

## The laboratory test utilization management toolbox

Geoffrey Baird

Department of Laboratory Medicine, University of Washington, Seattle, WA USA

\*Corresponding author: gbaird@u.washington.edu

### Abstract

Efficiently managing laboratory test utilization requires both ensuring adequate utilization of needed tests in some patients and discouraging superfluous tests in other patients. After the difficult clinical decision is made to define the patients that do and do not need a test, a wealth of interventions are available to the clinician and laboratorian to help guide appropriate utilization. These interventions are collectively referred to here as the utilization management toolbox. Experience has shown that some tools in the toolbox are weak and other are strong, and that tools are most effective when many are used simultaneously. While the outcomes of utilization management studies are not always as concrete as may be desired, what data is available in the literature indicate that strong utilization management interventions are safe and effective measures to improve patient health and reduce waste in an era of increasing financial pressure.

**Key words:** clinical laboratory services; utilization; utilization review

Received: December 12, 2013

Accepted: March 27, 2014

### Introduction

“(The) First Rule in government spending: Why build one when you can have two at twice the price?” - S.R. Hadden in *Contact*, 1997 (written by Carl Sagan).

S.R. Hadden, the wealthy industrialist in the Carl Sagan drama *Contact*, may have been poking fun at stereotypical government spending practices when he quipped the line above, but his statement was more profound than it sounds, and it is especially relevant to clinical laboratory test utilization. The health care payment systems operating today in many countries remove the clinicians who make utilization decisions from the financial consequences of their actions, with the result that some patients undergo at least twice the testing at twice the price, if not much more.

That test overutilization exists is no longer an item of debate, and it is clearly an international problem

(1). The reasons for overutilization are myriad, and may include as enabling factors even the technology we use to perform laboratory testing. There was once a time when it would be unthinkable and impractical for a hospital inpatient to get multiple laboratory tests *per day*, every day (2), but in the present day, high frequency testing is almost the rule in resource-rich settings, rather than the exception. Instrument vendors now advertise turn-around times in minutes rather than hours or days for clinical assays, automated platforms handle hundreds to thousands of samples *per hour*, and data systems distribute the results of these tests electronically, directly to caregiver’s electronic inboxes. In the outpatient care setting, testing may not occur daily for each patient, but the barrier to getting testing can be incredibly low; the standard routine “physical” that a healthy adult might receive at a doctor visit can be accompanied by a

panel of tests, perhaps only a blood count and electrolytes, but possibly also any one of a myriad of large so-called “wellness” panels offered by commercial laboratories that have financial incentives to drive frequent, high-cost testing. It does not simplify matters that additional technological advances, especially the genomic revolution, have exponentially increased the number of possible tests that are available. No one clinician can possibly know which novel test is appropriate in any given situation, and the problem compounds itself daily as additional genetic tests, small molecule mass spectrometry tests, or proteomic multivariate index assays are added to the global test menu. Absent any guidance, it appears that clinicians are apt to order whatever seems most familiar, any or all tests that sound like they might be appropriate, or whatever has been touted most heavily by ambitious sales representatives. Another problem unique to academic settings is that one must allow for the fact that physicians in training may need to order more than an optimal number of tests in order to learn how to use the results of this testing in the practice of medicine.

So, what is a laboratory director to do in the face of this growing adversity? How do we ensure that that patient’s interests come first, that neither inpatients nor outpatients get the wrong tests at all or the right tests too frequently, and that caregivers are able to select tests appropriately so that they can most easily interpret the results? Laboratory test utilization researchers, including this author, have often summed up the literature in lectures or in private by stating that, “30% of laboratory testing is likely wasteful,” and now this estimate has been supported by a thorough meta-analysis (3). However, it is has never been clear exactly how laboratory directors or clinicians should go about identifying those 30% of tests, especially given that the wasted testing is not evenly distributed across all patients or specific tests. What is needed, therefore, is a standardized approach for identifying malutilisation in daily clinical practice, and once it has been identified, a common set of tools should be available to fix the problem. This is the approach taken in this review.

The problem of identifying inappropriate laboratory test utilization is outlined below in the form of three “Rules,” all of which are intended to provide the general rationale and goals of a laboratory test utilization management program. Afterwards, the laboratory test utilization management toolbox, comprised of a collection of evidence-based tools available to the laboratorian and clinician alike, will be described, with emphasis on those tools that are most appropriate for specific types of inappropriate utilization.

## The three rules of laboratory test utilization

### Rule 1: “If you ask a stupid question, you get a stupid answer”

Rule 1, otherwise known to the statistician as Bayes’ Theorem, posits that the post-test probability of something being true is a product of the prior probability that the thing is true and the likelihood ratio provided by a test (4). In mathematical terms for laboratory testing, Bayes’ Theorem can be expressed as:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

This can be translated to the context of laboratory medicine to mean that the conditional probability of disease (A) given the result of a test (B) (also called the “positive predictive value”) is equal to the conditional probability of the result (B) given the disease (A) being present, multiplied by the probability of the disease (A) in the population (also called the “prevalence”) and divided by the probability of the test giving the result (B) in all members of the population.

One consequence of Rule 1 in laboratory testing is that ordering a test to rule in a condition when the prior probability is very low (i.e. a urine hCG test to assess pregnancy in an apparently male patient) is unlikely to be a fruitful endeavor, since negative results were already expected and positive results are most likely false positives. The same applies in the converse; one should not generally order a test

when it is extremely likely that the diagnosis in question is present. Rule 1 exists not just to discourage inappropriate ordering, but for the equally important reason of saving one from the onus of interpreting highly unlikely test results. The true power of testing in the setting of low pre-test probability is ruling OUT a diagnosis; negative results in these settings can be trusted, but positive results will always be confusing.

Rule 1 is not always easy to apply, however, for three reasons. One, prior probabilities of disease close to 0 or 1 or quite unusual, as clinical, radiologic, and historical findings do not always have enough evidentiary value to allow one to cease a workup without laboratory testing. The prior probability of disease could also be completely unknown, in which case the Rule helps not at all. Two, many physicians are not always able to estimate prior probabilities with enough savvy to allow an accurate application of the Rule. Clinical decision support and education can partly address this issue, but patients are all unique, so there can be no absolute guide to prior probability assessment for all situations. Three, human beings are curious, especially those who go into medicine. Curiosity manifests itself in the medical setting by exploration of unlikely possibilities, and while this may be an important educational activity, it should not be the basis for sound medical policy.

The three challenges of Rule 1 notwithstanding, the Rule can still be followed effectively by 1) using Bayesian thinking to assign at least relative, rather than absolute, likelihoods to various diagnoses to allow one to order sequential testing from most to least beneficial; 2) ceasing repetitive daily testing, or at least restricting testing to reflect our understanding of intraindividual biological variation, because asking "too many questions" has the same effect as asking a single "stupid question"; and 3) always developing a plan for test interpretation PRIOR to ordering a test. This third point cannot be stressed enough, especially during medical training, and even more critically in dealing with the most curious among us. Clinicians often construct large differential diagnoses in medicine to avoid missing the occasional rare presentation or to avoid anchoring bias, or perhaps simply

to impress or intimidate their peers (5), but we should emphasize to our trainees that testing for many things simultaneously just because it's possible or because "...I saw a patient like that once", is no reason to embark upon an unnecessarily costly diagnostic odyssey. Tests ordered out of pure curiosity have a funny way of presenting later as unexplainable positive results, and the cost of working up a false positive result is always larger than not ordering the test in the first place.

### **Rule 2: "Laboratory testing is for sick people"**

If Rule 2 sounds like a restatement of Rule 1, that is because it is. Indeed, all of laboratory test utilization management stems from Bayes' theorem, although different restatements of the theorem provide windows into distinct aspects of the utilization management problem. What Rule 2 focuses on is the well patient, and specifically, the notion of "Wellness Testing" (6,7). Wellness Testing commonly refers to single tests or panels of tests that are intended to be performed on well patients, usually self-selecting adults, with the goal of preventing unwanted later complications of disease. Put in this way, Wellness Testing is in fact a misnomer, in that the purpose of it is to discover that a patient is, or will become, unwell. In Bayesian terms, Wellness Testing poses a significant risk of diagnostic failure, or at least confusion, in the setting of low prior probabilities. In financial terms, however, Wellness Testing creates a highly attractive market for laboratory vendors that earn money on a fee for service basis.

Like all rules, however, Rule 2 has exceptions. Lipid testing, diabetes screening and newborn screening, for example, are reasonable uses of laboratory testing in ostensibly healthy people, or at least people with prior probabilities of disease that are equivalent to the general population risk. Like many exceptions, these exceptions prove the rule. For lipid testing, one might argue that the standard "healthy" patient in a population, for example a middle-aged male who is slightly overweight and exercises a tad too little, could reasonably be considered "sick" in terms of the risk that lipid-related disorders pose to his long-term health. More importantly, however, the reason why lipid testing

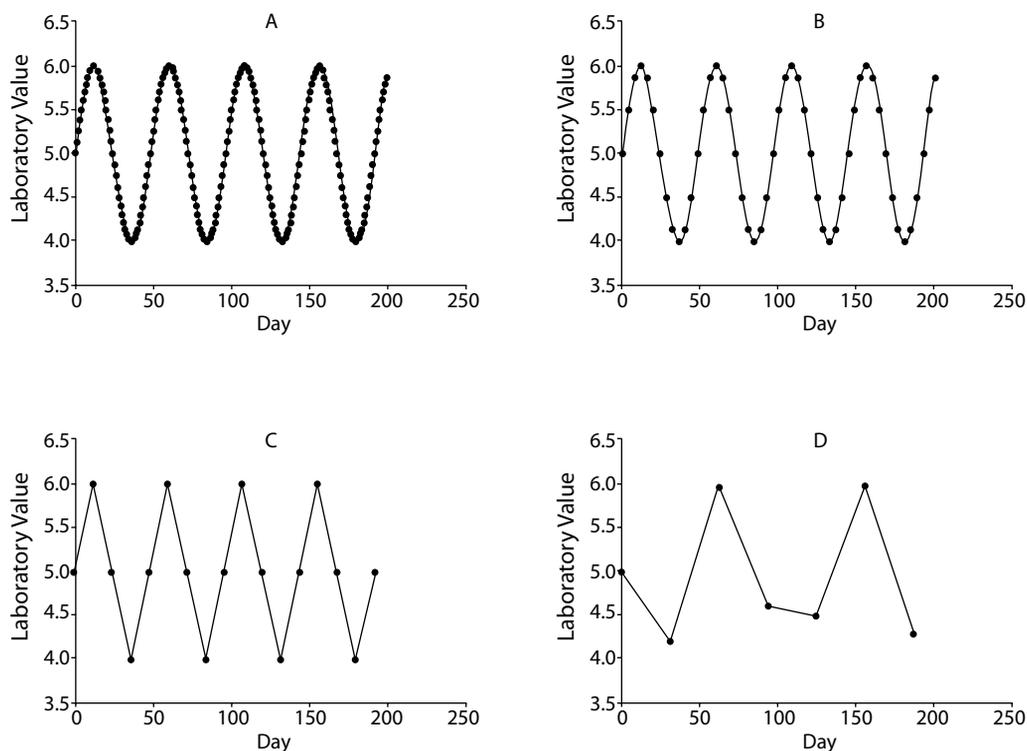
and newborn screening are exceptions to Rule 2 is that there exists evidence demonstrating a benefit of the testing, when coupled with appropriate downstream therapeutic interventions. In a lipid panel, we test for analytes (cholesterol, lipoproteins, triglycerides) that we know to be causally related to cardiovascular disease and that we know respond to therapy, and we understand much of the risk-benefit relationship of either measuring or not measuring lipids (8). Likewise, despite the fact that inborn errors of metabolism are exceedingly rare, and the fact that the positive results from many newborn screening tests are false positives, the testing allows us an opportunity to confirm true positives and initiate therapy to avert diseases with extremely high morbidity and mortality that would otherwise be devastatingly difficult and expensive to treat later in life. Numerous studies now indicate that newborn screening is cost-effective, as well (9-13). The WHO criteria for mass screening are a helpful resource for assessing the utility of any screening proposal (14).

The limits to Rule 2 have exceptions. Lipid testing is helpful when limited to those analytes that have been studied in large cohorts like the Framingham study (15). Newer expanded lipid panels, such as those including genetic testing (16,17), additional information about the size or chromatographic mobility of lipoprotein particles (18,19), or additional biochemical or proteomic biomarkers, run afoul of both Rule 1 and Rule 2. In some cases, these tests rely on evidence that is preliminary, limited, proprietary or otherwise insufficient to support widespread utilization, and in other cases these tests may in fact eventually find a place in appropriate lipid panels, but are currently waiting for evidence that shows us where to apply them most effectively. In the case of newborn screening, the calculus as to whether to include a test in a panel is actually fairly simple, as a test must be paired with a treatment that prevents either a disease or the sequel of that disease. If there is nothing to do with the result of a newborn screening test, then it does not make sense to do the test. That this truism extends to all laboratory testing should be obvious.

### **Rule 3: “Too many good tests are the same as one bad test”**

Although this idea has been hinted at in Rules 1 and 2, a separate rule should be reserved for the idea that repetitive tests or large panels of tests, even when comprised by individually reasonable tests, are a form of malutilisation. At the very least, too-frequent repetitive testing can come at odds with the Nyquist-Shannon theorem, a fundamental principle of information theory (20). The mathematical derivation of the Nyquist-Shannon theorem is beyond this discussion, but in words it can be said to define the appropriate relationship between how often one should sample a varying signal. In Figure 1, 4 potential situations are depicted, in which the same oscillating laboratory value (the “signal”) is measured at successively longer intervals (the “sampling”). Clearly, plots A and B indicate oversampling, where far more laboratory tests were obtained relative to what was needed to define the underlying signal, and plot D shows significant undersampling, leading to a misleading impression that the value is oscillating slower than it is in reality. Plot C, on the other hand, has precisely as many sampling points as is necessary to describe accurately the underlying signal. In clinical terms, the Nyquist-Shannon theorem tells us that repeated testing to assess a change in a laboratory parameter should occur at intervals on par with the expected time it would take the analyte value to change.

One obvious clinical manifestation of Nyquist-Shannon theorem in clinical laboratory is in Haemoglobin A1c (HbA1c) testing. Because the longevity of red blood cells allows haemoglobin molecules to remain in circulation for several months, it makes no sense to assess HbA1c on a scale of days to weeks, but rather it should be measured on a scale of 1-3 months to assess changes. Despite the inherent logic of this conclusion, there is still substantial variation amongst physicians in ordering HbA1c testing at rational intervals (21). It is also important to understand that undersampling is a significant risk for HbA1c testing, as too-infrequent tests (i.e. a year or years between determinations in a poorly-controlled diabetic) can cause a physician to miss clinically significant variation.



**FIGURE 1.** A hypothetical oscillating laboratory value (signal) with a period of approximately 48 days is measured once every 1 (A), 4 (B), 12 (C) and 31 (D) days.

A second problem with too-frequent testing derives from the statistics of repetitive applications of a single test. While the sensitivity for disease detection increases when testing repetitively, the specificity must concomitantly decrease. It is often pointed out, for example, that the chances of having one abnormal value in a panel or series of 14 tests whose reference ranges are defined by the central 95% percentile of the health population approaches 50%, even when the patient in question is totally normal (7). This counterintuitive result is of course not entirely accurate, as it assumes that all laboratory tests vary independently (they do not), but it drives home the point that the more testing is done, the more disease is discovered. This is not always a bad thing, though, and repetitive testing placed at appropriate intervals (i.e. for lipid testing) can be highly beneficial to populations. However, a disreputable laboratory vendor could also exploit this phenomenon to generate revenue from a very large panel of putative biomarker assays guaranteed to generate false posi-

tives that require follow-up testing. The phenomenon creeps even into highly complex and expensive genetic testing panels performed at reputable laboratories. For example, there are numerous genetic syndromes that are known to be each caused by a number of different possible mutations, and reference laboratories will often offer a panel of tests that probes many or all possible mutations, causative or otherwise. In such cases, it is often the case that one specific mutation is causative in the majority of cases, such that testing for a single mutation first with a cheaper screening methodology may obviate the need for the expensive full panel in a majority of patients, lowering the overall cost of testing (22). Reflexive testing algorithms that allow the providing laboratory to aid the clinician in defining the optimal sequential testing pathway are a key tool in the test utilization toolbox, and will be covered later.

Evidence that repetitive or daily testing is common abounds in the historical literature (23-31), yet fewer studies have shown how repetitive testing can

drive unnecessary and costly downstream activities. In a pair of large academic medical centres, for example, daily ionized calcium testing had become ingrained into the culture of the house staff and was ordered in nearly every patient. When daily ionized calcium testing was reduced by approximately 70% by introducing a reflexive testing strategy, doses of calcium administered by the pharmacy in the hospitals decreased by a very similar fraction, as did the diagnoses of "hypocalcemia" rendered by the providers (32). While not enumerated in the report, it is likely that the cost savings of these downstream effects far outstripped the laboratory-specific savings. No discernible difference in morbidity or mortality was discovered after this intervention, even amongst diagnoses known to be associated with significant hypocalcemia (tetany, seizures, myocardial infarction). In this case, a substantial burden of disease ("hypocalcemia") was being invented by daily ionized calcium testing practices, and it disappeared within a month as testing practices were altered, raising the question of how many other fictitious diagnoses and unnecessary treatments inpatients may receive when they undergo daily routine laboratory testing.

In another study that drew upon a large dataset comprising 4 million common outpatient tests performed in a large Canadian province (33), van Walraven *et al.* found that approximately 30% of testing for eight common analytes was repeated within a month. Potentially redundant testing, defined as testing repeated within defined time intervals, was estimated to have cost between \$13.9 and \$35.9 million (Canadian) annually. As this study only focused on 8 common tests with low unit costs, it is likely that the true cost of all redundant testing to large health care systems is probably much larger, and perhaps even orders of magnitude larger than what was observed in this study.

Guidelines have been recently published describing the "Minimum retesting interval", i.e. the minimum time before a test should be repeated(34). While this guideline is not comprehensive of all laboratory tests and all clinical situations, it provides a substantial, evidence-based list of recommendations that covers much of the current practice of medicine.

## Applying the rules

Applying the 3 rules of laboratory test utilization is a deceptively simple task, as identifying waste appears simple. However, the number of people who have decried the third of laboratory testing that is waste likely far outstrips the number of people who have ever successfully managed laboratory test utilization. Laboratory test malutilisation is thus like the weather, in that "...everyone talks about it, but no one does anything about it". How can this be changed?

First, in an institution desiring to curb inappropriate test utilization, there must be "buy-in" amongst the key stakeholders. There must be at least one member of the laboratory staff, a laboratory director, who is willing and able to expend the time needed to do the data analyses required to identify the problem areas. If this laboratory director is in academia, as often occurs, their efforts in managing utilization should be both pursued and rewarded as genuine scholarly activities, with all work resulting in publications that allow others to learn from their successes and failures. If an academic department values contributions to test utilization management below contributions to basic or applied sciences, then junior faculty with an interest in the field will simply not participate for fear of missing out on promotion, and the field will stagnate. Outside of laboratory staff, however, cooperative medical staff are needed in the remainder of the hospital, especially those who are willing and able to take the time and energy required to participate in utilization review and quality improvement activities. Second, as utilization management is a data-intensive activity, there must be adequate information technology resources available to the project leader. At a minimum, programmers and/or individuals savvy in building, managing and querying large databases are absolutely required to generate reports and analyses, or at the very least retrieve the raw data required to assess utilization. The ability to correlate laboratory test utilization data (orders) with test results and other patient data (metadata such as gender and age, or potentially diagnostic or outcome-related data) is key for assessing the opportunities and potential risks of utilization management interven-

tions, meaning that large relational healthcare information databases(35), although expensive, will be very valuable in future utilization studies.

### The laboratory test utilization management toolbox

The laboratory test utilization management toolbox is outlined in Table 1, with interventions that drive appropriate utilization sorted from strongest to weakest. For each tool, the intended target of the tool is listed, along with potential strengths and weaknesses, and finally a reference or example of how the tool can be used. The reference col-

umn is not intended to be an exhaustive list of all published examples of a particular tool being put to use, for there are too many to include in this table, but rather the cited example(s) are intended to demonstrate an important strength or weakness of the tool.

It is worth pointing out some tools that are not included in the toolbox because the available evidence suggests that they are not terribly effective. One of the common assumptions made about test utilization management is that clinicians would behave more rationally if they saw the cost of the laboratory tests they ordered. The results of interventions that actually did this are mixed (36-38),

**TABLE 1.** The laboratory test utilization management toolbox.

Strength	Tool	Target	Strengths	Weaknesses	Example/References
Strong	<b>Ban the test</b>	Obsolete tests, "Quack" testing, Legitimate tests used in inappropriate circumstances	This is the "Nuclear Option", as it ensures a complete cease to ordering	Only useful for tests with broad consensus as to lack of utility, which is unusual. Specific individuals may destroy consensus.	Bleeding time and other "Antiquated" tests (42).
Strong	<b>Laboratory test formulary</b>	All tests, especially those with utilization that is recognized, after analytics, to be above what is expected or justifiable.	A uniform policy across a system can be supported by a formulary, in the same way as a pharmacy formulary. Exceptions to formulary can be vetted by a committee or individual tasked with these decisions.	Requires authority and buy-in from multiple factions in a medical system, and likely participation by multiple specialties.	University of Michigan (43).
Strong	<b>Combined intervention</b>	Any test	By far the most effective, as the strengths of one intervention often complement the weaknesses of another.	Logistically complex, as many parties (the laboratory, clinicians, information services, payer systems, etc...) need to be involved.	Solomon meta-analysis (44), Massachusetts General Hospital Experience (45), hematology testing (46).
Strong	<b>Stop paying for unnecessary testing</b>	Any test	Similar to banning tests, this intervention is effective at nearly ceasing testing, depending on who decides to stop paying.	Depends on the payment system present in the medical system. Perceived as unfair, especially if the payer decides to stop paying for something without adequate justification. A physician may not know that a test will not be paid for, and the cost could be transferred to the patient.	Trends in reimbursement shown here (47), example of medical policy here (40).

**TABLE 1.** The laboratory test utilization management toolbox.

Strength	Tool	Target	Strengths	Weaknesses	Example/References
Strong	<b>Ban repetitive orders</b>	Daily inpatient tests	Powerful method of reducing automatic ordering that providers often do not even know is occurring.	Worry amongst some physicians that they might “miss something”. Actual risk of missing something clinically important if a clinically indicated repetitive test is disallowed (i.e. coagulation tests in patients on anticoagulants).	Make repeated orders difficult through computerized order entry (48), ban standing orders (49), limit tests to 24-hour period (50).
Strong	<b>Privilege ordering providers</b>	Complex single tests, high unit cost and/or difficult to interpret.	Limiting testing to physicians who know how to use a test increases the prior probability in the tested patients, increasing cost effectiveness and diagnostic yield.	Multiple physicians may want privileges, even in the absence of evidence that they deserve them.	Neurogenetic testing diagnostic yield ~30% for very rare diseases when expert providers order tests (51).
Strong	<b>Require high level approval</b>	Complex single tests, high unit cost and/or difficult to interpret.	Laboratory providers can have more insight into the utility of some tests than generalist providers.	Time consuming for laboratory staff or director, especially if there are no laboratory housestaff to take calls.	Large Genetics Sendout Testing Intervention (52).
Strong	<b>Change computerized order entry options</b>	Any test in a system with computerized ordering.	Computerized order changes can be made far more difficult to subvert than paper order form changes.	In the absence of a cultural change supporting modification of ordering practices, a complete stop to a specific order may increase provider abrasion. Unintended consequences can result if one is not careful in designing the intervention.	Reducing testing in coronary care unit (53). Change to routine testing menu (54).
Strong	<b>Offer reflexive testing.</b>	Any test where a cheaper screening test can be used before a more costly test.	Can work for computerized or paper ordering. Is a form of decision support that allows physician to follow correct testing algorithm with one order or click. Increases pre-test probability for more costly tests, making them more interpretable.	Requires an analyte for which a cheaper screening test exists. If using paper forms, one must realize that paper forms have a significant half-life in medical systems, and forms usually allow providers to “write in” tests that they cannot find on the form, thus allowing clinicians to subvert the intent of the reflexive panel.	Reflexive ionized calcium (32), coagulation panels (55).
Moderate/Strong	<b>Utilization report cards</b>	Routine outpatient panel testing, daily testing on inpatients.	Provides data on ordering to providers who may otherwise have no idea how they order tests, and thus may allow them to make informed decisions. Can be paired with reimbursement/financial penalties, or associated with peer feedback for added strength.	No one has to read the report card, especially if it is not associated with an incentive.	Outpatient report cards (56), intermittent feedback for physicians (57,58).

**TABLE 1.** The laboratory test utilization management toolbox.

Strength	Tool	Target	Strengths	Weaknesses	Example/References
Moderate	<b>Computerized reminders/decision support</b>	Selected tests with moderate volume and high likelihood of being misordered.	Can provide support in real time to physicians to increase prior probabilities.	“Pop-up fatigue” occurs if too many reminders are implemented, leading to provider abrasion. Providers will also cease to continue to read pop-ups after some time.	Magnesium intervention (41), 1,25 dihydroxy Vitamin D email reminder (59) [cited example also uses privileging].
Weak	<b>Post guidelines on paper order forms</b>	Selected tests with moderate volume and high likelihood of being misordered, but no computerized ordering available.	Can provide support in real time to physicians to increase prior probabilities.	As opposed to pop-ups on computerized forms, written guidelines on a paper are likely easier to ignore.	Redesigning test requisitions and promulgation of factsheets (46,60).
Weak	<b>Education alone/call for enhanced vigilance</b>	Any test	Required as a component of nearly all successful utilization management efforts. Interventions lacking an educational component risk failure due to lack of buy-in from interested parties who do not understand the purpose of the change.	Almost never works alone, or when it does, the effect wears off over time or completely disappears if new staff takes over (i.e. in a teaching hospital).	Example showing effect wearing off after time (38), mixed effects of reminding physicians of test costs (36-38).

however, and do not support whether or not provision of cost information has any true effect on ordering. While it does make sense that some information about cost might influence physician ordering behavior, the potential for a successful intervention based on this information is confounded by the fact that the true costs of laboratory tests are actually quite small, compared to more expensive items like radiologic scans, and perhaps more frustrating, it is often very difficult to determine what the actual cost of a laboratory test is in some health care delivery systems. In the United States, for example, specific laboratory tests are not usually reimbursed by third party payers for inpatient stays (adults with rare diseases and some children are notable exceptions), and the list prices of tests are often wildly inflated and/or kept secret. Therefore, while it is isn’t even clear what cost

should be shown to ordering physicians, the truth of the matter is that the most overutilised routine laboratory tests do not have high unit costs, as the variable cost for additional reagents needed to run a test on a large automated platform are miniscule compared to the difficult-to-assess fixed costs of maintaining that instrument and staffing required to run it. What we might want to show physicians at the time of order would be the potential downstream costs associated with following up unnecessary testing or the potential harm that could arise to their patients, but those costs and harms are difficult or impossible to tabulate.

Another financially-based tool that has been proposed is to simply pay physicians to stop ordering laboratory testing. This is a difficult tool to administer or evaluate ethically, but it has been done using trainees. In this study (39), medical residents

were paid in “book money” on a scale commensurate with the size of their collective percentage reductions in laboratory and radiology test ordering. Notably, residents were not individually rewarded for their individual performance, but rather for the performance of a group. There were overall modest reductions in test ordering through this intervention, but they were comparable to the control group, and they rebounded to pre-intervention levels after the study ceased. Residents who participated in simple chart reviews to study their own utilization patterns in this study had more significant and longer-lasting reductions in test utilization.

Whether or not tools based on provision of financial information or cash bonuses can be considered effective is perhaps not as important as the fact that third party payers or health care systems are already implementing interventions based on financial incentives (mostly negative incentives), and they will continue to implement more such incentives. While early targets are likely to be areas where the motivations for overutilization are primarily financially driven, i.e. fee-for-service reimbursement of tests with negligible or no clinical utility, US insurers are already starting to write medical policies that incorporate test utilization principles (40).

### Using the toolbox

While banning tests entirely or codifying a laboratory test formulary are listed in the toolbox as the strongest interventions, it must be restated that the most successful interventions in laboratory test utilization management are those that combine various tools. One can hardly expect an intervention that bans a physician’s favourite test to succeed without an educational component aimed at explaining to that physician why the test is no longer available. The same goes for computerized order entry changes. While these changes seem the least disruptive, i.e. a single check box might disappear, the interconnectedness of health information systems is such that small changes to a laboratory’s ordering interface can have unexpected and profound downstream changes. A computerized order entry-based intervention targeted at re-

ducing serum magnesium testing was found to paradoxically increase magnesium testing, for example, by inadvertently encouraging physicians to order serum magnesium together with serum calcium and phosphorus (41). The paradoxical effect found in this study was observed through constant utilization monitoring, allowing the investigators to institute a new, successful intervention, highlighting the fact that monitoring is a necessary part of all utilization management tools. Without ongoing monitoring of utilization, it is impossible to assess the effectiveness of any utilization management intervention. As was indicated in Rule 3 above, test utilization should be monitored on a timescale similar to how fast it is expected to change; it is not enough to measure utilization at the end of a project or the end of the year.

A final critical factor required for using the utilization toolbox is the development of a rapport between the laboratory and the rest of the hospital. Clinicians who order tests are rightfully wary of mysterious and draconian utilization management interventions forced upon them by unknown entities. Making things worse, in this author’s experience, clinicians are often surprised to learn that there are doctoral-level scientists and clinicians employed by the laboratory who are concerned with optimal utilization of laboratory tests. As education is a key tool in the toolbox, we in the laboratory must therefore embrace our roles as educators to ensure that the ideas we have are understood and appreciated by clinicians and patients alike. When done appropriately, utilization management interventions can lead to improved morale not just amongst laboratory staff, but amongst all parties involved in the process. Without all three entities (laboratory, clinician, patient) having adequate information and interest in participation, optimal utilization of laboratory testing cannot be achieved.

### Potential conflict of interest

Dr. Geoffrey Baird is a paid consultant and head of the clinical advisory board (paid positions) to Avalon Healthcare Solutions, a company based in the US that provides laboratory test utilization management services to insurance companies in the US.

## References

1. Smellie WS. Demand management and test request rationalization. *Ann Clin Biochem* 2012;49:323-36. <http://dx.doi.org/10.1258/acb.2011.011149>.
2. Griner PF, Liptzin B. Use of the laboratory in a teaching hospital. Implications for patient care, education, and hospital costs. *Ann Intern Med* 1971;75:157-63. <http://dx.doi.org/10.7326/0003-4819-75-2-157>.
3. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: A 15-year meta-analysis. *PLoS One* 2013;8:e78962. <http://dx.doi.org/10.1371/journal.pone.0078962>.
4. Efron B. Mathematics. Bayes' theorem in the 21st century. *Science* 2013;340:1177-8. <http://dx.doi.org/10.1126/science.1236536>.
5. Brancati FL. The art of pimping. *JAMA* 1989;262:89-90. <http://dx.doi.org/10.1001/jama.1989.03430010101039>.
6. Witte DL. Wellness testing. Design and experience of an established program. *Clin Lab Med* 1993;13:481-90.
7. Jørgensen LG, Brandslund I, Hyltoft Petersen P. Should we maintain the 95 percent reference intervals in the era of wellness testing? A concept paper. *Clin Chem Lab Med* 2004;42:747-51. <http://dx.doi.org/10.1515/CCLM.2004.126>.
8. Grundy SM, Cleeman JI, Merz CN, Brewer HB, Clark LT, Hunninghake DB, et al. Implications of recent clinical trials for the national cholesterol education program adult treatment panel iii guidelines. *J Am Coll Cardiol* 2004;44:720-32. <http://dx.doi.org/10.1016/j.jacc.2004.07.001>.
9. Tiwana SK, Rascati KL, Park H. Cost-effectiveness of expanded newborn screening in texas. *Value Health* 2012;15:613-21. <http://dx.doi.org/10.1016/j.jval.2012.02.007>.
10. Carroll AE, Downs SM. Comprehensive cost-utility analysis of newborn screening strategies. *Pediatrics* 2006;117:S287-95.
11. Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, et al. Neonatal screening for inborn errors of metabolism: Cost, yield and outcome. *Health Technol Assess* 1997;1:i-iv, 1-202.
12. Simpson N, Anderson R, Sassi F, Pitman A, Lewis P, Tu K, Lannin H. The cost-effectiveness of neonatal screening for cystic fibrosis: An analysis of alternative scenarios using a decision model. *Cost Eff Resour Alloc* 2005;3:8. <http://dx.doi.org/10.1186/1478-7547-3-8>.
13. Venditti LN, Venditti CP, Berry GT, Kaplan PB, Kaye EM, Glick H, Stanley CA. Newborn screening by tandem mass spectrometry for medium-chain acyl-coa dehydrogenase deficiency: A cost-effectiveness analysis. *Pediatrics* 2003;112:1005-15. <http://dx.doi.org/10.1542/peds.112.5.1005>.
14. Wilson J, Jungner G. Principles and practice of mass screening for disease. *WHO Chronicle* 1968;22.
15. Kannel WB, Castelli WP, Gordon T. Cholesterol in the prediction of atherosclerotic disease. New perspectives based on the framingham study. *Ann Intern Med* 1979;90:85-91. <http://dx.doi.org/10.7326/0003-4819-90-1-85>.
16. Roberts R. A customized genetic approach to the number one killer: Coronary artery disease. *Curr Opin Cardiol* 2008;23:629-33. <http://dx.doi.org/10.1097/HCO.0b013e32830e6b4e>.
17. Li Y, Iakoubova OA, Shiffman D, Devlin JJ, Forrester JS, Supperko HR. Kif6 polymorphism as a predictor of risk of coronary events and of clinical event reduction by statin therapy. *Am J Cardiol* 2010;106:994-8. <http://dx.doi.org/10.1016/j.amjcard.2010.05.033>.
18. Krauss RM. Lipoprotein subfractions and cardiovascular disease risk. *Curr Opin Lipidol* 2010;21:305-11. <http://dx.doi.org/10.1097/MOL.0b013e32833b7756>.
19. Kulkarni KR. Cholesterol profile measurement by vertical auto profile method. *Clin Lab Med* 2006;26:787-802. <http://dx.doi.org/10.1016/j.cll.2006.07.004>.
20. Shannon C. Communication in the presence of noise. *Proceedings of the IRE* 1949;37:10-21. <http://dx.doi.org/10.1109/JRPROC.1949.232969>.
21. Lyon AW, Higgins T, Wesenberg JC, Tran DV, Cembrowski GS. Variation in the frequency of hemoglobin A1c (HbA1c) testing: Population studies used to assess compliance with clinical practice guidelines and use of hba1c to screen for diabetes. *J Diabetes Sci Technol* 2009;3:411-7. <http://dx.doi.org/10.1177/193229680900300302>.
22. Miller C. Making sense of genetic tests. *Clinical Laboratory News* 2012;38:15.
23. Bates DW, Boyle DL, Rittenberg E, Kuperman GJ, Ma'Luf N, Menkin V, et al. What proportion of common diagnostic tests appear redundant? *Am J Med* 1998;104:361-8. [http://dx.doi.org/10.1016/S0002-9343\(98\)00063-1](http://dx.doi.org/10.1016/S0002-9343(98)00063-1).
24. Branger PJ, Van Oers RJ, Van der Wouden JC, van der Lei J. Laboratory services utilization: A survey of repeat investigations in ambulatory care. *Neth J Med* 1995;47:208-13. [http://dx.doi.org/10.1016/0300-2977\(95\)00100-8](http://dx.doi.org/10.1016/0300-2977(95)00100-8).
25. Baigelman W, Bellin SJ, Cupples LA, Dombrowski D, Coldiron J. Overutilization of serum electrolyte determinations in critical care units. Savings may be more apparent than real but what is real is of increasing importance. *Intensive Care Med* 1985;11:304-8. <http://dx.doi.org/10.1007/BF00273541>.
26. Dixon RH, Laszlo J. Utilization of clinical chemistry services by medical house staff. An analysis. *Arch Intern Med* 1974;134:1064-7. <http://dx.doi.org/10.1001/archinte.1974.00320240098012>.
27. Sox HC. Repeated testing. An overview and analysis. *Int J Technol Assess Health Care* 1997;13:512-20. <http://dx.doi.org/10.1017/S0266462300009983>.
28. Bülow PM, Knudsen LM, Staehr P. [duplication in a medical department of tests previously performed in a primary sector]. *Ugeskr Laeger* 1992;154:2497-501.
29. Valenstein P, Leiken A, Lehmann C. Test-ordering by multiple physicians increases unnecessary laboratory examinations. *Arch Pathol Lab Med* 1988;112:238-41.
30. Rix DB, Stump G. Is there duplication of diagnostic test results? *Can Med Assoc J* 1975;112:237-8, 41, 44.
31. Kwok J, Jones B. Unnecessary repeat requesting of tests: An audit in a government hospital immunology laboratory. *J Clin Pathol* 2005;58:457-62. <http://dx.doi.org/10.1136/jcp.2004.021691>.

32. Baird G, Rainey P, Wener M, Chandler W. Reducing routine ionized calcium measurement. *Clin Chem* 2009;55:533-40. <http://dx.doi.org/10.1373/clinchem.2008.116707>.
33. van Walraven C, Raymond M. Population-based study of repeat laboratory testing. *Clin Chem* 2003;49:1997-2005. <http://dx.doi.org/10.1373/clinchem.2003.021220>.
34. The association for clinical biochemistry and laboratory medicine, national minimum retesting interval project: a final report detailing consensus recommendations for minimum retesting intervals for use in clinical biochemistry. Available at: <http://www.acb.org.uk/docs/default-source/guidelines/acb-mri-recommendations-a4-computer.pdf?sfvrsn=2>, ed., Vol., 2013. Accessed October 25, 2013.
35. Plaisant C, Lam S, Lam SJ, Shneiderman B, Smith MS, Roseman D, et al. Searching electronic health records for temporal patterns in patient histories: A case study with microsoft amalgam. *AMIA Annu Symp Proc* 2008:601-5.
36. Cummings KM, Frisof KB, Long MJ, Hrynkiwicz G. The effects of price information on physicians' test-ordering behavior. Ordering of diagnostic tests. *Med Care* 1982;20:293-301. <http://dx.doi.org/10.1097/00005650-198203000-00006>.
37. Bates DW, Kuperman GJ, Jha A, Teich JM, Orav EJ, Ma'luf N, et al. Does the computerized display of charges affect inpatient ancillary test utilization? *Arch Intern Med* 1997;157:2501-8. <http://dx.doi.org/10.1001/archinte.1997.00440420135015>.
38. Miyakis S, Karamanof G, Liotos M, Mountokalakis TD. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J* 2006;82:823-9. <http://dx.doi.org/10.1136/pgmj.2006.049551>.
39. Martin AR, Wolf MA, Thibodeau LA, Dzau V, Braunwald E. A trial of two strategies to modify the test-ordering behavior of medical residents. *N Engl J Med* 1980;303:1330-6. <http://dx.doi.org/10.1056/NEJM198012043032304>.
40. Kansas BCBS. Medical policy: Testing for vitamin d deficiency. Available at: [http://www.bcbsks.com/customerservice/providers/MedicalPolicies/policies/policies/Testing\\_VitaminDDeficiency\\_2011-12-21.pdf](http://www.bcbsks.com/customerservice/providers/MedicalPolicies/policies/policies/Testing_VitaminDDeficiency_2011-12-21.pdf). Accessed October 25, 2013.
41. Rosenbloom S, Chiu K, Byrne D, Talbert D, Neilson E, Miller R. Interventions to regulate ordering of serum magnesium levels: Report of an unintended consequence of decision support. *J Am Med Inform Assoc* 2005;12:546-53. <http://dx.doi.org/10.1197/jamia.M1811>.
42. Wu AH, Lewandrowski K, Gronowski AM, Grenache DG, Sokoll LJ, Magnani B. Antiquated tests within the clinical pathology laboratory. *Am J Manag Care* 2010;16:e220-7.
43. Warren JS. Laboratory test utilization program: Structure and impact in a large academic medical center. *Am J Clin Pathol* 2013;139:289-97. <http://dx.doi.org/10.1309/AJCP4-G6UAUXCFTQF>.
44. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to improve physicians' use of diagnostic tests: A new conceptual framework. *JAMA* 1998;280:2020-7. <http://dx.doi.org/10.1001/jama.280.23.2020>.
45. Kim JY, Dzik WH, Dighe AS, Lewandrowski KB. Utilization management in a large urban academic medical center: A 10-year experience. *Am J Clin Pathol* 2011;135:108-18. <http://dx.doi.org/10.1309/AJCP4GS7KSDBACF>.
46. Bareford D, Hayling A. Inappropriate use of laboratory services: Long term combined approach to modify request patterns. *BMJ* 1990;301:1305-7. <http://dx.doi.org/10.1136/bmj.301.6764.1305>.
47. Shahangian S, Alspach TD, Astles JR, Yesupriya A, Dettwyler WK. Trends in laboratory test volumes for medicare part b reimbursements, 2000-2010. *Arch Pathol Lab Med* 2014;138:189-203. <http://dx.doi.org/10.5858/arpa.2013-0149-OA>.
48. Neilson E, Johnson K, Rosenbloom S, Dupont W, Talbert D, Giuse D, et al. The impact of peer management on test-ordering behavior. *Ann Intern Med* 2004;141:196-204. <http://dx.doi.org/10.7326/0003-4819-141-3-200408030-00008>.
49. Studnicki J, Bradham D, Marshburn J, Foulis P, Straumford J. Measuring the impact of standing orders on laboratory utilization. *Laboratory Medicine* 1992;23:24-8.
50. May TA, Clancy M, Critchfield J, Ebeling F, Enriquez A, Gallagher C, et al. Reducing unnecessary inpatient laboratory testing in a teaching hospital. *Am J Clin Pathol* 2006;126:200-6. <http://dx.doi.org/10.1309/WP59YM73L6CEGX2F>.
51. Edlefsen KL, Tait JF, Wener MH, Astion M. Utilization and diagnostic yield of neurogenetic testing at a tertiary care facility. *Clin Chem* 2007;53:1016-22. <http://dx.doi.org/10.1373/clinchem.2006.083360>.
52. Dickerson J, Cole B, Conta J, Wellner M, Wallace S, Jack R, et al. Improving the value of costly genetic reference laboratory testing with active utilization management. *Arch Pathol Lab Med* 2014;138:110-3. <http://dx.doi.org/10.5858/arpa.2012-0726-OA>.
53. Wang TJ, Mort EA, Nordberg P, Chang Y, Cadigan ME, Mylott L, et al. A utilization management intervention to reduce unnecessary testing in the coronary care unit. *Arch Intern Med* 2002;162:1885-90. <http://dx.doi.org/10.1001/archinte.162.16.1885>.
54. Shalev V, Chodick G, Heymann AD. Format change of a laboratory test order form affects physician behavior. *Int J Med Inform* 2009;78:639-44. <http://dx.doi.org/10.1016/j.ijmedinf.2009.04.011>.
55. Amukele TK, Baird GS, Chandler WL. Reducing the use of coagulation test panels. *Blood Coagul Fibrinolysis* 2011;22:688-95. <http://dx.doi.org/10.1097/MBC.0b013e32834b8246>.
56. Staff LE. Improving laboratory test utilization through physician report cards: An interview with dr. Kim riddell. *Laboratory Errors and Patient Safety* 2005;2:2-6.
57. Bunting PS, Van Walraven C. Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem* 2004;50:321-6. <http://dx.doi.org/10.1373/clinchem.2003.025098>.
58. Studnicki J, Bradham DD, Marshburn J, Foulis PR, Straumford JV. A feedback system for reducing excessive laboratory tests. *Arch Pathol Lab Med* 1993;117:35-9.
59. Dickerson JA, Jack RM, Astion ML, Cole B. Another laboratory test utilization program: Our approach to reducing unnecessary 1,25-dihydroxyvitamin d orders with a simple intervention. *Am J Clin Pathol* 2013;140:446-7. <http://dx.doi.org/10.1309/AJCPQS40FZTLTQDH>.
60. Emerson JF, Emerson SS. The impact of requisition design on laboratory utilization. *Am J Clin Pathol* 2001;116:879-84. <http://dx.doi.org/10.1309/WC83-ERLY-NEDF-471E>.