APPENDIX A1

A1 METHODS

A1.1 Survey content

A first draft of the benchmark survey questionnaire for piloting was prepared in consultation with the participating institutions in the IPM-sheep project.

A1.2 First pilot survey

The pilot questionnaire of 300 was sent out in May 2004 to four regions, including New England, Southern Queensland, Victoria and Western Australia. Addresses were chosen from a database of rural addresses selected randomly from Australian Federal Electoral Rolls. Addresses within this database were selected according to areas within each region identified as being within a 'sphere of influence' of the programs being run by regional IPM-sheep project managers. Postcodes deemed to fall within these areas provided the basis for the random selection of addresses from the Electoral Rolls.

A response rate of 24.5% (85 surveys) was achieved - this figure includes those who were ineligible (i.e. they had less than 500 sheep), as well as those who completed the survey. Eight completed surveys were received in total (response rate from 300 of 2.6% or 10% of those returned).

After four weeks a short form was sent out on 4 June to all addresses from which no response had been received. Those who had responded as either ineligible or RTS were not included in the mail-out. This abbreviated one-page survey aimed to provide information as to whether the low response rate was due to a low proportion of wool producers in the sampling frame, or to factors specific to the questionnaire content and format that were discouraging responses. In addition, a number of non–respondents in WA and Victoria were phoned shortly after the short survey was sent out. This revealed some issues that may have affected response rate. In particular, respondents in WA indicated that they were finalising their seeding operations and non-vital mail had not been looked at for several weeks. A similar situation occurred in Victoria, and it was also noted that several Victorian addresses had received two surveys from IRF in error - the other being one on foot-and-mouth preparedness, which being smaller was filled out in preference to the IPMS survey.

The short survey form achieved a response rate of 22% (48 of 218) by 25th June. Important feedback was received via e-mail from one respondent phoned as part of the pilot follow-up, and his comments were incorporated into the new version of the questionnaire.

A1.3 Analysis of first pilot survey

The completed surveys were relatively well filled in, with most responses indicating that the questions were easily understood, though some have required reworking (e.g. Q6, Q11). Several of the more detailed questions were frequently skipped or poorly answered (Qs 9, 10, 18, 26 & 34). There was no negative feedback regarding length or format of the survey, however the low response rate to the pilot was taken as an indication of this.

The response to the short survey suggested that the length and format of the full questionnaire was reducing response rates. This was indicated by several factors, including:

- the more immediate initial response to the short survey;
- the response of wool producers with well over 500 sheep to the short survey but not to the full questionnaire used in the pilot;
- indication from the same producers that they regarded IPM as being applicable to their property.

To reduce the perceived length of the questionnaire, the format was changed back to that originally specified by IRF, an A5 booklet. In consultation with the Board of Management, the survey content was altered with several questions that were too complex and time consuming to answer, removed. Other questions were rearranged to make them easier to read and answer. Further, approval was sought from AWI to use its levy-payers database. A request was placed on 21 June 2004 and the database was received on 23 August.

A1.4 Second pilot survey

The second pilot using the new questionnaire content and formatting in A5 booklet form was sent out to 300 sheep farmers using the AWI database from 27 August 2004. This second pilot achieved a response of 36 completed surveys in the first two weeks. On the basis of this relatively quick response compared to the first pilot, and without analysis of the results, it was decided to proceed with the main survey. Time was a factor affecting the decision to proceed, as well as the knowledge that the AWI database was being used and it was assumed that the target audience was being achieved. The prompt response indicated that the new format was not a problem. An initial analysis of the first 25 completed surveys confirmed that most respondents were able to understand the questions (by filling them in correctly) and that most questions were not problematic (since a majority were answered by most respondents). A total of 36 completed surveys were eventually received.

A1.5 Main survey

The addresses provided in the AWI database were from a list of postcodes provided to AWI. These postcodes were selected, as before, on the basis of the regions of influence indicated by the IPM-sheep regional project managers. Addresses were sorted by State and region basis (QLD, New England, NSW, VIC, SA & WA), then assigned random numbers. Due to there being less than 1500 addresses (the target number per state) in QLD (383), SA (751) and New England (728), all addresses provided by AWI were used in these areas. In NSW, VIC and WA the first 1500 addresses were selected from the randomised list (excluding any addresses used in the pilot). A total of 6362 addresses were selected.

The first surveys were sent out from late September over a period of several weeks, with surveys being sent to WA addresses later in the period. Reminders were sent out during the week beginning 25 October 2004 to New England, QLD, NSW, VIC and SA, with reminders sent to WA addresses the week after. A short one page letter and questionnaire (short survey) was developed in consultation with the board of management members and sent out from 25 November 2004 to those who had not responded at this time. This was to encourage non-responders to answer just a few key questions from the main questionnaire so that it was possible to analyse the extent to which there was non-response bias in the data from the full questionnaire.

Data from the surveys received up until 10 February 2005 was included in the analysis. Surveys received after this date were entered into the survey database and the data will be used in the analysis and report that follows the second report.

Figures for responses received up until 10 February 2005 are shown in Table A1.1. The total number of geographically locatable responses from respondents with 500 or more sheep in 2003 or in a typical year was 1342 full surveys and 961 short surveys.

A1.6 Coding of text answers

The full questionnaire contained 77 questions or parts of questions where the respondent could provide a text answer (rather ticking a box, or providing a numerical answer or numerical rating). In many cases, questions with tick boxes or numerical ratings of a series of items were followed by a space with “Other, please describe”. This provided a check that the series of items had not omitted something that was important to respondents. Where a small number of text answers were provided, and it could be inferred from these answers that no important item had been omitted, the test answers were used as a check on the answers to the items preceding the “Other, please describe” space.

Table A1.1. Survey response rates. Response rate is calculated as follows: the number of producers with 500+ sheep in the original mailout is estimated using the proportion of returned questionnaires with <500 sheep and 500+ sheep. The response rate is given by the number of completed questionnaires with 500+ sheep as a percentage of the estimated number of producers with 500+ sheep in the original mailout (allowing for questionnaires returned as not deliverable by Australia Post due to the addressee having left the address or not being known at the given address).

Region	No. Mailed Out	Mailed Out Less RTS	Full surveys returned 500+ sheep	Full surveys returned <500 Sheep	Short surveys returned 500+ sheep	Short surveys returned <500 sheep	Estimate of No. in Mail Out with >500 Sheep	Response Rate (full survey) (%)	Response Rate (full and short surveys) (%)
New Eng.	728	719	181	101	105	19	506	35.7	56.5
QLD	383	374	88	49	47	8	263	33.5	51.3
NSW (rem)	1500	1472	319	212	245	32	1027	31.0	54.9
VIC	1500	1472	357	215	222	24	1042	34.3	55.6
SA	751	729	202	95	104	11	541	37.3	56.5
WA	1500	1460	218	122	235	40	1075	20.3	42.1
TOTAL	6362	6226	1365	794	958	134	4456	33.6	52.1

There was only one question where text answers indicated that an item important to respondents had been omitted (question 21, concerning incidence of fly strike). In this case, the text answers were used to create another item in the list of types of strike (pizzle strike) in the survey dataset. The remaining questions with text answers required analysis in their own right and coding schemes for each question were developed in close consultation with the project participants.

A1.7 Data quality control

Data was analysed using SPSS and R (SPSS Inc, 2001;R Development Core Team, 2004). Frequency distributions of all variables in the dataset were examined (the dataset comprised a rectangular array of numbers with a row for each respondent and a column or columns for each question – each row is termed a case, and each column is termed a variable). Where values outside the expected range of values were encountered, the data was checked against the returned questionnaires for misreading or keystroke errors and corrections made where necessary. Where out-of-range values were not due to either misinterpretation of the question by the respondent or an error by the data entry operator, these were noted as possible outliers and given further consideration as to their inclusion or exclusion at the appropriate stage of the analysis.

A number of questions required specific quality control procedures. These are described in the subsections below

A1.7.1 Property area

The total property area reported by the respondent was compared with the sum of the areas under various land uses, viz. area grazed, area cropped, cropping area grazed as stubble, cropping area grazed as green and 'Other'. For 53.3 per cent of respondents the sum of areas under various land uses was equal to the area given as total property area. In these cases, it is assumed that respondents provided the land uses on the property at a particular point in time. Consequently, the figures reported under "Area grazed", "Cropping area grazed as stubble" and "Cropping area grazed as green" were summed to give the overall area grazed on the property. Similarly, the three land uses: "Area cropped", "Cropping area grazed as stubble" and "Cropping area grazed as green", were summed to provide a figure for the area cropped.

The sum of the areas of the various types of land use was greater than the total property area for 33.0 per cent of respondents. Four of these respondents had obviously made errors in reporting their total property area, possibly leaving off some digits from their answer. In these cases the total property area was set to the sum of areas and the adjustments described in the previous paragraph made.

In the remaining cases where the areas of the various types of land use was greater than the total property area, the areas entered under "Area grazed", "Area cropped", "Cropping area grazed as stubble" and "Cropping area grazed as green" referred to all or part of the same area of land, i.e. the respondent had provided figures typical of land use over time, such that there was an element of double counting, resulting in the sum of areas exceeding the total property area. Inspection of individual responses suggested that the commonest form of double counting was when "Area cropped", "Cropping area grazed as stubble" and "Cropping area grazed as green" referred to all or part of the one area of land. Consequently, "Area cropped" was let stand, while the overall area grazed was obtained by adding "Area grazed" to the greater of "Cropping area grazed as stubble" and "Cropping area grazed as green".

The remaining 13.9 per cent of respondents provided a total property area that was greater than the sum of areas. In several cases, this disparity was due to a total property area in acres being written in the space for total property area in hectares and these cases were corrected. For the remaining respondents, it appears that the cause of the disparity was the omission of some land uses from the figures provided. For this reason, the total property area provided by the respondent was taken as the total property area. Similar to the approach taken where the sum of land uses equalled the total property area, the figures reported under "Area grazed", "Cropping area grazed as stubble" and "Cropping area grazed as green" were summed to give the overall area grazed on the property. The three land uses: "Area cropped", "Cropping area grazed as stubble" and "Cropping area grazed as green", were summed to provide a figure for the area cropped.

A1.8 Non-response bias

The responses to the full and short surveys were compared for the set of questions common to both surveys to assess the extent of non-response bias in the full survey responses. The rationale for this is that, if those who responded to the full survey were systematically different in some way from those who did not respond, then the generalisation of the survey results to the overall producer population will not be valid. For example, if those who do not respond tend to have smaller flocks, then the estimate of flock size calculated from the returned questionnaires will be biased upwards.

If it is assumed that those who responded to the short survey are representative of all those who did not respond to the full survey, then comparison of the responses to the full and short surveys provides an indication of the existence of non-response bias. If there are significant differences between the full and short surveys on particular questions, then the magnitude of these differences can be used to calculate weighting factors to adjust the findings from the full survey, so that the influence of non-response bias is reduced as much as possible.

The questions for which there was a significant ($p<0.01$) difference between the full and short survey responses are shown in the tables below. The tables are presented in the order in which the questions appeared in the short survey. As the weighting procedure requires that respondents be grouped according to their responses to the questions that were common to the full and short surveys, sheep numbers were used to divide respondents into quartiles. In the case of cattle numbers, slightly over 50 per cent of respondents had no cattle and the remaining respondents were divided into three approximately equal groups according their cattle numbers. In the tables below, the numbers of respondents varies from table to table as respondents can miss answering particular questions or parts of questions.

A1.8.1 Cattle numbers

Those who did not fill in the full survey, but responded to the short survey, had significantly more cattle.

Table A1.2. Difference in cattle numbers between the full and short surveys.

Responders to ...	Proportion of respondents with cattle numbers in the ranges below (%)			
	No cattle	Less than 50	50 – 149	150 or more
Full survey	62.1	15.0	12.6	10.2
Short survey	48.1	8.6	17.3	26.0

Chi-squared test: $\chi2=128.09$, d.f.=3, $p<0.00005$, n=2274.

A1.8.2 Drench resistance test

Those who did not fill in the full survey, but responded to the short survey, were more likely to have tested for drench resistance in their flock.

Table A1.3. Difference in testing for drench resistance between the full and short surveys.

Responders to ...	% who had tested for drench resistance
Full survey	43.7
Short survey	49.8

Fisher's Exact Test, $p=0.005$, n=2272

A1.8.3 Ranking of factors important in deciding when to drench ewes

Those who did not fill in the full survey, but responded to the short survey, appear to be less convinced about the importance of faecal egg counts when deciding when to drench ewes.

Table A1.4. Difference between the full and short surveys in respondents' ranking of the importance of faecal egg count results in deciding when to drench ewes.

Responders to ...	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Full survey	59.0	17.9	9.1	14.0
Short survey	48.8	24.8	13.2	13.1

$\chi2=24.71$, d.f.=3, $p<0.00005$, n=1723.

Those who did not fill in the full survey, but responded to the short survey, also appear to be less convinced about the importance of the time of year when deciding when to drench ewes.

Table A1.5. Difference between the full and short surveys in respondents' ranking of the importance of the time of year in deciding when to drench ewes.

Responders to ...	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Full survey	54.0	32.7	9.2	4.1
Short survey	46.3	40.1	9.9	3.6

$\chi2=14.29$, d.f.=3, $p=0.003$, n=2074.

A similar pattern of response differences between the full and short survey is evident in the ranking of the importance of seasonal weather conditions in deciding when to drench ewes.

Table A1.6. Difference between the full and short surveys in respondents' ranking of the importance of seasonal weather conditions in deciding when to drench ewes.

Responders to ...	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Full survey	30.5	32.7	23.6	13.2
Short survey	23.8	38.9	24.2	13.2

$\chi 2=13.01$, d.f.=3, $p=0.005$, n=1934.

Those who did not fill in the full survey, but responded to the short survey, appear to rank pasture quality slightly higher than those who responded to the full survey.

Table A1.7. Difference between the full and short surveys in respondents' ranking of the importance of pasture quality in deciding when to drench ewes.

Responders to ...	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Full survey	16.6	34.0	27.0	22.4
Short survey	17.6	38.7	28.3	15.4

$\chi 2=15.01$, d.f.=3, $p=0.002$, n=1832.

Those who did not fill in the full survey, but responded to the short survey, appear to rank the presence of daggy sheep in the mob more highly as a factor in deciding when to drench ewes.

Table A1.8. Difference between the full and short surveys in respondents' ranking of the importance of the presence of daggy sheep in the mob in deciding when to drench ewes.

Responders to ...	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Full survey	23.7	28.3	31.9	16.1
Short survey	27.4	33.7	30.0	8.9

$\chi 2=27.62$, d.f.=3, $p<0.00005$, n=1957.

A1.8.4 Grazing strategy

Those who did not fill in the full survey, but responded to the short survey, were less likely to be following a set stocked grazing strategy.

Table A1.9. Difference between the full and short surveys in the proportion of respondents with a set stocked grazing strategy.

Responders to ...	% with set stocking grazing strategy
Full survey	55.4
Short survey	46.0

Fisher's Exact Test, $p<0.00005$, n=2223

Those who did not fill in the full survey, but responded to the short survey, were more likely to be following a grazing strategy that involved alternating between sheep and crop stubble.

Table A1.10. Difference between the full and short surveys in the proportion of respondents with a grazing strategy that involved alternating between sheep and crop stubble.

Responders to ...	% with a grazing strategy that involved alternating between sheep and crop stubble
Full survey	27.0
Short survey	38.7

Fisher's Exact Test, p<0.00005, n=2218

A1.8.5 Treatment for blowfly strike

Those who did not fill in the full survey, but responded to the short survey, were less likely to indicate that they typically treated blowfly strike by treating individual sheep that become struck.

Table A1.11. Difference between the full and short surveys in the proportion of respondents who indicated that they typically treated blowfly strike by treating individual sheep that become struck.

Responders to ...	% treating individual sheep
Full survey	75.6
Short survey	65.9

Fisher's Exact Test, p<0.00005, n=2241

A1.9 Derivation of weights for non-response bias

The preceding tables show that there are some significant differences between those who filled in the full survey and those who filled in the short survey, suggesting that estimates of the characteristics of the population of sheep producers derived from the full survey sample may be affected by non-response bias. This bias may be corrected by weighting procedures based on the differences in the tables above. However, where there are differences across a relatively large number of survey questions, the numbers of full survey respondents in the groups to which particular weighting factors are applied may become unduly small. Large weighting factors applied to small groups of respondents may introduce other biases that are not apparent from the subset of questions common to the full and short surveys. For this reason, it is necessary to rank the tables listed in the preceding section according to the magnitude of the differences exhibited and examine the size of respondent groups and weighting factors as the number of tables included in the calculation is increased to include tables with smaller differences (Table A1.12).

It can be seen from Table A1.12, that as the number of questions included in the calculation of weighting factors increases, there is also an increase in the incidence of small respondent groups with relatively large weighting factors. As might be expected, the small respondent groups are those with relatively larger cattle numbers who are pursuing a grazing strategy that involves alternating between sheep and crop stubbles. With two questions included in the calculation of weighing factors, there are only 21 respondents with 150 or more cattle and pursuing the above grazing strategy. These 21 would be multiplied by a weighting factor of 3 if the full survey data was to be adjusted for non-response bias using cattle numbers and the grazing strategy of alternating between sheep and crop stubbles. This was judged as attributing too much weight to a relatively small group of respondents. Accordingly, non-response weights were based solely on cattle numbers.

Table A1.12. Table of respondent groups and calculated weighting factors based on including one, two or three questions in the calculation. Cattle numbers show the greatest difference between the full and short surveys, followed by a grazing strategy that involves alternating between sheep and crop stubble, followed by blowfly treatment that typically involves treating individuals in the mob that become struck.

No of tables in weighting calculation	Cattle numbers	Alternating between sheep and crop stubble	Typically treat individuals that become struck	Number of respondents to full survey	Calculated weighting factor
1	No cattle			816	0.84
	Less than 50			197	0.70
	50-149			166	1.26
	150 or more			134	2.08
2	No cattle	No		532	0.68
	Less than 50	No		145	0.65
	50-149	No		128	1.13
	150 or more	No		107	1.94
	No cattle	Yes		238	1.21
	Less than 50	Yes		49	0.77
	50-149	Yes		31	1.76
	150 or more	Yes		21	3.00
3	No cattle	No	No	123	0.84
	Less than 50	No	No	39	0.76
	50-149	No	No	37	1.15
	150 or more	No	No	19	3.72
	No cattle	Yes	No	50	1.64
	Less than 50	Yes	No	9	1.75
	50-149	Yes	No	6	2.65
	150 or more	Yes	No	6	3.59
	No cattle	No	Yes	392	0.62
	Less than 50	No	Yes	101	0.61
	50-149	No	Yes	86	1.16
	150 or more	No	Yes	82	1.59
	No cattle	Yes	Yes	182	1.07
	Less than 50	Yes	Yes	40	0.54
	50-149	Yes	Yes	25	1.53
	150 or more	Yes	Yes	15	2.61

A1.10 Significance of weighted distributions

Using the weighting factors in the top four rows of Table A1.12, above, i.e. those based solely on cattle numbers, weighted frequency distributions were calculated for a selection of the questions common to the full and short surveys. The weighted and unweighted frequency distributions are shown in the tables below.

Table A1.13. Difference in sheep number (typical year) estimates with and without weighting for non-response bias.

Basis	Proportion of respondents with sheep numbers in the ranges below (%)			
	500-1499	1500-2999	3000-4999	5000 or more
Unweighted	24.4	28.2	22.5	24.9
Weighted	21.4	26.9	22.7	28.9

Chi-squared goodness-of-fit test: $\chi 2$=13.97, d.f.=3, p=0.003, n=1342.

Table A1.14. Difference in cattle number estimates with and without weighting for non-response bias.

Basis	Proportion of respondents with cattle numbers in the ranges below (%)			
	No cattle	Less than 50	50 – 149	150 or more
Unweighted	62.2	15.0	12.6	10.2
Weighted	52.3	10.6	15.9	21.2

Chi-squared goodness-of-fit test: $\chi 2$=133.65, d.f.=3, p<0.00005, n=1313.

Table A1.15. Difference between unweighted and weighted estimates of the proportion of sheep producers testing for drench resistance.

Basis	% who had tested for drench resistance
Unweighted	43.7
Weighted	45.9

Binomial test, p=0.116, n=1326.

Table A1.16. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of faecal egg count results in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	59.0	17.9	9.1	14.0
Weighted	60.7	17.4	8.7	13.3

Chi-squared goodness-of-fit test: $\chi 2$=1.08, d.f.=3, p=0.782, n=900.

Table A1.17. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of the time of year in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	54.0	32.7	9.2	4.1
Weighted	52.6	34.2	9.1	4.1

Chi-squared goodness-of-fit test: $\chi 2$=1.23, d.f.=3, p=0.745, n=1159.

Table A1.18. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of seasonal weather conditions in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	30.5	32.7	23.6	13.2
Weighted	29.8	33.7	23.6	12.9

Chi-squared goodness-of-fit test: $\chi 2=0.53$, *d.f.=3, p=0.911, n=1054.*

Table A1.19. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of pasture quality in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	16.6	34.0	27.0	22.4
Weighted	15.7	33.9	27.3	23.0

Chi-squared goodness-of-fit test: $\chi 2=0.67$, *d.f.=3, p=0.880, n=969.*

Table A1.20. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of the presence of daggy sheep in the mob in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	23.7	28.3	31.9	16.1
Weighted	22.5	27.4	33.4	16.6

Chi-squared goodness-of-fit test: $\chi 2=1.90$, *d.f.=3, p=0.594, n=1067.*

Table A1.21. Difference between the unweighted and weighted estimates of the proportion of respondents with a set stocked grazing strategy.

Basis	% with set stocking grazing strategy
Unweighted	55.4
Weighted	56.2

Binomial test, p=0.573, n=1283

Table A1.22. Difference between the unweighted and weighted estimates of the proportion of respondents with a grazing strategy that involved alternating between sheep and crop stubble.

Basis	% with a grazing strategy that involved alternating between sheep and crop stubble
Unweighted	27.0
Weighted	25.4

Binomial test, p=0.199, n=1279

Table A1.23. Difference between the unweighted and weighted estimates of the proportion of respondents who indicated that they typically treated blowfly strike by treating individual sheep that become struck.

Basis	% treating individual sheep
Unweighted	75.6
Weighted	75.9

Binomial test, p=0.770, n=1297

The preceding tables show that, apart from the estimates of sheep and cattle numbers, there is no significant difference between unweighted and weighted estimates from a range of questions about grazing and sheep parasite management. It can be concluded from this that, although sheep producers with larger numbers of cattle are significantly under-represented in the full survey sample, there appears to be little difference in grazing and sheep parasite management between those with relatively more and those with fewer cattle. Consequently, adjustment for the under-representation of sheep producers with larger numbers of cattle has no significant effect on the estimates of characteristics associated with grazing and parasite management.

However, these findings then raise the question, if weighting was based on one or more of the questions about grazing and parasite management, whether the adjustment for non-response biases shown by these questions would lead to weighted estimates that were significantly different from unweighted estimates. Table A1.24 shows the size of respondent groups and weighting factors for the three questions about grazing and parasite management that showed the greatest differences between the full and short surveys. The possibility of using a fourth question was investigated, however, because the next question in the sequence had four categories, this resulted in unsatisfactorily small respondent groups.

Table A1.24. Table of respondent groups and calculated weighting factors based on including one, two or three questions relating to grazing and sheep parasite management in the calculation. A grazing strategy that involves alternating between sheep and crop stubble shows the greatest difference between the full and short surveys, followed by blowfly treatment that typically involves treating individuals in the mob that become struck, and a set stocked grazing strategy.

No of tables in weighting calculation	Alternating between sheep and crop stubble	Typically treat individuals that become struck	Set stocked grazing strategy	Number of respondents to full survey	Calculated weighting factor
1	No			934	0.89
	Yes			345	1.31
2	No	No		225	0.68
	Yes	No		73	0.65
	No	Yes		676	1.76
	Yes	Yes		266	3.00
3	No	No	No	89	1.30
	Yes	No	No	45	2.21
	No	Yes	No	268	0.87
	Yes	Yes	No	154	1.22
	No	No	Yes	136	1.00
	Yes	No	Yes	28	1.38
	No	Yes	Yes	408	0.78
	Yes	Yes	Yes	112	1.00

Table A1.24 shows that three grazing and parasite management questions can be used to calculate weighting factors, without resulting in unduly small respondent groups or unduly large weighting factors.

Using the weighting factors in the lower eight rows of Table A1.24, above, i.e. those based on the three grazing and parasite management questions with the greatest difference between the full and short surveys, weighted frequency distributions were calculated for a selection of the questions common to the full and short surveys. The weighted and unweighted frequency distributions are shown in the tables below.

Table A1.25. Difference in sheep number (typical year) estimates with and without weighting for non-response bias.

Basis	Proportion of respondents with sheep numbers in the ranges below (%)			
	500-1499	1500-2999	3000-4999	5000 or more
Unweighted	24.4	28.2	22.5	24.9
Weighted	22.5	26.6	28.3	22.5

Chi-squared goodness-of-fit test: $\chi 2=22.86$, d.f.=3, $p<0.00005$, n=1342.

Table A1.26. Difference in cattle number estimates with and without weighting for non-response bias.

Basis	Proportion of respondents with cattle numbers in the ranges below (%)			
	No cattle	Less than 50	50 – 149	150 or more
Unweighted	62.2	15.0	12.6	10.2
Weighted	62.9	15.3	12.4	9.5

Chi-squared goodness-of-fit test: $\chi 2=0.95$, d.f.=3, $p<0.812$, n=1313.

Table A1.27. Difference between unweighted and weighted estimates of the proportion of sheep producers testing for drench resistance.

Basis	% who had tested for drench resistance
Unweighted	43.7
Weighted	44.9

Binomial test, p=0.408, n=1326.

Table A1.28. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of faecal egg count results in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	59.0	17.9	9.1	14.0
Weighted	58.0	17.9	10.2	14.0

Chi-squared goodness-of-fit test: $\chi 2=1.17$, d.f.=3, p=0.761, n=900.

Table A1.29. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of the time of year in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	54.0	32.7	9.2	4.1
Weighted	54.7	32.8	8.9	3.7

Chi-squared goodness-of-fit test: $\chi 2=0.77$, d.f.=3, p=0.857, n=1159.

Table A1.30. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of seasonal weather conditions in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	30.5	32.7	23.6	13.2
Weighted	29.3	32.8	24.4	13.6

Chi-squared goodness-of-fit test: $\chi 2=0.88$, d.f.=3, p=0.831, n=1054.

Table A1.31. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of pasture quality in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	16.6	34.0	27.0	22.4
Weighted	15.9	34.1	27.4	22.6

Chi-squared goodness-of-fit test: $\chi 2=0.40$, d.f.=3, p=0.941, n=969.

Table A1.32. Difference between the unweighted and weighted estimates of respondents' ranking of the importance of the presence of daggy sheep in the mob in deciding when to drench ewes.

Basis	Proportion of respondents indicating the ranks below (%)			
	Very important	Important	Somewhat important	Not important
Unweighted	23.7	28.3	31.9	16.1
Weighted	23.4	27.8	32.8	16.0

Chi-squared goodness-of-fit test: $\chi 2=0.40$, d.f.=3, p=0.939, n=1067.

Table A1.33. Difference between the unweighted and weighted estimates of the proportion of respondents with a set stocked grazing strategy.

Basis	% with set stocking grazing strategy
Unweighted	55.4
Weighted	48.7

Binomial test, p<0.00005, n=1283

Table A1.34. Difference between the unweighted and weighted estimates of the proportion of respondents with a grazing strategy that involved alternating between sheep and crop stubble.

Basis	% with a grazing strategy that involved alternating between sheep and crop stubble
Unweighted	27.0
Weighted	35.3

Binomial test, p<0.00005, n=1279

Table A1.35. Difference between the unweighted and weighted estimates of the proportion of respondents who indicated that they typically treated blowfly strike by treating individual sheep that become struck.

Basis	% treating individual sheep
Unweighted	75.6
Weighted	68.6

Binomial test, p<0.00005, n=1297

Tables A1.25 – A1.35 show weighting based on the three grazing and parasite management questions with the greatest difference between the full and short surveys results in four frequency distributions that are significantly different from the unweighted distributions, viz., the distribution of flock size (in a typical year) and the three questions on which the weighting was based: whether or not producers used a grazing strategy involving alternation between sheep and crop stubbles, whether or not producers typically treated individual sheep that become struck, and whether or not producers used a set stocked grazing strategy. For other aspects of parasite management, such as the ranking of the importance of various factors to be considered when deciding when to drench ewes and testing for drench resistance, there was no significant difference between the weighted and unweighted distributions.

Overall, the investigation of non-response bias suggests that there are not major and systematic differences between the full and short surveys that extend across the full range of questions common to both surveys. There appears to be some minor non-response biases with respect to particular respondent characteristics, however there are not sufficiently strong relationships between these and other characteristics to warrant universal weighting of the findings based on these biases.

For example, producers with 150 or more cattle are under-represented in the full survey by a factor of around 2.5 (Table A1.2). Examination of the relationship between cattle numbers and drench resistance testing shows that 59.3 per cent of producers with 150 or more cattle had tested for drench resistance in their sheep flock, compared to 44.1 per cent of producers who had no cattle. However, producers with 150 or more cattle comprise only 16.8 per cent of producers, so that weighting of the data from the full survey to compensate for the under-representation of producers with 150 or more cattle results in only a small and non-significant increase in the estimate of the proportion of producers who have tested for drench resistance, from 43.7 per cent to 45.9 per cent (Table A1.15).

While universal weighting of the findings appears not to be warranted, there may be grounds for simple adjustment of the findings for each of the small number of questions for which there were significant differences between the full and short surveys. Given that the questions common to the full and short surveys were chosen for their central relevance to informing the extension phase of the IPM-sheep project, it is worth using the data from the short survey to provide the best possible estimates of the producer characteristics which these questions are concerned. It was also decided that, for reasons of consistency, the findings from the remaining questions common to both surveys (those for which there was not a significant difference between the two surveys) would also be presented as estimates adjusted to take account of the data from both full and short surveys.

For example, suppose a question has a proportion of x per cent giving a certain answer in the full survey and y per cent giving the same answer in the short survey. If N respondents answered the question in the full survey and M answered the question in the short survey and P did not respond to either, then the adjusted estimate of the percentage giving the particular answer to the question, x_{adj} is:

$$x_{adj} = \frac{(x \times N) + (y \times (M + P))}{(N + M + P)}$$

This assumes that y per cent of those who did not respond to either survey would have given the particular answer if they had responded.

A1.11 Explanation of Tables

Tables presented in the main body of the report and in Appendix A2 fall into four main types. The types of statistics presented varies according to the type of table.

A1.11.1 Continuous variables

As described in the explanation of tables in section 3.1 of the main body of the report, the summary tables for continuous variables report the sample size (n), the minimum, median and maximum values, the mean and the 95% confidence interval on the estimate of the mean are provided.

As the focus of the benchmark survey was on the characteristics of parasite management rather than the characteristics of the national flock in the regions of interest, weighted means were calculated only in a small number of cases, viz. in question 8 where respondents provided wool cut and fibre diameter for each of three classes of sheep. The mean wool cut and mean fibre diameter for each respondent was calculated by weighting the figure given for each class of sheep by the number of sheep in that class. Mean wool cuts and fibre diameters for regions, or for all regions, were *not* weighted by the flock size of each respondent. Consequently these figures represent the mean wool cut and fibre diameter being achieved by producers in each region, not the mean wool cut and fibre diameter for the flock in each region.

A small histogram of the frequency distribution is also provided for each region in the table. Within any one table, the histograms have the same range on the horizontal axis, so that visual comparisons can be made between regions. However, the histograms are scaled to be of the same height, so that the histograms for regions with a small number of responses are not unduly small and difficult to discern. The class limits for the histogram bars are provided under each table. Histogram counts are the number of respondents with values greater than the lower class limit and less than or equal to the upper class limit. For example, for the class limits 100-260-420-580-740-900-1060-1220-1380-1540-1700, the count of respondents represented by the left-most histogram bar is the number of respondents with values greater than 100 and less than or equal to 260. The count for the next histogram bar is the number of respondents with values greater than 260 and less than or equal to 420, and so on.

Below the histogram class limits at the base of each table, basic statistics are provided for a test of whether there are significance differences in the mean between regions. The test most commonly used was analysis of variance. In a number of cases, however, the variables are strongly bi-modal, with the bulk of responses at the minimum and maximum values of the range. In these cases, the distributions depart substantially from that assumed in analysis of variance, and significance values may be in error. In particular, care should be taken in the interpretation of significance values close to 0.05 when the distributions of the variable of interest in the regions are strongly bi-modal or skewed.

Some ordinal variables were treated as continuous and had means reported for them, although they had four categories or less, which is below the threshold at which ordinal variables can be treated as continuous. This approach was followed where the variable had a relatively small number of integer values, such as the number of treatments in a year for lice control, and where a mean value would be a more convenient way of summarising the data than presenting percentages of respondents with each of the integer values. In these cases, the non-parametric Kruskal-Wallis test rather than an analysis of variance was used to test the hypothesis that there was no difference in means between regions.

For other ordinal variables, such as rankings on the importance of factors used in deciding when to drench ewes, where a mean would have relatively little meaning, the reporting format described in the next section was used.

A1.11.2 Ordinal and nominal variables from single choice questions

For tables reporting proportions for ordinal and nominal data, and where space permits, the upper and lower 95% confidence limits on the estimates of proportions are provided in greyed text to the left and right of the proportion. These confidence limits should be treated with care when the proportion to which they belong is close to zero or 100 per cent, as the approximation of the sample estimate of a proportion by the normal distribution, upon which the calculation of the confidence limits depends, becomes poorer as the proportion nears zero or 100 per cent.

Results of chi squared tests of independence of factors are shown below tabulations of ordinal and nominal variables by regions. When the number of cells with expected frequencies less than five falls below that generally regarded as acceptable for the chi squared test (about 20 per cent of cells), this information is not reported below the table. In these cases, cells in an individual region with values significantly greater than the value for all regions taken together (as indicated by a standardised residual greater than 2.0) are highlighted by bolding and underlining. Where cells in an individual region have values significantly less than the value for all regions taken together (as indicated by standardised residuals less than -2.0), this is denoted by bolding only.

In tables where there are 20 per cent or more of cells with expected frequencies less than 5, this figure is reported in addition to the chi squared statistics. As low expected frequencies of this magnitude compromise the validity of the probability (p value) for the test, no figures in cells in the table are bolded and underlined, or just bolded, even if the p value is less than 0.05.

Likewise, when the chi square test of independence of factors is not significant, no figures in the table are bolded and underlined, or bolded.

Lastly, where tables for ordinal or nominal variables involved only two categories (apart from the 10 region categories), such as when the respondent indicated whether or not they had tested for drench resistance, only the percentages for one category are shown, as the percentages for the other category can readily be obtained by subtracting from 100.

A1.11.3 Nominal variables from multiple choice questions

A number of questions in the questionnaire give rise to multi-level data, i.e. data where there are several levels that could be chosen as the unit of analysis. For example, respondents reporting on their lice control practices could nominate one or more practices (such as plunge dip). For each practice they could nominate one or more years between 2001 and 2003 when they undertook the practice. For each year they could nominate one or more products used in that year. Such a data structure can be analysed with respondents as the unit of analysis (and for example, aggregating products across years and practices), or with treatments as the unit of analysis (aggregating products within years or within types of practices), or with products as the unit of analysis. A further complication with this type of data structure is that the same product can be validly named several times (for example, where it is used each year for three years). In this situation, a table of proportions based on counts of respondents may have a cell in which the proportion is greater than 100 per cent. While the figure is quite correct given the structure of the data and the proportion based on counts of respondents, it is cognitively discomforting to comprehend the meaning of a statement that, for example, 125 per cent of respondents used product X. The best that can be inferred from such a statement is that a fair few people must have used product X and quite a lot probably used it several times.

To avoid this type of problem the following guidelines were followed in reporting from multiple choice and multi-level questions. Where the nature of the multiple choice question was such that the same category could not be indicated more than once by the respondent, the percentages in the table reporting on the question were expressed in terms of respondents. In this situation, no single cell can be more than 100 per cent, but the sum of a row of cells may exceed 100 per cent, due to the question allowing multiple choices. Where the latter is a possibility, this is noted in the footnote to the table.

Where the nature of the multiple choice question was such that the same category could be indicated twice (as in the example of the same lice control product used in consecutive years), the percentages in the table were expressed in terms of products, or treatments, or whatever it was that the respondent could validly name several of the same category.

Since multiple choice and multi-level data is generally ill fitted to the assumptions behind the statistical tests used on data from single choice questions, no statistics are presented in the tables from multiple choice or multi-level questions.

A1.12 Cluster Analysis

The form of cluster analysis used was "partitioning around medoids" ("pam"), as implemented in the R statistical package (R Development Core Team, 2004) This method is similar to the well known k-means iterative re-allocation method (Hartigan and Wong, 1979), but has the advantage of greater robustness and a derived silhouette coefficient which provides guidance as to the number of clusters that best represent the structure in the data (Kaufman and Rousseeuw, 1987). Where "pam" was used, the silhouette coefficient was calculated for 2 to 8 cluster solutions and the solution with the maximum silhouette coefficient accepted. Silhouette coefficients were interpreted following the guidelines provided by Kaufman and Rousseeuw (1987), shown below.

Silhouette coefficient	Interpretation
0.71 – 1.00	A strong structure
0.51 – 0.70	A reasonable structure
0.26 – 0.50	A weak structure, possibly an artefact.
0.00 – 0.25	No structure

Only cluster solutions with a silhouette coefficient greater than 0.50 have been reported. The coefficient obtained and other details are given in the table below.

Cluster analysis	Section of main report	No of clusters with maximum silhouette coefficient	Silhouette coefficient
Q3 – sheep and wool income	3.2.3	2	0.55

A1.13 Calculation of DSEs

Where stock numbers have been converted to DSEs, the conversion factors used were taken from McLaren (1997). Attwood provides conversion factors based on daily energy requirements for a number of classes of livestock at two liveweights and, in some case, at different rates of weight gain. As the survey questionnaire did not collect information on liveweight or weight gain, conversion factors in the middle of the range given by Attwood were used. The conversion factors used are shown in the table below.

Livestock type in questionnaire	Factor for conversion to DSEs
Q5 – Cows	12.0
Q5 – Heifers (weaning – 2 years)	7.0
Q5 Steers (weaning – sale)	7.0
Q5 – Bulls	12.0
Q5 – Other	Factor chosen according to description
Q6 – Merino ewes	1.2
Q6 Other ewes	1.2
Q6 – Wethers	1.0
Q6 – Merino weaners	1.3
Q6 – Other weaners	1.3
Q6 – Rams	1.0

A1.14 Calculation of Mean Wool Cut and Mean Fibre Diameter for Adult Sheep

In Q8 of the survey questionnaire, respondents provided data on the number of sheep shorn, wool cut and fibre diameter for adult breeding ewes and adult dry ewes and wethers. To provided a single figure for adult sheep, a weighted mean was calculated for each respondent by multiplying the wool cut or fibre diameter figure by the number of sheep to which the figure applied, adding the products so obtained, and dividing by the total number of adult sheep shorn.

As it was assumed that almost all producers would provide wool clip figures on a greasy wool basis, the survey questionnaire did not specify whether the wool cut was to be provided on a clean or greasy wool basis. A small number of producers may have provided figures on a clean wool basis, in which case the estimates based on the responses to this question may slightly under-estimate the wool cut in kg/head.

A1.15 References

Fidler, F. and Cumming, G. 2005. Teaching confidence intervals: problems and potential solutions. Paper presented at the 55th session of the International Statistics Institute, Sydney, Australia, April 5-12, International Association for Statistical Education Session 49: Research in Statistical Education. Paper available from: www.stat.auckland.ac.nz/~iase/publications/13/Fidler-Cumming.pdf.

Hartigan, J.A. and M.A. Wong 1979. A k-means clustering algorithm, *Applied Statistics*, 28:100–108.

Kaufman, L. and P.J. Rousseeuw 1987. Clustering by means of medoids. In: Dodge, Y. (ed.) *Statistical Data Analysis based on the L1 Norm* North-Holland, Amsterdam 405– 416.

McLaren, C. 1997. *Dry Sheep Equivalents for Comparing Different Classes of Livestock*. Information Note AG0590. Victorian Department of Primary Industries, Melbourne.

R Development Core Team 2004. *.R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

SPSS Inc. 2001. SPSS 11.0 Syntax Reference Guide. SPSS Inc., Chicago. ISBN 0-13-034842-2.