Bureau of Health Information

Guide to interpreting differences in patient survey results

October 2015

Please note that there is the potential for minor revisions of data in this report. Please check the online version at www.bhi.nsw.gov.au for any amendments.

bureau of
**health
information**

Introduction

This document provides guidance for interpreting survey results from the NSW Patient Survey Program (NSW PSP), specifically when comparing results of hospitals, LHDs or NSW as a whole. This program is managed by the Bureau of Health Information (BHI) on behalf of the NSW Ministry of Health and the local health districts (LHDs). For information on the numerator and denominator used in calculations for each question, please refer to the Technical Supplement for each survey.

Overview

- **Results are for a sample of patients**

Patient surveys are a quantitative method used to estimate what people think about the care they received without having to ask every person who received care. This is done through the use of sampling. Sample sizes for the NSW PSP are optimised to provide high precision while keeping costs as low as possible. Detail on how to estimate the precision of the results is presented below under 'Accuracy of results'.

- **Results show the range of results for each question**

Survey results are shown in Healthcare Observer, BHI's online reporting portal. These show survey results in a number of ways. Some webpages demonstrate the variation in results between healthcare organisations, for example, between LHDs compared to the NSW average or between hospitals within an LHD or peer group. Others compare the differences in how patients from different demographic groups respond, for example, by age, sex, education or language spoken at home.

These graphics present the range of results for each question, however, they do not include statistically testing of differences, i.e. whether the results were truly different and not just due to chance. Detail on how to estimate whether differences may be more than what could be expected by chance is presented below under 'How to compare results'.

- **Results reflect the experience of care for patients from each facility**

Most surveys from the NSW PSP surveys sample using a method known as "random stratified sampling". This means that patients within each demographic strata have the same chance of being included in the survey. However, the likelihood of responding to the survey can differ between groups and the composition of patients responding to the survey can be different to the composition at the hospital. The survey program adjusts the results using a process known as weighting to reduce the gap between the respondent composition and the patients that attended hospital during the sampling period.

Results are weighted to the number of patients in each stratum (for example, the Adult Admitted Patient Survey weights for two age groups and for the stay type of the patient (same day vs. overnight), at each hospital). This means that each facility contributes to the LHD, peer group and NSW results in proportion to the number and demographic of patients they provided care for.

It is important to note that the results have not been adjusted for direct comparisons between hospitals, i.e. results have not been standardised to account for differences in each hospital's patient mix or to a NSW average.

Accuracy of results

When looking at the result for a single question for any hospital, LHD or for NSW, the percentages accurately reflect the responses expected in the patient population based on the sample of patients who responded to the survey. The precision of a result (what sort of range in the percentage we might obtain if we repeated the survey many times in the broader population of patients), depends on the number of respondents and the actual result (estimated percentage) obtained from the survey. A more precise result will have a narrower range in the expected results. In general:

- Precision is higher for larger numbers of patients responding; and
- Precision is lowest when patients' responses are split evenly between response options (50:50) and increases as the response split approaches either 0% or 100%.

Table 1 provides guidance on the precision of individual results in Healthcare Observer. The method used to create the values in the table is presented in Appendix 1.

Table 1: 95% confidence interval around estimated percentages, by number of respondents

Respondent number	Average result split reported in *Healthcare Observer* (%)				
	10% or 90%	20% or 80%	30% or 70%	40% or 60%	50%
30000	+/- 0.5%	+/- 0.6%	+/- 0.7%	+/- 0.8%	+/- 0.8%
20000	+/- 0.6%	+/- 0.8%	+/- 0.9%	+/- 0.9%	+/- 0.9%
10000	+/- 0.8%	+/- 1.1%	+/- 1.2%	+/- 1.3%	+/- 1.3%
5000	+/- 1.1%	+/- 1.5%	+/- 1.7%	+/- 1.9%	+/- 1.9%
2000	+/- 1.8%	+/- 2.4%	+/- 2.7%	+/- 2.9%	+/- 3.0%
1000	+/- 2.5%	+/- 3.4%	+/- 3.9%	+/- 4.2%	+/- 4.2%
750	+/- 2.9%	+/- 3.9%	+/- 4.5%	+/- 4.8%	+/- 4.9%
500	+/- 3.6%	+/- 4.8%	+/- 5.5%	+/- 5.9%	+/- 6.0%
300	+/- 4.6%	+/- 6.2%	+/- 7.1%	+/- 7.6%	+/- 7.7%
200	+/- 5.7%	+/- 7.6%	+/- 8.7%	+/- 9.3%	+/- 9.5%
100	+/- 8.0%	+/- 10.7%	+/- 12.3%	+/- 13.1%	+/- 13.4%
50	+/- 11.4%	+/- 15.2%	+/- 17.4%	+/- 18.6%	+/- 19.0%

An example of the use of this table is as follows: For NSW, 64% of patients rated their 'overall care' as 'very good'. This result was based on 35,000 respondents. According to Table 1, our confidence in the precision of results (for a 95% confidence interval) would ± 0.7%. We would be 95% certain that the true result is between 63.3% and 64.7%.

How to compare results

When comparing survey results for a measure between two hospitals and/or LHDs, or with NSW, the difference that can be expected by chance depends on the:

- Number of respondents in each of the two groups being compared; and
- Estimated percentages (results) of the two groups being compared.

Figure 1 demonstrates how different results need to be, compared with the NSW result, for these to be 'statistically significant', for different number of respondents in the comparison group.

Figure 1: Differences between two percentages required to detect statistical significance
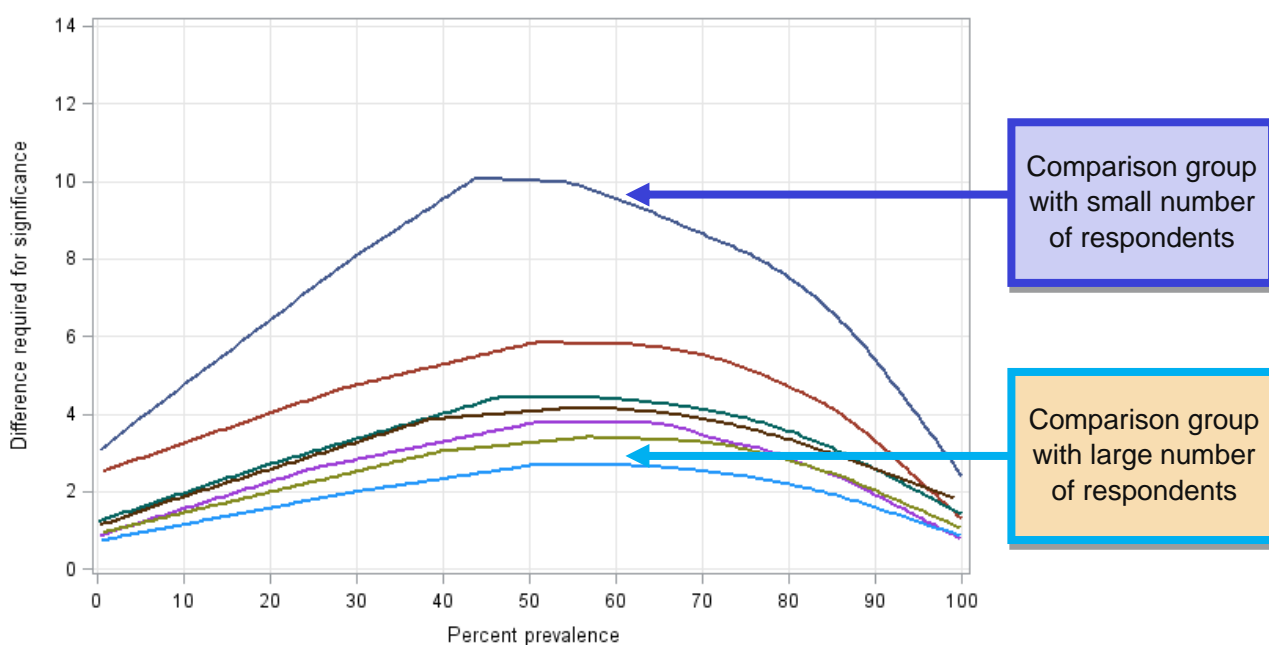
This graph was designed to show the general effect of the number of respondents in the comparison group, assuming that NSW was the other group – that is, if one of the groups has a large, consistent number of respondents. If comparison between LHDs or hospitals is required, testing for statistical significance needs information about the respondent number in both groups.

Table 2 shows how different results have to be in order to be statistically different (i.e. not due to chance) when two groups of different respondent number are compared. Because the size of the confidence interval also depends on the results, the value closest to the average of the two estimated percentages from the results is used. Details on the calculations underlying these analyses are provided in Appendix 1.

Table 2: Difference between two estimated percentages that must be exceeded for the difference to be statistically significant (at 5% level of significance)

Approximate respondent number of first group being compared	Approximate respondent number of second group being compared	Average of the two estimated percentages				
		10% or 90%	20% or 80%	30% or 70%	40% or 60%	50%
4000+	4000+	2%	2%	3%	3%	3%
	2000	2%	3%	3%	4%	4%
	1000	3%	4%	4%	5%	5%
	750	3%	4%	5%	5%	5%
	500	4%	5%	6%	6%	6%
	200	6%	8%	9%	10%	10%
	100	8%	11%	12%	13%	14%
	50	11%	15%	17%	19%	19%
2000	2000	3%	3%	4%	4%	4%
	1000	3%	4%	5%	5%	5%
	750	3%	5%	5%	6%	6%
	500	4%	5%	6%	7%	7%
	200	6%	8%	9%	10%	10%
	100	8%	11%	13%	13%	14%
	50	12%	15%	18%	19%	19%
1000	1000	4%	5%	5%	6%	6%
	750	4%	5%	6%	6%	6%
	500	4%	6%	7%	7%	7%
	200	6%	8%	10%	10%	10%
	100	8%	11%	13%	14%	14%
	50	12%	16%	18%	19%	19%
750	750	4%	6%	6%	7%	7%
	500	5%	6%	7%	8%	8%
	200	6%	9%	10%	10%	11%
	100	9%	11%	13%	14%	14%
	50	12%	16%	18%	19%	20%
500	500	5%	7%	8%	8%	8%
	200	7%	9%	10%	11%	11%
	100	9%	12%	13%	14%	15%
	50	12%	16%	18%	19%	20%
200	200	8%	11%	12%	13%	13%
	100	10%	13%	15%	16%	16%
	50	13%	17%	19%	21%	21%
100	100	11%	15%	17%	19%	19%
	50	14%	19%	21%	23%	23%
50	50	16%	21%	25%	26%	27%

Below are two examples of how this table can be used to compare results statistically:

- For the question on cleanliness of the bathrooms used while in hospital, 60% of respondents in the Central Coast LHD said that they were 'very clean' compared with 57% of respondents across all of NSW. The NSW result is based on over 4,000 responses, while the Central Coast figure is based on 1,083 responses. The average of these two percentages is 59% so the closest value provided (60%) is used. When one figure is based

on over 4,000 observations and the second is based on 1,000, the difference has to be greater than 5% for it to be more than by chance. A difference of 3% could therefore be explained by chance, so the conclusion is that the level of cleanliness of bathrooms in hospitals in Central Coast LHD is not statistically different to the average across the state.

- Conversely, 67% in the Southern NSW LHD classified bathrooms as 'very clean', 10% more than for NSW, based on 2,556 responses at Southern NSW LHD. The average of these two percentages is 62%. Using the 60% column, if the difference is over 4% it is sufficiently large that it is unlikely to be by chance. Therefore it can be concluded that bathrooms in Southern NSW LHD are considered by patients to be cleaner than on average in NSW.

Throughout Healthcare Observer the results (estimated percentages) are accompanied by the number of respondents for that question.

Appendix 1

The values in Table 1 are based on the standard error of a proportion, adjusted for the design effect (DEFF). More information on the DEFF can be found here:
http://surveyanalysis.org/wiki/Design_Effects_and_Effective_Sample_Size

Confidence interval for a proportion (at $\alpha = 5\%$) $\approx 1.96 \times SE(p) \times \sqrt{DEFF}$

$$\approx 2 \times \sqrt{\left(\frac{(p)(1-p)}{n}\right)} \times \sqrt{DEFF}$$

Where p = the survey-based percentage
                  n= respondent number
                  $DEFF$ = design effect. An average value of 1.8 has been used in the calculation for Table 1
                  a = the level of significance. Using $a = 5\%$ leads to the resultant difference being based on a 95% confidence interval.

The values in Table 2 are based on standard error of the difference between two proportions, adjusted for the DEFF, caused by using a stratified sampling design. It assumes independence of samples.

Difference for significance (at $\alpha = 5\%$) $\approx 1.96 \times SE(p_1 - p_2) \times \sqrt{DEFF}$

$$\approx 2 \times \sqrt{(p)(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \times \sqrt{DEFF}$$

Where DEFF = an average design effect value, set at 1.8 for Table 2.

Where p = the average of the two survey-based percentages for the groups being compared
                  n_1, n_2 = the number of respondents for the two groups being compared
                  $DEFF$ = design effect. An average value of 1.8 has been used in the calculation for Table 2
                  a = the level of significance. Using $a = 5\%$ leads to the resultant difference being based on a 95% confidence interval.