

Choosing quality problems wisely: identifying improvements worth developing and sustaining

Christine Soong ¹, Hyung J Cho ^{2,3}, Kaveh G Shojania⁴

¹GIM, Mount Sinai Hospital, Toronto, Ontario, Canada

²Quality and Safety, NYC Health and Hospitals, New York, New York, USA

³Medicine, NYU Grossman School of Medicine, New York, New York, USA

⁴Medicine, University of Toronto Faculty of Medicine, Toronto, Ontario, Canada

Correspondence to

Dr Christine Soong, GIM, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; christine.soong@utoronto.ca

Accepted 17 April 2020

Published Online First

29 April 2020

The years since launch of the Choosing Wisely Campaign^{1 2} have seen an increase in studies reporting interventions aimed at reducing low-value care, from unindicated imaging^{3 4} and laboratory tests^{3 4} to prescriptions for medicines^{5–7} that deliver no net benefit. Many describe use of some combination of the usual suspects of intervention types: education,^{5 8} performance feedback data (sometimes described as audit-feedback, social comparison or peer comparison), policy changes (eg, restricting release of blood products to 1 unit at a time based on a haemoglobin cut-off in non-haemorrhage situations)⁹ and computer provider order entry (CPOE)-based modifications (eg, alerts) or restrictions.¹⁰

In this issue of *BMJ Quality and Safety*, Ambasta and colleagues examined the impact of a social comparison and education intervention on routine blood test utilisation at a single academic medical centre.³ Trainees and attending physicians each received their own performance feedback in comparison with a group aggregate. Compared with controls, the intervention groups ordered fewer routine laboratory tests (incidence rate ratio 0.89; 95% CI 0.79 to 1.00; $p=0.048$) with an associated cost savings of \$68 877 in Canadian dollars ($p=0.020$).

Before commenting on the intervention itself, the value of controlled comparisons bears noting. The statistical process control charts shown by Ambasta *et al* clearly demonstrate special cause variation, with an obvious reduction in the weekly mean number of target laboratory tests per patient-day. In addition, the reduction is temporally associated with the roughly 4.5-month intervention period and remains sustained during

a 1-year post-intervention period. Yet, the three control sites show the same pattern. The results may thus constitute a case of secular trends, with a ‘rising tide lifting all boats’.¹¹ The rising tide here presumably owes its origin to the widespread interest in eliminating low-value care, as with the Choosing Wisely Movement.

The authors argue against this interpretation with their difference-in-difference analysis that the decrease in testing at the intervention site included an 11% reduction beyond that observed at the control sites. How to interpret this small incremental reduction remains unclear. First, the control sites had slightly higher baseline testing rates than the intervention sites and the post-intervention differences look very similar. It is thus difficult to know if the additional reduction suggested by the analysis reflects a real effect or some degree of intrinsic differences between the control and intervention sites. Second, while difference-in-difference analyses have enjoyed long-standing use, experts have increasingly called attention to the potential for bias due to regression to the mean^{12 13} and looked for newer methods for drawing inferences from non-randomised comparisons.¹⁴

Even if the effect is attributable to the intervention and statistically significant, the ‘clinical significance’ remains questionable. An argument can be made about preventing potential downstream testing and treatment. But on the face of it, saving about \$69 000 Canadian dollars (about £40 000 or US\$49 000) over a year seems on the small side. Moreover, the costs here do not take into account the difference between fixed and incremental costs.¹⁵ The authors state that they obtained the costs of the laboratory



► <http://dx.doi.org/10.1136/bmjqs-2019-010118>



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Soong C, Cho HJ, Shojania KG. *BMJ Qual Saf* 2020;**29**:790–793.

tests targeted by their study from laboratory services provider to their provincial health system. Such numbers typically reflect dividing the costs for the machines and reagents and personnel by the number of tests run in a given time period. Yet, once the laboratory exists, the marginal costs from running more tests are typically vanishingly small. True savings do not accrue until one eliminates enough tests to pay one less technician or order one less batch of reagents.

COMMON PROBLEMS WITH QUALITY IMPROVEMENT INTERVENTIONS

The issues encountered by Ambasta and colleagues highlight recurring themes among quality improvement work: small effect sizes, unintended consequences and scale/sustainability.

1. Small effect sizes

In pointing these issues out, we do not single out the thoughtful intervention developed and reported by Ambasta and colleagues. Small effect sizes represent the rule not the exception when it comes to this sort of work. The social comparison they employed falls into the general category of audit and feedback. While this strategy can modify clinician behaviour,¹⁶ a Cochrane review of 140 randomised trials reported a median improvement in desired clinician behaviour of only 4%.¹⁷ A similar analysis of computerised decision support interventions showed a strikingly similar result.¹⁸ Across 32 trials involving various improvement targets (eg, ordering venous thromboembolism prophylaxis for inpatients or influenza vaccination for outpatients), computer reminders and more complex forms of decision support improved the proportion of patients who received the target processes of care by a median of only 4.2%. In other words, if 30% of patients in the control group received the desired element of care, 34% of patients in the intervention group would receive it. One of us (KGS) has recently updated this meta-analysis (currently under peer review). Across 100 trials, computerised reminders and decision support interventions delivered a median absolute improvement of 7%—that is, whatever percentage of patients were receiving the desired care at baseline, the typical intervention increased the percentage of patients who received that desired care by only 7%.

Lest these comments be interpreted as implying that something is fundamentally wrong with improvement science, it's worth remembering that most clinical research produces null results. One telling analysis showed that, once trial registration became mandatory, the proportion of large cardiovascular trials reporting positive results dropped from 57% to 8%.¹⁹ Most areas of biomedical research produce small, incremental gains, with major improvements in patient outcomes few and far between. And, this

same pattern of small effects (not to mention serious side effects) for new clinical treatments exists despite massive public and private investments in basic and clinical research—many orders of magnitude more than the resources invested in the science and practice of improving healthcare.

2. Unintended consequences

Not only do these common types of improvement interventions—computerised decision support, audit and feedback—typically achieve small improvements, they have downsides and require non-trivial resources to implement. For computerised reminders, clinicians develop alert fatigue and frustration over pop-up screens interrupting their workflow, especially as the interruptions do not relate to the task at hand. For instance, no clinician wants to see a reminder to screen for thromboembolism risk while trying to order intravenous antibiotics for a patient with sepsis. Social comparisons, such as report cards and other feedback type interventions, can also produce unintended effects.²⁰ In other contexts, we certainly recognise that receiving performance critiques can produce negative reactions. In a qualitative analysis, social comparison elicited emotions among clinicians such as frustration/irritation (67%), resentment (33%) and embarrassment/shame (17%).²¹ In addition, the vast majority (83%) of clinicians in the study reflected on how they felt unfairly penalised. Furthermore, clinicians may suffer 'scorecard overload', or a type of informational feedback fatigue. The larger question remains, how many performance metrics can a clinician interpret and manage well? Answering this question likely requires weighing a number of factors, including the magnitude of the improvement produced by feeding back the performance reports, the degree to which the effect persists over time, the resources required to sustain performance feedback, and the risks for adverse effects such as provider burnout or frustration.

3. Scale and sustainability

While all of the aforementioned factors deserve careful consideration, we have found scale and sustainability in particular to receive the least attention. Working in hospital settings, our experience is that interventions frequently target a single geographical area (eg, a hospital ward) or single clinical service (eg, the general surgery service) and tend to not spread to other wards or services. In addition, how long should an intervention like performance feedback continue to achieve a sustained effect? Within the published literature, the feedback delivered varies widely in frequency and rarely persists for more than 1 year.¹⁷ There is a paucity of data evaluating the effect after feedback ceases. Study designs focusing on active withdrawal of the intervention would prove invaluable for institutions with limited resources. Experts have proposed practical tips to optimise success when using practice

feedback interventions although overall effectiveness remains unknown.²²

MOVING FORWARD: CHOOSE TARGET PROBLEMS AND INTERVENTIONS WISELY

In summary, interventions aimed at changing clinician behaviour typically achieve small improvements, often carry unintended consequences²³ and require resources to maintain. In that context, we need to think not just about ‘choosing wisely’ when we decide when to order a test or treatment for a patient, but when and how we intervene to improve care. The first question to answer is whether the problem is worth the investment (‘Is the juice worth the squeeze?’). Prior to the decision to intervene, we should carefully consider the following factors: Does addressing the problem significantly impact on any elements in the quadruple aim²⁴ (population health, individual experience of care, per capita cost of healthcare, and provider well-being)? How likely it is that the intervention will produce a clinically worthwhile effect? Is the required effort (eg, direct, indirect and opportunity costs, time) justified? And is it possible to mitigate against unintended consequences?

After answering these questions and carefully choosing to move forward, additional factors to consider are scale and sustainability. For instance, it would seem sensible to maximise results by spreading an effective intervention beyond one single area of overuse such as inpatient laboratory testing. In particular, the approach of targeting each specific Choosing Wisely recommendations in isolation misses an opportunity for a broader, more impactful approach. Cited barriers to providing high-value care^{25 26} are often similar among a range of low-value practices, so applying a comprehensive approach across a number of areas would seem logical rather than focusing on each area as a discreet problem each with its own unique intervention. For example, CPOE-enabled changes are often used to reduce inpatient laboratory testing¹⁰ which can be scaled up to also reasonably guide appropriate medication prescribing (eg, antimicrobials²⁷) and diagnostic testing (eg, chest radiographs²⁸), and the impact of the sum could be greater than the individual components, making the investment worthwhile. Broadly sweeping systems-based interventions that have capacity to influence behaviour change across multiple areas have demonstrated success.²⁹ When tackling low-value care, it is time to move away from ‘one-offs’ to a systematic approach that leverages carefully chosen interventions scaled across the health system in a sustainable fashion.

Twitter Christine Soong @christinesoong and Hyung J Cho @HyungChoMD

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iDs

Christine Soong <http://orcid.org/0000-0002-1397-9671>

Hyung J Cho <http://orcid.org/0000-0002-1737-5701>

REFERENCES

- 1 Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *JAMA* 2012;307:1801–2.
- 2 Levinson W, Kallewaard M, Bhatia RS, *et al.* ‘Choosing Wisely’: a growing international campaign. *BMJ Qual Saf* 2015;24:167–74.
- 3 Ambasta A, Ma IWY, Woo S, *et al.* Impact of an education and multilevel social comparison-based intervention bundle on use of routine blood tests in hospitalised patients at an academic tertiary care hospital: a controlled pre-intervention post-intervention study. *BMJ Qual Saf* 2020;29:825–32.
- 4 Strauss R, Cressman A, Cheung M, *et al.* Major reductions in unnecessary aspartate aminotransferase and blood urea nitrogen tests with a quality improvement initiative. *BMJ Qual Saf* 2019;28:809–16.
- 5 Bruno C, Pearson S-A, Daniels B, *et al.* Passing the acid test? evaluating the impact of national education initiatives to reduce proton pump inhibitor use in Australia. *BMJ Qual Saf* 2020;29:365–73.
- 6 Fan-Lun C, Chung C, Lee EHG, *et al.* Reducing unnecessary sedative-hypnotic use among hospitalised older adults. *BMJ Qual Saf* 2019;28:1039–45.
- 7 Dreischulte T, Donnan P, Grant A, *et al.* Safer prescribing—a trial of education, informatics, and financial incentives. *N Engl J Med* 2016;374:1053–64.
- 8 Soong C, Shojania KG. Education as a low-value improvement intervention: often necessary but rarely sufficient. *BMJ Qual Saf* 2020;29:353–7.
- 9 Gob A, Bhalla A, Aseltine L, *et al.* Reducing two-unit red cell transfusions on the oncology ward: a choosing wisely initiative. *BMJ Open Qual* 2019;8:e000521.
- 10 Eaton KP, Levy K, Soong C, *et al.* Evidence-based guidelines to eliminate repetitive laboratory testing. *JAMA Intern Med* 2017;177:1833–9.
- 11 Chen Y-F, Hemming K, Stevens AJ, *et al.* Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Qual Saf* 2016;25:303–10.
- 12 Daw JR, Hatfield LA. Matching and regression to the mean in difference-in-differences analysis. *Health Serv Res* 2018;53:4138–56.
- 13 Daw JR, Hatfield LA. Matching in difference-in-differences: between a rock and a hard place. *Health Serv Res* 2018;53:4111–7.
- 14 Walkey AJ, Bor J, Cordella NJ. Novel tools for a learning health system: a combined difference-in-difference/regression discontinuity approach to evaluate effectiveness of a readmission reduction initiative. *BMJ Qual Saf* 2020;29:161–7.
- 15 Rauh SS, Wadsworth EB, Weeks WB, *et al.* The savings illusion—why clinical quality improvement fails to deliver bottom-line results. *N Engl J Med* 2011;365:e48.
- 16 Jamtvedt G, Young JM, Kristoffersen DT, *et al.* Does telling people what they have been doing change what they do? A

- systematic review of the effects of audit and feedback. *Qual Saf Health Care* 2006;15:433–6.
- 17 Ivers N, Jamtvedt G, Flottorp S, *et al.* Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;7:1–229.
 - 18 Shojania KG, Jennings A, Mayhew A, *et al.* Effect of point-of-care computer reminders on physician behaviour: a systematic review. *Can Med Assoc J* 2010;182:E216–25.
 - 19 Kaplan RM, Irvin VL. Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS One* 2015;10:e0132382.
 - 20 Navathe AS, Emanuel EJ. Physician peer comparisons as a nonfinancial strategy to improve the value of care. *JAMA* 2016;316:1759–60.
 - 21 Payne VL, Hysong SJ. Model depicting aspects of audit and feedback that impact physicians' acceptance of clinical performance feedback. *BMC Health Serv Res* 2016;16:260.
 - 22 Brehaut JC, Colquhoun HL, Eva KW, *et al.* Practice feedback interventions: 15 suggestions for optimizing effectiveness. *Ann Intern Med* 2016;164:435–41.
 - 23 Shojania KG. Are increases in emergency use and hospitalisation always a bad thing? Reflections on unintended consequences and apparent backfires. *BMJ Qual Saf* 2019;28:687–92.
 - 24 Sikka R, Morath JM, Leape L. The quadruple aim: care, health, cost and meaning in work. *BMJ Qual Saf* 2015;24:608–10.
 - 25 Zikmund-Fisher BJ, Kullgren JT, Fagerlin A, *et al.* Perceived barriers to implementing individual Choosing Wisely® recommendations in two national surveys of primary care providers. *J Gen Intern Med* 2017;32:210–7.
 - 26 Bishop TF, Cea M, Miranda Y, *et al.* Academic physicians' views on low-value services and the choosing wisely campaign: a qualitative study. *Health Care* 2017;5:17–22.
 - 27 Rittmann B, Stevens MP. Clinical decision support systems and their role in antibiotic stewardship: a systematic review. *Curr Infect Dis Rep* 2019;21:29.
 - 28 Sy E, Luong M, Quon M, *et al.* Implementation of a quality improvement initiative to reduce daily chest radiographs in the intensive care unit. *BMJ Qual Saf* 2016;25:379–85.
 - 29 Lee VS, Kawamoto K, Hess R, *et al.* Implementation of a value-driven outcomes program to identify high variability in clinical costs and outcomes and association with reduced cost and improved quality. *JAMA* 2016;316:1061–72.