



# Games People Play: Lessons on Performance Measure Gaming from New Zealand

## Comment on “Gaming New Zealand’s Emergency Department Target: How and Why Did It Vary Over Time and Between Organisations?”

Lisa M. Lines<sup>1,2\*</sup>

### Abstract

For decades, observers have noted that gaming of performance measurement appears to be both endemic and endlessly creative. A recent study by Tenbenschel and colleagues provides a detailed look at gaming of a health system performance measure—emergency department (ED) wait times—within four hospitals in New Zealand. Combined, these four hospitals handled more than 25% of the ED visits in the country each year. Tenbenschel and colleagues examine whether the New Zealand ED wait time target was set appropriately and whether we can trust any performance measure statistics that are not independently verified or audited. Their thought-provoking examination is relevant to anyone working in quality improvement and provides a valuable set of tools for detecting gaming in performance measurement.

**Keywords:** Gaming, Performance Measurement, Emergency Departments, New Zealand, Healthcare Quality

**Copyright:** © 2020 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Citation:** Lines LM. Games people play: lessons on performance measure gaming from New Zealand: Comment on “Gaming New Zealand’s emergency department target: how and why did it vary over time and between organisations?” *Int J Health Policy Manag.* 2020;x(x):x–x. doi:10.34172/ijhpm.2020.41

### Article History:

Received: 30 January 2020

Accepted: 5 March 2020

ePublished: 18 March 2020

### \*Correspondence to:

Lisa M. Lines

Email: [llines@rti.org](mailto:llines@rti.org)

In *Gaming New Zealand’s Emergency Department Target: How and Why Did It Vary Over Time and Between Organisations?*, Tenbenschel and colleagues provide a detailed look at gaming a health system performance measure—emergency department (ED) wait times—within four hospitals in New Zealand.<sup>1</sup> Those hospitals saw more than 25% of the ED visits in New Zealand between 2006 and 2012.

### Defining the Target

When individuals arrive at an ED, they are typically triaged (assessed for how urgent their condition is and how quickly they must be seen), diagnosed, and then either treated, transferred, admitted, or discharged. Measures for describing time spent during ED visits may refer to *visit lengths* or lengths of stay ([LOS] – total time spent in the ED) or *wait time* (time until being seen by a provider). These concepts are similar, but wait time is a subset of total LOS. Once triaged and seen by the initial provider team in the ED, overall LOS may be determined by factors outside of the ED’s control, such as the availability of specialists, imaging equipment, or beds at another unit or facility. Patients waiting in the ED for resources outside the ED has been cited as the primary cause of ED overcrowding in New Zealand, although the Ministry of Health also cites problems with triage processes, insufficient

ED beds, and inadequate ED staffing.<sup>2</sup>

### A Hard Target to Hit?

The target set by the New Zealand Ministry of Health for ED *wait times*, defined as number of minutes between when a person arrives at the ED and when that person is treated by a provider, was 6 hours or less for at least 95% of patients. This target may have been difficult for hospitals to reach. At baseline, the four hospitals studied had wildly varying performance on this measure, with anywhere from 56% to 81% of ED visits with wait times less than 6 hours. After the target was introduced in 2009, this increased to 85 to 98% of ED wait times being less than 6 hours in those same four hospitals.<sup>1</sup> According to the latest government data, the average is 85% across New Zealand.<sup>3</sup> As for the effects, according to one observer:

“...the target has worked to reduce overcrowding of patients in ED by moving them on much faster to other parts of the acute hospital, or through speedier discharge from the ED. The working environment for ED staff improved as a consequence of the target...”<sup>4</sup>

Nevertheless, compared with other countries’ wait times, the achievements might seem rather poor. According to a 2010 study, at the median hospital in the United States, 87% of ED visits lasted less than 4 hours, and 93% lasted less

than 6 hours.<sup>5</sup> In the United Kingdom, the National Health System set a policy in 2000 to reduce ED *visit lengths*.<sup>6</sup> Through concerted efforts, in 2008, 98% of ED visit lengths in the United Kingdom were 4 hours or less.<sup>7</sup> Many, however, have observed that the targets in the United Kingdom were sometimes achieved without improving patient care—and in fact, may have worsened quality.<sup>8,9</sup> Providers may have cut visits short or transferred patients inappropriately, known as “hitting the target, but missing the point.”<sup>1,6,10</sup>

### Lies, Damned Lies, and Statistics

Tenbenschel and colleagues examine whether we can trust the statistics above. Gaming is endemic, yet research into variation is rare. Unfortunately, there may be as many ways to game a performance measure as there are providers.

Decades of observers have pointed out potentially problematic reactions to performance measures. Back in 1956, Ridgeway made the following observation in the journal *Administrative Science Quarterly*:

*“Quantitative performance measurements—whether single, multiple, or composite—are seen to have undesirable consequences for over-all organizational performance. The complexity of large organizations requires better knowledge of organizational behavior...”*<sup>11</sup>

Hospitals are indeed complex systems in and of themselves, and national healthcare systems more complex yet. More recently, Braithwaite, writing in the *British Medical Journal*, noted the following:

*“Policy-mandated change is never given the same weight as clinically driven change. ... change is always unpredictable, hard won, and takes time, it is often tortuous, and always needs to be tailored to the setting.”*<sup>12</sup>

Gaming is not even the only potential hazard associated with performance measures. Writing about the UK’s national, extensive efforts to set targets and benchmarks, Mannion and Braithwaite observed 20 possible hazards, which they divided into four categories:

*“These are poor measurement (measurement fixation, tunnel vision, myopia, ossification, anachronism and quantification privileging), misplaced incentives and sanctions (complacency, silo-creation, overcompensation, undercompensation, insensitivity and increased inequality), breach of trust (misrepresentation, gaming, misinterpretation, bullying, erosion of trust and reduced staff morale), and politicisation of performance systems (political grandstanding and creating a diversion).”*<sup>10</sup>

Another ED-related example cited by Mannion and Braithwaite is the introduction of “hello nurses” in some British EDs—nurses hired to greet patients within the prescribed time frame and nothing more, thereby increasing costs but not providing any actual clinical benefit.<sup>10</sup> Also fitting within Mannion and Braithwaite’s taxonomy are the ways that staff and line management dealt with the intense pressure to meet the target in the four case study hospitals described by Tenbenschel and colleagues. The authors describe in detail how hospitals try to appear to have reached the target, from sending patients into “black holes,” to fudging the numbers, to increasing use of short stay and observation units.<sup>1</sup>

Recent increases in the incidence and lengths of observation

stays among patients in the United States<sup>13,14</sup> have been largely explained as a result of providers trying to delay or avoid hospital admissions, whether because of lack of space on a desired inpatient unit,<sup>15</sup> attempts to reduce (game) hospitalization and/or readmission rates,<sup>16-18</sup> or legitimate clinical reasons.<sup>19</sup> Informed by that and other research, many analysts and evaluators now analyze observation visits and outpatient ED visits separately from ED visits resulting in a hospital stay.

Beyond these kinds of ad hoc, after-the-fact adjustments, it is important to have independent verification and audits. Tenbenschel and colleagues used many tools that could and should be applied elsewhere to detect implausible patterns in the data. Particularly notable is their analysis of terminal digit preference bias among the four hospitals studied. For this measure, they looked only at visits with a recorded length of stay of between 360 and 369 minutes (since the target was 6 hours, or 360 minutes). Mathematically, roughly 10% of visits in that range should have had a last digit of 0 (in other words, a recorded length of stay of 360 minutes). Tenbenschel and colleagues found that terminal digit preference bias showed up after the introduction of the ED target at all four case study hospitals, with rates ranging from 11% (about what would be expected mathematically) to 38%. The higher the percentage, the more gaming. Tenbenschel and colleagues’ paper plots these bias estimates in informative ways. This analysis and similar analyses should be the norm whenever analysts and policy-makers look at performance measure data.

Performance measure developers, healthcare providers and administrators, policy-makers, and researchers in the field would do well to be both humbled and encouraged by this research. Process improvement benefits have ceiling effects, and even the best measure can be improved. What does it mean for gaming to have *increased* after the benefits were realized? Would a lower target have achieved the same benefits? These and other questions are hard to answer. In the end, we are still where Ridgeway was in 1956<sup>11</sup>: more research is needed.

### Acknowledgements

The author would like to thank the anonymous peer reviewers for their helpful feedback and Claire Korzen, editor at RTI International, for editing assistance.

### Ethical issues

Not applicable.

### Competing interests

Author declares that she has no competing interests.

### Author’s contribution

LML is the single author of the paper.

### References

1. Tenbenschel T, Jones P, Chalmers L, Ameratunga S, Carswell P. Gaming New Zealand’s emergency department target: how and why did it vary over time and between organisations? *Int J Health Policy Manag.* 2020;9(4):152-162. doi:10.15171/ijhpm.2019.98
2. Tenbenschel T, Chalmers L, Jones P, Appleton-Dyer S, Walton L, Ameratunga S. New Zealand’s emergency department target - did it reduce ED length of stay, and if so, how and when? *BMC Health*

- Serv Res.* 2017;17(1):678. doi:10.1186/s12913-017-2617-1
3. Ministry of Health NZ. How is My DHB Performing? <https://www.health.govt.nz/new-zealand-health-system/health-targets/how-my-dhb-performing-2017-18>. Accessed January 29, 2020. Published 2020.
  4. Chalmers LM. Inside the Black Box of Emergency Department Time Target Implementation in New Zealand [dissertation]. New Zealand: University of Auckland; 2014.
  5. Horwitz LI, Green J, Bradley EH. US emergency department performance on wait time and length of visit. *Ann Emerg Med.* 2010;55(2):133-141. doi:10.1016/j.annemergmed.2009.07.023
  6. Mason S, Weber EJ, Coster J, Freeman J, Locker T. Time patients spend in the emergency department: England's 4-hour rule—a case of hitting the target but missing the point? *Ann Emerg Med.* 2012;59(5):341-349. doi:10.1016/j.annemergmed.2011.08.017
  7. Howell E. *The Key Findings Report for the 2008 Emergency Department Survey*. Oxford: Picker Institute Europe; 2009.
  8. Mason S. Keynote address: United Kingdom experiences of evaluating performance and quality in emergency medicine. *Acad Emerg Med.* 2011;18(12):1234-1238. doi:10.1111/j.1553-2712.2011.01237.x
  9. Boyle A, Mason S. What has the 4-hour access standard achieved? *Br J Hosp Med (Lond).* 2014;75(11):620-622. doi:10.12968/hmed.2014.75.11.620
  10. Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Intern Med J.* 2012;42(5):569-574. doi:10.1111/j.1445-5994.2012.02766.x
  11. Ridgway VF. Dysfunctional consequences of performance measurements. *Adm Sci Q.* 1956;1(2):240-247. doi:10.2307/2390989
  12. Braithwaite J. Changing how we think about healthcare improvement. *BMJ.* 2018;361:k2014. doi:10.1136/bmj.k2014
  13. Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. *Health Aff (Millwood).* 2012;31(6):1251-1259. doi:10.1377/hlthaff.2012.0129
  14. Wright B, O'Shea AM, Ayyagari P, Ugwi PG, Kaboli P, Vaughan Sarrazin M. Observation rates at veterans' hospitals more than doubled during 2005-13, similar to Medicare trends. *Health Aff (Millwood).* 2015;34(10):1730-1737. doi:10.1377/hlthaff.2014.1474
  15. Delia D, Cantor JC. Emergency department utilization and capacity. *Synth Proj Res Synth Rep.* 2009;(17):45929.
  16. Himmelstein D, Woolhandler S. Quality Improvement: 'Become Good at Cheating and You Never Need to Become Good at Anything Else.' *HealthAffairsBlog*;2015. doi:10.1377/hblog20150827.050132
  17. Martin GP, Wright B, Ahmed A, Banerjee J, Mason S, Roland D. Use or abuse? a qualitative study of emergency physicians' views on use of observation stays at three hospitals in the united states and england. *Ann Emerg Med.* 2017;69(3):284-292.e282. doi:10.1016/j.annemergmed.2016.08.458
  18. Wright B, Martin GP, Ahmed A, Banerjee J, Mason S, Roland D. How the availability of observation status affects emergency physician decisionmaking. *Ann Emerg Med.* 2018;72(4):401-409. doi:10.1016/j.annemergmed.2018.04.023
  19. Wright B, Zhang X, Rahman M, Kocher K. Informing Medicare's two-midnight rule policy with an analysis of hospital-based long observation stays. *Ann Emerg Med.* 2018;72(2):166-170. doi:10.1016/j.annemergmed.2018.02.005