

Measuring overuse: a deceptively complicated endeavour

Christine Soong,^{1,2,3} Scott M Wright⁴

¹Division of General Internal Medicine, Sinai Health, University of Toronto, Toronto, Ontario, Canada

²Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

³Center for Quality Improvement and Patient Safety, University of Toronto, Toronto, Ontario, Canada

⁴Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Correspondence to

Dr Christine Soong, 19.105-600 University Avenue, Toronto M5G 1X5, Canada; Christine.Soong@utoronto.ca

Accepted 29 June 2021

Published Online First

19 July 2021

Choosing Wisely is a campaign aimed at reducing unnecessary tests, procedures and treatments.¹ The goals of the campaign have been to both reduce health-care expenditures and to prevent harms associated with inappropriate care—such as adverse effects of medications or radiation exposure from unwarranted imaging. The premise is clear and rather simple—do not order something the patient does not need. Yet, when applying a rigorous scientific improvement lens to reducing overuse, measurement nuances make evaluating this phenomenon anything but simple. Appropriateness is specific to clinical scenarios rather than being a representative property intrinsically coupled with individual tests or procedures. Assigning a label of ‘unnecessary’ to a test requires defining the target denominator, which can include a defined patient population, indication and/or test (ie, it is unnecessary to perform imaging for patients with low-back pain without risk factors). When quantifying appropriate use of a test or procedure, differences in the characterisation of the denominator will naturally affect the findings; this makes comparisons challenging when there is a lack of standardisation across measurement methodologies.

In this issue of *BMJ Quality & Safety*, Müskens and colleagues present a systematic review of studies published up to February 2020 reporting the prevalence of low-value diagnostic testing.² Studies were conducted in both ambulatory care and hospital settings; the findings yield prevalence estimates for the overuse of tests according to relevant guidelines such as Choosing Wisely, ‘Do not do recommendations’, the English National Institute for Health and Care Excellence and those from professional societies. As part of their analysis, Müskens and colleagues classify studies based on the category used for the denominator (referred to

as ‘lenses’). Two categories or lenses are identified: (1) the service-centric lens that calculates the proportion of test indications defined as low value and (2) the patient-centric lens that determines the proportion of patients who undergo low-value tests. The patient-centric lens can be further subdivided into a patient-indication lens and a patient-population lens. For example, ‘low-value low-back imaging’ can be defined as the number of low-value low-back imaging scans performed as a percentage of all low-back imaging scans (service-centric) or the number of patients receiving a low-value imaging scan as a proportion of all patients who presented with low-back pain (patient-centric). A different lens with a different denominator will answer a different question.

Across the studies, the median prevalence of low-value or unnecessary diagnostic testing was found to be 11%. At first glance, this estimate does not appear large enough to signal a significant quality problem that deserves attention or concern. However, the authors discovered a very wide range across studies—from 0.1% to 98%. This expansive range in the prevalence of overuse is notable in two ways. First, studies are using different denominators to measure use when looking at the same test; for example, various studies analysing imaging for low-back pain consider the patient-indication, patient-population and service-based denominators resulting in overuse estimates ranging from 4.4% to 86.2%. The different perspectives taken into account in the denominators also necessitate distinctive methodologies and measurements. Second, even when different studies use the *same* denominator when evaluating a single test, wide ranges of estimates are noted; studies examining imaging for low-back pain using a patient-indication denominator report an



► <http://dx.doi.org/10.1136/bmjqs-2020-012576>



© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Soong C, Wright SM. *BMJ Qual Saf* 2022;**31**:8–10.

unnecessary test range of 4.4%–63.1%. Distinctions across patient populations or those tied to community expectations or institutional norms may contribute to wide disparities in the proportion of patients with low-back pain undergoing unnecessary testing.

Whether the wide range in overuse is attributable to real versus measurement-related causes, the studies suggest that it is at least in part driven by variations in practice patterns, thereby justifying the need for practice standardisation. This is particularly evident in studies where physician-level ordering patterns are reported.^{3,4} Drilling down to these specific areas and contexts of overuse can more decisively provide granular data on where and how to invest resources to intervene. The root causes of the high rate of overuse may be similar across settings or there may be unique local factors amenable to specifically targeted interventions.

Despite increased characterisation of low-value care, several foundational elements are holding back efforts to advance and ameliorate overuse measurement. First, a lack of target benchmarks in medical overuse precludes our ability to gauge the success of improvement efforts. Because the optimal use rate has not been agreed on and aiming for zero low-value testing would likely result in ‘underuse’ and could increase diagnostic errors, a target benchmark might be useful to guide quality improvement investigators. Second, agreement and consistency in low-value measurement definitions are currently lacking. For example, each research team measures overuse in different way, resulting in disparate measures and units—even when studying the same tests. Using different denominators or lenses when assessing the data naturally affects the magnitude of overuse quantification. While one metric may not always be superior to others, decision-makers must bear in mind the unit of measurement when making comparisons across studies or populations and settings. Third, overuse measurements for any test or therapy are usually context-specific (eg, low-risk patients being screened must be considered differently from those who are acutely ill or symptomatic). This reality requires clinical information such as clinical indication and patient condition; these details and specificity are well beyond what is captured in most administrative datasets. Such considerations underscore the problem of aggregating all-cause use data for any particular test. Fourth, when data are pulled for overuse measurement from distinct sources (such as administrative data vs chart review), the assessments of appropriateness may differ because of specifics including timing, location of the patient, sequence of testing and other factors. For example, in a study to reduce unnecessary sedative initiation among inpatients, it was imperative to review the charts to both identify patients who were not taking sedatives prior to hospitalisation *and* were prescribed a sedative in the hospital for the indication of insomnia.⁵ This nuanced clinical information requires meticulous data extraction and is certainly

not accessible from most administrative databases. Fifth, many reports do not include sufficient detail for readers to critically appraise or reproduce data collection and measurement. In a review of health service overuse, the authors found that only 37 of 160 measures had specified definitions of numerators, denominators and exclusions.⁶ Finally, defining and quantifying unintended consequences of measuring and reducing overuse (such as patient anxiety related to unnecessary testing and awaiting results, or diagnostic cascades related to further work-up of ‘incidentalomas’⁷ are underemphasised in the literature. These outcomes may actually have significant negative effects.⁸ All of these elements hamper efforts to quantify and address overuse reliably. Ultimately, advances in the research agenda for medical overuse may be necessary to establish valid methods of measurement and standardisation in the reporting of results. This will allow for better comparison and interpretation of data across studies

The study by Müskens and colleagues in this issue of the journal helps to characterise the prevalence of overuse from a number of angles, suitable to answer different questions. This information can guide improvement teams to appropriately align the intervention target and question at hand with the best suited unit of measurement. For example, if an intervention is directed at providers, such as academic detailing aiming to reduce antimicrobial prescribing in a clinic setting, the outcome measures should focus on patient or population denominators, such as the proportion of clinic patients prescribed unnecessary antibiotics. If a radiology department implements a new ordering process incorporating validated risk assessment tools for venous thromboembolism aiming to improve use of chest CT, the outcome measure should use a service denominator such as proportion of all chest CTs ordered to rule out pulmonary embolism. Explicitly defining the outcome measures and the denominators, aligned with the research question, will also allow for ‘apples to apples’ comparisons across studies.

In looking forward to gain clarity about optimal methods for characterising overuse, the research agenda must also seek to clarify the factors that are consequential in promoting overuse—particularly exploring the roles of clinical uncertainty and cognitive biases. The field needs to innovate with creative strategies to substantively mitigate overuse. A research agenda for medical overuse has been proposed,⁹ with many of the goals appropriately focused on validity and quality. A combined research and quality improvement agenda could include

- ▶ Use of agreed-upon appropriateness criteria, preferably with a consistent denominator for overuse measurement.
- ▶ Balancing measures so as to minimise unintended harms or consequences such as underuse, health inequities and diagnostic errors.⁸ It is necessary to also consider patient

experience measures and cost savings (both direct and indirect).

- ▶ Presentation of reliable and valid overuse data clearly and meaningfully—making this information accessible and transparent for interested users (especially clinicians and administrators).
- ▶ Identify major drivers of overuse and study targeted interventions to counteract their impact.
- ▶ Leverage systems-focused interventions that push evidence-based practice and behaviours (eg, inability to order a glycosylated haemoglobin test where a result has been reported in the last 3 months).

Going forward, a more precise and systematic approach to overuse measurement will serve to more accurately characterise the scope of the problem. This will allow for the tracking of meaningful advances in the field while also providing direction about where future efforts should be focused. After nearly a decade of ‘Choosing Wisely’, it is time to focus on refining indicators to preside over the next phase of this initiative. Without a thoughtfully committed approach, it will be difficult to understand the real impact of any of the resource stewardship efforts.

Twitter Christine Soong @christinesoong

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.

REFERENCES

- 1 Choosing wisely: an initiative of the ABIM foundation. Available: <https://www.choosingwisely.org> [Accessed 25 May 2021].
- 2 Müskens JLJM, Kool RB, van Dulmen SA. Overuse of diagnostic testing in healthcare: a systematic review. *BMJ Qual Saf* 2022;31:54–73.
- 3 Colla CH, Sequist TD, Rosenthal MB, *et al.* Use of non-indicated cardiac testing in low-risk patients: choosing wisely. *BMJ Qual Saf* 2015;24:149–53.
- 4 Verma AA, Guo Y, Jung HY, *et al.* Physician-level variation in clinical outcomes and resource use in inpatient general internal medicine: an observational study. *BMJ Qual Saf* 2021;30:123–32.
- 5 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.
- 6 Chan KS, Chang E, Nassery N, *et al.* The state of overuse measurement: a critical review. *Med Care Res Rev* 2013;70:473–96.
- 7 Oren O, Kebebew E, Ioannidis JPA. Curbing unnecessary and wasted diagnostic imaging. *JAMA* 2019;321:245–6.
- 8 Mathias JS, Baker DW. Developing quality measures to address overuse. *JAMA* 2013;309:1897–8.
- 9 Morgan DJ, Brownlee S, Leppin AL, *et al.* Setting a research agenda for medical overuse. *BMJ* 2015;351:h4534.