

From stoplight reports to time series: equipping boards and leadership teams to drive better decisions

James Mountford,^{1,2} Doug Wakefield³

¹UCLPartners, London, UK

²Royal Free London NHS Foundation Trust, London, UK

³Center for Health Care Quality (CHCQ), Department of Health Management and Informatics, University of Missouri, Columbia, Missouri, USA

Correspondence to

Dr James Mountford,
UCLPartners, 3rd Floor, 170
Tottenham Court Road, London
W1T 7HA, UK; james.
mountford@uclpartners.com

Accepted 7 March 2016

Published Online First

31 March 2016

One of us was shown a letter received by a hospital infection control leader from the CEO congratulating her on an excellent monthly performance—for the previous month MRSA infections had decreased from 4 to 2 cases. A couple of months later the same CEO sent a letter expressing serious concern, asking for an explanation of why the monthly MRSA cases had doubled from 2 to 4. Implicit in the CEO's letter is an all too common misunderstanding when using point-to-point data comparisons that every data point is a signal of meaningful change. Absent any information about or understanding of the nature and extent of the underlying variation of the process or event type being analysed, in point-to-point comparisons the only thing one can be sure of is that the second data point will likely be either higher or lower than the preceding data point.

Common to board members, corporate-suite executives, directors and managers is the need to rapidly interpret key data and to decide what if any actions are needed. Two papers in this edition highlight the critical need to ensure that such data presentations do not lead decision-makers astray. In the first paper by Schmidtke *et al*,¹ analysing data presented to Boards of English NHS Trusts, control charts are offered as an effective and efficient tool to distinguish results due to chance variation from results due to significant changes. The Anhøj *et al*² paper from Denmark critiques the use of the seemingly ever present 'red, amber, green' stoplight reports, and also endorses the need for longitudinal analyses to detect trends and meaningful data shifts rather than looking at individual data points in

isolation. Together these two papers are useful contributions to a literature about what forms of data decision-making groups should see in order to focus attention on the most pressing areas, to understand the causes that underpin what the data show, and determine what action should follow. The central question is: how to get data to decision-makers in a form which drives the most useful decision-making?

Anhøj *et al* make the striking claim that red, amber, green management reporting is at best useless and at worst harmful. These reports rely on the simple colour-coded heuristic of 'green is good... proceed as is', 'yellow or amber is warning...proceed with caution' and 'red is bad...stop and take action'. We think their critique is a bit too stark: there are situations when application of the stoplight type reporting may be appropriate. For example, in situations in which process reliability should be 100%—for example, as with never events—each data point can represent a meaningful signal. Likewise for well understood, tightly controlled processes with little inherent variation, stoplight reports may be of value. The primary advantage of stoplight reports is their simplicity and ease with which a large amount of information can be quickly presented.

The problem with stoplight reports is not this inherent simplicity, but rather how their use has been overextended beyond their limitations, and perhaps a lack of awareness of the limitations. It is important to remember that these reports have been adapted from an origin as useful road-traffic controls. When driving, a stoplight is a useful, real-time decision aid for the driver—which links to a clear desired and immediate action



► <http://dx.doi.org/10.1136/bmjqs-2015-004967>

► <http://dx.doi.org/10.1136/bmjqs-2015-004951>



To cite: Mountford J, Wakefield D. *BMJ Qual Saf* 2017;**26**:9–11.

for the driver to take: proceed or prepare to stop. They also signal trajectory requiring an immediate decision: if green is followed by yellow, then you know red is coming and vice versa. In contrast, stoplight reports might better be compared with looking in the rear view mirror to see where you have been and then using this information to decide what to do at the next intersection.

There are several important reasons why stoplights in organisational datapacks have important limitations. First, the inability to reflect the trajectory of what has happened: are the results stable, worsening or improving relative to the desired standard? Each of these trajectories requires different potential actions. Green does not necessarily mean you are doing well and that no additional attention is needed. Similarly, red does not necessarily mean there is a problem which requires senior management intervention. A series of green data points could mask the reality of a steady deterioration in performance, while a series of red values could equally mask the reality of steady improvement towards the standard.

A related problem with stoplight reports is how the standard or threshold values are selected and defined. They may be arbitrary and reflect neither what is 'acceptable' performance nor the level of performance which may be achievable. For example, in England when the national emergency department 4 hour access target was relaxed from 98% to 95%, many organisations 'improved' in stoplight status despite no changes on the ground—and no change in patients' care. In such instances 'green' may lead to a false reassurance and inadvertently inhibit the drive toward better performance.

Given that we live in a world of stoplight reporting, we need to get better at using an imperfect tool. We can usefully enquire what a knowledgeable user should ask about the data, and suggest asking the following questions:

- ▶ What is the purpose of the stoplight report? What key information is it supposed to be communicating? What types of decisions are expected to be made with it?
- ▶ How were the performance standards selected and how were the red–yellow–green threshold values operationally defined?
- ▶ Does absence of information about the 'trajectory' of the reported results matter?
- ▶ How much management (and local staff) attention does this require? Does it represent a major risk?

As both papers argue, organisational leaders can glean more useful information through run and control charts—but the use of these is sadly the exception rather than the rule in many healthcare organisations. Such time series representations, especially when equipped with statistical limits can better answer questions such as:

- ▶ Is the mean (performance level) acceptable? And is the level of variation expected or acceptable?

- ▶ Does the time period between longitudinal measurements allow for timely action given the requirement for multiple data points to reveal a trend? If long time periods such as quarterly data are used, is supplemental monthly data available that could be used to enhance understanding and decision-making? Could/should data be gathered more frequently?
- ▶ What do the data tell us about the stability or changing trajectory of the underlying process's performance? What is the nature and extent of variation of the underlying data being reflected by the stoplight report?
- ▶ Are the process changes we made working, even if process performance has not reached the desired target?
- ▶ Is this an area which needs greater or different focus, and/or resource commitment to achieve the desired performance level?

All this matters: we know that what boards and leadership teams do matters in terms of organisational performance,³ and boards' focus is influenced by the data they see. What boards choose to prioritise in turn directs organisational attention and resources. Consequences can be severe, including: failure to focus on areas where greatest attention is required; undue focus on areas where additional attention is not required—with negative consequences stemming from a failure to differentiate special cause variation from common cause variation. And by appearing disconnected from the root causes of performance, boards may undermine their perceived legitimacy and reduce staff morale.

Deming wrote: 'If I had to reduce my message for management to just a few words, I'd say it all had to do with reducing variation'.⁴ Stoplights alone offer little basis for understanding and managing variation. We need progressively to move to time series data with control limits highlighting variation and trend. This requires better information infrastructure within organisations, which will require appropriate investment. But there is also a skills and knowledge angle: board members and other senior leaders need to understand effective data use and the advantages and limitations of different representations in order to ask for the most useful information. A question for healthcare executives leading organisations to consider seriously is whether sufficient educational investments have been made to ensure its board members' analytic capabilities and sophistication are sufficient, particularly in the evolving era of 'big data'. A question for board members is whether the information they are presented with by executives leading their organisations provides sufficient information to make the best decisions.

Analogous to the '5-rights' of medication administration, one might well ask whether the right information is being provided to the right decision-makers, in the right manner, in the right amount and at the right time. Alongside better representation of data and understanding how best to drive insight and action

from data, we also need more relevant metrics to inform groups' decision-making: we need meaningful and actionable metrics which capture what matters most to patients and populations across pathways of care, and metrics linking quality to resource use. Information systems' capabilities matter here too, but equally important is boards' willingness to challenge whether the metrics they are offered are as relevant as they could be to improving results for those the organisation exists to serve.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- 1 Schmidtke, *et al.* Considering chance in quality and safety performance measures: an analysis of performance reports by boards in English NHS trusts. *BMJ Qual Saf* 2017;26: 60–8.
- 2 Anhøj, *et al.* The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures. *BMJ Qual Saf* 2017;26:80–3.
- 3 Tsai TC, Jha AK, Gawande AA, *et al.* Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Aff (Millwood)* 2015;34:1304–11.
- 4 Neave HR. *The deming dimension*. Knoxville, TN: SPC Press, 1990.