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,4 with advanced diagnostic imaging leading the way and diagnostic laboratory testing a close second.5 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 any 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.12

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.19 For acute dizziness in the emergency department (ED), the number of patients seeking care has more than doubled over the past two decades.13 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.’20 For example, the fraction of ED patients with dizziness undergoing CT scans rose steadily from 9% in 1995 to over 40% currently, 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 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

Care.52 Risk aversion may be a particularly important incentive (box 3), with mixed results for optimal patient care driven by a mix of incentives and disincentives (peer review, local standards for test-ordering (peer practices),32 patient preferences and fear of litigation).33 Non-traditional symptoms, with atypical and non-traditional symptoms vs 64% of strokes presenting with traditional symptoms,63 result in diagnostic error (eg, false positives) or underuse.64 When an appropriate diagnostic test has been ordered but 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,55 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,22 despite the fact that relatively simple, non-invasive, bedside physical examination tests have been shown repeatedly to identify more than 99% of strokes.25 27 56

Estimates of diagnostic error-associated adverse events are 0.1% of primary care visits57 and 0.4% of hospital admissions,58 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%.9 Rates appear to vary by specialty (eg, 5% in radiology vs 12% in emergency medicine),59 by disease (eg, 2% of myocardial infarctions vs 9% of strokes),60 and especially by clinical presentation (eg, 4% of strokes presenting traditional symptoms vs 64% of strokes presenting non-traditional symptoms),61 with atypical and nonspecific presentations (such as dizziness) increasing the risk of misdiagnosis dramatically.60 62

Misdiagnoses, when they relate to test ordering (~15% of misdiagnoses55), generally result from underuse.63 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.64 Test overuse, however, can also result in diagnostic error (eg, false positives) or
Box 3 Incentives and disincentives that drive diagnostic test-ordering behaviour by physicians

Incentives

 principio oriented
  - Seeking best outcomes of care
  - Responding to patient and family requests
 principio oriented
  - Practice guidelines or care pathways/standing orders
  - Required diagnostic performance metrics
 principio cultural
  - Tradition
  - Small area practice variation
 principio efficiency-driven
  - Productivity targets or reduced visit length (since it is faster to ‘just order the test’ or order multiple tests ‘in parallel’ rather than ‘in series’)
  - Unavailability of prior test results (faster to repeat a test than track down results)
  - Purely financial
  - Greatest profit for individual physicians, hospitals, or health systems in a fee-for-service payment system
  - Risk aversive
  - Desire to avoid ‘chagrin/regret or uncertainty
  - Fear of medicolegal liability—‘defensive medicine’

Disincentives

 principio oriented
  - Medical futility
  - Risks of complications from testing
 principio oriented
  - Preapproval requirements
  - Practice guidelines
 principio cultural
  - Peer pressure
  - Physician responsibility for healthcare stewardship
  - Public awareness campaigns (eg, ‘Choosing wisely’)
 principio efficiency-driven
  - Unavailable locally
  - Difficult to order
  - Long delay for result
  - Purely financial
  - Greatest profit for individual physicians, hospitals, or health systems in a bundled or capitated payment system
  - Cost-containing
  - Payer policies

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 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).

EXPANDED NOTION OF A ‘TEST’: BEDSIDE EXAMS, CLINICAL PATHWAYS, AND DOCTORS AS ‘TESTS’

A medical ‘diagnostic test’ often connotes paraclinical tests such as blood chemistries or imaging procedures. However, every piece of information acquired during the diagnostic process can be considered a diagnostic test. Each element of history taking and physical examination is a separate diagnostic test with measurable test properties, such as sensitivity, specificity and reproducibility. For continuous (eg, duration of symptoms) or ordinal (eg, severity of cardiac murmur) outcomes, the specific threshold for considering the result abnormal is somewhat arbitrary, and tradeoffs between sensitivity and specificity can be represented by a receiver operating characteristic (ROC) curve. Also, like laboratory and radiographic tests, bedside ‘tests’ have an associated cost in physician and patient time and are integral to clinical billing and reimbursement schemas.

When combined together in bundles or sequences, bedside tests are sometimes called clinical decision rules, or clinical algorithms. These multicomponent assessments, if routinely used together, can be treated mathematically as individual tests. For diagnosis of stroke in acute dizziness, a battery of three non-invasive tests of eye movement function known as ‘HINTS’ has been shown to be 99% sensitive and 97% specific. This bedside decision rule substantially improves accuracy over the current best available diagnostic alternative (early MRI) in the first 72 h after symptom onset (early MRI sensitivity ~86%), with comparison to delayed, confirmatory MRI.

Similarly, the routine diagnostic practices of individual physicians or groups of physicians can be assessed for their accuracy in terms of sensitivity, specificity and costs in terms of overall diagnostic bedside assessment and test usage behaviour. Representing these clinical performance characteristics in the form of ROC curves makes it clear that the goal of efforts to improve diagnostic accuracy should be to move physician ROC curves towards the upper left corner (ie, maximise the area under the curve, creating ‘Deft Diagnosticians’) rather than to force physicians to slide upwards along the curve by sacrificing specificity in favour of sensitivity (expensive ‘Nervous Nellies’), or the reverse (dangerous ‘Crazy Cowboys’) (figure 1). Improving diagnostic performance might be achieved through improved diagnostic education combined with low-technology (eg, test indication curves) or high-technology (eg, computerised diagnostic decision support) tools. For example, ED physicians may soon use a novel ‘eye ECG’ approach to diagnose stroke in dizziness.

WHAT IS DIAGNOSTIC QUALITY? APPROPRIATE USE AND VALUE IN DIAGNOSIS

High-quality diagnosis is accurate, timely, impactful, patient-centred, ethical and efficient. The importance of accuracy and timeliness to high-quality diagnosis is self-evident. Impact is crucial since ‘diagnosis sake,’ per se, offers no direct health benefit. Top-quality diagnosis considers individual patient preferences in making difficult decisions about the risks and rewards of resolving diagnostic uncertainty (ie, ‘shared decision making’). The ethics of care for specific patients and professional responsibility for stewardship of finite societal healthcare resources should also help shape diagnostic decisions. Education and public awareness campaigns may play an important role in conveying these core values.

Efficiency includes parsimony (conceptual efficiency), speed (temporal efficiency), and cost-effectiveness (financial efficiency). Parsimony (eg, taking a ‘least moves’ strategy to arriving at a correct diagnosis, or choosing not to pursue rare untreatable diagnoses) and speed are attributes we naturally associate with good diagnosticians and good diagnostic process, but it may initially be less intuitive why costs and quality are inseparable for diagnosis. It is our view that profile-gate, non-parsimonious, inefficient diagnosis, even if accurate, cannot be considered high-quality diagnosis. A physician who orders every imaginable test for every patient with a given symptom would not be considered a high-quality diagnostican—this would indicate an inability to properly judge pretest probabilities at the bedside. By contrast, for the parsimonious physician evaluating a patient with acute dizziness or vertigo, this means the ability to rapidly assess the risk of stroke at the bedside, rather than referring all patients for neuroimaging.

GETTING ‘BANG FOR OUR BUCK’ IN DIAGNOSIS: THE ROLE OF ECONOMIC ANALYSIS

Generally missing from categorical ‘appropriateness’ definitions are explicit considerations of the value (in health or dollars) of making a correct diagnosis, the incremental benefit of one diagnostic strategy over another, and the societal opportunity cost of recommending a particular diagnostic test be performed when total healthcare resources are finite. Also usually missing from these ‘go’-‘no go’ assessments of average appropriateness are individual patient disease probabilities or personal preferences (utilities) for specific health outcomes, the psychological impact (positive or negative) of the search for a diagnosis (vs watchful waiting), or the knowledge of the diagnosis itself, and explicit estimation of the effects of uncertainty or risk of bias in the evidence base underlying the overall recommendation.
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 se\textsuperscript{77}) and, consequently, the link between improved diagnosis of a condition and improved health is more uncertain (eg, overdiagnosis of cancer\textsuperscript{64}). 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.\textsuperscript{77} Furthermore, there may be benefits to ‘knowing’ about a diagnosis even if there are no immediate treatment implications\textsuperscript{89}; and there may also be harms (as with a progressive, untreatable disorder such as Huntington’s disease).\textsuperscript{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.\textsuperscript{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,\textsuperscript{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)\textsuperscript{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).

**HOW ECONOMIC ANALYSIS MIGHT HELP IMPROVE DIAGNOSTIC SAFETY: OUR CASE STUDY REVISITED**

In the case of acute dizziness\textsuperscript{95} (box 1), where diagnostic accuracy is low, leading to incorrect management and patient harm, it seems intuitive that fixing the quality problem should make economic sense. If new bedside techniques of such high sensitivity and specificity are available, one might just assume that dissemination strategies will improve quality and reduce costs (ie, be cost beneficial). Such an assumption would seem justified since there is strong evidence of unwanted practice variation\textsuperscript{32} including both...
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 stroke 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.

**Table 1** Current and projected ED and hospital resource utilisation with routine VOG use

<table>
<thead>
<tr>
<th>Resource usage (imaging, admission)</th>
<th>Current (2013 US national†*)</th>
<th>Projection with ED VOG use</th>
</tr>
</thead>
<tbody>
<tr>
<td>All ED dizziness CT rate (%)</td>
<td>41.2</td>
<td>10.3</td>
</tr>
<tr>
<td>All ED dizziness MRI rate (%)</td>
<td>2.4</td>
<td>3.0</td>
</tr>
<tr>
<td>All ED dizziness admission rate (%)</td>
<td>18.8</td>
<td>17.2</td>
</tr>
<tr>
<td>Total ED/hospital workup costs (US$)</td>
<td>9 242 624 941</td>
<td>8 198 729 820</td>
</tr>
<tr>
<td>Projected annual US healthcare saving (US$)</td>
<td>--</td>
<td>1 043 895 121</td>
</tr>
</tbody>
</table>

Newman-Toker, unpublished. See online supplement for parameters and sources. These numbers are purely resource-use based, and do not consider the societal and personal benefits from lives saved due to early stroke interventions, or reduced morbidity from improved treatment of inner ear disorders.

**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/](http://creativecommons.org/licenses/by-nc/3.0/)

**REFERENCES**


73 Jaeschke R, Guyatt GH, Sackett DL. Users’ guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994;271:703–7.
90 Newman-Toker DE, Hsieh YH, Camargo CA Jr, et al. Spectrum of dizziness visits to US emergency departments:


