who consented at some stage. Thus, some clients visit both participating and non-participating agencies and some clients do not always consent or always refuse. In addition, depending on the stratification used, it is quite likely that clients with multiple support periods will visit more than one stratum. As a consequence:

- N'_r will be an overestimate of the number of clients as it does not allow for clients who visit both participating and non-participating agencies;

- N' will overestimate the total number of clients as it does not allow for clients who visit more than one stratum;

- the estimated average number of visits will be too low, both because the underlying assumptions do not allow for clients to visit both participating and non-participating agencies and because they do not allow clients both to consent and refuse.

The level of overestimation will depend on the extent of clients visiting more than one stratum, the percentage of support periods without consent that are repeat support periods for consenting clients, and the actual distribution of number of support periods received by clients.

Methods of allowing for mixed consent, cross-use of agencies and stratum overlap are described below. Mixed consent and cross-use of participating and non-participating agencies are similar problems and so are dealt with together.

A.3 Modifying weights for client behaviour

The notation used in the derivation of the final weights is set out below.

Variables:

g = the proportion of non-consent support periods that relate to consenting clients (based on a 'sample' of 12,500 support periods)

R = estimate of observed support periods without consent that relate to clients who consent at some stage

$\tilde{R}$ = estimate of all support periods without consent that relate to clients who consent at some stage

s = observed support periods with consent in participating agencies only

S = all observed support periods in participating agencies only

c = number of consenting clients in participating agencies

C = total number of clients in participating agencies

29

N = total number of clients in all agencies

T = all support periods in all agencies

ν = support periods, both with and without consent (imputed), for clients who consent at some stage (that is, all support periods for consenting clients)

$\bar{o}$ = observed average number of non-consent support periods per client with mixed consent in 'sample'

$\bar{e}$ = average number of non-consent support periods per client with mixed consent

n = random number for support period between 0 and 1

α = S/T', estimated participation rate in terms of support periods

β = proportion of support periods in non-participating agencies that relate to clients who also visited participating agencies

γ = s/T', estimated overall consent rate

δ = indicator function for mixed consent

$\bar{z}$ = mean number of support periods per client

Identifiers (or subscripts):

i = non-consent support period in participating agency

u = support period with consent in participating agency

k = client with some consent in participating agency

j = stratum used to adjust for non-consent

d = grouping used to adjust for mixed consent

q = participating agencies

ξ = non-participating agencies

1 = single-stratum clients

m = multistratum client

Mixed consent and cross-use of participating and non-participating agencies

As mentioned above, there are two reasons for trying to adjust for mixed consent and cross-use of agencies by clients:

- To ensure that the number of clients is not overestimated. Ignoring mixed consent assumes that all support periods without consent relate to clients other than those who have given consent. Ignoring cross-use implies a similar assumption for support periods in participating and non-participating agencies.

- To provide better estimates of the number of support periods that clients have. If there is mixed consent, or cross-use of agencies, then the average number of visits estimated using support periods with consent will underestimate the average number of visits per client.

During 1996–97, for nearly 12,500 support periods valid alpha codes were obtained even though consent was not given. Using these cases it is possible to get an indication of the extent of mixed consent and, consequently, the number of support periods without consent

that relates to clients who at some stage have given consent. It is also possible to get adjusted estimates of the distribution of numbers of visits by clients.

Note that, because alpha codes were given by mistake for some support periods without consent and not consistently for all support periods, the data cannot be used to estimate the probability of clients consenting on particular occasions of support given that they have or have not consented previously.

Method

The method used to adjust for mixed consent is outlined below. The weights that correspond to this approach are then presented. For ease of discussion, the initial description assumes that adjustment is to be done at the national level. However, to increase accuracy, a finer level of adjustment was actually used, as is seen in the weights derivation.

Adjusting the data for mixed consent involves four steps.

Step 1

Using the 'sample'[10] of support periods with valid alpha codes but without consent, the percentage that relate to clients who have given consent at some time can be derived. By applying this to the total number of support periods without consent, the number of support periods without consent that relate to consenting clients can be estimated.

Of the 12,483 support periods with valid alpha codes but without consent, some 37% (4,570) related to clients who had also given consent at some stage. This suggests that, in 1996–97, of the observed 53,988 support periods without consent, some 19,765 related to clients who had given consent. Conversely, 34,223 periods without consent related to clients who had never given consent.

Step 2

Also needed is the number of clients to whom the 19,765 support periods relate so that adjustments can be made to numbers of visits. Ostensibly, this number can be estimated from our 'sample' by estimating the average number of visits by clients corresponding to the 4,570 support periods without consent that related to consenting clients. However, because we have only a sample of support periods without consent that have valid alpha codes, it is possible that other support periods without consent will relate to some of the same clients. That is, the average number of additional support periods without consent that relate to consenting clients from our 'sample' will tend to underestimate the true average number of additional support periods without consent that relate to consenting clients.

An estimate of the average number of all additional support periods without consent that relate to consenting clients was obtained using simulation techniques (Appendix D). These simulations suggest that the average number of additional support periods per client with mixed consent was 1.57 in 1996–97. Assuming that 95.4% of these additional support periods were in participating agencies indicates that the estimated 19,765 support periods relating to clients who had given consent were used by 13,180 clients. Consequently, of the 60,878 clients identified as having support periods through valid alpha codes with consent, 22% had additional support periods without consent, that is, they had mixed consent.

---

10. The 'sample' referred to here was a set of 12,483 support periods without consent but with valid alpha codes. The word 'sample' is put in quotes since the support periods in the set used became available due to respondent error.

In addition, the 60,878 consenting clients had an estimated 94,885 + 19,765 support periods, giving an average number of support periods per client of 1.88 rather than the observed average of 1.56. Consequently, the estimated number of clients visiting all SAAP agencies in the year was around 156,000/1.88 = 83,000, rather than the 148,873/1.56 = 95,430 obtained using data for support periods with consent in participating agencies.

Step 3

Having identified the distribution of additional support periods per consenting client with mixed consent, a method is required to adjust the observed distribution of number of support periods with consent per client to allow for mixed consent. Since it is not generally known which clients have mixed consent and how many additional periods they have, this adjustment is done using randomisation techniques:

- Clients with consent are randomly assigned as having mixed consent using the proportion found in step 2 (around 20% of consenting clients overall).
- Distributions of additional support periods are fitted using simulations (see Appendix D).
- The distribution of number of additional support periods given an observed number with consent is then used to impute the number of additional support periods for a client assigned as having mixed consent.

Using the results from the above, estimates of the numbers of support periods per client for subsets of the population can be derived. Remember that the use of randomisation techniques adds another layer of uncertainty to the estimates, and that for small groups the resulting estimates of support periods per client should be used with care.

Step 4

Although the proportion of support periods in non-participating agencies relating to clients in participating agencies is unknown, an estimate can be obtained using observed support periods with valid alpha codes and subsampling agencies to simulate agency non-participation. Simulating agency non-participation indicates that in 1996–97, on average, one-third of support periods in 'non-participating' agencies related to clients of 'participating' agencies.

Using this information, in this last step, a broad level adjustment is used. It is assumed that one-third of support periods in non-participating agencies related to clients with support periods in participating agencies.

Details of weight adjustment for mixed consent and agency cross-use

The weights derived previously were of the form

$$
\begin{aligned}
\text{Weight for client in group } r &= \frac{\text{Total number of support periods in participating agencies in stratum } j}{\text{Number of support periods with consent in participating agencies in stratum } j} \times \frac{\text{Total number of agencies in grouping } t}{\text{Number of participating agencies in grouping } t} \\
&= \left(\begin{array}{c} \text{Adjustment for non-consent in} \\ \text{participating agencies} \end{array} \right) \times \left(\begin{array}{c} \text{Adjustment for agency non-} \\ \text{participation} \end{array} \right)
\end{aligned}
$$

Adjusting the non-consent factor for mixed consent

In this stage the non-consent adjustment is modified to allow for mixed consent in participating agencies.

Because there are different rates at which non-consent support periods relate to clients who consent, the mixed consent adjustment is done at a subgroup level. For example, although it is estimated that, overall, 37% of support periods without consent relate to clients who have given consent, among support periods for young people at crisis agencies the rate is about 48%.

The subgroups used to allow for mixed consent are defined in terms of:

- primary target group (6 classes)
- whether or not the client was accommodated (2 classes)
- State/Territory (9 classes, including a proclaimed place indicator for New South Wales)
- service model (8 classes).

(See Appendix C for derivation of this stratification.)

The number of non-consent support periods in stratum j in participating agencies that relate to consenting clients is estimated by randomly assigning non-consent support periods in subgroup d as relating to consenting clients. The probability used to assign mixed consent randomly is g_d, the estimated proportion of non-consenting support periods in group d that relate to consenting clients. That is, using random allocation[11] to assign 'mixed' consent, we have:

$$\delta_i = \begin{cases} 1 & \text{if } n_i \leq g_d \text{ so that period } i \text{ is imputed to relate to a consenting client} \\ 0 & \text{otherwise so that period } i \text{ is imputed not to relate to a consenting client} \end{cases} \quad \text{for } i \text{ in } d$$

When deriving the weight for clients, all support periods relating to consenting clients are then treated as if they were with consent. Consequently, using the logic in the basic derivation of the weights, the modified non-consent part of the weight is given by:

$$\frac{\text{Total number of support periods in participating agencies in stratum } j}{\text{Number of support periods relating to consenting clients, in participating agencies in stratum } j}$$

$$= S_j / v_j \quad \text{where } v_j = R_j + s_j \quad \text{with } R_j = \sum_{i \in j} \delta_i$$

The client weights modified for non-consent and mixed consent only are then:

$$w_k = (S_j / v_j) \quad \text{where client } k \text{ is in stratum } j$$

Adjusting for cross-use of participating and non-participating agencies

Having adjusted for mixed consent, we now adjust for agency non-participation, including cross-use of agencies.

There are no direct data on cross-use of participating and non-participating agencies by clients. However, an estimate of the extent of cross-use can be obtained by simulating cross-use using data from participating agencies. For 1996–97, this was done by randomly allocating 4.6% (the 1996–97 participation rate) of agencies with data as being

---

11. This random allocation allows different stratifications to be used for non-consent and mixed consent adjustment.

non-participating and seeing how many support periods related to clients in the other 95.4% of agencies. (Note that only support periods with valid alpha codes can be used in this analysis.) By repeating the simulation a number of times, an average estimate for the extent of cross-use can be obtained.

Using 30 simulations, it was found that on average one-third (34%) of support periods in 'non-participating' agencies related to clients of 'participating' agencies. The actual rate depended on which agencies were said to be non-participating, with individual estimates ranging from 23% to 50%.

We have from before the estimated total number of support periods:

$$T' = \sum_r T'_r$$
$$= \sum_{r=j!} s_r (S_J / s_j)(A_t / A_{qt})$$

Also, we know S, the total number of support periods in participating agencies.

If there are C_q clients in participating agencies, then these have $\{S + \beta(T - S)\}$ support periods where β is the proportion of support periods in non-participating agencies used by consenting clients. This gives an estimated mean number of support periods per client of $\bar{z}$ = $\{S + \beta(T - S)\}/C_q$, and consequently

$C_\xi = (1 - \beta)(T - S)/\bar{z}$ additional clients.

The total number of clients is then

$$N = C_q + C_\xi$$
$$= C_q + (1 - \beta)(T - S)/\bar{z}$$
$$= C_q + C_q (1 - \beta)(T - S)/\{S + \beta(T - S)\}$$
$$= C_q \{S + \beta(T - S) + (1 - \beta)(T - S)\}/\{S + \beta(T - S)\}$$
$$= C_q T/\{S + \beta(T - S)\}$$

Estimating T by T' and putting $\alpha = S/T'$ gives:

$$N = C_q/\{\alpha + \beta(1 - \alpha)\}$$

$1/\{\alpha + \beta(1 - \alpha)\}$ can therefore be used to adjust the client weights derived from participating agencies for non-participation.

Now C_q is derived using the weights $w_k = (S_j/v_j)$. The client weights modified for cross-use of participating and non-participating agencies as well as mixed consent are then:

$$w_k = (S_j/v_j)/\{\alpha + \beta(1 - \alpha)\} \text{ where } k \text{ is in stratum } j$$

In words, this equates to:

$$\text{Weight for client in stratum } j = \frac{\text{Total number of support periods for clients in participating agencies in stratum } j}{\text{Number of support periods in participating agencies in stratum } j \text{ relating to clients who consent at some time}} \times \frac{\text{Estimated total number of support periods}}{\text{Estimated number of support periods relating to clients that visit participating agencies}}$$

The adjustment for agency non-participation for clients is at the national level rather than an agency group level as is used for support periods. This is both because of lack of data and because clients cannot inherently be assigned to particular agency groups.

To give an example of the adjustment, consider the case where $\alpha = 0.95$ and $\beta = 1/3$ (similar to 1996–97), then:

$$1/\{\alpha + \beta(1 - \alpha)\} = 1.034$$

Thus instead of increasing the number of clients by 5% for non-participation, the number of clients is increased by 3.4%.

The agency non-participation adjustment factor also has implications for the distribution of number of visits by clients. A method for allowing for this as well as for mixed.consent is included when adjusting the number of support periods per client.

Adjusting the number of support periods per client

Data from the 'sample' suggest that the number of additional support periods (without consent) for a consenting client has an exponential distribution. Using a range of exponential distributions for additional support periods, several sets of support periods were set up. For these sets of support periods, whether or not a period related to a consenting client and how many related to the same client were modelled. Random samples of support periods were then drawn from these populations matching the estimated sample fraction (20%) that the sample of 12,483 was of all non-consenting support periods including those from non-participating agencies. By comparing the average number of 'observed' additional support periods relating to consenting clients from these simulated sets to the average number of observed additional support periods relating to consenting clients in our 'sample', an appropriate distribution underlying the number of additional support periods for consenting clients was identified (see Appendix D for more details).

Now, there are $R_d = \sum_{i \in d} \delta_i$ support periods without consent in group d in participating agencies that relate to clients with consent. (The expected value of R_d is $g_d(S_d - s_d)$). In addition, there are some consenting clients with support periods in non-participating agencies. The number of support periods involved can be estimated as follows. Of the estimated $(T' - S)$ support periods in non-participating agencies, it is expected that $\beta(T' - S)$ are for clients of participating agencies. Assuming similar proportions relate to consenting clients as did for participating agencies, of these support periods an estimated proportion $R/(S - s)$ are for consenting clients. Again assuming that support periods from participating and non-participating agencies have similar patterns, the expected number of support periods from non-participating agencies that are for consenting clients in group d is

$$\beta(T' - S) \times R/(S - s) \times R_d/R = \beta(T' - S)R_d/(S - s)$$
$$= R_d \beta(1 - \alpha)/(\alpha - \gamma)$$

where $\gamma = s/T'$ is the estimated overall consent rate for support periods (60%).

The total number of non-consenting support periods in group d that are for consenting clients is then given by:

$$\tilde{R}_d = R_d + R_d \beta(1 - \alpha)/(\alpha - \gamma)$$
$$= R_d\{1 + \beta(1 - \alpha)/(\alpha - \gamma)\}$$

(As an example, if $\beta = 1/3$, $\alpha = 0.95$ and $\gamma = 0.6$ then $\tilde{R}_d = 1.048R_d$).

The number of clients $\tilde{c}_d$ that these $\tilde{R}_d$ support periods relate to is estimated as:

$$\tilde{c}_d = \tilde{R}_d / e'_d \qquad \text{where} \quad e'_d \text{ estimates } \bar{e}_d .$$

35

e'_d is obtained from the 'sample' estimate $\bar{o}_d$ of observed average number of additional support periods per client in group d,[12] adjusted for non-'sample' non-consent support periods. The adjustment is found using the exponential distribution of 'additionals' fitted to the 'additionals' observed from the 'sample' (mentioned above), and, when deriving it, it is assumed that each group d has the same distribution of support periods with consent per client as all clients with consent (see Appendix D). The adjustment is the same for all d and is:

$$\frac{\text{Fitted actual mean number of additionals per client with mixed consent}}{\text{Fitted 'observed' mean number of additionals per client with mixed consent}}$$

so that $e'_d = \bar{o}_d \times \bar{e}/\bar{o}$ ($\bar{e}/\bar{o} = 1.573/1.135$, see Table D.2).

In the same way that support periods without consent were imputed to relate to consenting clients, clients with consent are imputed to have additional support periods. The probability used to impute the presence of additional support periods is $\tilde{c}_d/c_d$ where c_d is the number of clients with consent in group d.

The numbers of additional support periods for clients imputed to have mixed consent are also imputed. In this case, the number of additional support periods is imputed depending on the number of support periods with consent that have been observed. This is done using exponential distributions based on the 'sample' distribution (Appendix D). The imputation is not carried out within the subgroups used to adjust for mixed consent because the 'sample' used to derive the distributions was not large enough to support such fine level estimation of distributions. For the same reason, only one distribution of additional support periods was used for clients with seven or more support periods with consent.

Stratum overlap

As shown in the main part of this paper, the number of clients is overestimated if no allowance is made for clients who have support periods in more than one non-consent adjustment stratum. That is, the estimates will overstate the number of clients if stratum overlap is ignored. The problem of stratum overlap can be avoided by ensuring that clients do not have support periods in different strata.

Overlap can be eliminated by collapsing strata until no more overlap occurs. However, since some clients visit agencies in more than one State or Territory (5% in 1996–97), some use agencies in different target groups (7% in 1996–97), and so on, collapsing strata in this way will lead to a much reduced stratification which may not be able to adjust for the differing profile of periods with and without consent. An alternative method is to assign all clients with support periods in more than one stratum to a stratum of their own. However, this also can reduce the effectiveness of the stratification.

A combination of both approaches is adopted here. First, strata with similar non-consent rates are combined to reduce the incidence of stratum overlap (see Appendix B). Second, if overlap still occurs, clients with support periods in more than one stratum ('multistratum' clients) are put into a stratum of their own. The resulting weights are derived below.

---

12. Clients are allocated to group d according to their last support period with consent.

Single-stratum clients

Using the same arguments as before, the weight for single-stratum clients should be:

$$
\text{Weight for single stratum clients in stratum } j = \frac{\text{Total number of support periods for single-stratum clients in participating agencies in stratum } j}{\text{Number of support periods in participating agencies in stratum } j \text{ relating to single-stratum clients who consent at some time}} \times \frac{\text{Estimated total number of support periods in all agencies}}{\text{Estimated number of support periods relating to clients who visit participating agencies}}
$$

Using the notation from before, this is written as:

$$
w_{j1} = \frac{S_{j1}}{v_{j1}} \times \frac{1}{\{\alpha + \beta(1 - \alpha)\}}
$$

Assuming that the ratio of support periods for all consenting clients to those for single-stratum consenting clients is the same as the ratio of all support periods in the stratum to all those relating to single-stratum clients, that is, $v_j/v_{j1} = S_j/S_{j1}$, gives an estimate for S_{j1}/v_{j1}. The weight is then:

$$
w_{j1} = \frac{S_j}{v_j} \times \frac{1}{\{\alpha + \beta(1 - \alpha)\}}
$$

That is, the weight for single-stratum clients remains unchanged.

Multistratum clients

Similarly, the weight for multistratum clients should be:

$$
\text{Weight for multistratum clients} = \frac{\text{Total number of support periods for multistratum clients in participating agencies}}{\text{Number of support periods in participating agencies relating to multistratum clients who consent at some time}} \times \frac{\text{Estimated total number of support periods in all agencies}}{\text{Estimated number of support periods relating to clients who visit participating agencies}}
$$

That is:

$$
w_m = \frac{S_m}{v_m} \times \frac{1}{\{\alpha + \beta(1 - \alpha)\}}
$$

Again assume that the ratio of support periods for all consenting clients to those for multistratum consenting clients is the same as the ratio of all support periods in the stratum to all those relating to multistratum clients, that is, that $v_j/v_{jm} = S_j/S_{jm}$. This gives an estimate for S_{jm}. The total number of support periods in participating agencies relating to multistratum clients can then be estimated by:

$$
S_m = \sum_j \dot{S}_{jm}
$$
$$
= \sum_j v_{jm} \times S_j/v_j
$$

37

where $v_{jm} = v_j - v_{jl}$ and v_{jl} is derived using imputed 'additionals' for mixed consent clients. The non-consent adjustment section of the weight for multistratum clients is then:

$$S_m / v_m = S_m / \sum_j v_{jm}$$

$$= \sum_j (v_{jm} \times S_j / v_j) / \sum_j v_{jm}$$

$$= \sum_j (S_j / v_j)(v_{jm} / \sum_j v_{jm})$$

This is the weighted average of the single-stratum 'non-consent' client weights, where the weighting is proportional to the number of support periods in stratum j relating to multistratum consenting clients.

The full weight for multistratum clients is then:

$$w_m = \frac{\sum_j S_j v_{jm} / v_j}{\sum_j v_{jm}} \times \frac{1}{\{\alpha + \beta(1 - \alpha)\}}$$

A.4 Summary

The final weights to be used to obtain estimates adjusted for both agency non-participation and client non-consent are as follows.

Support periods u with consent in subset $r = j:t$

$$w_u = w_r = \frac{S_j}{s_j} \times \frac{A_t}{A_{qt}}$$

Single-stratum clients k with consent in subset j

$$w_k = w_{jl} = \frac{S_j}{v_j} \times \frac{1}{\{\alpha + \beta(1 - \alpha)\}}$$

Multistratum clients k with consent

$$w_k = w_m = \frac{\sum_j S_j v_{jm} / v_j}{\sum_j v_{jm}} \times \frac{1}{\{\alpha + \beta(1 - \alpha)\}}$$

Tables A.2 and A.3 list the average full weights for support periods by non-consent adjustment stratum and the final weights for clients. Averages are presented for the 46 non-consent adjustment strata for support periods for brevity. In fact, there are 175 different support period weights, corresponding to the 46 non-consent strata by six primary target groups used for agency non-participation adjustment. (Some combinations do not occur, resulting in 175 rather than 276 distinct weights.)

To obtain estimates of numbers of support periods or clients at an aggregated level, weights from contributing support periods with consent or clients with some consent are simply

38

summed. For example, to estimate the total number of support periods for women aged 21, the weights of support periods with consent for women aged 21 are added together:

$$T'_{women,21} = \sum_{u=women,21} w_u \qquad \text{where } w_u = w_r \text{ for support period with consent } u \in r$$

To estimate the number of women clients aged 21 we use:

$$N'_{women,21} = \sum_{k=women,21} w_k$$

$$\text{where } w_k = \begin{cases} w_{j1} & \text{for single - stratum clients with some consent, } k \in j \\ w_m & \text{for multistratum clients with some consent} \end{cases}$$

For variables where an average (or total) is required, weighted averages (or totals) are derived. For example, consider average length of support. Denoting the length of support period u as x_u, then $\bar{x}$, the average length of support, is:

$$\bar{x} = \sum_u x_u w_u \Big/ \sum_u w_u \qquad \text{where } w_u = w_r \text{ for support period with consent } u \in r.$$

Table A.2: Average full support period weights, by stratum for non-consent adjustment, 1996–97

Non-consent adjustment stratum[a]	Number of support periods with consent	Average support period weight[b]
0	14	109.74
6	53	14.66
10	67	9.98
12	278	8.02
14	22	6.71
16	663	6.10
18	188	5.41
20	362	4.87
22	86	4.85
24	477	4.15
26	11	4.02
28	634	3.73
30	439	3.42
32	347	3.10
34	554	3.08
36	783	2.83
38	361	2.66
40	643	2.54
42	1,114	2.45
44	2,154	2.30
46	1,298	2.16
48	2,402	2.15
50	1,412	2.04
52	333	1.96
54	622	1.91
56	2,249	1.82
58	1,101	1.79
60	3,360	1.73
62	1,583	1.66
64	4,946	1.63
66	4,453	1.58
68	4,303	1.52
70	4,962	1.49
72	4,988	1.44
74	2,422	1.39
76	9,077	1.37
78	6,878	1.33
80	2,989	1.30
82	2,345	1.26
84	1,331	1.24
86	2,881	1.20
88	1,522	1.18
90	3,659	1.17
92	5,858	1.14
94	7,835	1.12
98	826	1.06
Total	94,885	1.65

(a) See Table B.1 for consent rates within stratum.
(b) Weights vary with primary target group within strata.

Table A.3: Client weights, by stratum for non-consent adjustment, 1996–97

Non-consent adjustment stratum[a]	Number of clients with some consent	Client weight[b]
0	7	3.91
6	45	4.95
10	41	2.80
12	217	3.06
14	21	2.96
16	504	2.94
18	131	2.01
20	193	2.03
22	68	2.68
24	321	2.33
26	10	2.12
28	492	2.43
30	349	1.99
32	253	2.22
34	383	2.08
36	542	1.82
38	260	1.95
40	479	1.79
42	761	1.70
44	1,773	1.82
46	825	1.57
48	1,154	1.37
50	1,005	1.65
52	203	1.60
54	399	1.48
56	1,363	1.43
58	693	1.44
60	1,829	1.33
62	864	1.31
64	3,351	1.40
66	2,943	1.33
68	3,043	1.32
70	3,225	1.30
72	3,271	1.25
74	1,333	1.20
76	5,413	1.20
78	2,212	1.11
80	1,903	1.17
82	1,102	1.12
84	620	1.12
86	1,669	1.12
88	619	1.09
90	1,152	1.07
92	2,549	1.07
94	1,015	1.05
98	297	1.04
Multistratum	9,976	1.35
Total	60,878	1.37

(a) See Appendix B for description of strata.
(b) Based on a single imputation (see Appendix E).

Appendix B: Deriving stratification for non-consent adjustment

For best results, in terms of accuracy of weighted estimates, we want to define a stratification that allows for the factors that influence consent and the characteristics of support periods. The method used to identify these factors and to define the stratification used to adjust for non-consent is described below.

Variables important in determining whether or not consent was obtained for a support period were identified using classification tree modelling. The classification tree method divides the population into two groups which are most diverse in terms of the variable of interest (in this case, consent). The resulting two groups are then themselves divided into two in the same way, and so on until the whole population is divided up into a number of groups based on the variable of interest. (Various stopping rules are used to decide when splitting should cease.) Usually, the resulting splits are represented using a bifurcation diagram, or tree (dendrogram), hence the name of the method. The variables used to divide the population up give an indication of the variables that influence the variable of interest.

Because the results of any statistical method can be influenced by random effects, classification tree models were fitted to three distinct subsamples of about 50,000 support periods. In this way, variables used consistently as the basis for splitting in all three classification trees could reasonably be considered influential in determining whether or not consent was obtained.

B.1 Steps in deriving stratification for non-consent adjustment

A reasonable number of variables are obtained for all support periods from participating agencies. Most of them relate to whether or not a particular service was needed and provided. To avoid the confusion that can arise by having a large number of predictors, variables relating to whether or not a particular service was needed were aggregated into six summary variables. Variables relating to whether or not a particular service was provided were not used. A total of sixteen variables were considered for predicting whether or not consent was obtained:

- State/Territory (variable name *state*)
- primary target group (variable name *tgrp1*)
- service model (variable name *servmode*)
- reference source (variable name *refsrc*)
- presenting unit (variable name *famtype*)
- length of support period (variable name *length*)
- number of accompanying children (variable name *kids*)
- whether or not the client was accommodated (variable name *accmclnt*)
- whether or not the agency was a proclaimed place (variable name *procplac*)

- whether a general or high volume form was completed for the client (variable name *datasrc*)
- number of types of housing assistance needed (variable name *hsg*)
- number of types of financial assistance needed (variable name *fin*)
- number of types of counselling assistance needed (variable name *coun*)
- number of types of specialist assistance needed (variable name *spec*)
- number of types of general assistance needed (variable name *gen*)
- number of types of other assistance needed (variable name *oth*).

The allocation of types of assistance to the last six variables corresponds to published groupings (AIHW 1997: 48).

The steps in deriving the stratification to be used to adjust for non-consent were as follows.

1. Classification trees were derived for the three samples. The different samples all gave slightly different results (see attached dendrograms, Figures B1–B3). However, the first three splits were always using the variables:

    - primary target group
    - length of support period
    - whether or not the client was accommodated.

    Other variables used early in the trees (in rough order of precedence) included:

    - State/Territory
    - whether or not the agency was a proclaimed place
    - number of types of special assistance needed
    - number of types of general assistance needed
    - service model
    - number of types of housing assistance needed
    - number of types of financial assistance needed
    - reference source.

2. Using the above information, five variables were chosen for use in the stratification:

    - primary target group (6 classes)
    - length of support period (0, 1–3, 4–7, 8+ days, and unknown)
    - whether or not the client was accommodated (2 classes)
    - State/Territory (8 classes)
    - whether or not the agency was a proclaimed place.

    Of these variables, all except State/Territory could be expected to affect the characteristics of support periods, as well as the probability of consent. An additional variable was added to the stratification solely because of its importance in determining the type of assistance provided to clients. This variable was:

    - whether a general or high volume form was completed for the client.

    Consequently, a total of six variables were used to define an initial stratification to adjust for non-consent. Further variables were not included because of problems with small sample sizes in resulting strata.

3. The above stratification resulted in 559 strata with at least one support period. Overall, 196 cells had fewer than 30 support periods, including 32 cells with only one support period. Since the purpose of the stratification is to provide estimates of probabilities of consent so that periods with consent can represent periods without consent, cells with such small sample sizes do not allow sufficiently accurate (if any) estimates of either these probabilities or of characteristics of periods within the stratum. Therefore, strata with a small number of support periods were amalgamated.

- Support periods in cells with fewer than 30 support periods were reallocated to collapsed strata based on primary target group, length of support period and whether a general or high volume form was completed for the client.

The first two variables were chosen because of their importance in determining consent rates while the third was chosen because of its effect on the type of support provided. Thus, strata with small samples were collapsed across State/Territory, whether or not the agency was a proclaimed place and whether or not the client was accommodated. Small samples meant that some combining of length of support categories was also necessary, as was combining support periods using general and high volume forms for those target groups with few high volume agencies.

4. The above reallocation to new strata resulted in a total of 386 cells, all with at least 26 support periods.

5. To minimise the problem of overlap (that is, clients having support periods in more than one strata), a stratification with a small number of cells is required. Since the purpose of the stratification is to group support periods with like probabilities, if the above 386 cells can be put into groups with similar prevalences of consent the number of strata could be reduced.

Such a reduction was achieved by assigning cells to broad strata according to the estimated prevalence of consent in the cell. Strata identified in steps 1 to 4 above were assigned to broad groups based on their consent rate, with each consent rate group having a width of two percentage points for the consent rate. In addition, since estimates of prevalence at the extremes are less accurate than those in the mid-ranges (for a given sample size), to increase the accuracy of probabilities two groups with larger ranges were used at either end of the prevalence scale, resulting in a final stratification of 46 cells. Using 1996–97 data, all but two such strata had samples of more than 380. The average sample size was 3,236, with a minimum of 41 and a maximum 11,840). Table B.1 shows the sample sizes and consent rates for the final stratification for support periods for 1996–97 data.

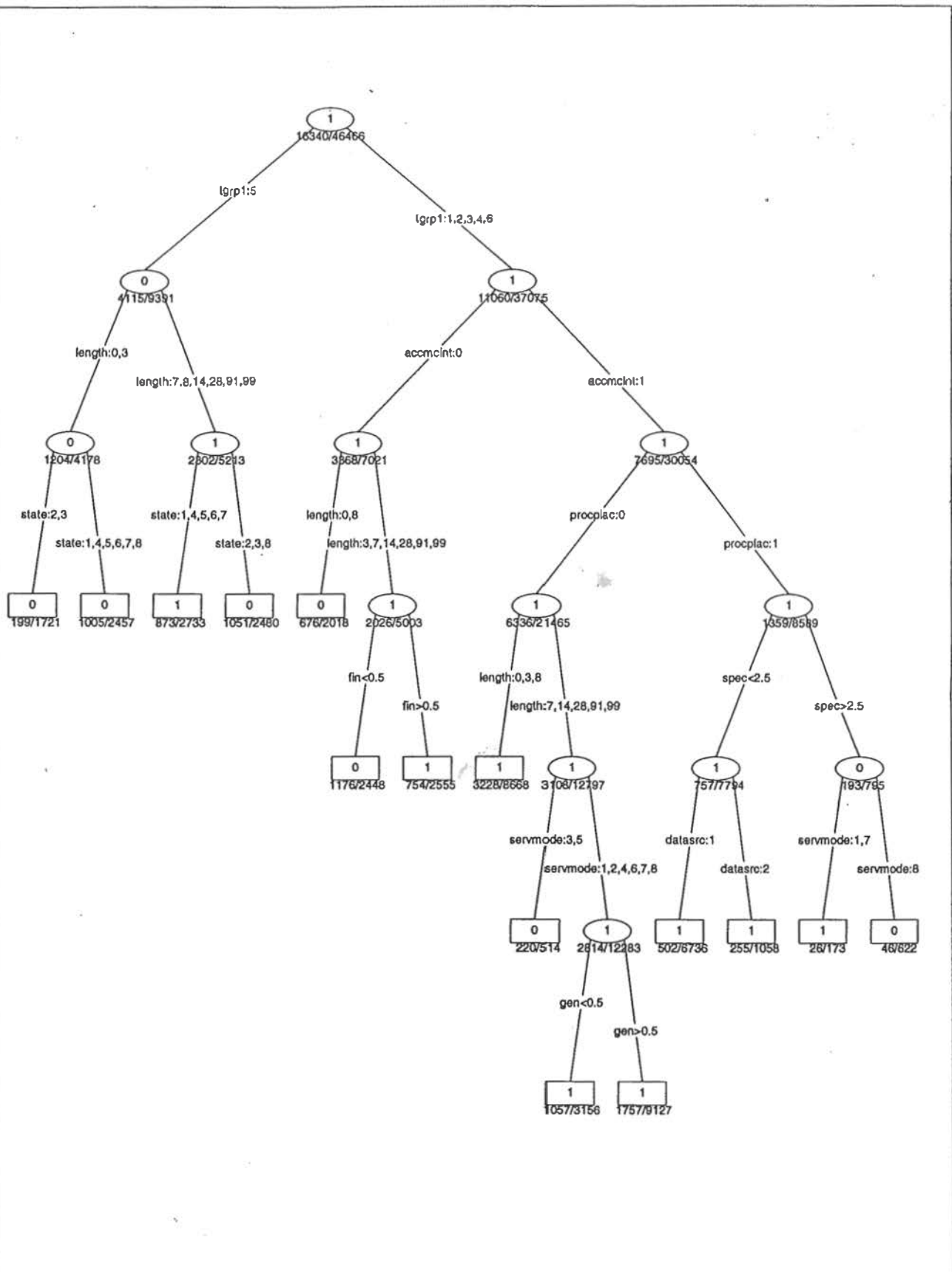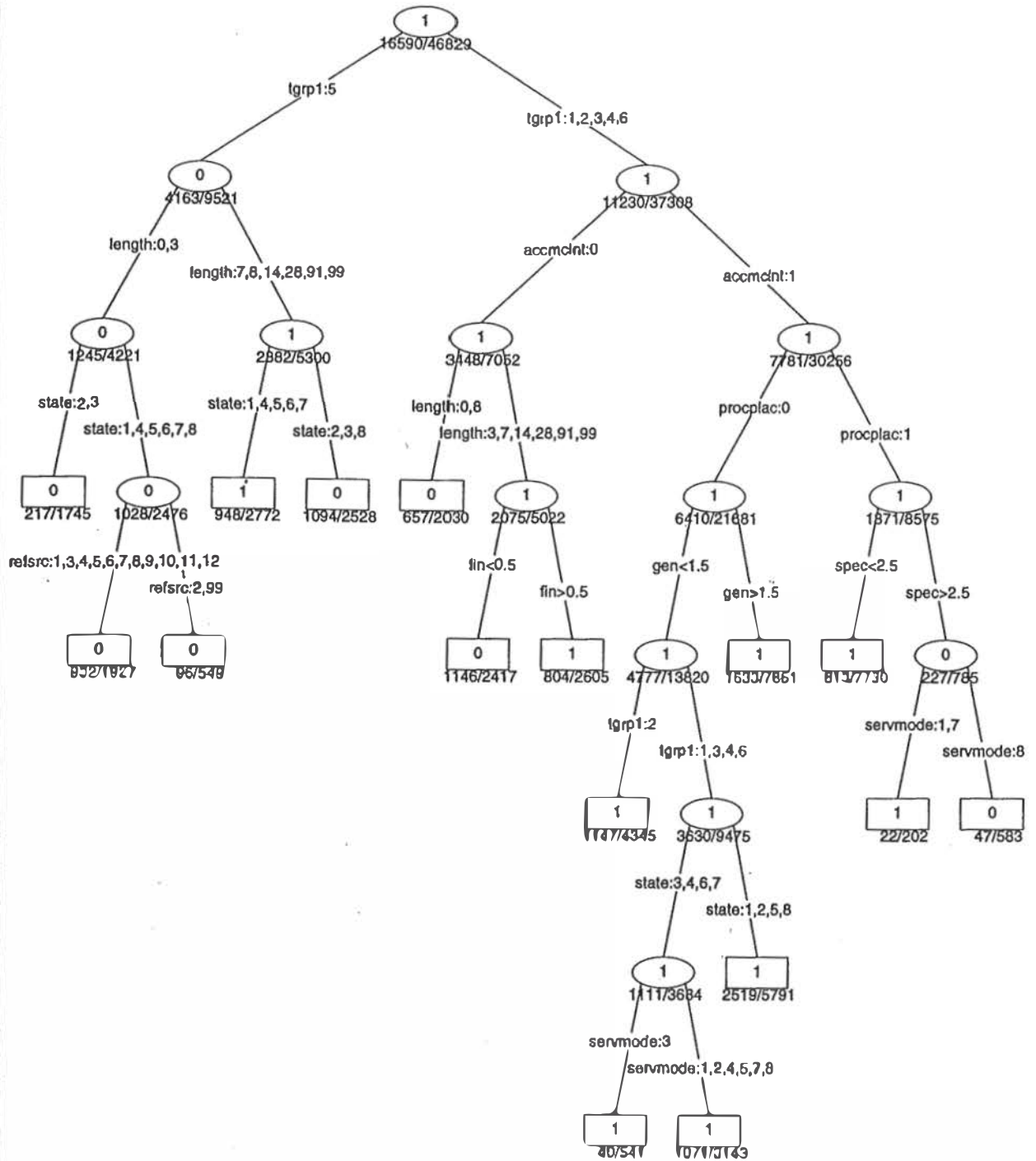**Figure B.1: Classification tree for consent, best 15 end nodes, first subsample**

Figure B.2: Classification tree for consent, best 15 end nodes, second subsample
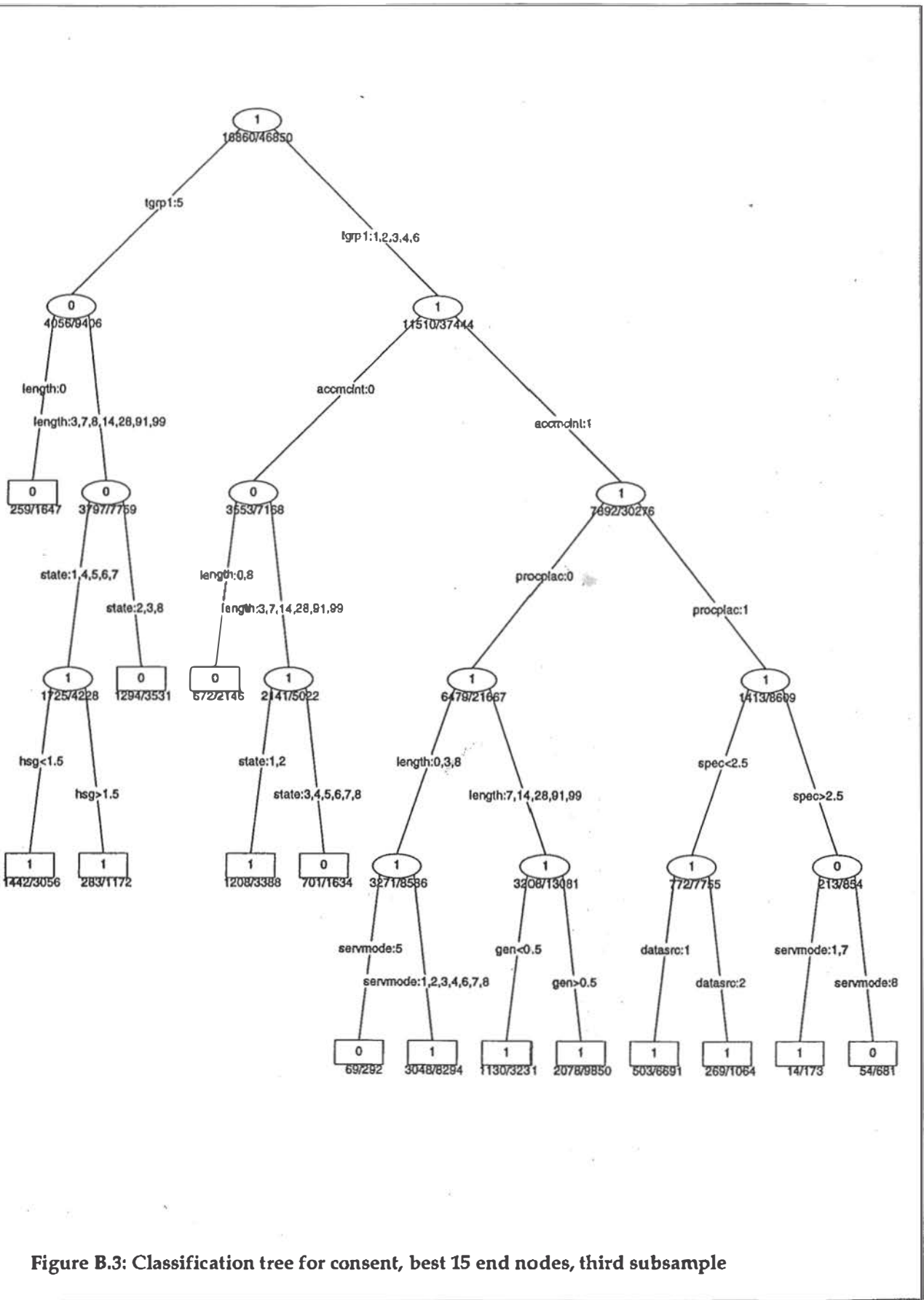
Figure B.3: Classification tree for consent, best 15 end nodes, third subsample

Table B.1: Sample size and consent rates in grouped strata for non-consent adjustment, participating agencies 1996–97

Non-consent adjustment stratum	Consent rate range for grouping fine strata (%)	Sample size	Consent rate in grouped strata
0	0 up to 6	1,478	0.009
6	6 up to 10	752	0.070
10	10 up to 12	637	0.105
12	12 up to 14	2,173	0.128
14	14 up to 16	143	0.154
16	16 up to 18	3,900	0.170
18	18 up to 20	967	0.194
20	20 up to 22	1,707	0.212
22	22 up to 24	387	0.222
24	24 up to 26	1,910	0.250
26	26 up to 28	41	0.268
28	28 up to 30	2,217	0.286
30	30 up to 32	1,424	0.308
32	32 up to 34	1,045	0.332
34	34 up to 36	1,593	0.348
36	36 up to 38	2,098	0.373
38	38 up to 40	922	0.392
40	40 up to 42	1,567	0.410
42	42 up to 44	2,585	0.431
44	44 up to 46	4,820	0.447
46	46 up to 48	2,727	0.476
48	48 up to 50	4,945	0.486
50	50 up to 52	2,767	0.510
52	52 up to 54	628	0.530
54	54 up to 56	1,137	0.547
56	56 up to 58	3,937	0.571
58	58 up to 60	1,877	0.587
60	60 up to 62	5,509	0.610
62	62 up to 64	2,499	0.633
64	64 up to 66	7,592	0.651
66	66 up to 68	6,678	0.667
68	68 up to 70	6,212	0.693
70	70 up to 72	6,968	0.712
72	72 up to 74	6,869	0.726
74	74 up to 76	3,220	0.752
76	76 up to 78	11,840	0.767
78	78 up to 80	8,671	0.793
80	80 up to 82	3,678	0.813
82	82 up to 84	2,813	0.834
84	84 up to 86	1,570	0.848
86	86 up to 88	3,315	0.869
88	88 up to 90	1,706	0.892
90	90 up to 92	4,027	0.909
92	92 up to 94	6,317	0.927
94	94 up to 98	8,168	0.959
98	98 up to 100	837	0.987
Total		148,873	0.637

Appendix C: Deriving stratification for adjusting for mixed consent

Since mixed consent was more prevalent for some groups than others, a stratification was required to refine the mixed consent adjustment. As when deriving a stratification for non-consent adjustment, classification tree models were used to derive a stratification suitable for mixed consent adjustment.

The purpose in this case was to determine which variables influenced whether or not a support period without consent is likely to belong to a client who consents at some stage. To find this out, the 12,500 cases which had valid alpha codes even though they did not have consent were used to fit the models. Initially, models were fitted to three roughly equal random subsamples so that variables which consistently influence mixed consent could be identified. However, the resulting small samples within categories led to a lot of variation in the trees. Therefore, the variables to be used for mixed consent adjustment were chosen by examining the classification tree based on the full data set.

Fourteen variables were considered for predicting how likely it was that a support period without consent related to a client who gave consent at some stage:

- State/Territory (variable name *state*)
- primary target group (variable name *tgrp1*)
- service model (variable name *servmode*)
- length of support period (variable name *length*)
- number of accompanying children (variable name *kids*)
- whether or not the client was accommodated (variable name *accmclnt*)
- whether or not the agency was a proclaimed place (variable name *procplac*)
- whether a general or high volume form was completed for the client (variable name *datasrc*)
- number of types of housing assistance needed (variable name *hsg*)
- number of types of financial assistance needed (variable name *fin*)
- number of types of counselling assistance needed (variable name *coun*)
- number of types of specialist assistance needed (variable name *spec*)
- number of types of general assistance needed (variable name *gen*)
- number of types of other assistance needed (variable name *oth*).

1. From Figures C.1 and C.2[13] it can be seen that variables used early in the tree (in rough order of precedence) included:

   - whether or not the client was accommodated
   - primary target group
   - length of support period
   - State/Territory

---

13. The sample size shown in the diagrams is less than 12,483 due to the exclusion of cases with missing data from the analysis.

- service model
- whether a general or high volume form was completed for the client.

2. Using the above information, a cross-classification using the following variables was used to define an initial stratification (length of support period was not used as it was giving confusing results, sometimes keeping non-adjacent period lengths together):

   - primary target group (6 classes)
   - whether or not the client was accommodated (2 classes)
   - State/Territory (9 classes, including a proclaimed place indicator for New South Wales)
   - service model (8 classes).

   Additional variables were not used because of problems with small sample sizes in resulting strata.

3. The above stratification resulted in 229 strata with at least one support period. Overall, 148 cells had fewer than 30 support periods. However, the purpose of the stratification is to provide estimates of probabilities of mixed consent to be used to estimate the total number of non-consent support periods relating to consenting clients. Cells with small sample sizes do not allow sufficiently accurate estimates of these probabilities; therefore, cells with fewer than 30 support periods were reallocated to strata. The reallocation was based on two variables used in early splits:

   - primary target group (6 classes)
   - whether or not the client was accommodated (2 classes).

   That is, strata with fewer than 30 support periods were collapsed across State/Territory, service model and whether or not the agency was a proclaimed place.

4. Strata were further collapsed if they contained fewer than 25 clients (based on their last support period). The same rule as above was applied. This was done because the same stratification is used when estimating how may clients correspond to support periods without consent that relate to consenting clients (see Appendix A). Strata, therefore, needed to contain reasonable numbers of clients as well as support periods.

5. The above reallocations to new strata resulted in a total of 50 cells. All but three of these had at least 25 clients (Table 3.1).

6. Note that overlap (see Appendix A) is not an issue when considering mixed consent, and so no further reduction in the number of strata for adjusting for mixed consent was carried out.

In the above stratification, the average samples were 250 support periods and 74 clients. Mixed consent rates ranged from 10% to 70% of support periods without consent.
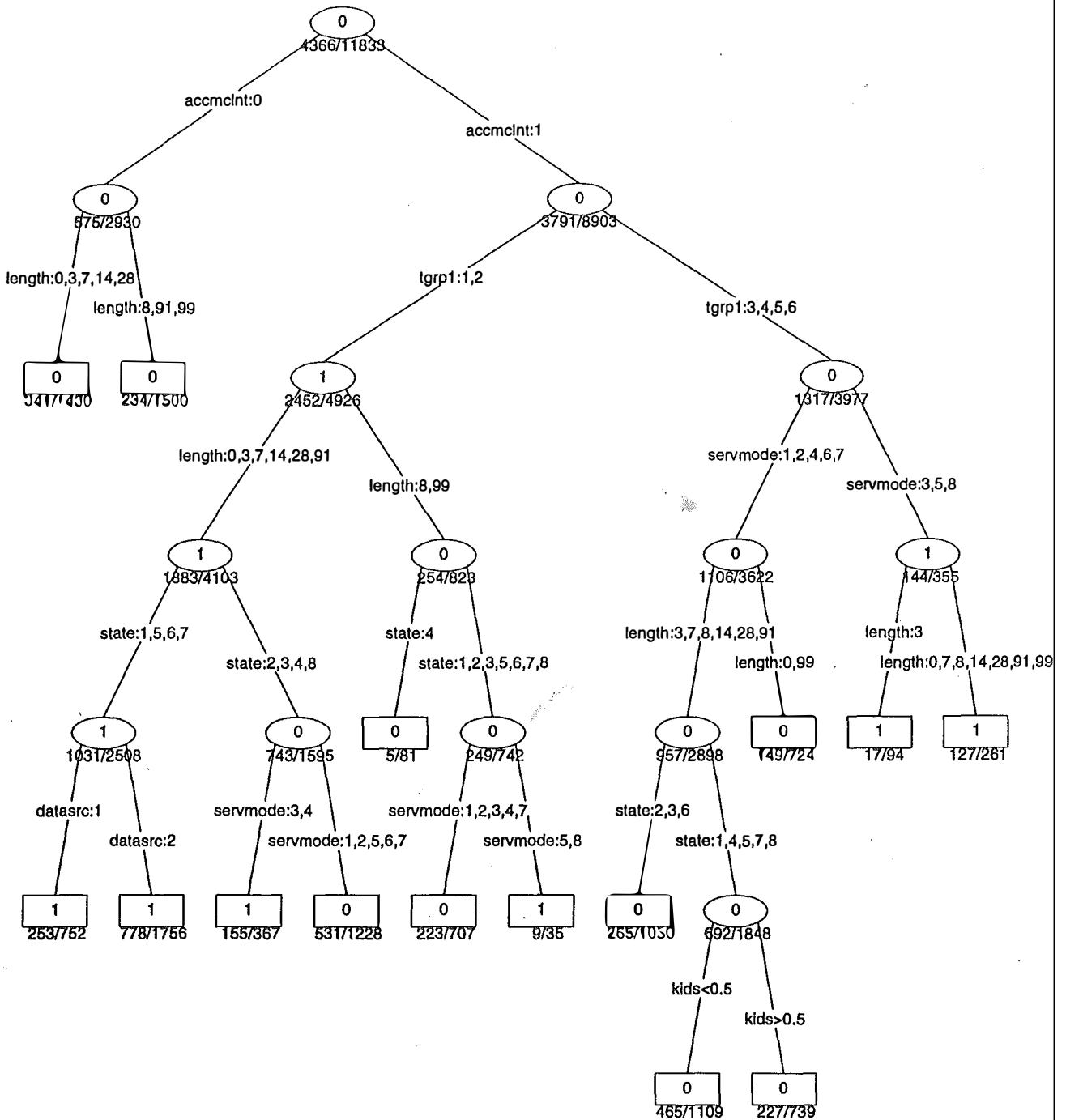
50

Figure C.1: Classification tree for mixed consent, best 15 end nodes
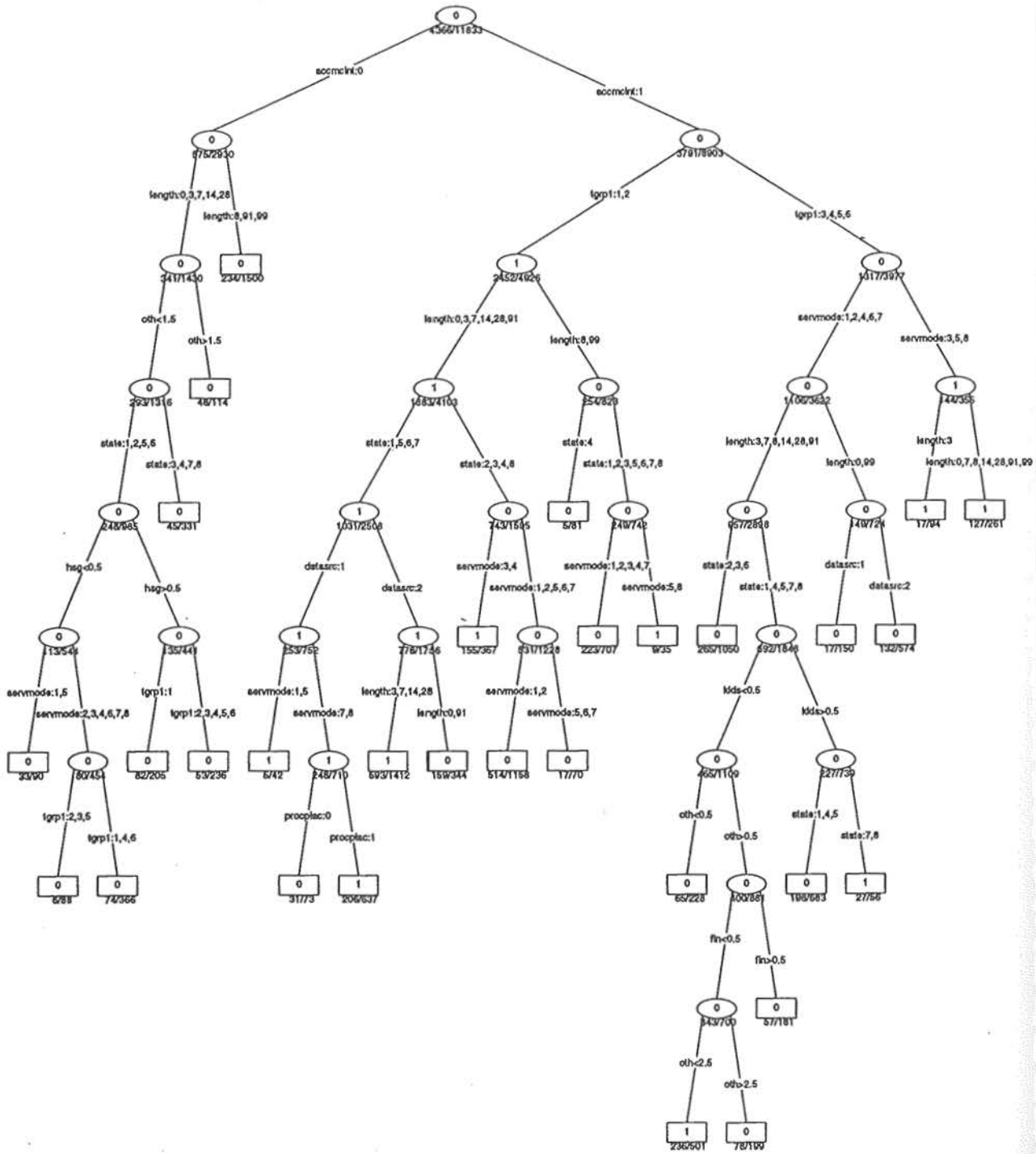
Figure C.2: Classification tree for mixed consent, best 30 end nodes

Table C.1: Sample size and mixed consent rates in stratification for mixed consent adjustment, using 12,483 support periods without consent but with valid alpha codes, 1996–97

Stratum	Number of support periods	Mixed consent rate for support periods	Number of clients	Stratum	Number of support periods	Mixed consent rate for support periods	Number of clients
1022	474	0.268	25	3199	200	0.320	37
1099	981	0.197	38	4099	303	0.116	15
1111	478	0.573	84	4199	385	0.252	87
1112	113	0.416	60	5099	382	0.165	28
1117	110	0.545	79	5111	776	0.331	163
1122	299	0.338	81	5131	160	0.206	36
1123	437	0.545	92	5141	543	0.315	147
1131	143	0.392	34	5151	91	0.275	27
1132	68	0.456	43	5181	88	0.432	37
1157	74	0.324	36	5199	380	0.266	124
1199	643	0.359	189	6022	466	0.174	109
2099	83	0.361	5	6099	281	0.199	78
2107	370	0.616	61	6107	31	0.290	30
2108	482	0.687	106	6108	158	0.709	199
2111	328	0.494	79	6111	91	0.363	86
2117	77	0.390	113	6117	47	0.298	31
2121	313	0.380	32	6122	344	0.238	148
2122	72	0.403	40	6123	34	0.441	91
2131	247	0.510	91	6131	177	0.254	89
2132	81	0.543	26	6133	61	0.426	86
2151	178	0.697	83	6137	41	0.098	39
2157	101	0.634	27	6141	247	0.466	102
2167	142	0.514	55	6153	93	0.559	77
2199	408	0.456	114	6177	47	0.468	57
3099	93	0.151	10	6199	262	0.298	167
				Total	12,483	0.366	3,693

Notes

1. The four digit stratum code is as follows
    1st digit   primary target group indicator
    2nd digit   if accommodated indicator
    3rd digit   State/Territory indicator (0 indicates proclaimed place in New South Wales; 9 signifies collapsing across all regions)
    4th digit   service model indicator (9 signifies collapsing across all models).

2. Mixed consent rate refers to the proportion of support periods without consent that relate to consenting clients.

Appendix D: Adjusting the distribution of number of support periods per client

The existence of mixed consent means that the actual average number of support periods per client is higher than observed. In addition, the spread of number of visits per client is different from that obtained using only support periods with consent.

For 1996–97, a 'sample' of 12,483 support periods without consent but with valid alpha codes was available to examine the extent of mixed consent and the number of additional support periods used by clients with mixed consent.[14] Of these 12,483 support periods, some 4,570 corresponded to 3,693 clients who consented at some stage. At first it would seem that this 'sample' could be used to estimate the distribution and average number of visits without consent for clients with mixed consent. However, because we have only a sample of support periods without consent (with valid alpha codes), it is possible that other support periods without consent also related to the same clients. That is, the average number of additional support periods without consent from our 'sample' that relate to consenting clients will tend to underestimate the true average number of additional support periods without consent for clients with mixed consent.

Table D.1: Distribution of additional support periods for clients with mixed consent, observed from 'sample', 1996–97

Number of support periods with consent	Number of support periods without consent							Client sample	
	1	2	3	4	5+	Total	Mean	Total	
	% of clients						Support periods	%	Clients
1	91.6	7.2	0.9	0.2	0.1	100	1.102	44.5	1,642
2	87.6	9.7	2.2	0.1	0.4	100	1.167	19.8	731
3	81.8	15.1	2.6	0.2	0.2	100	1.219	11.5	424
4	78.1	15.7	4.6	1.2	0.4	100	1.306	6.6	242
5	73.6	17.6	3.9	2.2	2.8.	100	1.467	4.9	182
6	66.4	22.1	7.7	1.0	2.9	100	1.577	2.8	104
7+	58.4	22.3	11.1	5.4	2.7	100	1.758	10.0	368
Total 'sample' (%)	83.9	11.6	3.0	0.9	0.7	100	1.239	100	
Sample (clients)	3098	428	109	33	25				3,693
Estimated for all clients with consent[a] (%)	89.6	8.4	1.4	0.3	0.2	100	1.118		60,878

(a) Based on the distributions of additional support periods for a given number of support periods with consent (given in this table, from the 'sample'), weighted together using the distribution of support periods with consent obtained from all clients with consent (see Table 9).

---

14. The word 'sample' is put in quotes since the support periods in the set used became available due to respondent error.

In order to overcome this bias, simulation techniques which allow for the sampling effect were used. Data from the 'sample' suggested that the number of additional support periods for a consenting client has an exponential distribution (see Table D.1, also Figure D.1). Using a range of exponential distributions for additional support periods, several sets of support periods without consent were set up. For these sets, whether or not a period related to a consenting client and how many related to the same client were modelled.

From each of the sets of modelled non-consent support periods, a 20% sample was taken. This proportion corresponds to the proportion of support periods that the 12,483 were of all support periods without consent, estimated to be 61,600 support periods. The 20% samples provide estimates of distributions of 'observed' additional support periods for clients with mixed consent for given underlying distributions. By knowing which set of modelled support periods a sample came from, the actual distribution of additional support periods can be identified. The models to be used in imputation were then chosen by comparing the average number of 'observed' additional support periods relating to consenting clients from these simulated sets with the average number of observed additional support periods relating to consenting clients in our 'sample'.

A number of distributions of 'observed' additional support periods and their corresponding actual additional support periods from these modelled populations are given in Table D.2. The distributions presented are those used to impute the number of additional support periods for clients said to have mixed consent. Different distributions were used depending on the number of support periods with consent that a client actually had. The distributions were chosen by finding that model with the mean number of 'observed' additionals closest to the mean number of observed additionals from our original 'sample', given in Table D.1. The model chosen for the overall distribution of additional support periods is compared with that originally obtained from the 'sample' in Figure D.1. The distribution of the number of actual additional support periods under the chosen model is also shown.

Having identified appropriate distributions for additional support periods for consenting clients with mixed consent, the number of additional support periods for a particular client was then imputed using the distributions in the second half of Table D.2. For example, clients with mixed consent with one support period with consent had a 65% chance of being imputed to have one additional support period, a 23% chance of being imputed to have two additional support periods, and so on.
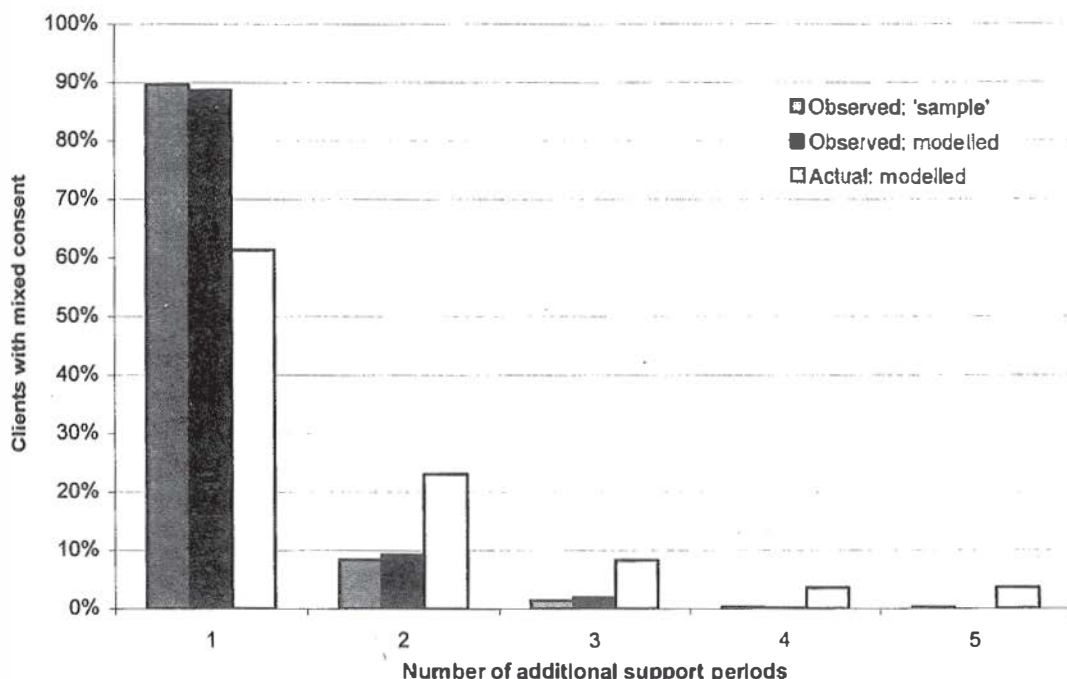
Source: Tables D1 and D2.

Figure D.1: Distributions of additional support periods for clients with mixed consent, from modelled population and 1996–97 'sample'

Table D.2: Distribution of observed and actual additional support periods for clients with mixed consent, from simulations

Support periods with consent[a]	Additional support periods without consent for consenting clients									Average number
	1	2	3	4	5	6	7	8	All	
	Proportion with number of additional support periods									
Modelled observed additional support periods										
1	0.911	0.074	0.015	0.000	0.000	0.000	0.000	0.000	1.00	1.104
2	0.853	0.131	0.015	0.001	0.000	0.000	0.000	0.000	1.00	1.165
3	0.819	0.147	0.028	0.006	0.001	0.000	0.000	0.000	1.00	1.224
4	0.747	0.207	0.040	0.004	0.001	0.000	0.000	0.000	1.00	1.304
5	0.666	0.237	0.073	0.019	0.005	0.001	0.000	0.000	1.00	1.463
6	0.589	0.269	0.105	0.032	0.005	0.001	0.000	0.000	1.00	1.596
7+	0.489	0.328	0.136	0.041	0.006	0.001	0.000	0.000	1.00	1.749
Total[b]	0.886	0.093	0.019	0.001	0.000	0.000	0.000	0.000	1.00	**1.135**
Modelled actual additional support periods										
1	0.652	0.231	0.076	0.029	0.007	0.002	0.002	0.001	1.00	1.506
2	0.554	0.247	0.105	0.054	0.027	0.006	0.004	0.002	1.00	1.752
3	0.478	0.255	0.126	0.067	0.034	0.017	0.008	0.014	1.00	1.915
4	0.385	0.230	0.148	0.093	0.057	0.035	0.020	0.032	1.00	2.194
5	0.258	0.189	0.148	0.106	0.071	0.060	0.050	0.119	1.00	2.549
6	0.178	0.150	0.114	0.099	0.084	0.061	0.059	0.255	1.00	2.841
7+	0.092	0.081	0.080	0.070	0.058	0.057	0.052	0.509	1.00	2.891
Total[b]	0.613	0.231	0.083	0.036	0.013	0.006	0.004	0.012	1.00	**1.673**

(a) Found by matching modelled 'observed' average number of additional support periods to the observed average number of additional support periods derived from the sample of 12,483 support periods (in Table D.1).

(b) Based on the distributions of additional support periods for a given number of support periods with consent (from the 'sample', given in Table D.1), weighted together using the distribution of support periods with consent obtained from all clients with consent (see Table 9).

Appendix E: Accuracy of estimates

It is very difficult to ascertain the accuracy of the adjusted estimates presented in this paper. This is because of the way in which data was, or was not, provided. In surveys it is generally possible to estimate the accuracy of results because the sampling process is controlled by the data collection agency. However, in the current situation, those support periods and agencies which gave data were not a controlled sample of all support periods and agencies. In addition, imputation techniques are used when deriving client weights.

In the following discussion, two sources of error are discussed: error due to having data for only a sample of support periods and error caused by the use of imputation techniques (in client estimates only). Biases in estimates due to the non-random nature of the sample may also be present. However, such biases cannot be measured from the sample, but need additional information on non-consenting support periods to be estimated. Such data are not currently available for the SAAP collection and so these biases are not discussed here.

E.1 Sampling error

In order to obtain adjusted estimates, the weights have been derived assuming that support periods with consent were drawn at random from all support periods within particular strata, and that participating agencies were drawn at random from among all agencies within primary target groups. Although it is not possible to provide reliable data on the accuracy of estimates of the number of support periods or clients, an indication can be obtained by looking at the estimated standard (or sampling) errors of estimates, assuming that the data came from a simple random sample of support periods and agencies. These estimates cannot provide an indication of any biases that may be present, but they are useful for identifying estimates which are likely to have relatively large errors.

Even where there is a controlled sampling process, it requires considerable effort to derive standard errors for all estimates from a sample with a complex design. In the current situation, not only do we have two levels of stratification — for agencies and for support periods — but we have the complexity added by modifications required to reflect client behaviour. That is, the accuracy of estimates is affected by the adjustments included to allow for mixed client consent, cross-use of participating and non-participating agencies, stratum overlap by clients and the imputation of additional number of support periods without consent used by consenting clients.

Because the sampling process was not controlled, there is no statistically rigorous way in which we could estimate actual standard errors. As a result, trying to derive standard errors which reflect in detail the complexity of the weighting system may only add mathematical difficulty to the problem. It is not clear that using more complicated estimation methods would give any more useful estimates of error than using a simpler method. Therefore, in the following discussion of standard errors of estimates, it is assumed that the estimates are based on unstratified simple random samples without replacement. The resulting standard errors should be considered indicative only, measuring a single source of error, and do not reflect either any biases that may be inherent in the sample or any uncertainty in the weights.

The notation used is as follows:

Variables:

A = number of agencies

f = sampling fraction

c = number of consenting clients in participating agencies

C = total number of clients in participating agencies

N = total number of clients in all agencies

P = proportion of population

s = number of support periods for consenting clients only in participating agencies

S = number of support periods for all clients in participating agencies

T = number of support periods for all clients in all agencies

σ^2 = estimated population variance (with denominator of {sample − 1})

z = number of support periods per client

Identifiers (or subscripts):

t = primary target group

q = participating agency

a = agency

g = group of interest

For support period broad level estimates

For estimates obtained using non-participation weights only (for example, estimates of the number of support periods at the national, State/Territory and primary target group level), non-participation of agencies is the sole source of sampling error. Therefore, indicative standard errors are derived assuming that the estimates were obtained from a simple random sample of agencies.

Estimates are of the form:

$$T'_g = A_g / A_{qg} \sum_a S_a$$

$$= \sum_{a \in g} S_a / f_g \qquad \text{where } f_g \text{ is the sampling fraction for agencies in group } g$$

and under simple random sampling without replacement, the variance of T'_g is given approximately by:

$$Var(T'_g) = (1 - f_g) \times A_g \times \sigma_g^2 / f_g^2$$

where σ_g^2 is the variance of the number of support periods in agencies in g.

The standard error relative to the estimate (or relative standard error) is then:

$$RSE(T'_g) = \sqrt{Var(T'_g)} / T'_g \times 100\%$$

Table E.1 gives indicative relative standard errors derived using this formula for a number of broad level estimates for 1996–97. The large variation of numbers of support periods per

agency results in sizeable estimated standard errors even though the participation rates are high. This is particularly noticeable for agencies for single men and agencies in New South Wales.

Table E.1: Indicative relative standard errors for estimates of number of support periods by State/Territory and by primary target group, 1996–97

	Participating agencies	Participation rate	Mean number of support periods per agency	Standard deviation of number of support periods per agency	Total support periods	Indicative relative standard error
State/Territory	Number	%	Number	Number	000s	%
New South Wales	350	94.2	157	602	57.9	5.0
Victoria	283	96.9	123	221	36.8	1.8
Queensland	168	94.4	142	265	25.1	3.4
Western Australia	91	96.8	127	139	12.0	2.0
South Australia	67	97.1	162	247	11.4	3.1
Tasmania	40	95.2	122	109	5.2	3.1
Australian Capital Territory	33	97.1	92	146	3.2	4.6
Northern Territory	28	93.3	165	210	4.8	6.4
Primary target group						
Young people	396	94.8	81	113	33.7	1.6
Single men only	98	95.2	377	666	38.8	3.9
Single women only	41	95.3	74	107	3.2	4.9
Families	93	98.9	79	74	7.4	1.0
Women escaping domestic violence	242	97.6	127	195	31.5	1.5
Cross-target/ multiple/general	190	92.8	205	707	41.9	6.7
Australia	**1060**	**95.4**	**140**	**389**	**156.5**	**1.8**

For support period detailed estimates

For estimates of support periods relying on data provided only if consent is given, it is reasonable to think that the primary source of error would be non-consent. Therefore, indicative standard errors can be derived assuming that we have a simple random sample of support periods. Under this assumption we have:

$$RSE(T'_g) = RSE(T'_g / T')$$

$$= RSE(P'_g) \qquad \text{with } P'_g = T'_g / T' \text{ the population proportion in group } g$$

$$= \sqrt{Var(P'_g)} / P'_g \times 100\%$$

$$= \sqrt{\left\{\frac{(1-f)}{(fT''-1)}\right\}(1-P'_g)/P'_g} \times 100\% \quad \text{where } f \text{ is the overall sampling fraction } s/T'$$

As can be seen, the indicative relative standard errors derived using this formula are not large for most estimates (Table E.2). This is because the sampling fraction is relatively high at about 60% of the estimated 156,500 support periods.

Table E.2: Indicative relative standard errors for detailed estimates, 1996–97

Estimated proportion (as %)	Estimated number of support periods	Estimated number of clients	Indicative relative standard error
	Number	Number	%
0.25	400	200	4.1
0.5	800	400	2.9
0.75	1,200	600	2.3
1	1,600	800	2.0
2	3,100	1,700	1.4
3	4,700	2,500	1.2
4	6,300	3,300	1.0
5	7,800	4,200	0.9
6	9,400	5,000	0.8
7	11,000	5,800	0.7
8	12,500	6,700	0.7
9	14,100	7,500	0.6
10	15,600	8,300	0.6
15	23,500	12,500	0.5
20	31,300	16,700	0.4
25	39,100	20,800	0.4
30	46,900	25,000	0.3
40	62,600	33,300	0.2
50	78,200	41,700	0.2
60	93,900	50,000	0.2
70	109,500	58,300	0.1
80	125,200	66,600	0.1
90	140,800	75,000	0.1

As reflected in Table E.2, in general the fewer the number of support periods that fall into a category the less reliable is the estimate for that category. Comparing Table E.2 with Table E.1 suggests that the effect of agency non-participation may be a more important source of error in detailed estimates than random non-consent. Unfortunately, due to resource constraints it is not possible at this stage to explore further the accuracy of detailed estimates. However, the small relative standard errors in Table E.2 would seem to overstate the accuracy of these estimates.

For client estimates

The accuracy of estimates of the number of clients is even harder to gauge than estimates of the number of support periods. This is because of the complication caused by the random imputation of mixed consent. Sampling as a source of error is discussed below. The effects of imputation are examined in the following section.

It is not possible to estimate sampling errors for client estimates using the method applied for broad level support period estimates. This is because clients often use more than one agency. One method of obtaining an estimate of accuracy is to consider the number of clients as the ratio of two other estimates: the number of support periods and the mean number of support periods per client. A rough estimate of the relative standard error can then be derived as follows.

For $N' = T'/\bar{z}$, where $\bar{z}$ is the estimated average number of support periods per client, we have the well-known approximation:

$$RSE(N') \cong \sqrt{\{RSE(T')\}^2 + \{RSE(\bar{z})\}^2}$$

$$RSE(N') \cong \sqrt{\frac{Var(T')}{T'^2} + \frac{Var(\bar{z})}{\bar{z}^2} - \frac{2Cov(T',\bar{z})}{T'\bar{z}}} \times 100\%$$

Assuming that the covariance term is positive, that is, as the mean number of support periods per client increases so too does the total number of support periods, an overestimate of the relative standard error is given by:

$$RSE(N') \cong \sqrt{\{RSE(T')\}^2 + \{RSE(\bar{z})\}^2}$$

From before we have $RSE(T') \cong 1.8\%$. $RSE(\bar{z})$ can be estimated roughly from the sample as:

$$RSE(\bar{z}) \cong \sqrt{\frac{(1-f)\sigma_z^2}{c}} / \bar{z} \times 100\%$$

where f is the client 'sampling fraction' and σ_z^2 is the variance of support periods per client.

Now the 'sampling' fraction f is 73%, or 60,878 out of an estimated 83,200 clients, and estimates for $\bar{z}$ and σ_z^2 based on the sample are 1.874 and 19.9 respectively. Thus we have:

$$RSE(\bar{z}) \cong \sqrt{\frac{(1-f)\sigma_z^2}{c}} / \bar{z} \times 100\%$$

$$= 0.5\%$$

Consequently :

$$RSE(N') \cong \sqrt{\{RSE(T')\}^2 + \{RSE(\bar{z})\}^2}$$

$$= \sqrt{1.8^2 + 0.5^2} \%$$

$$= 1.9\%$$

The above result suggests that the estimated relative standard errors for estimates of the number of support periods at the broad level can be used to provide an indication of the accuracy of the corresponding estimates of number of clients. Table E.3 summarises indicative relative standard errors for broad level estimates of both support periods and clients.

For estimates of the number of clients at finer levels, very crude estimates can be obtained by using a method similar to that used for detailed estimates of number of support periods. In this case, the sampling fraction is 73% of an estimated 83,200 clients. Relative standard errors are then approximated by:

$$RSE(N_R') = RSE(N_R' / N')$$

$$= RSE(P_R')$$

$$= \sqrt{Var(P_g')} / P_g' \times 100\%$$

$$= \sqrt{\left\{\frac{(1-f)}{(fN'-1)}\right\} (1-P_g')/P_g'} \times 100\%$$

where $P'_g = N'_g / N'$ is the population proportion in group g and f is the overall sampling fraction c/N'.

Since in this equation there is very little difference between the first term for support period and client relative standard errors, Table E.2 can also be used to give crude estimates of relative standard errors for client estimates. Again, the effect of agency non-participation and the effect of imputing mixed consent by clients are likely to be more important sources of error in these estimates than random non-consent. The indicative relative standard errors presented in Table E.3 suggest that the estimates in Table E.2 might lead to overconfidence in the accuracy of fine level estimates.

Table E.3: Indicative relative standard errors by State/Territory and by primary target group, 1996–97

	Estimated number of support periods[a]	Estimated number of clients[b]	Indicative relative standard error
State/Territory[c]	000s	000s	%
New South Wales	57.9	25.4	5.0
Victoria	36.8	22.9	1.8
Queensland	25.1	13.9	3.4
Western Australia	12.0	7.4	2.0
South Australia	11.4	6.4	3.1
Tasmania	5.2	3.0	3.1
Australian Capital Territory	3.2	1.9	4.6
Northern Territory	4.8	2.3	6.4
Primary target group[c]			
Young people	33.7	19.7	1.6
Single men only	38.8	18.7	3.9
Single women only	3.2	2.0	4.9
Families	7.4	5.4	1.0
Women escaping domestic violence	31.5	17.2	1.5
Cross-target/multiple/ general	41.9	20.2	6.7
Australia	**156.5**	**83.2**	**1.8**

(a) Figures have been weighted to adjust for agency non-participation.

(b) Figures have been weighted to adjust for agency non-participation and client non-consent. Estimates are based on a single set of imputations for mixed consent.

(c) For client estimates, this is as on first support period in the financial year.

E.2 Imputation effects

The imputation used in the course of obtaining client weights affects the accuracy of client-based estimates. The reason for this is that, each time client weights are derived, different clients may be allocated as having mixed consent and different numbers of additional support periods per client may be imputed.

Imputation affects two types of estimates: estimates of numbers of clients and estimates of support periods per client. The effect of randomly assigning mixed consent to clients and then imputing their number of additional support periods can be gauged by repeating the imputation a number of times. Using 100 repetitions of the imputation procedures, the variation in estimates in these two areas are discussed below.

Numbers of clients

Over 100 iterations, the average estimate of the total numbers of clients was 83,261 (Table E.4). There was remarkably little variation in the estimates with all estimates lying between 82,933 and 83,469 — only a 0.6% difference. In general, for nearly all the groups in Table E.4 there was less than a 1% difference between the smallest and largest estimates over 25 imputations.

Table E.4: Estimates of numbers of clients from different imputations, 1996–97

	Number of imputations	Mean	Minimum	Maximum	Range difference	
Gender[a]		Number			Number	% mean
Female	25	42,907	42,764	43,008	244	0.57%
Male	25	39,856	39,793	39,924	131	0.33%
Age[b, c]						
Under 15	25	1,767	1,762	1,770	8	0.45%
15–19	25	18,054	18,011	18,074	63	0.35%
20–24	25	13,401	13,370	13,421	51	0.38%
25–44	25	38,566	38,465	38,631	166	0.43%
45–64	25	9,805	9,783	9,818	35	0.36%
65+	25	1,622	1,618	1,624	6	0.37%
State/Territory[c]						
New South Wales	25	25,443	25,390	25,487	97	0.38%
Victoria	25	22,881	22,805	22,933	128	0.56%
Queensland	25	13,925	13,872	13,976	104	0.75%
Western Australia	25	7,356	7,332	7,384	52	0.71%
South Australia	25	6,379	6,359	6,404	45	0.71%
Tasmania	25	3,008	2,990	3,028	38	1.26%
Australian Capital Territory	25	1,935	1,932	1,940	8	0.41%
Northern Territory	25	2,341	2,329	2,353	24	1.03%
Primary target group[c]						
Young people	25	19,684	19,603	19,745	142	0.72%
Single men only	25	18,721	18,673	18,754	81	0.43%
Single women only	25	2,034	2,024	2,043	19	0.93%
Families	25	5,370	5,339	5,387	48	0.89%
Women escaping domestic violence	25	17,289	17,141	17,389	248	1.43%
Cross-target/multiple/general	25	20,147	20,065	20,217	152	0.75%
Australia	100	83,261	82,933	83,469	536	0.64%

(a) Excludes 380 unweighted cases with gender missing.
(b) Excludes 30 unweighted cases with age missing.
(c) As on first support period in the financial year.

Support periods per client

In terms of percentage difference, estimates of support periods per client are more variable across imputations than estimates of number of clients. For the 100 imputations, the average number of support periods per client was 1.873. The variation in the average number of support periods per client was larger than that for the estimated number of clients, with individual estimates ranging from 1.865 to 1.882 — a 0.9% difference (Table E.5). At the finer level, over 25 imputations there was usually no more than a 4% difference between the

lowest and highest estimates for the classifications considered. For a few groups, the effect was somewhat higher, with estimates of support periods per client having the greatest variation at nearly 7% for clients aged under 15.

Table E.5: Estimates of average number of support periods per client from different imputations, 1996–97

	Number of imputations	Mean	Minimum	Maximum	Range difference	
Gender[a]			Number		Number	% mean
Female	25	1.580	1.568	1.589	0.021	1.33%
Male	25	2.191	2.179	2.207	0.028	1.28%
Age[b,c]						
Under 15	25	1.651	1.600	1.708	0.108	6.54%
15–19	25	1.722	1.709	1.739	0.030	1.74%
20–24	25	1.673	1.649	1.690	0.041	2.45%
25–44	25	1.893	1.886	1.900	0.014	0.74%
45–64	25	2.270	2.253	2.302	0.049	2.16%
65+	25	2.590	2.542	2.635	0.093	3.59%
State/Territory[c]						
New South Wales	25	2.431	2.419	2.448	0.029	1.19%
Victoria	25	1.503	1.490	1.514	0.024	1.60%
Queensland	25	1.705	1.679	1.719	0.040	2.35%
Western Australia	25	1.631	1.604	1.660	0.056	3.43%
South Australia	25	1.762	1.740	1.789	0.049	2.78%
Tasmania	25	1.682	1.646	1.712	0.066	3.92%
Australian Capital Territory	25	1.690	1.659	1.733	0.074	4.38%
Northern Territory	25	1.947	1.910	1.991	0.081	4.16%
Primary target group[c]						
Young people	25	1.730	1.720	1.739	0.019	1.10%
Single men only	25	2.350	2.337	2.367	0.030	1.28%
Single women only	25	1.555	1.510	1.587	0.077	4.95%
Families	25	1.319	1.307	1.332	0.025	1.90%
Women escaping domestic violence	25	1.627	1.614	1.644	0.030	1.84%
Cross-target/multiple/general	25	1.957	1.943	1.970	0.027	1.38%
Australia	**100**	**1.873**	**1.865**	**1.882**	**0.017**	**0.91%**

(a) Excludes 380 unweighted cases with gender missing.
(b) Excludes 30 unweighted cases with age missing.
(c) As on first support period in the financial year.

Note that the estimated average number of support periods per client is not derived by dividing the number of support periods by the number of clients. Rather, it is derived as the weighted average of the number of support periods per consenting client. At the national level the weighted average gives a mean estimate of 1.873, while the ratio of the two totals gives an estimate of 156,490/83,261 = 1.880. The difference is caused by the complex methods used to estimate the number of clients and to impute the number of non-consent support periods by consenting clients. Note also that it is not currently possible to determine the accuracy of the distribution of number of support periods per client.

E.3 Conclusion

It is very difficult to ascertain the accuracy of the adjusted estimates presented in this paper. This is because those support periods and agencies which gave data were not a controlled sample of all support periods and agencies.

In the above discussion, rough estimates of the sampling error of the estimated number of support periods or clients have been obtained by assuming that the data came from a simple random sample of support periods and agencies. In addition, the uncertainty in client estimates due to imputation has also been examined.

The resulting estimates of errors are indicative only of the reliability of weighted estimates. They do not measure any of the biases that may be present, but they are useful for identifying estimates which are likely to be relatively unreliable.
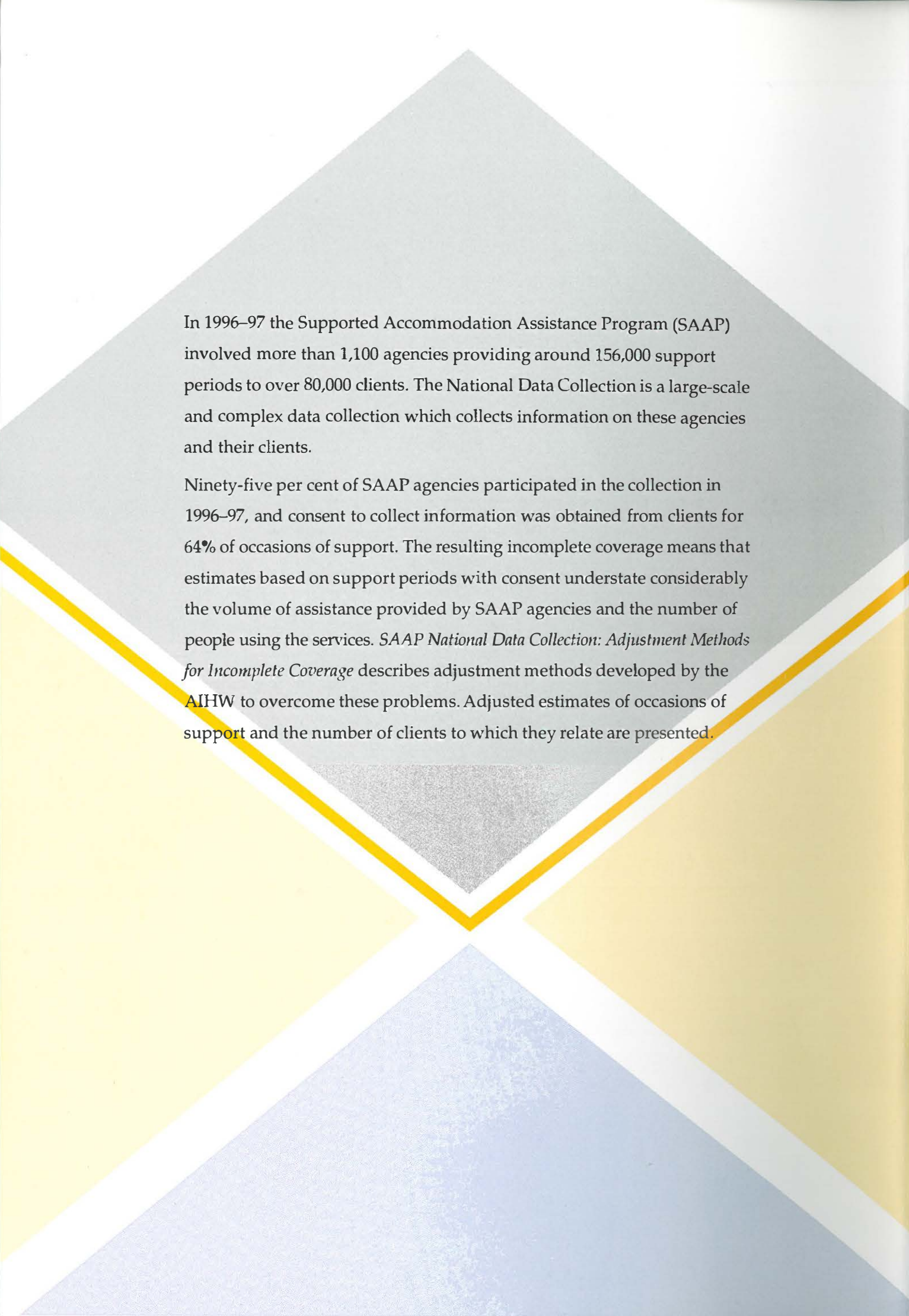
References

Australian Bureau of Statistics (ABS) 1997. Housing occupancy and costs, Australia 1995–96. ABS Cat. No. 4130.0. Canberra: ABS.

Australian Institute of Health and Welfare (AIHW) 1997. SAAP national data collection annual report 1996–97 Australia. (SAAP NDCA report. Series 2). AIHW Cat. No. HOU 10. Canberra: AIHW.

Cochran W G 1977. Sampling techniques, 3rd ed. New York: John Wiley and Sons.

Hoel Paul G 1971. Introduction to mathematical statistics. New York: John Wiley and Sons.

In 1996–97 the Supported Accommodation Assistance Program (SAAP) involved more than 1,100 agencies providing around 156,000 support periods to over 80,000 clients. The National Data Collection is a large-scale and complex data collection which collects information on these agencies and their clients.

Ninety-five per cent of SAAP agencies participated in the collection in 1996–97, and consent to collect information was obtained from clients for 64% of occasions of support. The resulting incomplete coverage means that estimates based on support periods with consent understate considerably the volume of assistance provided by SAAP agencies and the number of people using the services. *SAAP National Data Collection: Adjustment Methods for Incomplete Coverage* describes adjustment methods developed by the AIHW to overcome these problems. Adjusted estimates of occasions of support and the number of clients to which they relate are presented.