



# Improve the Design and Implementation of Metrics From the Perspective of Complexity Science

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



Junqiao Chen\*

### Article History:

Received: 9 February 2020  
Accepted: 24 March 2020  
ePublished: 4 April 2020

### \*Correspondence to:

Junqiao Chen  
Email:  
junqiao.chen@thecarevoice.com

### Abstract

From the perspective of complexity science, this commentary addresses Tenbenschel and colleagues’ study, which reveals varied gaming behaviours to meet the New Zealand Emergency Department (ED) metric. Seven complexity-informed principles previously published in this Journal are applied to formulate recommendations to improve the design and implementation of metrics. (1) Acknowledge unpredictability. When designing a metric, policy-makers need to leave room for flexibility to account for unforeseen situations. When implementing a metric, they need to promote sense-making of relevant stakeholders. (2) Sense-making shall be encouraged because it is a social process to understand a metric, align values and develop a coherent strategy. Sense-making is important to (3) cope with self-organised gaming behaviours and to (4) facilitate interdependencies between ED and other departments as well as organisations. (5) We also need to attend to the relationship between senior management and frontline staff. Additionally, to address one of the methodological weaknesses in Tenbenschel and colleagues’ study, (6) adaptive research approach is needed to better answer emerging questions. (7) Conflict should be harnessed productively. I hope these recommendations could limit gaming in future metrics and encourage stakeholders to view inevitable gaming as an improvement opportunity.

**Keywords:** Emergency Room, Policy Implementation, Gaming Behaviours, Complexity Science, New Zealand

**Copyright:** © 2021 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:** Chen J. Improve the design and implementation of metrics from the perspective of complexity science: 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.* 2021;10(5):273–276. doi:10.34172/ijhpm.2020.47

Tenbenschel and colleagues’ paper (the lead paper) reports the varied gaming behaviours across time and place with regards to the New Zealand Emergency Department (ED) target (95% of patients being admitted, discharged or transferred from an ED within 6 hours).<sup>1</sup> This commentary will attempt to explain and advise the design and implementation of such metrics from the perspective of complexity science.

This perspective is chosen because both this Journal and the lead authors have track records of complexity-informed publications (for example, see Kitson and colleagues’ paper<sup>2</sup> from this Journal and Tenbenschel’s other papers<sup>3,4</sup>). These enable me to take advantage of the audience’s familiarity with complexity to dive deep into this case study. While many previous complexity-informed discussions only provided high-level ideas, I am able to draw nuanced recommendation here thanks to the rich information from this case and the larger research that it is a part of. Since a unified complexity perspective does not exist,<sup>5</sup> I will sidestep academic disputes<sup>6</sup> and choose the following ‘rules of thumb’ presented by a recent paper of this Journal<sup>7</sup> as my pragmatic framework.<sup>8</sup>

### Acknowledge Unpredictability

First, acknowledge unpredictability in the design of metric. The arrival of patients to the ED and their needs are unpredictable. There will be situations in which even a very efficient ED cannot meet the target. A good metric design should provide a ‘way out,’ otherwise ‘the only means at their disposal’<sup>1</sup> of frontline staff is to game the metric. For example, in United States’ clinical quality measurement schema, flexibility in denominator is allowed for medical, patient, and system reasons to account for unforeseen situations.<sup>9</sup> The technical term for this is ‘exception.’ When I volunteered in an emergency medical service in the United States, we encountered a homeless patient visiting multiple EDs in a day despite no urgent treatment was needed. Aware of the situation, one ED allowed him to overstay to meet his basic needs, including having a sandwich and fully charging his phone. This ED might fail a time-centric target but what it did actually benefited the entire system. Better metric design might allow this case to be reported as an exception (for patient reasons) and removed from the denominator. Allowing flexibility in metric design could encourage clinicians to do

the right things in special circumstances.

Another way to account for unpredictability is to select core populations for whom certain timely action is non-negotiable, and to allow more flexibility for other less urgent populations and actions. An example of urgent action on a core population is the interpretation of imaging for stroke patients. In the United States, a nationally-endorsed measure requires that for stroke patients who arrive at the ED within 2 hours of symptom onset, they need to have a head computed tomography or magnetic resonance imaging scan interpreted within 45 minutes of arrival.<sup>10</sup> Otherwise, it is a critical failure. This idea bears resemblance to the ED triage process.<sup>11</sup> It also suggests that microscopic metrics are needed to complement the main metric. Because 'quality' is a multifaceted and multi-layered concept<sup>12</sup> (or 'ontologically emergent'<sup>13</sup>), we need multi-layered metrics, each with different ranges of flexibility to handle unpredictability. Multi-layered metrics will be summarised in another rule of thumb 'Facilitate Interdependencies.'

Unpredictable gaming might also occur during metric implementation.<sup>14</sup> The lead article reports many interwoven factors that could influence behaviours, resulting in unexpected levels of gaming across study sites.<sup>1</sup> Traditionally, audit is the approach to address gaming. However, audit might not work in complex adaptive systems due to its reductionist and determinist nature.<sup>15</sup> Instead, I recommend a peer-to-peer social approach, which will be elaborated in the next rule of thumb 'Encourage Sense-making.'

### Encourage Sense-Making

It was appraised that "unlike the British experience, the call for this target was 'bottom up,' lifted by the passion of concerned clinicians,"<sup>16</sup> so 'the passion persists and is unlikely to tolerate gaming on its patch.' However, the target was actually recommended by an advisory group, made up of senior ED clinicians and managers.<sup>17</sup> Since this is not entirely 'bottom up,' it needs to be socialised from 'top down' and understood by frontline clinicians ('sense-making'). This resembles the classical metaphor of clinical guidelines versus 'mindlines,' the latter of which is 'a collective sense-making by which knowledge, both explicit and tacit, is negotiated, constructed, and internalised in routine practice.'<sup>18</sup> Sense-making activities have been seen in national quality initiatives such as the Australian Health Service Safety and Quality Accreditation Scheme,<sup>19</sup> which is a complexity-informed initiative that undertook hundreds of consultations with health organisations to facilitate knowledge transfer and stakeholder engagement.<sup>20</sup>

By sense-making, various stakeholders (not just ED clinicians) could be more aligned on the goals, necessary resources and potential consequences of a metric. The objective is to engender a culture in which participants ask questions, admit ignorance, explore paradoxes and exchange different viewpoints,<sup>7</sup> such that each organisation has a coherent value stance and business strategy toward a metric. As echoed in the lead paper, such coherence was not always seen within each of its study sites.

Stakeholders should also be encouraged to brainstorm

possible gaming behaviours across the healthcare sector, and to apply local expert judgement on whether a certain behaviour is gaming or not in one's local context.<sup>21</sup> There is no right or wrong judgement, no blaming or retaliation (such as auditing). The objective is to increase awareness of varying local contexts, respect local expertise, and eventually strengthen professional code of conduct by holding each other accountable for fairness. Fairness is 'a plurality of social values, perspectives and interests.'<sup>22</sup> The goal of interorganisational sense-making is not to reach consensus on fairness, but to make the underline assumptions explicit and *ex ante*. A strengthened professional conduct from increased consciousness of fairness could limit 'wild gaming'<sup>14</sup> of a metric.

In brief, policy-makers, hospitals managers and clinicians need to encourage sense-making within an organisation to create a coherent business strategy, and encourage interorganisational sense-making to help learning, reflection and fairness across the industry. The next rule of thumb further analyses this.

### Recognise Self-organisation

Self-organisation is the emergence of new pattern through the activity of microscopic units that do not have access to global patterns but follow local rules. These rules could be explicit policies or customary procedures.<sup>23,24</sup>

In the case of our lead paper, a new pattern (skewed distribution of ED wait times) emerged through staff who, without knowledge of the global distribution, followed customary procedures of the people around them or senior staff,<sup>25</sup> to modify data entry (rounding terminal digits, for example). There was no written policy to guide staff modification. How often and to what extent modification occurred was based on a staff's own comfort level and instinct, without assessment of modifications by peers. Each person self-organised his/her gaming behaviour without coordination. (If they had coordinated, the pattern of skewness might be different.) One staff might be comfortable with rounding 369 minutes down to 360, while another staff might only be comfortable if it is within 365 minutes, for example.

To cope with this, we again need to encourage sense-making. By making the implicit rules more explicit and re-focusing on professional conduct<sup>26</sup> in the discourse, wild gaming could be reduced. What is left is the 'inevitable,' which should be viewed as improvement opportunity for policy-makers (how to measure what matters<sup>27</sup>) and healthcare providers (how to provide what matters).

### Facilitate Interdependencies

'Interdependent work across units is germane to the ED'<sup>28</sup> as ED clinicians must manage challenging constantly-changing dynamics at the boundaries of the ED and other departments and organisations. This is evident in Hospital 4 in the lead paper.

To facilitate interdependencies, there needs to be metrics and resources established towards the whole system, rather than the parts.<sup>29</sup> Silos and resistance are inevitable when a

systems approach was absent (eg, Hospital 3). Whole-system targets encourage generative relationships among various stakeholders and provoke creative ideas.<sup>29</sup> A hypothetical global metric could require patient care to be handed over to appropriate provider within six hours from the onset of the need for emergency care. Whole systems design could unite all components in the acute care journey, including pre-hospital, ED, inpatient, and post-discharge services.

The impact of whole-system targets on gaming is unknown. Based on the 'Motives, Opportunities and Means' framework in the lead paper, my hypothesis is that it will not reduce the opportunities or means to game, but might lower the motives. This is because now changing one data point is less likely to have a decisive system-wide effect. To effectively game a target, it will require explicit orchestrated acts from multiple players, which will be much harder to do operationally and ethically.

If this principle is viewed in combination with the first one ('Acknowledge unpredictability'), it is advised that metrics on three levels are needed in order to promote the best outcome: one on the overall acute care, one on the care within ED, and one on a core population and/or intervention.

### Attend to Human Relationship

The lead paper provides novel findings on how relationships between frontline staff and senior management can influence gaming. In public view and research, there is a strong lack of sympathy and attention toward managers.<sup>30,31</sup> The lead paper stresses managers' varied values and struggles over time and place. It is via their interactions with frontline staff that a policy could be implemented. I add to the lead paper that not only 'strategic behaviours' of managers are worth studying, also their nuanced interpersonal interactions with staff, which might be best observed by ethnography (that is how Google studied their managers<sup>32</sup>). Even in 'bottom-up' policy development, it is paramount to have managers involved to provide management's perspectives.

### Develop Adaptive Capabilities

This rule of thumb does not really apply to ED clinicians, who are already very adaptive to bridge the gap between a target ('work as imaged') and what could be done.<sup>33</sup> Gaming is the last resort of adaptation to meet a metric.

Adaptive capabilities are needed more for policy evaluation. Studying complex adaptive systems requires 'theoretically grounded, methodologically pluralistic, flexible and adaptive study design.'<sup>34</sup> The lead paper is part of a larger research with a rigorous mixed-method protocol.<sup>17</sup> From the formulation of new research questions to the secondary data analysis, the lead paper already brings new ideas on top of the original protocol, and I am appreciative to the lead authors' pursuit of knowledge. However, a limitation is inherent in secondary use of data, that the qualitative analysis does not augment, converge with or get embedded in the quantitative analysis, or vice versa.<sup>35</sup> For example, the quantitative analysis identifies terminal digit preference bias, but the qualitative analysis is not able to explain how exactly it occurred. In the lead paper, these two methods basically answer different questions,

because the data of one type was not collected after being informed by the analysis of another type. For the sake of argument, rather than calling it a mixed-method study, I will call it a study with two parts, each using different methods.

This goes back to the inflexibility of the original protocol, which does not allow future work when needed. As complexity thinkers comment: 'Research protocols consisting entirely of pre-ordained work packages arranged around a boxes-and-arrows diagram accompanied by tight milestones and timelines may be the stuff that funding panels' dreams are made of, but ... less likely to generate meaningful findings than studies which engage pragmatically with the multiple uncertainties involved and offer a flexible and emergent approach to exploring them.'<sup>34</sup> Large research projects should hold back some budget to allocate as and when it is needed in the future.<sup>36</sup>

### Harness Conflict Productively

The conflict between *business efficiency* and *integrity of care* in ED was well studied ethnographically.<sup>26</sup> The different levels of gaming are by-products from different levels of the conflict. Such tension could never be resolved,<sup>37</sup> as 'local-level discretion at the site of service delivery, or street-level bureaucracy, can never be fully managed out by such rational administrative logic.'<sup>26</sup> To make the best of this situation, we need to build relationship, maintain dialogue, mutually respect each profession's norm and value, and muddle through.<sup>7</sup> So to answer the lead paper's question 'whether or not gaming is a necessary evil... in the pursuit of improved health service,' my response is: no, it is not a necessary evil, but a *catalyst* for improvement if handled appropriately.

In conclusion, policy-makers, hospitals managers and clinicians need to work together to improve the design and implementation of metrics. Complex systems are inherently unpredictable, so we need to allow flexibility in metric design and encourage sense-making in implementation. Within an organisation, sense-making can help to cope with self-organised gaming behaviours, to facilitate interdependencies, and to develop a coherent strategy. Interorganisational sense-making could help to promote learning, reflection and fairness across the industry. It is important to view conflict as a catalyst for improvement, and to reserve research capacity to respond to emerging questions as we muddle through.

### Acknowledgements

Comments from Ms. Joanna Wu and an anonymous reviewer are gratefully appreciated, especially during a time of COVID-19 pandemic.

### Ethical issues

Not applicable.

### Competing interests

Author declares that she has no competing interests.

### Author's contribution

JC is the single author of the paper.

### References

1. Tenbenschel T, Jones P, Chalmers LM, 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. Kitson A, Brook A, Harvey G, et al. Using Complexity and Network Concepts to Inform Healthcare Knowledge Translation. *Int J Health Policy Manag.* 2018;7(3):231-243. doi:10.15171/ijhpm.2017.79
  3. Grudniewicz A, Tenbenschel T, Evans JM, Steele Gray C, Baker GR, Wodchis WP. 'Complexity-compatible' policy for integrated care? Lessons from the implementation of Ontario's Health Links. *Soc Sci Med.* 2018;198:95-102. doi:10.1016/J.SOCSCIMED.2017.12.029
  4. Tenbenschel T. Bridging complexity theory and hierarchies, markets, networks, communities: a 'population genetics' framework for understanding institutional change from within. *Public Manag Rev.* 2018; 20(7):1032-1051. doi:10.1080/14719037.2017.1364409
  5. Hannigan B. Connections and consequences in complex systems: Insights from a case study of the emergence and local impact of crisis resolution and home treatment services. *Soc Sci Med.* 2013;93:212-219. doi:10.1016/j.socscimed.2011.12.044
  6. Churruca K, Pomare C, Ellis LA, Long JC, Braithwaite J. The influence of complexity: a bibliometric analysis of complexity science in healthcare. *BMJ Open.* 2019;9(3):e027308. doi:10.1136/bmjopen-2018-027308
  7. Greenhalgh T. Bridging the 'two cultures' of research and service: can complexity theory help? Comment on "experience of health leadership in partnering with university-based researchers in Canada – a call to 're-imagine' research." *Int J Health Policy Manag.* 2020;9(2):87-88. doi:10.15171/ijhpm.2019.89
  8. Long KM, McDermott F, Meadows GN. Being pragmatic about healthcare complexity: Our experiences applying complexity theory and pragmatism to health services research. *BMC Med.* 2018;16(1):94. doi:10.1186/s12916-018-1087-6
  9. Centers for Medicare & Medicaid Services. CMS Web Interface Measure Documentation. Centers for Medicare & Medicaid Services. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharesavingsprogram/Downloads/cms-web-interface-measure-documentation.pdf>. Published 2019. Accessed February 9, 2020.
  10. Centers for Medicare & Medicaid Services. Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke Patients who Received Head CT or MRI Scan Interpretation within 45 minutes of ED Arrival. National Quality Forum. [https://cmit.cms.gov/CMIT\\_public/ViewMeasure?MeasureId=918](https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=918).
  11. New Zealand Ministry of Health. Emergency department triage. New Zealand Government. <https://www.health.govt.nz/our-work/hospitals-and-specialist-care/emergency-departments/emergency-department-triage>. Accessed March 24, 2020. Published June 30, 2011.
  12. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. *Bull World Health Organ.* 2017;95:368-337. doi:10.2471/BLT.16.179309
  13. O'Connor T, Wong HY. Emergent Properties. In: Stanford Encyclopedia of Philosophy. Spring 202. Metaphysics Research Lab, Stanford University; 2015. <https://plato.stanford.edu/archives/spr2020/entries/properties-emergent/>.
  14. Ariely D. How honest people cheat. *Harv Bus Rev.* 2008;86(2):24.
  15. Geyer R. Can complexity move UK policy beyond 'evidence-based policy making' and the 'audit culture'? Applying a 'complexity cascade' to education and health policy. *Polit Stud.* 2011;60(1):20-43. doi:10.1111/j.1467-9248.2011.00903.x
  16. Ardagh M. How to achieve New Zealand's shorter stays in emergency departments health target. *N Z Med J.* 2010;123(1316):95-103.
  17. Jones P, Chalmers L, Wells S, et al. Implementing performance improvement in New Zealand emergency departments: the six hour time target policy national research project protocol. *BMC Health Serv Res.* 2012;12(1):45. doi:10.1186/1472-6963-12-45
  18. Gabbay J, le May A. Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care. *BMJ.* 2004;329(7473):1013.
  19. Braithwaite J, Churruca K, Long JC, Ellis LA, Herkes J. When complexity science meets implementation science: A theoretical and empirical analysis of systems change. *BMC Med.* 2018;16(1). doi:10.1186/s12916-018-1057-z
  20. Greenfield D, Hinchcliff R, Banks M, et al. Analysing "big picture" policy reform mechanisms: The Australian health service safety and quality accreditation scheme. *Health Expect.* 2015;18(6):3110-3122. doi:10.1111/hex.12300
  21. Muller JZ. *The Tyranny of Metrics*. Princeton: Princeton University Press; 2018.
  22. Giuseppe M. Dealing with Fairness in Public Policy Analysis: A Methodological Framework. Luxembourg; Publications Office of the EU: 2017. doi:10.2760/75185
  23. Paley J, Eva G. Complexity theory as an approach to explanation in healthcare: A critical discussion. *Int J Nurs Stud.* 2011;48(2):269-279. doi:10.1016/j.ijnurstu.2010.09.012
  24. Yackinous WS. Overview of an Ecological System Dynamics Framework. In: Yackinous WS, ed. *Understanding Complex Ecosystem Dynamics: A Systems and Engineering Perspective*. Boston: Academic Press; 2015:83-91. doi:10.1016/B978-0-12-802031-9.00005-X
  25. Jiang DY, Lin YC, Lin LC. Business moral values of supervisors and subordinates and their effect on employee effectiveness. *J Bus Ethics.* 2011;100(2):239-252.
  26. McCann L, Granter E, Hassard J, Hyde P. "You Can't Do Both—Something Will Give": Limitations of the Targets Culture in Managing UK Health Care Workforces. *Hum Resour Manage.* 2015;54(5):773-791. doi:10.1002/hrm.21701
  27. Harris M, Tayler B. Don't Let Metrics Undermine Your Business. *Harv Bus Rev.* 2019;97(5):62-69.
  28. Nugus P, Carroll K, Hewett DG, Short A, Forero R, Braithwaite J. Integrated care in the emergency department: A complex adaptive systems perspective. *Soc Sci Med.* 2010;71(11):1997-2004. doi:10.1016/j.socscimed.2010.08.013
  29. Plsek PE, Wilson T. Complexity science: Complexity, leadership, and management in healthcare organisations. *Br Med J.* 2001;323(7315):746-749. doi:10.1136/bmj.323.7315.746
  30. Mark L. Managerialism and public attitudes towards UK NHS managers. *J Manag Med.* 1997;11(4):214-221. doi:10.1108/02689239710177314
  31. Faruk M. NHS managers' views of their culture and their public image: The implications for NHS reforms. *Int J Public Sect Manag.* 2003;16(7):549-563. doi:10.1108/09513550310500409
  32. Garvin DA. How Google Sold Its Engineers on Management. *Harv Bus Rev.* 2013;91(12):74-82.
  33. Back J, Ross AJ, Duncan MD, Jaye P, Henderson K, Anderson JE. Emergency department escalation in theory and practice: a mixed-methods study using a model of organizational resilience. *Ann Emerg Med.* 2017;70(5):659-671. doi:10.1016/j.annemergmed.2017.04.032
  34. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med.* 2018;16(1):95. doi:10.1186/s12916-018-1089-4
  35. Wisdom J, Creswell JW. *Mixed Methods: Integrating Quantitative and Qualitative Data Collection and Analysis While Studying Patient-Centered Medical Home Models*. Rockville, Maryland; 2013. <https://pcmh.ahrq.gov/page/mixed-methods-integrating-quantitative-and-qualitative-data-collection-and-analysis-while>.
  36. Engineering and Physical Sciences Research Council. *Management of Large Research Activities and Centres*. Swindon; 2006. <https://epsrc.ukri.org/newsevents/pubs/management-of-large-research-activities-and-centres/>.
  37. Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ.* 2001;323(7313):625. doi:10.1136/bmj.323.7313.625