4.3 Missing data for closed questions

While the proportion of data missing increases in relation to the number of questions asked, this cannot automatically be assumed to be a fatigue response. An alternative scenario is that the more questions asked, the more likely the respondent is to be interrupted, particularly where surveys are completed before appointments or structured activities. Review of standard deviations scores shows minimal change in the variability of responses, suggesting that fatigue does not affect the quality of the data that is provided.

Completion of each question in the survey was voluntary. The proportion of data that is missing gives some indication of how easy clients found the survey to complete and the relevancy of questions. There was a total combination of 7770 possible responses to closed questions (experience, overall and demographic questions) (Table 5). Overall, 9% of possible responses were missing. The proportion of missing data was higher for iPad completion (15%) than paper (6%) which may reflect a level of interest in the technology rather than the survey. Similarly, the proportion of missing data was higher for inpatients (12%) than community clients (3%).

Table 5: Missing data (sample level)

Missing Data Category	Total sample (n=222)	Inpatient (n=123)	Community (n=99)	iPad (n=74)	Paper (n=148)
Missing responses	704	500	204	387	317
Possible responses	7770	4305	7140	2590	5180
% data missing	9%	12%	3%	15%	6%

The proportion of missing data by question ranged from 13% to 5% (Table 6). Generally, less data was missing from experience questions (9%) than either overall or demographic questions (both 13%). It is important to note that the experience questions constitute the first 22 questions, and this difference may reflect fatigue and order effects.

The missing data was further examined based on the position of questions in the survey (Figure 1). It is clear that later questions have a higher rate of missing data than earlier questions.

Figure 1: Missing data based on question position in survey (with linear trend line)
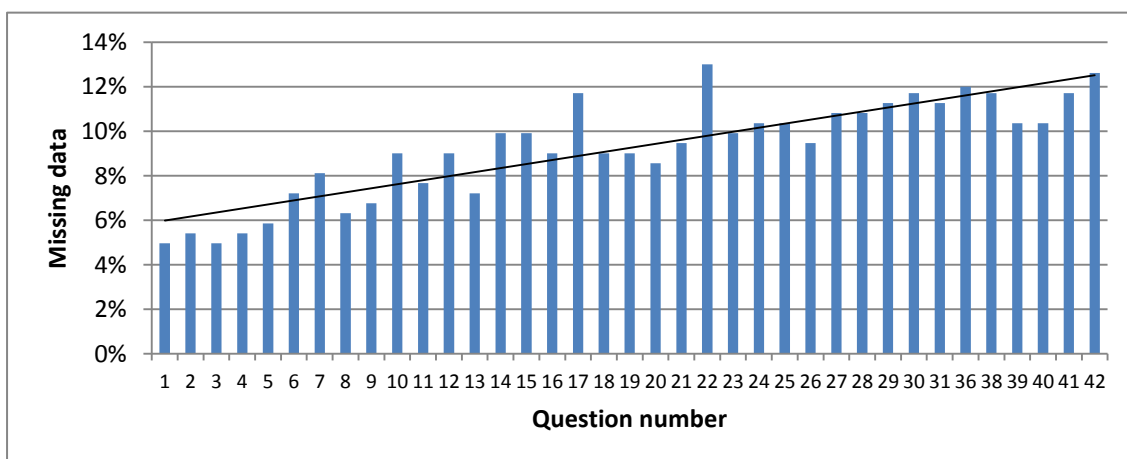
Table 6: Missing data (by question)

Question	Total (n=222)	Inpatient (n=123)	Community (n=99)	iPad (n=74)	Paper (n=148)
Q22. You had things to do that were meaningful for you	**13%**	13%	-	16%	11%
Q42. At any point during the last 3 months were you receiving involuntary treatment?…	**13%**	14%	11%	19%	9%
Q36. What is your gender?	**12%**	15%	9%	22%	7%
Q17. Staff worked as a team in your care and treatment…	**12%**	15%	8%	15%	10%
Q30. The effect the service had on your hopefulness for the future	**12%**	14%	9%	16%	9%
Q38. Are you of Aboriginal or Torres Strait Island origin?	**12%**	14%	9%	20%	7%
Q41. How long have you been receiving care from this service on this occasion?	**12%**	12%	11%	20%	7%
Q29. The effect the service had on your ability to manage your day to day life	**11%**	14%	8%	16%	9%
Q31. The effect the service had on your overall well-being	**11%**	14%	8%	16%	9%
Q27. Development of a care plan with you that considered all of your needs…	**11%**	15%	6%	18%	7%
Q28. Overall, how would you rate your experience of care with this service in the last 3 months?	**11%**	15%	6%	18%	7%
Q24. Convenience of the location for you…	**10%**	14%	6%	19%	6%
Q25. Explanation of your rights and responsibilities	**10%**	12%	8%	15%	8%
Q39. What is your age?	**10%**	11%	9%	18%	7%
Q40. Is this the first time you have been a consumer of this service?	**10%**	11%	9%	18%	7%
Q14. You were involved in planning your future care	**10%**	13%	6%	16%	7%
Q15. You had opportunities to discuss your progress with the staff caring for you	**10%**	13%	6%	16%	7%
Q23. Access to peer support…	**10%**	13%	6%	15%	7%
Q21. Staff showed hopefulness for your future	**9%**	12%	6%	15%	7%
Q26. Information given to you about this service…	**9%**	12%	6%	15%	7%
Q10. You were able to do the things that were important to you while using this service…	**9%**	10%	8%	14%	7%
Q12. Your individuality and values were respected…	**9%**	11%	6%	14%	7%
Q16. Staff showed respect for how you were feeling	**9%**	12%	5%	15%	6%
Q18. Staff ensured you understood the effects of your treatment options…	**9%**	12%	5%	15%	6%
Q19. You felt safe using this service	**9%**	12%	5%	15%	6%
Q20. Your privacy was respected	**9%**	11%	5%	15%	5%
Q7. You had access to a range of other professional services if you needed…	**8%**	11%	4%	16%	4%
Q11. Staff caring for you took the time to get to know you as a person	**8%**	11%	3%	15%	4%
Q6. You had access to your treating doctor or psychiatrist when you needed	**7%**	11%	3%	16%	3%
Q13. You were listened to in all aspects of your care and treatment	**7%**	11%	3%	14%	4%
Q9. The facilities and environment met your needs …	**7%**	10%	3%	12%	4%
Q8. You felt welcome at this service	**6%**	9%	3%	12%	3%
Q5. You were able to get in contact with this service when you needed	**6%**	9%	2%	12%	3%
Q2. Your opinions about the involvement of family or friends in your care were respected	**5%**	7%	3%	8%	4%
Q4. Staff made an effort to see you when you wanted	**5%**	7%	3%	9%	3%
Q1. You had opportunities for your family and carers to be involved in your treatment and care if you wanted	**5%**	7%	3%	8%	3%
Q3. You felt safe to ask questions, provide feedback or make a complaint if you wanted	**5%**	7%	2%	9%	3%

Qualitative feedback from consumer workers revealed that some respondents returned incomplete surveys due to factors such as being called in to an appointment or activity. So a longer survey has an opportunity cost - the more questions the higher the likelihood of interruption, particularly when surveys are completed before appointments. We also know that some respondents verbally reported the survey was too long. Observations reported by consumer workers of completion time ranged from 10 30 minutes. We would expect, if missing data results from fatigue, rather than interruption, that the standard deviations of mean scores would increase with the question load. That is, there would be more variability in answers later in the survey. However, plotting of standard deviation scores (Figure 2) shows that there is minimal increase in standard deviation as a function of question number.

Figure 2: Standard deviation by question number* (with linear trend line)
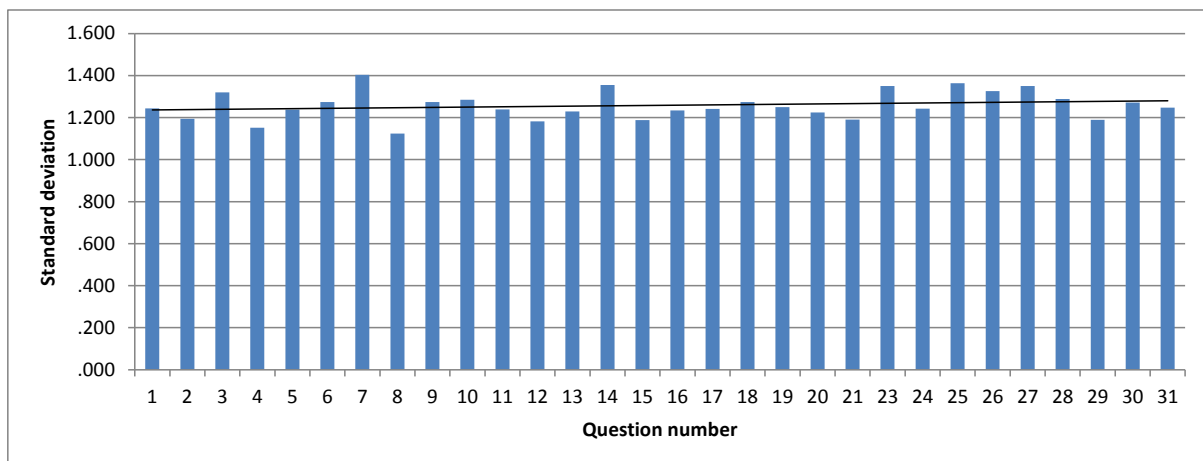
* As Question 22 was only asked of inpatients it has been removed from this analysis as the smaller sample size affects the variability in the data.

4.4 Rating scales

As the scale response options were already positively weighted, no changes to the scales are recommended. Where skewness is problematic for analysis, transformations can be used.

While the distribution of inpatient scores on the performance scales demonstrated some characteristics of kurtosis, this was not found to impact on analyses.

The availability of *Not Applicable* for a subset of questions worked well. The availability of *Not Applicable* did not affect the proportion of questions left blank, suggesting that the option filled a different need.

The analysis suggests that the scales can be assigned numeric properties. As these properties are very close to equal distance, future research should explore the impact of using the score from the Ipsos general population online poll with equal distance values (for example, reviewing changes in rank order or questions or services).

This section examines the performance of the rating scales in the Proof of Concept Trial. The *Experiences of Care Survey* includes two rating scales:

A frequency scale for questions 1 to 22 – *Never, Rarely Sometimes, Usually, Always*

A performance scales for questions 23 to 31 – *Poor, Fair, Good, Very Good, Excellent*

As previous research demonstrated that consumer ratings of mental health services are positively skewed, both rating scales used in the survey were also positively skewed (three positive points, two negative points) to move the data towards a normal distribution. Two tests of normality were used to analysis the distribution of the data, skewness and kurtosis. These analyses have been conducted for the total sample, inpatients and community service users.

4.4.1 Skewness

Skewness is a measure of the asymmetry of a distribution from the mean. For a normal distribution the skewness is 0.

The analysis found that responses on the frequency scale were skewed towards positive results (Data Appendix A1). This demonstrates that the services in the sample were considered high performing by their clients. Responses on the performance scale were generally within the normal distribution, demonstrating that the outcome and recovery questions measured on this scale are harder to achieve. Specifically, the number of questions skewed by sample was (Figure 3):

- 24 total sample
- 18 inpatient sample
- 21 community sample.

Figure 3: Skewness by question

4.4.2 Kurtosis

Kurtosis is a measure of the shape of the distribution to determine the level of volatility. For a normal distribution the kurtosis is 0.

The analysis (Appendix A1) found that the frequency questions generally fell within the normal range, though there were a few exceptions for different samples (Figure 4). However, for inpatients responses on the performance scale generally fell outside of the normal range.

Figure 4: Kurtosis by question
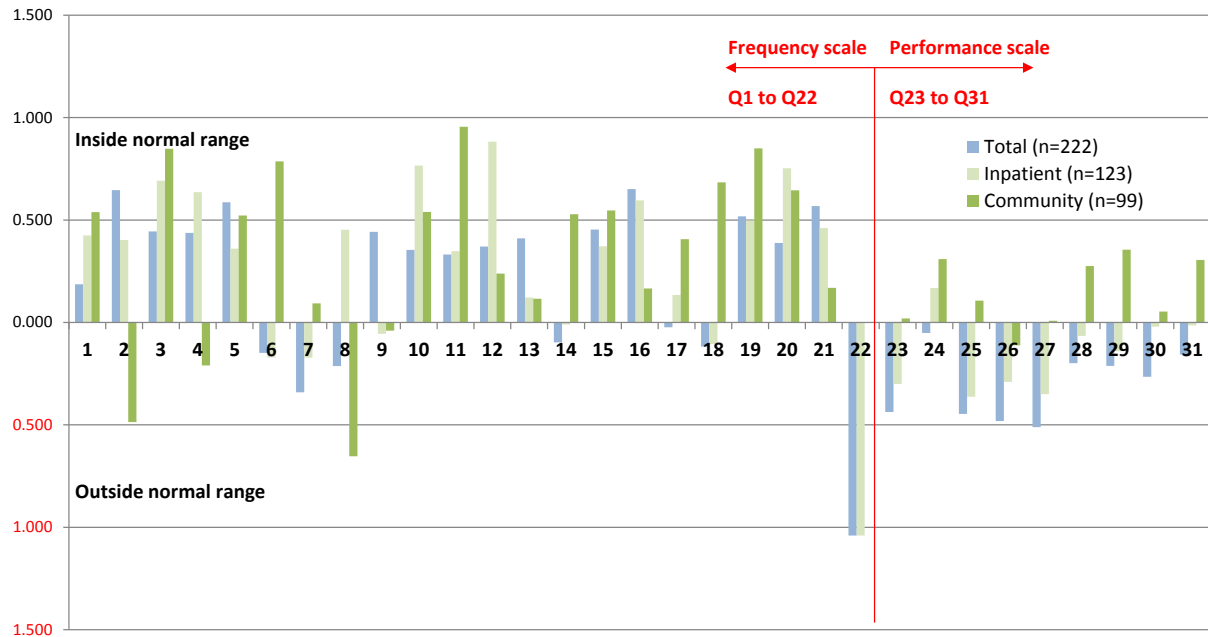
To identify the impact of kurtosis for the performance scale, the shape of normal probability plots of residuals from the linear regression models were reviewed. The shape revealed a linear relationship (rather than this S shape associated with kurtosis) (Data Appendix A2).

**4.4.3 Rating scales with *Not Applicable* option**

The survey included eight experience questions with a *Not Applicable* option. This option was included on rating questions that might not be relevant to all respondents to allow people a way of moving through the survey. It is important to note that respondents could also leave questions blank if they wished. The questions with a *Not Applicable* option were:

> *1. You had opportunities for your family and carers to be involved in your treatment and care if you wanted*

> *2. Your opinions about the involvement of family or friends in your care were respected.*

> *3. You felt safe to ask questions, provide feedback or make a complaint if you wanted.*

> *4. Staff made an effort to see you when you wanted.*

> *5. You were able to get in contact with this service when you needed.*

> *6. You had access to your treating doctor or psychiatrist when you needed.*

> *7. You had access to a range of other professional services if you needed (such as dietary advice, talking therapies, skill development, etc).*

> *23. Access to peer support (such as information about peer workers, referral to consumer programs, advocates, etc).*

21

The proportion of missing data in for these questions was reviewed to see if the availability of a *Not Applicable* option reduced the proportion of results that were missing (Figure 1). The rate of missing data reflected the questions' position in the survey.

The pattern of *Not Applicable* responses was reviewed against the proportion of missing data to see if there was a relationship (Figure 5). For example, an inverse relationship would suggest that the two options are interchangeable. This was not found to be the case. There was no discernible pattern between the proportion of missing data and *Not Applicable* responses.

Figure 5: Use of *Not Applicable* in rating scales
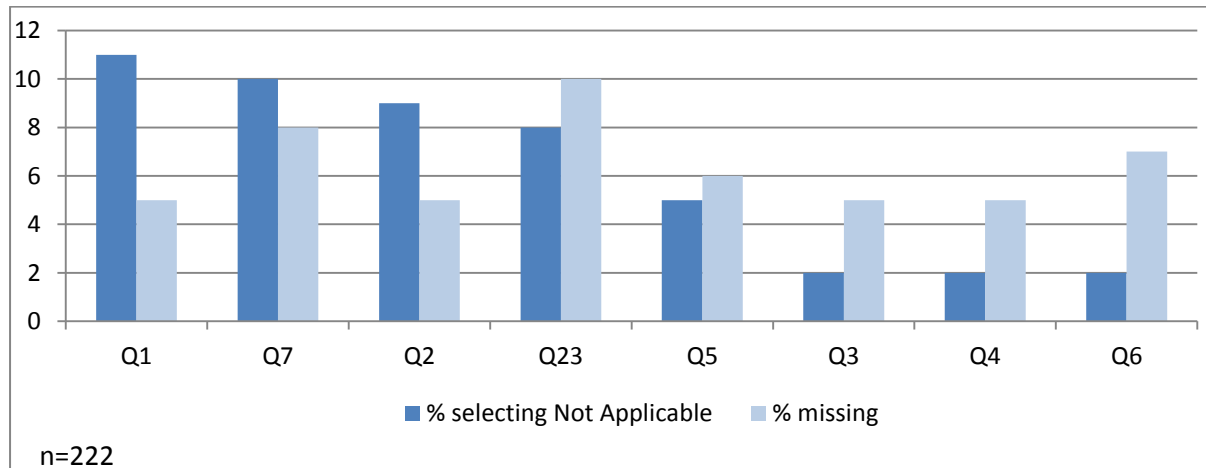
Three questions had few responses (2%) to the *Not Applicable* option:

> *3. You felt safe to ask questions, provide feedback or make a complaint if you wanted.*

> *4. Staff made an effort to see you when you wanted.*

> *6. You had access to your treating doctor or psychiatrist when you needed.*

4.4.4 Distance between points on scales

Two semantic scales were tested for this trial:

- A frequency scale (Never, Rarely, Sometimes, Usually, Always)
- A performance scale (Poor, Fair Good, Very Good, Excellent)

In order to be able to produce aggregate scores, the semantic scales must be converted into numeric values. The psychometric properties of these scales are largely unknown (and not tested on a population of mental health consumers).

At the end of the iPad survey, consumers were asked to indicate how close or far apart the three inside points of the scales (eg, Rarely, Sometimes, Usually) were from one another and the two extreme points (eg. Never, Always).

Consumers used a slider rule to adjust the three inside points of each scale as they saw fit. The mean actual distance was measured between all points to provide the mean psychological distances between the points of each scale.

The task proved difficult for consumers to complete. The level of participation was low (27 for the frequency scale and 17 for the performance scale). While the results were consistent with those found

by Rohrman[1] reported in the literature (Table 7), the Rohrman study did not present scales in sets but tested the intrinsic value of individual words divorced of context. So this study may not be directly comparable.

To provide additional information on the properties of the scales in context, in April 2013 Ipsos conducted an online survey of 1,024 Australians. The same method employed in the iPad trial was employed, with the exception that the three movable points were aligned left (that is, in a neutral position).

When filtered to respondents that adjusted the scale and removing outliers, the results showed that respondents considered the points on the frequency scale to be very close to equal distance apart (Table 7). This is despite only being giving an example of a skewed scale in the introduction to the test.

Table 7: Numeric values of the frequency scale

Sources of data	Never	Rarely	Some-times	Usually	Always
Rohrman (2007) n=100	0	1.3	3.6	7.4 *	10
NCEoC PoC (extremes and non-users of the scale removed (n=26)	0	1.4	4.4	6.9	10
Ipsos online poll n=887	0	2.1	5.0	7.7	10

* Rohrman tested *Frequently* not *Usually*

For the performance scale, the results are consistent with prior research and demonstrate no difference between the general population and mental health consumers (Table 8). Fair, Good and Very Good each sit slightly above an equal distance position.

Table 8: Numeric values of the performance scale

Sources of data	Poor	Fair	Good	Very Good	Excellent
Rohrman (2007), n=100	1.5	5.2	7.2	8.5	10
Proactive Insight (2002)[2], n=417	2.0	4.8	6.8	n.a.	9.0
Proactive Insight (1996)[3], n= unknown	2.5	5.0	7.0	n.a.	9.1
NCEoC PoC, n=17	0	3.0	5.5	8.0	10
Ipsos online poll n=876	0	2.7	5.4	7.9	10

---

[1] Rohrman Bernd *Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data, University of Melbourne, Jan 2007*

[2] *Proactive Insight, Scale research report, 2002, Unpublished*

[3] *Proactive Insight, 1996, reported in Proactive Insight, 2002, Unpublished*