Reporting of service performance of mental health services

Tom Trauer

St Vincent's Hospital, Melbourne

University of Melbourne

Monash University

1

Background

The 4th National Mental Health Plan provides for

"*measuring and reporting progress*",

for which the indicator is

"*The proportion of services publicly reporting performance data*".

There are however no clues as to what should be reported, or how.

2

Rationale

Public performance reporting is relatively new for mental health, but has a history in cardiac surgery (USA) and in education (UK).

The rationale for public reporting is that is may:

- Enable service users (patients, purchasers) to make value-based choices,

and/or

- Stimulate quality improvement activities within services.

3

Reactions

Public performance reporting has proved contentious. The literature suggests that:

Managers tend to favour it.

Doctors tend to be sceptical.

Patients and the general public

- are unaware of public reports
- value different types of information.

Services generally respond to public reporting with increased activity in the areas reported.

4

Issues

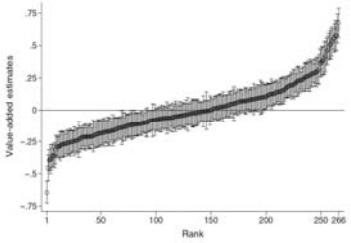
Public reporting of performance needs to have regard to a number of issues:

- adjust for factors outside the control of the service (eg. casemix),
- take into account the natural variability of the measurements,
- present the results in a way that is easy to understand,
- avoid the production of “league tables”, which can become a focus for unhelpful controversy.

5

League Tables

Involve ranking provider organizations on their scores on some indicator.

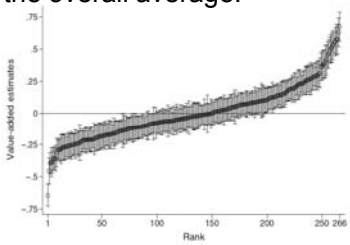


The graph shows 2007 results of 266 UK schools, after adjustment for several covariates, with 95% confidence intervals.

6

League Tables

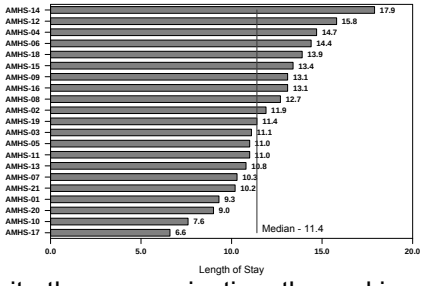
37% of schools were not significantly different from the overall average.



Also, most schools were not significantly different from most of their neighbours.

7

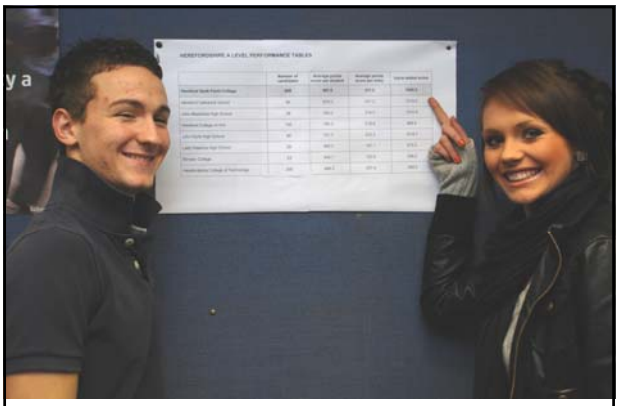
From a review of KPIs in Victoria in 1999



Despite the anonymization, the ranking clearly implies “better” and “worse”.

Yet who can say if 17.9 days is “too long”, and may not 6.6 days be “too short”?

8



League tables can promote gloating

There are fears that national data will lead inevitably to league tables 11

Relative v Absolute judgements

Another issue is whether services should be judged relative to each other or relative to some agreed standard.

Relative to each other is what "league tables" are good for.

Relative to some standard allows judgments of competency, adequacy, or excellence.

12

Relative v Absolute judgements

Imagine a set of services where the performance on some indicator improves 5% every year.

A service that is average will stay average each year.

This could result in a misleading and possibly demoralizing conclusion that this service is not changing.

This suggests that any reporting scheme should show how a service compares to a standard.

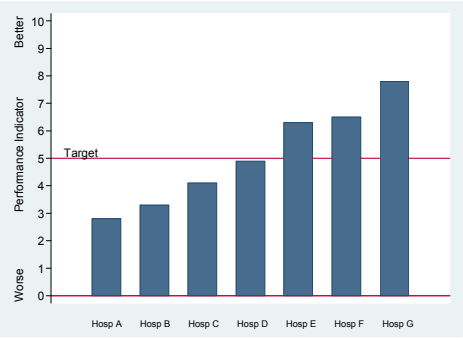
13

A possible scheme

There follows a schematic approach that ticks some of the boxes.

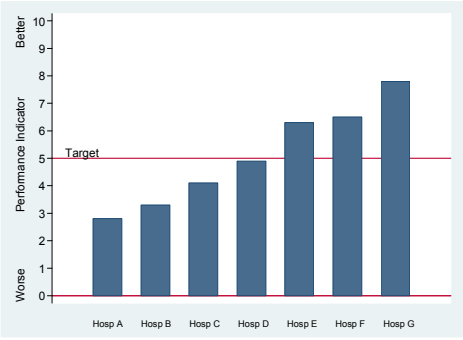
14

Imagine 8 services, with mean scores on an indicator, and a target of 5



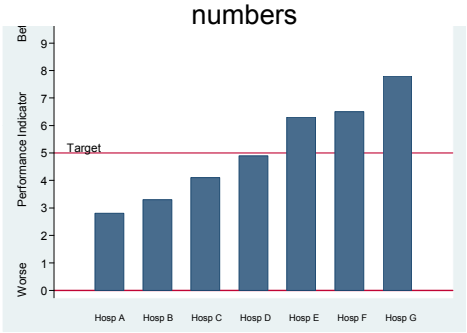
15

Limitations of this way of presenting



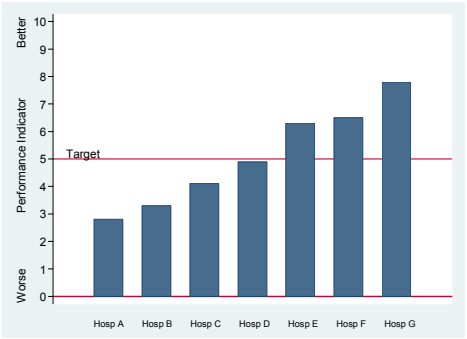
16

There is no consideration of the stability of the averages; those based on small numbers will be less stable than those based on large numbers



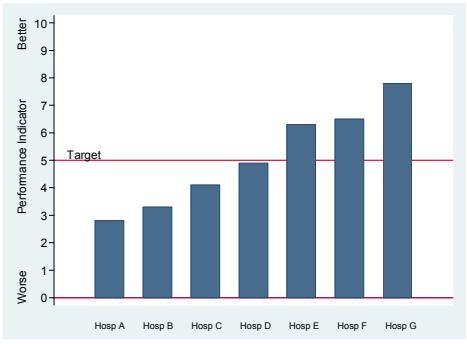
17

The ordering encourages comparison with neighbours, not against the standard



18

Invites possibly invalid inferences, like hospital F is really better than hospital E



19

Acknowledging uncertainty

We begin with the initial performance scores, which are usually averages or percentages

For example, a service treats 1000 consumers, and we have an average improvement of 4.5 units on some measure for 100 of them

Let us say that the standard deviation of the 100 obtained scores is 1.8

20

Acknowledging uncertainty

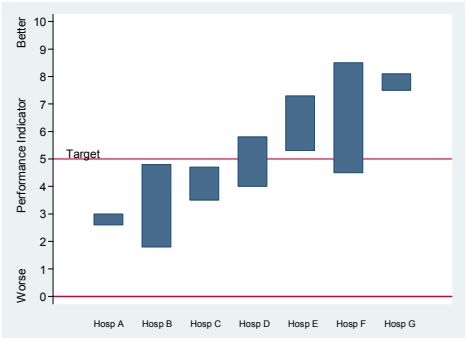
This information would allow us to calculate confidence intervals around the average of 4.5

In this case, the 95% confidence intervals would be 4.1 to 4.9

That is, using the information available, we could be 95% confident that the mean of the 1000 lay between 4.1 and 4.9

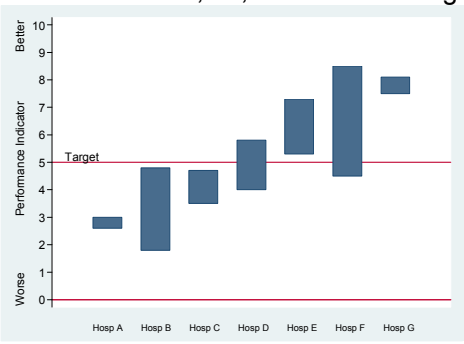
21

Next, we replace each mean with its 95% CIs



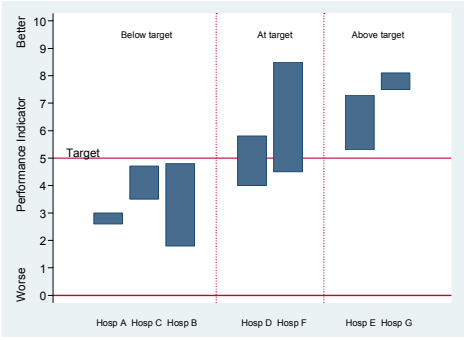
22

Next we see whether each hospital meets, exceeds, or falls short of the target, by seeing if the CIs are above, on, or below the target



23

And we group the hospitals according to whether they meet, exceed, or fall short of the target



24

Advantages of doing it this way

- Acknowledges the variability around simple hospital means
- Reduces unhelpful focus on relative positions of similar hospitals
- Hospitals with overlapping CIs cannot be said to be significantly different
- Refocuses attention on whether hospitals exceed, meet, or fall short of a standard
- Inhibits presentations that look like league tables

25