

Benchmarking patient satisfaction (2): Time to take stock

It is exactly one year ago since the Hospital Monitoring Directorate (HMD, then known as CCMAU) published the 100 page "Patient Satisfaction Survey Guidelines". The report, which embodied the collaborative effort of several CCMAU/HFA staff and a number of public hospital Quality Managers and Customer Services personnel, described the newly proposed Inpatient and Outpatient questionnaires and explained in great detail the "Best Practice" methodology that should be used by all New Zealand public hospitals so that they would be able to monitor patient satisfaction accurately and reliably.

In an accompanying letter, the Hon Minister of health Ms Annette King said that (these guidelines) would ...

- Improve the statistical robustness of survey results and the consistency with which HHSs can apply them
- Expand the base of the patient populations being surveyed
- Focus the questions asked on the key determinants of patient satisfaction, from the patients' perspective

The question we have to ask ourselves now is, has the implementation of the new survey gone to plan? And have we been able to increase the statistical robustness and the usefulness of the statistics obtained? More to the point, and keeping in mind the issues we raised in Zwiier & Clarke (1999; 2001), are we now in a position to use the data to better understand and/or increase patient satisfaction?

In endeavouring to answer this question, we analysed the entire database compiled by the HMD on the basis of the survey results submitted by each DHB¹ to the HMD. This database, which contains 22,648 inpatient and 25,014 outpatient records from 21 New Zealand DHBs, incorporates patient satisfaction ratings on 15 (inpatient) and 17 (outpatient) items respectively. It presents us with a treasure trove of information, both from the perspective of statistical analyses as from the potential use that can be made of it to further improve our patient satisfaction ratings.

The present overview is divided into two separate sections:

1. an assessment of the reliability and validity of the questionnaire and
2. a preliminary analysis of the results of the survey data using ESPRI software

This unavoidably sketchy overview is concluded with a recommendation regarding future requirements.

1. How reliable and valid is the data?

While a thorough analysis of the data must necessarily be deferred to a later date, it may be of some interest to carry out a preliminary investigation into the reliability and validity of the present survey. Because if we found it to be severely lacking, a lot of effort is made to no avail. The public could rightly accuse us of wasting good public hospital money.

Reliability

Across the board, and on a scale where 1=very poor and 5=very good, average patient satisfaction ratings for inpatient services range from 3.79 (quality of hospital food) to 4.51 (treating the patient with dignity and respect). For outpatient services, the scores range from 3.73 (informing the outpatient about how long they would have to wait) to 4.52 (treating the patient with dignity and respect).

The scores are well distributed and have relatively large standard deviations of 20%. These large standard deviations demonstrate that there is considerable variability across the 21 DHBs on the 32 measures of quality.

To determine the reliability of the inpatient and outpatient questionnaires, we calculated the most commonly used measure of internal consistency: a statistic called "Cronbach alpha". The value of alpha can range between 0 and 1 and we are told that if a set of items has an alpha above .60, it is usually considered to be internally consistent. If it goes above .80, it signifies a very high reliability.

Following Nelson et al (1989), who assessed the reliability and validity of the 68-item "Patient Judgement System"

1 Since the beginning of this year, all HHSs have been renamed and reformed as District Health Boards (DHB)

2 This Patient Survey questionnaire is developed and used by the Hospital Corporation of America which owns and/or operates 378 hospitals in the USA and abroad

3 A few DHBs were not included in the sample because data was missing or incorrectly coded.

(PJS)², we also measured the alpha statistic of the New Zealand Inpatient & Outpatient Questionnaires. In addition, we evaluated reliability by using a variant of the test-retest method; i.e. by comparing the means of two consecutive quarterly periods to ascertain whether the difference between the two periods is statistically significant.

Although our New Zealand questionnaires were not constructed to assess patient satisfaction on a set of dimensions (as does the 68-item PJS), our results show that on measures that gauge satisfaction with specific aspects of treatment such as communication (i.e. providing explanation and information), adopting a personal approach and facets of organising patient care, alpha levels of 0.87 and 0.84 were achieved. These are very high reliability coefficients (see table 1)

Table 1

"Construct"	All ages	75+ years old
<i>Inpatients</i>		
Communication (items 3,4,6)	0.868	0.893
Personal approach (items 5,8,9)	0.835	0.839
Organisational (items 10,11,12)	0.843	0.843
<i>Outpatients</i>		
Explanation (items 6,7,8)	0.864	0.857
Personal approach (items 8,10,11)	0.841	0.826

Table 1 also shows that equally high levels of alpha were obtained when the analysis was restricted to a subgroup for whom, according to Nelson et al, one could expect reliability to be low: i.e. respondents 75 years or older. Further comparisons revealed that confining the analysis to Maori patients made no difference to these alpha levels. This suggests that one can safely analyse the data at subgroup level.

Another method by which one can assess the reliability of a survey instrument is to perform a test-retest reliability analysis. Test-retest reliability estimates are obtained by repeating the measurement using the same questionnaire under as nearly equivalent conditions as possible. However, as it is not possible to re-administer the questionnaire to the same patient three months later, we

compare the average absolute value of the difference between the two means of two consecutive periods (as an example we have used inpatient data for quarter 4, 2000 and quarter 1, 2001).

Table 2

Average of	Mean scores		Absolute Difference
	2000 (4)	2001 (1)	
Q1	3.81	3.81	0.009
Q2	4.04	4.05	0.013
Q3	4.32	4.30	0.016
Q4	4.19	4.18	0.007
Q5	4.38	4.35	0.031
Q6	4.41	4.37	0.038
Q7	4.40	4.38	0.020
Q8	4.28	4.21	0.069
Q9	4.53	4.52	0.010
Q10	4.40	4.42	0.013
Q11	4.20	4.16	0.042
Q12	4.24	4.26	0.014
Q13	4.30	4.27	0.026
Q14	4.40	4.37	0.026
Q15	3.79	3.79	0.003
Q16	4.51	4.51	0.002
Q17	4.46	4.42	0.032
		Average	0.022

The results shown in table 2 show extremely small changes in the average scores from one period to the next. If we compare the entire sample in this manner, the difference is less than 0.5%. Without even taking into account the possibility that some of these differences are caused by actual changes in the delivery process, this stability of measurement provides further support for the reliability of the measures.

Validity

Further analyses focussing on the annual period ending March 2001 show that there is substantial variability across the DHBs on all items in both questionnaires. These statistically significant differences between the DHBs (many at $p < .01$, all at $p < .05$) provides some support for the validity of the items used.

In the absence of a set of different scales all measuring the same construct, the best example of convergent validity must be the way in which all items are in some way or another associated with the one general validity indicator

variable, namely an item which inquires directly to the patient's overall satisfaction with his or her treatment.

The results indicate that, among inpatients, the "overall satisfaction" item is highly correlated with items such as staff availability (item 13; $r=0.68$) and being treated with dignity (item 9; $r=0.68$). Among outpatients, overall satisfaction is most strongly correlated with items inquiring about the quality of information given to patients on how to manage their condition after the visit to the clinic (item 14; $r=0.66$) and organising care with other healthcare providers such as the patient's GP or midwife (item 12; $r=0.64$).

It is interesting to note that the highest correlations were found between items that measured clearly related aspects of patient care. For instance, among inpatients, information given by ED staff on (a) the patient's condition and (b) length of waiting time (items 1 and 2) were strongly correlated ($r=0.77$). Among outpatients, the very high correlation ($r=0.70$) between (a) approval of the effort exerted by staff to make an appointment time that suited the patient and (b) satisfaction with the appointment time itself (items 1 and 2) was most revealing.

Conversely, discriminant validity of the nation-wide Inpatient Survey is shown by the very low correlations between items such as satisfaction with the quality of hospital food and informed consent ($r=0.27$) or level of cleanliness and being kept informed of the wait ($r=0.27$). Similarly, in the Outpatient survey, the lowest correlation was evident between the item measuring satisfaction with cleanliness and suitability of appointment time ($r=0.27$).

As the survey clearly distinguishes between items that ought to correlate with one another and items between which one would not expect to find a strong association, these findings provide additional empirical support for the validity of these items

Response rates and Representativeness

Even if the nationwide questionnaire is reliable and valid, the survey must be based on an acceptable number of individuals returning the questionnaire, high response rates, error-free data input rates and a truly representative sample of patients.

However, even though the Patient Satisfaction Survey Guidelines clearly specified that these details needed to be forwarded to the HMD along with the quarterly patient scores, this has not occurred. This means that the data that has been submitted may be based on incorrect response codes, too small a sample, and/or suffer from systematic bias. For instance, a lower response rate from younger and non-European patients could well have

resulted in an under-representation of these age and ethnic groups. This produces a sample that is biased in favour of older and European patients.

2. What do the results tell us?

Before we draw any quick conclusions from a cursory look at the data, it is important to assess the characteristics of the sample.

Age and sex

Fig 1

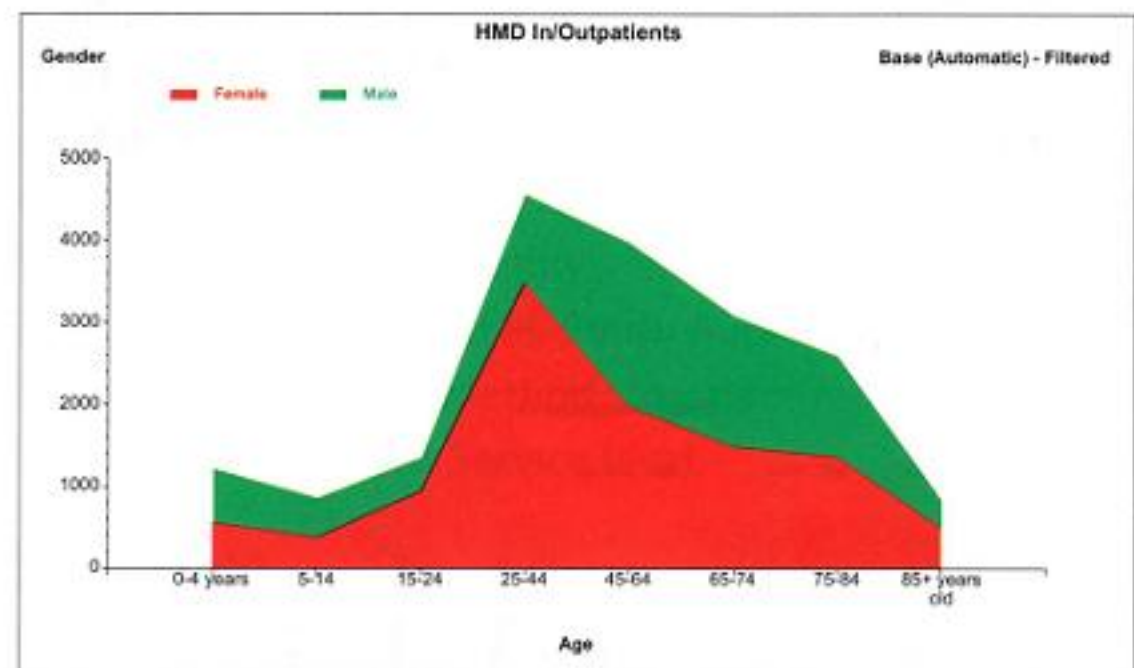
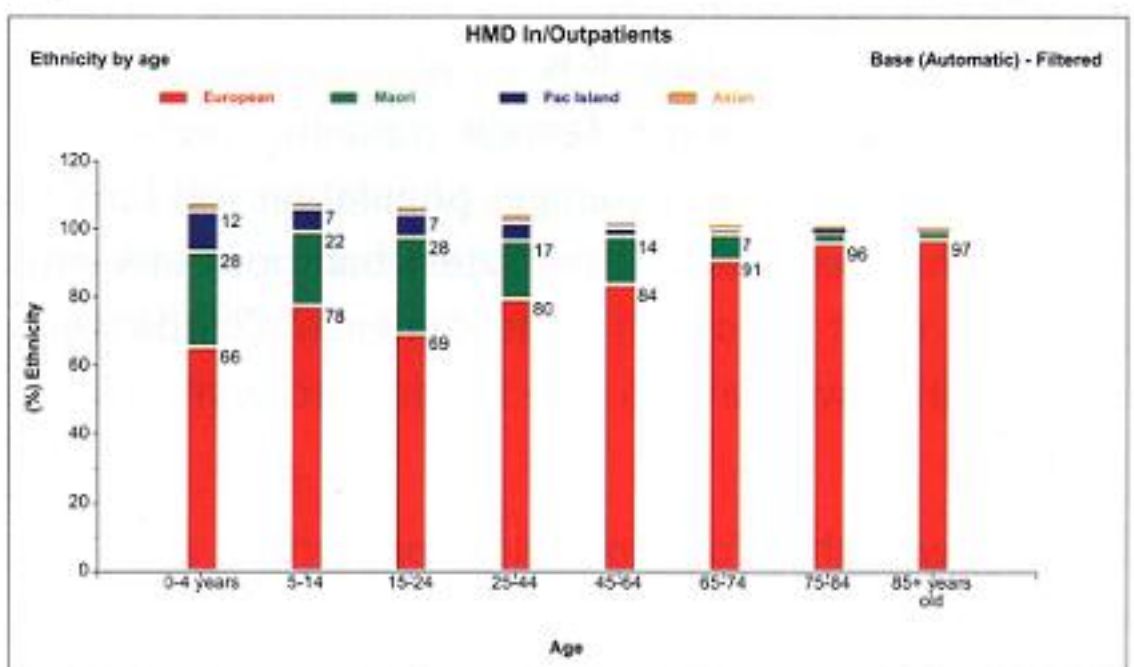


Fig 1 shows that the sample of 18,423 inpatients³ can be separated in 10,630 female patients, many of whom of childbearing age, and 7,793 male patients. The distribution of sex and age among outpatients is not dissimilar but slightly more skewed in favour of the 45-64 age band.

Ethnicity

Fig 2



Across the board, 84% of inpatients are European, 14% are Maori, 4% are Pacific Island and 2% are of Asian origin (some patients indicate more than one ethnicity). Fig 2 shows that Maori and Pacific Island patients are disproportionately represented in the lower age bands. This is similar to the distribution of age and ethnicity found among outpatients.

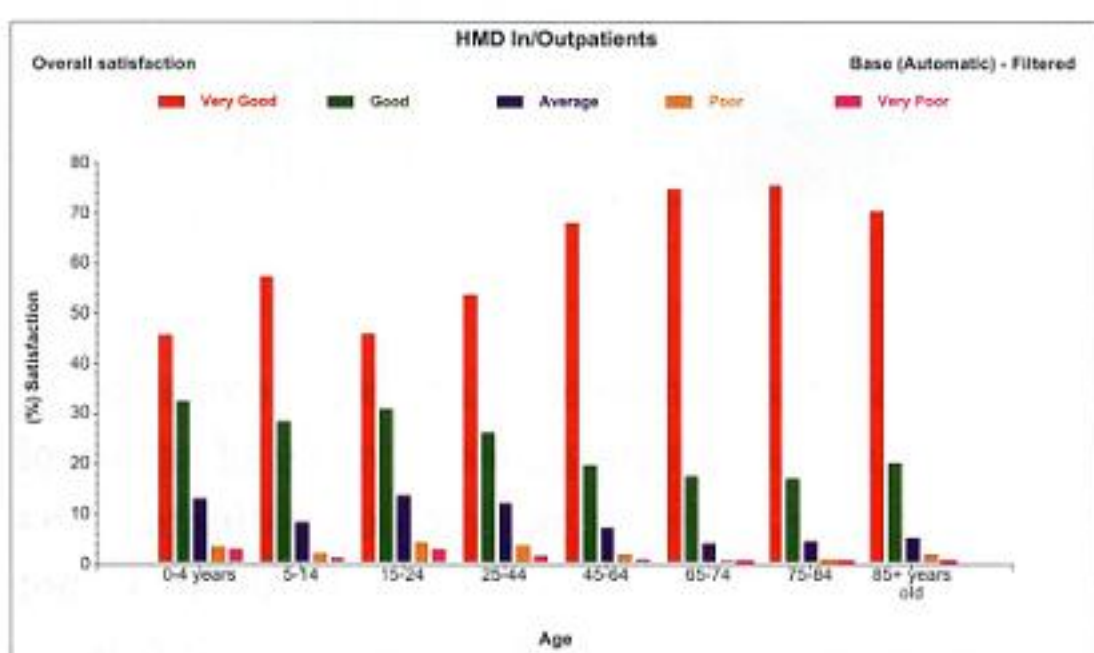
Satisfaction

Now we can try and answer the questions: “How satisfied are New Zealand patients?” and “How does patient satisfaction relate to these demographic variables?”

Our ESPRI analysis of the 18,661 inpatients who answered the general “overall satisfaction” question shows that 64% are “very satisfied” and an additional 23% are “satisfied”. Only 5% of inpatients express overall dissatisfaction. Among outpatients, the percentages are very similar: 65% and 24%. Only 3% are dissatisfied.

However, as expected, patient satisfaction rates are a function of age, sex and ethnic group. For instance, fig 3 shows that age is strongly correlated with satisfaction: older patients are more likely to express satisfaction than are younger patients ($p < .01$).

Fig 3



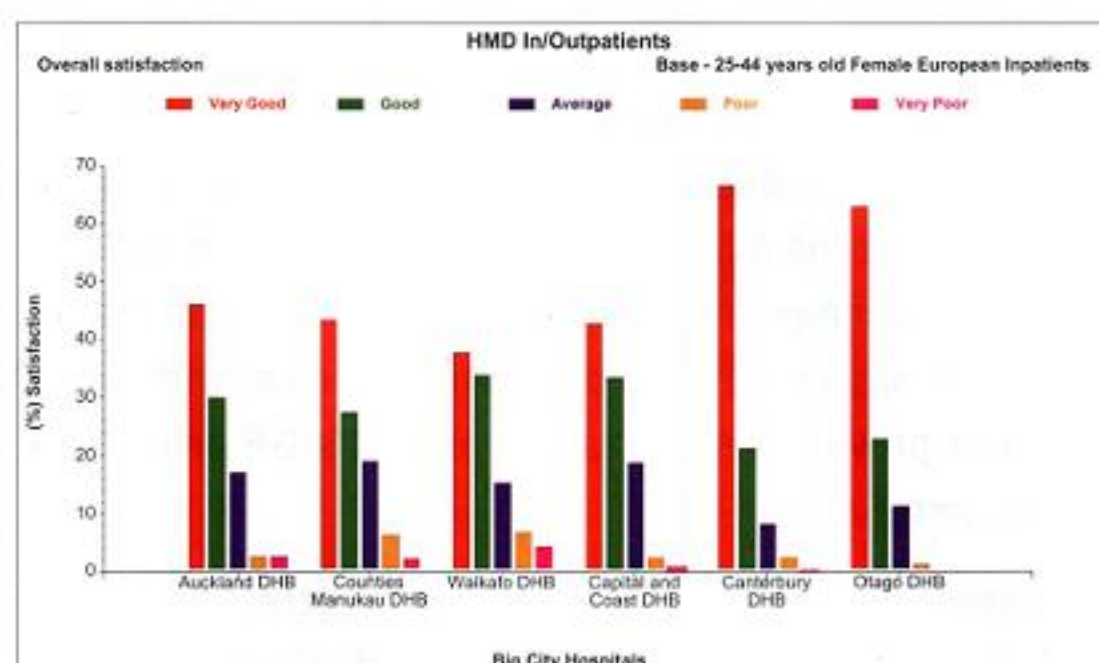
Similarly, patient satisfaction correlates with the patient’s sex (males are more likely to express satisfaction; $p < .01$) and ethnicity (European patients are more likely to report satisfaction; $p < .01$). Thus it is no surprise that hospitals with proportionally more female patients, more non-European patients and a younger population will tend to have lower patient satisfaction rates than hospitals with more and older European male patients. Comparisons between DHBs will have to take this into account to be of any use.

The best way therefore to make appropriate and valid comparisons is either to apply a post-stratification weighting method (i.e. weighting each response using inverted selection probabilities x the ratios of expected to observed counts) as suggested by Waring and Meyer (1999), or by confining one’s analysis to a subset of the database, e.g. a specific age or ethnic group or sex.

For instance, when we compare the big city hospitals selecting only 25-44 year old female European inpatients ($n=1288$), we find that there is no statistically significant difference in overall satisfaction between the north island

DHBs, but that both Canterbury and Otago DHBs enjoy greater overall satisfaction in this subgroup (see fig 4).

Fig 4



Unlike Draper and Hill (1996) who express reservations about benchmarking patient satisfaction at national level, we believe that a very careful comparison between hospitals on all 32 measures of patient satisfaction has the potential to generate extremely valuable information that can be used to increase patient satisfaction throughout New Zealand.

However, to make this a reality, it is imperative that we can rely on the data being accurate. Without an audit that would verify that DHBs are using the correct sampling method, have adequate sample sizes, check that statistical representativeness has been achieved and that error checking procedures are in place, we simply will not know whether the data submitted to the HMD is in any way an accurate reflection of patient satisfaction either in that DHB or, when combined with the other DHBs, in New Zealand as a whole.

The fact that the HMD is planning to audit the survey procedures not until October this year along with the other components of the “Balanced Score Card” means that we will have to wait until then to find out whether or not the comparisons we make and the conclusion we draw are in fact justified.

Gerard Zwier PhD, Managing Director, Health Services Consumer Research Limited

David Clarke ME MComm MBA, CEO South Auckland Health

The inpatient questionnaire asked:

Please rate the Emergency Department staff on:

- (1) ...keeping you informed of how long you could expect to wait
- (2) ...telling you how the Emergency Department would treat your problem

Please rate our staff on:

- (3) ...explaining what was wrong with you
- (4) ...informing you about different treatment options
- (5) ...asking your permission to treat you
- (6) ...listening to you
- (7) ...involving your family/whanau, as much as you wanted
- (8) ...offering choices specific to your culture
- (9) ...treating you with dignity and respect
- (10) ...organising your care with other departments in the hospital
- (11) ...preparing you for leaving hospital
- (12) ...organising your care with other health care providers (such as your doctor or midwife)

Please rate:

- (13) ...if staff were around when you needed them
- (14) ...how clean your ward/unit was
- (15) ...how much you liked the food we gave you
- (16) ...how safe and secure you felt in the hospital environment

Most importantly:

- (17) Overall, how satisfied were you with our service?

The outpatient questionnaire asked:

- (1) How well did your appointment time suit you?

Please rate our staff on:

- (2) ...their effort to make an appointment time that suited you
- (3) ...providing clear information to prepare you for your appointment
- (4) ...making you feel welcome when you arrived
- (5) ...telling you how long you would wait, when you arrived
- (6) ...explaining what was wrong with you
- (7) ...informing you about different treatment options
- (8) ...asking your permission to treat you
- (9) ...listening to you
- (10) ...offering choices specific to your culture
- (11) ...treating you with dignity and respect

- (12) ...organising your care with other health care providers (such as your doctor or midwife)

Please rate:

- (13) ...how clean our facilities were
- (14) ...the information we gave you on how to manage your condition after your visit

Most importantly:

- (15) Overall, how satisfied were you with our service?

A 5-point Likert scale (Very good, good, average, poor, very poor) provided the response categories for each item. Questionnaires were despatched by mail to a random sample of patients who had recently been discharged from hospital or who had recently visited an outpatient clinic. Some of the DHBs (such as South Auckland Health) use a stratified sampling method to also monitor patient satisfaction ratings at Service level.

References

1. CCMAU. Patient Satisfaction Survey Guidelines. Unpublished Crown Company Monitoring Advisory Unit Report June 2000
2. Draper M, Hill S. Feasibility of national benchmarking of patient satisfaction with Australian hospitals. *International Journal for Quality in Health Care*, 1996; 8: 457-466
3. Nelson, E C, Hays R D, Larson C, Batalden P B. The Patient Judgment System: Reliability and Validity QRB June 1989
4. Waring M, Meyer D. Report to Crown Company Monitoring Advisory Unit in Respect of the Proposal to Revise the Patient Satisfaction Survey. Unpublished document. Albany Campus. Massey University May 1999, p6
5. Zwier G, Clarke D. How well do we monitor patient satisfaction? Problems with the nation-wide patient survey. *New Zealand Medical Journal* 1999; 112 (1097); 371-5
6. Zwier G, Clarke D. Benchmarking Patient Satisfaction: Do we have a level playing field? *New Zealand Health and Hospital*, January-February 2001, 21-24