How much diagnostic safety can we afford, and how should we decide?
A health economics perspective

David E Newman-Toker,1 Kathryn M McDonald,2,3 David O Meltzer4

INTRODUCTION
Controlling the costs of healthcare, which now exceed US$2.7 trillion, is an economic imperative.1–3 Costs of diagnostic testing probably account for more than 10% of all healthcare costs, and that fraction is rising rapidly over time,1 with advanced diagnostic imaging leading the way and diagnostic laboratory testing a close second.3 Molecular diagnostic testing for ‘personalised medicine’ may fuel disproportionate rises.6 7 Some diagnostic tests are misused or overused, with waste from diagnostic imaging alone estimated at more than US$25 billion.8

At the same time, diagnostic errors are frequent and often result in death or disability,9 with recent estimates suggesting more than a million a year harmed by diagnostic error in the USA.10 For those harmed, direct costs accrue from failure to treat the true condition, inappropriate testing and treatments for the incorrectly diagnosed one, and medicolegal costs or payments.11 Indirect costs also arise from defensive medicine, increased medical liability premiums, and downstream effects.11 The annual costs of ‘defensive medicine’ alone—mostly unnecessary diagnostic tests obtained to guard against malpractice law suits—are at least US$45–60 billion12 13 and perhaps hundreds of billions.14

Given these facts, public awareness campaigns (eg, ‘Choosing Wisely’) have sought to foster dialogue between doctors and patients about potential ways to improve the safety and efficiency of diagnosis.17 Nevertheless, it remains challenging to determine whether diagnostic tests are being overused or underused and when ‘more’ diagnosis is not ‘better’ diagnosis. In this article, we leverage a case study example (box 1) to explore complex inter-relationships between diagnostic test characteristics, appropriate use, actual use, diagnostic safety and cost effectiveness. We frame our discussion around the question, ‘How much diagnostic safety can we afford?’ to assess the role of economic analysis and suggest areas for future research related to the public health imperative of better value and safety in diagnosis.

FINITE RESOURCES, INFINITE DEMAND, AND WASTE IN DIAGNOSTIC TEST USE
The demand for healthcare and technological advances in diagnosis is effectively infinite.29 For acute dizziness in the emergency department (ED), the number of patients seeking care has more than doubled over the past two decades.18 The continuous upward spiral in what we ‘can do’ in medicine through technological advances propels a progressively widening gap between what we ‘actually do’ and what we ‘should do.’30 For example, the fraction of ED patients with dizziness undergoing CT scans rose steadily from 9% in 1995 to over 40% currently,18 but doing so has not increased the yield of stroke or other neurologic diagnoses.31 32 Some advances produce societal benefits exceeding their costs, while others do not.33 More importantly, higher costs are often unrelated to health outcomes—in other words, healthcare resources are often spent inefficiently.34 Some frequently used tests appear to offer no diagnostic benefit at all,31 never mind a downstream health benefit. Use of diagnostic tests under such circumstances is generally considered inappropriate use (box 2).

Most diagnostic decisions are influenced by factors other than maximising individual patient outcomes or even total societal benefit. In our case study example, CT scans are grossly overused in...
an effort to ‘rule out’ stroke. Studies showing small area practice variation in use of CT scans for dizziness confirm that factors other than optimal patient care are at play in decisions to order a CT. Diagnostic test use by physicians is driven by a mix of incentives and disincentives (box 3), with mixed results for optimal patient care. Risk aversion may be a particularly important factor contributing to wasteful diagnostic test overuse. In the case of CTs for dizziness, major contributing factors include knowledge gaps regarding best evidence, local standards for test-ordering (peer practices), patient preferences and fear of litigation for missed stroke.

Another incentive often mentioned by physicians is reduced time for patient interaction53 from crowding out by documentation tasks. It is often faster for a physician to order a test than to think through its appropriateness. Tackling such practical barriers will be important for reducing test overuse—for example, measuring and rewarding diagnostic quality and efficiency could provide incentives that counteract needless test overuse.

### Box 2 Definitions for inappropriate use of diagnostic tests

**Underuse**—The failure to provide a diagnostic test when it would have produced a favourable outcome for a patient. An example would be failure to provide pap smears to eligible patients.

**Overuse**—Providing a diagnostic test in circumstances where the potential for harm exceeds the potential for benefit. An example would be conventional cerebral angiography to rule out brain aneurysm in a patient with typical, uncomplicated migraine-type headaches and a normal neurologic examination.

**Mispuse**—When an appropriate diagnostic test has been selected but a preventable complication occurs and the patient does not receive the full potential benefit of the test. An example would be pulmonary CT angiography to diagnose pulmonary embolus in a patient with dyspnoea who has a known contrast dye allergy but receives no pretreatment for a possible allergic reaction.

Modified from AHRQ Web M&M35

**DIAGNOSTIC SAFETY: DIAGNOSTIC TEST USE AND DIAGNOSTIC ERROR**

Eliminating all diagnostic errors is impossible, since diagnosis occurs under uncertainty. Efforts to reduce uncertainty toward zero result in increasing marginal costs with diminishing marginal returns for patient safety. Nevertheless, potential exists to improve diagnostic safety. In the ED dizziness/vertigo case, an estimated 35% of underlying strokes are missed, despite the fact that relatively simple, non-invasive, bedside physical examination tests have been shown repeatedly to identify more than 99% of strokes. Estimates of diagnostic error-associated adverse events are 0.1% of primary care visits and 0.4% of hospital admissions, but the diagnostic error rate is likely much higher, since harm does not invariably occur. Ballpark estimates place the overall diagnostic error rate in the range of 10–15%. Rates appear to vary by specialty (eg, 5% in radiology vs 12% in emergency medicine), by disease (eg, 2% of myocardial infarctions vs 9% of strokes), and especially by clinical presentation (eg, 4% of strokes presenting traditional symptoms vs 64% of strokes presenting non-traditional symptoms), with atypical and non-specific presentations (such as dizziness) increasing the risk of misdiagnosis dramatically. Misdagnoses, when they relate to test ordering (~15% of misdiagnoses), generally result from underuse. For example, this appears to be the case for young patients presenting dizziness to the ED whose strokes may be missed, sometimes with devastating consequences. Test overuse, however, can also result in diagnostic error (eg, false positives) or

overdiagnosis (ie, correct but unnecessary diagnoses). Overdiagnosis includes conditions deliberately being sought, but too mild to warrant treatment, and unrelated ‘incidentalomas’ that may beget further inappropriate testing or treatment. Tests may not be appropriate when the expected probability of disease is too low or too high for the test to resolve diagnostic uncertainty in a way that positively influences clinical decision making. Since overuse often occurs in patients with expected low disease prevalence, the risk of false positives is particularly high. Sometimes blind obedience to false positive or false negative test results from advanced diagnostic technology can lead to error—this appears to be the case with CT for ‘ruling out’ stroke in acute dizziness.

### TESTS PROPERTIES AND TEST-ORDERING: ACCURACY, ACTION THRESHOLDS AND DECISION ANALYSIS

We assume most clinicians intuitively understand the notion that tests are imperfect, and, because of the risk of both false positives (imperfect specificity) and false negatives (imperfect sensitivity), that they should be judicious in choosing and interpreting tests. Physicians probably understand the basic concept of how a test influences an estimate of pretest probability to yield a post-test probability (Bayesian logic) and that obtaining a diagnostic test whose post-test probability could not affect management (ie, could not result in crossing a subsequent test or treatment decision threshold) is usually unjustified. Clinicians also presumably grasp that the relative health value of possible downstream outcomes of care following diagnostic tests (health utility) and the probabilities of each potential outcome vary for different patients. They are intuitively aware that these utilities and probabilities can be combined conceptually for sensible decision making about whether or not to test (expected utility analysis for the diagnostic test in that patient). In practice, however, it is not clear whether these decision-analytic concepts are fully understood or applied by practicing clinicians. For example, errors overestimating the capacity of CT to ‘rule out’ stroke (sensitivity only ~16% in the first 24 h after stroke onset) drive overuse in dizziness.

While the real-world behaviour of clinicians may sometimes reflect so-called ‘irrational’ psychology in decision making, the complexity of such diagnostic decisions should not be underestimated. Scientific evidence about diagnostic tests rarely goes beyond the diagnostic accuracy level or immediate effect on diagnostic reasoning or therapeutic decisions, even though patient-centred outcomes would allow more direct inferences about overall test utility. There is often combined uncertainty in the estimates of disease probability, test sensitivity and specificity, efficacy of treatment options, and probability and health benefit of the outcomes. This degree of uncertainty usually makes a thorough, quantitative determination of the optimal decision-analytic choice (to test or not to test) during the patient encounter problematic. Future interventions to optimise test
choices for frequently occurring important decisions may need to be mostly preprepared (eg, practice guidelines or computer-based tools; in the dizziness case, perhaps prepackaged decision support using device-based physiologic diagnosis).
Economic modelling and related analytic techniques (cost–benefit, cost–effectiveness, estimated value of information) offer a robust alternative to assess the societal value of medical diagnostic testing, although some special considerations are required. The economic valuation of therapy is relatively straightforward— if a treatment improves health outcomes, its added value can be weighed against its costs in dollars and adverse effects. The value proposition for diagnosis is usually less transparent— diagnosis is more remote to the desired outcome (ie, ‘better health’ not ‘better diagnosis’, per se77) and, consequently, the link between improved diagnosis of a condition and improved health is more uncertain (eg, overdiagnosis of cancer64). For practical reasons, scientific evidence backing the use of diagnostic tests is also usually indirect, requiring a two-step inference that generally assumes, given a correct diagnosis, that the application of correct treatment will result in better outcomes.77 Furthermore, there may be benefits to ‘knowing’ about a diagnosis even if there are no immediate treatment implications89; and there may also be harms (as with a progressive, untreatable disorder such as Huntington’s disease).90 These attributes lead to greater complexity in analytically assessing the value of ordering a diagnostic test, even if familiar qualitative estimates (eg, the ‘chagrin’ of making the wrong choice) are substituted for less familiar quantitative calculations of net benefit.44 Conceptual complexities notwithstanding, we believe that as long as the psychological value (benefit or harm) of uncertain, correct and incorrect diagnoses is considered for its impact on health-related quality of life,91 then the standard measure of health effect used in economic analyses of medical treatments, the ‘quality-adjusted life year’ (QALY), is an appropriate measure of diagnostic test outcomes. Similarly, the cost per QALY or incremental cost-effectiveness ratio (ICER) is an appropriate measure of societal value in diagnosis. Within this framework, individual diagnostic tests or overall diagnostic strategies which lead to health benefits at a cost of <US$100 000 per QALY (or any societally sanctioned alternative threshold)92–94 would be considered cost effective. In a comparative effectiveness (or relative cost effectiveness) framework, diagnostic interventions offering the greatest number of QALYs per healthcare dollar spent should be endorsed at a societal level. We propose that high-value targets are those with a high burden of harm from misdiagnosis and a low cost of reducing misdiagnosis, while those with the opposite profile are of low value (figure 2).

**Figure 1** Tradeoffs versus improvements in diagnostic performance as illustrated by the ROC curve. (A) Performance tradeoffs (sliding along the ROC curve). In this scenario, clinicians alter their thresholds for seeking a particular diagnosis, but do not improve their overall diagnostic performance. Incentives, such as fear of malpractice litigation, drive physicians toward being ‘Nervous Nellies’, while incentives, such as productivity pressures, drive them towards being ‘Crazy Cowboys.’ (B) Performance improvements (moving the ROC curve). In this scenario, clinicians alter their diagnostic performance with regard to a particular diagnosis, rather than merely altering their threshold for decision making. Only incentives that focus on cost-effectiveness or ‘value’ (ie, combine ‘quality’ and ‘productivity’ measures) can drive clinicians to become ‘Deft Diagnosticians.’ Dx, diagnosis; ROC, receiver operating characteristic.
overuse and underuse of tests, particularly neuroimaging. The situation suggests a compelling opportunity for resource realignment to reclaim value in diagnosis. Yet, a more nuanced economic model results in a different approach to improving diagnostic safety in dizziness than might initially be expected.

Dizziness patients at high risk for stroke would be a reasonable group to target given advances in understanding bedside stroke diagnosis and the life-threatening, time-dependent nature of the disease. The approximately 15% subgroup of patients with acute, continuous, dizziness or vertigo can be readily identified at the bedside and have about a 25% risk of stroke. Modelling different diagnosis options reveals that better bedside diagnosis (with device-based eye movement interpretation using video-oculography (VOG) equipment) could produce cost-effective quality improvements (<US$8000 per QALY), but would not be directly cost saving relative to current practice (figure 3). Diagnostic testing is less expensive in the new care approach (neuroimaging would be reduced from ~45% to <15%), but caring for stroke patients costs money in order to save lives—so correctly identifying more strokes costs more than current practice. In the cost-effectiveness framework, the shift in diagnostic strategy is not cost ‘saving’ since the net economic benefit to society of these saved or improved lives (estimated at about US$600 million per year in the USA) is not incorporated. Although perhaps counterintuitive, increased diagnostic safety will only rarely produce direct healthcare cost savings when the condition being diagnosed is expensive to treat, even if costs of diagnostic testing decline.

If one also considers, however, the impact of improved diagnosis on patients with benign vestibular disorders, a different cost–benefit picture emerges. This population might initially seem less important to target, since lives are not usually at stake when patients with self-limited inner ear conditions are missed. Nevertheless, these ‘benign’ conditions do reduce quality of life for patients, and treatments reclaim these losses. More importantly, from an economic perspective, the societal costs of unnecessary diagnostic tests or admission for ‘stroke workups’ in these patients are enormous. With appropriate reductions in CT, modest increases in MRI, and slight decreases in overall admissions, we have estimated that total healthcare savings for ED dizziness amount to more than US$1 billion annually in the USA alone (additional material online table 1).

This finding suggests that an intervention targeting the broader population of patients at risk for both stroke and vestibular disorders would save lives and money. Thus, our economic analysis points to a different population (all acute dizziness or vertigo) than might initially have been targeted based on common-sense approaches (subset at high risk). Before pursuing a research study to prove and disseminate these techniques, it would be possible to use estimated value of perfect information analyses to measure the expected value to society of any future research efforts to prove and disseminate the new approach.

CONCLUSIONS: SOCIETAL VALUE PRIORITISATION OF DIAGNOSTIC SAFETY AND QUALITY EFFORTS

Using the case example of acute dizziness, we have illustrated the potential benefit of economic analysis for guiding quality improvement approaches targeting...
reduced diagnostic error. From a societal value perspective, the most sensible approach to improving diagnostic safety and quality might be to identify the diagnostic failures (misdiagnoses and inappropriate test use) with the greatest total economic burden for society and target these first for quality assurance or improvement initiatives. While doing this across the spectrum of all conceivable problems and conditions seems daunting, conducting economic analyses for a finite set of important problems (eg, 10 most common presenting symptoms in primary and emergency care; 10 leading causes of morbidity and mortality) could help prioritise high-yield targets.

Economic analyses will only inform diagnostic safety and quality if we can define the necessary parameters to construct the analyses. As diagnostic techniques evolve, it will be critical to study not only diagnostic test properties, but the impact of different diagnostic strategies on health outcomes. Future research should seek to explicitly measure the rates of diagnostic error for common symptoms and important diseases, as well as misdiagnosis-related harms and associated costs. Standards for applying estimated value of perfect information analyses and other advanced techniques to diagnostic problems should be developed to help guide funders in determining the potential societal value of solving a particular diagnostic problem. Stakeholders, including research funding agencies, should make economic analyses priority topics for scientific inquiry related to diagnosis and diagnostic errors.

**Collaborators** Newman-Toker DE, Butchy GT, Lehmann HP, Aldrich EM, Channugam A, Frick KD.

**Contributors** DEN-T had full access to all the information in the study and takes responsibility for the integrity of the information and the accuracy of the content. DEN-T: conceived manuscript concept; drafted manuscript; reviewed and critically edited the manuscript; approved the final version. KM: helped conceive manuscript concept; reviewed and critically edited the manuscript; approved the final version. DM: helped conceive manuscript concept; reviewed and critically edited the manuscript; approved the final version.

**Funding** Agency for Healthcare Research and Quality (Grant #R13HS019252).

**Competing interests** None.

**Provenance and peer review** Commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

**REFERENCES**


