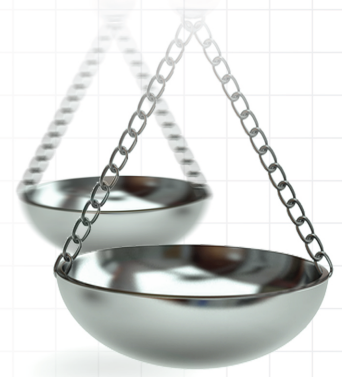




Early cost–effectiveness modeling for better decisions in public research investment of personalized medicine technologies

Daphne I Ling^{*,1,2}, Larry D Lynd^{1,2}, Mark Harrison^{1,2}, Aslam H Anis^{1,3} & Nick Bansback^{1,3}

¹Centre for Health Evaluation & Outcome Sciences, St Paul's Hospital, Vancouver, British Columbia, Canada

²Collaboration for Outcomes Research & Evaluation, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada

³School of Population & Public Health, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

*Author for correspondence: lingd@hss.edu

Millions of dollars are spent on the development of new personalized medicine technologies. While these research costs are often supported by public research funds, many diagnostic tests and biomarkers are not adopted by the healthcare system due to lack of evidence on their cost–effectiveness. We describe a stepwise approach to conducting cost–effectiveness analyses that are performed early in the technology's development process and can help mitigate the potential risks of investment. Decision analytic modeling can identify the key drivers of cost effectiveness and provide minimum criteria that the technology needs to meet for adoption by public and private healthcare systems. A value of information analysis can quantify the added value of conducting more research to provide further evidence for policy decisions. These steps will allow public research funders to make better decisions on their investments to maximize the health benefits and to minimize the number of suboptimal technologies.

First draft submitted: 16 April 2018; Accepted for publication: 29 October 2018; Published online: 10 December 2018

Keywords: cost–effectiveness • decision analysis • economic evaluation • health technology assessment • modeling • personalized medicine • value of information

In April 2013, a presentation on the clinical and cost–effectiveness of the carbon-13 urea breath test (13C UBT) was made before the Ontario Health Technology Advisory Committee (OHTAC), whose mandate is to provide recommendations on health interventions to Ontario practitioners, health system leaders and the Ministry of Health [1]. The 13C UBT is a noninvasive test to detect *Helicobacter pylori* in patients with dyspepsia, a condition defined by chronic pain or discomfort in the upper GI tract [1] (Table 1).

A meta-analysis of 21 studies (containing 4536 patients) performed by staff at Health Quality Ontario found that while sensitivity was similar between the 13C UBT (95%) and ELISA serology (93%, the first-line diagnostic test), the specificities were 91 and 71%, respectively, indicating that 20% more patients would be falsely positive on serology and unnecessarily treated [1]. In the economic evaluation, these test accuracy measures were used to inform a decision analytic model to compare three different testing strategies from the health system perspective: ELISA serology, 13C UBT and a two-step algorithm where only positive serology results are confirmed by the 13C UBT [2]. This analysis showed that while both the 13C UBT and two-step strategy were more cost-effective compared with ELISA serology for cases of misdiagnosis avoided (false positive and false negative results), the 13C UBT was dominated (i.e., it was more costly and less effective) by the 2-step strategy. More importantly, the budget impact analysis indicated that the province of Ontario would have to spend \$8 million to implement the 13C UBT and \$5 million for the two-step strategy.

In their decision-making framework, OHTAC uses a decision determinants tool that considers four main criteria: overall clinical benefit (measure of the net health benefit of a technology to diagnose or manage a disease, condition

Table 1. Sample of articles that are most relevant for economic modeling of personalized medicine technologies.

Author (Year)	Title	Journal	Description	PubMed ID	Ref.
Masucci (2013)	Cost-effectiveness of the carbon-13 urea breath test for the detection of <i>Helicobacter pylori</i> : an economic analysis	<i>Ont. Health Technol. Assess. Ser.</i>	Example of a test that has been developed and evaluated to not be cost effective	24228083	[2]
Hall (2010)	Health economics in drug development: efficient research to inform healthcare funding	<i>Eur. J. Cancer</i>	Article argues for economic evaluations to be performed earlier in the drug development pathway	20655197	[5]
Buisman (2016)	The early bird catches the worm: early cost-effectiveness analysis of new medical tests	<i>Int. J. Technol. Assess. Health Care</i>	Article provides general steps and advice to test developers for early economic evaluations	27002226	[6]
Phillips (2014)	The economic value of personalized medicine tests: what we know and what we need to know	<i>Genet. Med.</i>	Systematic review finds that economic evaluations are usually performed after tests are already adopted	24232413	[8]
Najafzadeh (2012)	Cost-effectiveness of using a molecular diagnostic test to improve preoperative diagnosis of thyroid cancer	<i>Value Health</i>	Example of how economic modeling can be used to evaluate hypothetical tests	23244801	[15]
Bansback (2009)	Statin therapy in rheumatoid arthritis: a cost-effectiveness and value-of-information analysis	<i>PharmacoEconomics</i>	Example of how value of information analysis can be used within the context of economic evaluations	19178122	[37]
McKenna (2016)	Methods to place a value on additional evidence are illustrated using a case study of corticosteroids after traumatic brain injury	<i>J. Clin. Epidemiol.</i>	Example of how value of information analysis can be used within the context of economic evaluations	26388041	[38]
Bryan (2014)	Breaking the addiction to technology adoption	<i>Health Econ.</i>	Article argues for greater emphasis to be placed on technology management rather than adoption	24590701	[53]
Van Gestel (2012)	The role of the expected value of individualized care in cost-effectiveness analyses and decision making	<i>Value Health</i>	Methodological work on the impact of heterogeneity in evaluating personalized medicine	22264967	[40]
Basu (2007)	Value of information on preference heterogeneity and individualized care	<i>Med. Decis. Making</i>	Methodological work on the impact of heterogeneity in evaluating personalized medicine	17409362	[39]

or healthcare related issue), consistency with societal and ethical values (may include measured preferences or ethical principles relevant to the use of the technology), value for money (measure of the net cost or efficiency of the health technology compared with available alternatives) and feasibility of adoption (measure of the ease with which a technology can be adopted into the healthcare system through the identification of specific issues likely to arise from implementation) [3]. On this basis, OHTAC recommended the continued use of ELISA serology as the first-line diagnostic test because false-positive results were considered less of a concern than false-negative results, which can allow progression to ulcers or gastric cancer. In addition, the cost for scale-up of the 13C UBT was not trivial; the test requires specialized laboratory equipment and is itself more costly than ELISA serology (\$75 vs \$14) [2]. The recommendation not to roll out the 13C UBT represents a nightmare scenario for test developers and serves as an example of the ‘last-mile problem’, in which a new technology is not adopted by the healthcare system despite million-dollar investments in R&D before the product is launched on the market [4].

While there are finite budgets for health systems, both private companies and the public sector also have finite resources for funding research. While private test developers are driven by profit motives and have to answer to their shareholders, research funders such as the Canadian Institutes of Health Research (CIHR) and Genome Canada are also motivated to maximize the return on public research investments. By viewing the research that they fund as investment portfolios, public funders also have to prove the impact of their research to society. They report to the government who provides funding through taxpayer money, so that it is important for them to invest in research that demonstrates health and/or economic benefits. With the rapidly growing pipeline of personalized medicine technologies, public research funders are increasingly aware that research funding may be invested in technologies that ultimately will be too expensive to be adopted by the healthcare system. Among a pool of emerging technologies, only a fraction will be clinically effective and, of those, a fraction will be cost-effective.

Hall *et al.* have proposed early-phase economic models for drug development that would synthesize the evidence upstream in the pathway in order to inform downstream reimbursement decisions [5]. A similar issue exists for personalized medicine technologies, which we define as tools to measure biomarkers or identify genetic mutations. In turn, these technologies are developed into diagnostic tests that can be evaluated with test accuracy measures (i.e., sensitivity, specificity and predictive values) that are dependent on the test thresholds used. In particular, companion diagnostic tests are pivotal to the personalized medicine movement. These types of tests are administered before a particular drug is given and incorporate molecular biomarkers to predict safe and effective responses to guide treatment selection in patients. Buisman *et al.* have developed a framework with general steps for conducting early cost–effectiveness analyses of medical tests from the perspective of test developers for internal decision-making on hitting profit-driven targets [6]. There is also a need for methodological approaches that can reduce the premarket uncertainty related to public investment in research and, furthermore, to prioritize which proposals should be funded and which technologies should be developed into diagnostic tests [7].

A review of personalized medicine diagnostics has shown that cost–effectiveness analyses of these tests are usually performed after they are adopted [8]. In many jurisdictions, the diagnostic test is already implemented even if it does not provide good value for money, due to the fact that economic evaluations are not usually required for regulatory approval of medical tests. This means that scarce resources are taken away from other tests that would provide more benefits to the population. Increasingly, countries are starting to consider opportunity costs and, like OHTAC, will not pay for tests that are not cost effective [9–11]. Furthermore, there is the issue of service capacity and whether the health system can afford to introduce the diagnostic test.

There are several areas in which research investments are worthwhile from an economic perspective because of the potential cost savings associated with more informed decisions. Many chronic diseases present promising avenues for a personalized medicine approach to disease management, including rheumatoid arthritis and chronic obstructive pulmonary disease (COPD). Biologic agents are used to treat rheumatoid arthritis when conventional therapy has failed. These biologics can cost up to \$25,000 per year per patient, but only 60% of patients will achieve a satisfactory response at 6 months with the first biologic prescribed [12]. A biomarker that can predict treatment response and guide the choice of biologics would likely be cost effective. Similarly, there are no biomarkers to predict COPD exacerbations, the leading cause of hospitalization and mortality among COPD patients [13], although this is an active area of development [14]. In the area of oncology, one study has shown that a molecular test used to improve the diagnosis of thyroid cancer would be cost effective compared with using fine-needle aspiration biopsy alone [15].

We describe methods that can help guide which health technologies make better research investments based on early evaluations of its likelihood of being adopted into routine practice, even before development begins. This approach will help research funders to better manage their risks of investment and maximize the impact of public funds to achieve health and economic benefits for society.

A stepwise approach

A four-phase clinical trial system is well-established for evaluating the safety and efficacy of new drugs. In contrast, there is an absence of a strong framework or clear international standards for medical tests [16,17]. Recent efforts by the Institute of Medicine have proposed a process for development and evaluation of genomic tests that consists of analytical validity in the laboratory, clinical/biological validity in a blinded sample, then followed by clinical utility after completion of the discovery phase [7]. These steps of the evaluation process are interdependent. During the clinical utility phase, in which tests are evaluated for their intended use in clinical practice, potential study designs include a hybrid prospective–retrospective design or prospective clinical trial in which the test may or may not direct patient management. In the hybrid design, previously archived specimens from a past clinical trial are linked to existing data on clinical outcomes of the test-directed treatment [7].

If a health technology assessment is performed for reimbursement purposes, a common pathway is shown in Figure 1. Once a test or biomarker meets the technical requirements in a laboratory setting, its applications in a clinical setting are assessed, starting with its diagnostic accuracy. The test then goes through the regulatory approval process, and cost–effectiveness studies are conducted for reimbursement decisions. Often, pragmatic trials or observational data are required to obtain further information before decisions can be made. Once a test has been licensed, however, further research becomes more difficult, as the justification of clinical equipoise is hard to demonstrate and both patients and physicians are less willing to engage in research [5]. On the other hand, if a test is beneficial, then the delay in implementation represents missed opportunities in improving health outcomes.

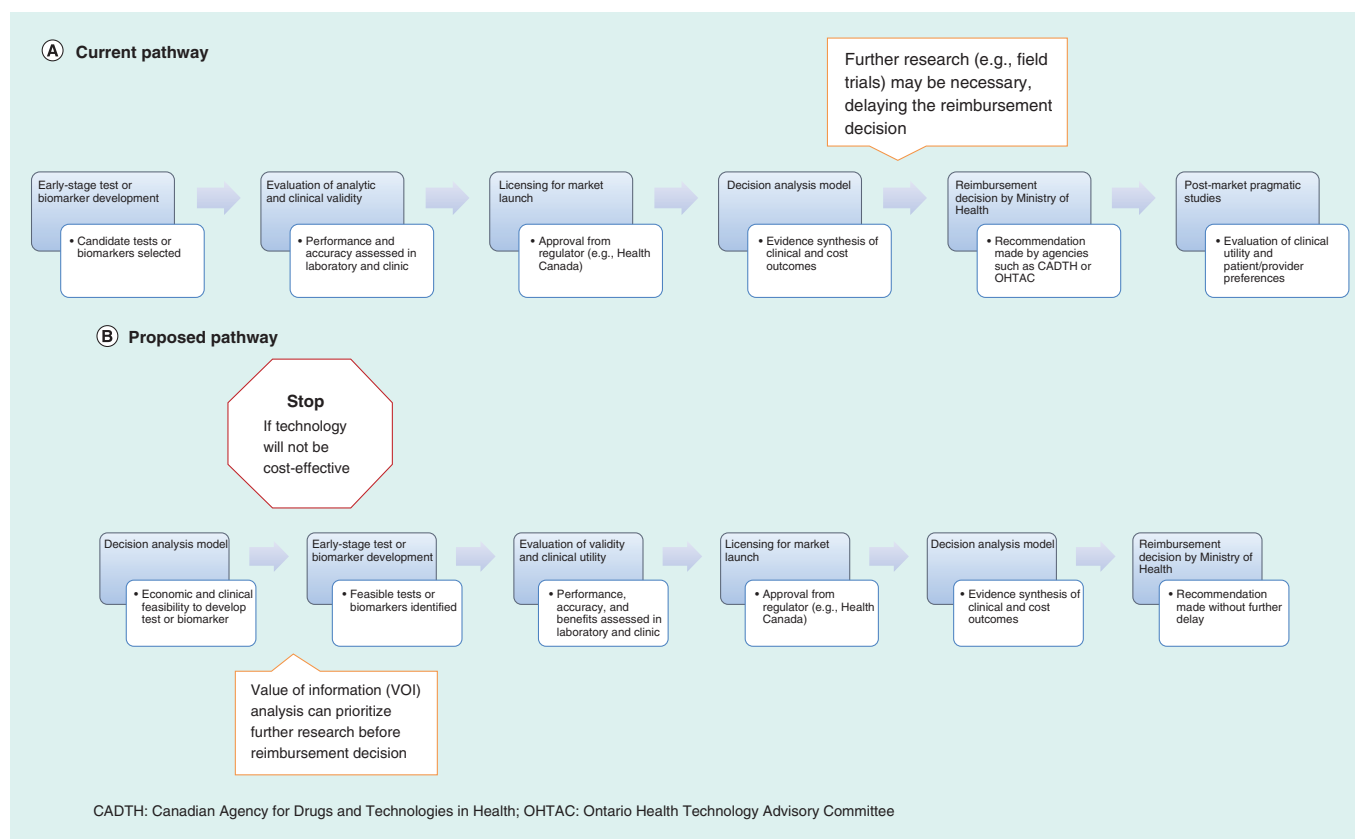


Figure 1. The current and proposed technology development pathways. In the proposed pathway, the decision analysis model is moved upstream before development begins to assess feasibility. There is a no/no-go decision to be made at this stage depending on whether the technology will be cost effective. Furthermore, the model can help determine whether the new technology will have any clinical utility according to patient and provider preferences. Subsequently, value of information (VOI) analysis can inform which research studies should be conducted earlier so there is no delay in the reimbursement decision. Adapted with permission from [5,13,58].

By moving economic evaluations further upstream and identifying research gaps earlier, the pathway of new technologies from development to adoption becomes more efficient.

Another important difference between drugs and diagnostic tests is that tests themselves are generally not useful if their results are not actionable. There are exceptions, however, such as the end of the ‘diagnostic odyssey’ for rare diseases or the nonhealth benefits of ‘empowerment’ when accessing clinical genetic services for incurable conditions [18]. The clinical utility of a new test has to be demonstrated in terms of whether it influences diagnostic and treatment decisions and how they, in turn, improve downstream patient outcomes [19,20]. Even in the early stages of development, it becomes important to consider if and how a new technology will be used according to both provider and patient preferences. Currently, studies on clinical utility are usually reserved for the postmarket stage (shown in Figure 1A).

There are numerous examples of tests that are available on the market but that are not used in clinical practice. In one study, most children who had negative results on a follow-up blood test for tuberculosis were given preventive therapy because they had been in close contact with active tuberculosis cases [21]. Physicians were reluctant to withhold treatment because there was still the risk of progression to active disease in spite of the negative test result. Similarly, a trial evaluating a pharmacogenetic test for azathioprine found that clinicians did not use the results from TPMT genotyping to guide their decision making; even in patients identified to be at low risk of profound neutropaenia, physicians chose not to prescribe higher drug doses [22]. Recently, studies have also evaluated the accuracy of noninvasive screening tests for Down’s syndrome in the first trimester of pregnancy [23,24]. While these blood-based tests and markers have been shown to be highly accurate compared with standard screening tests, the difficult decisions to be made following a positive test may not be acceptable to all women due to individual values and preferences.

Randomized controlled trials (RCTs) can be used to evaluate a test's impact on patient outcomes. However, unlike therapeutic interventions, this type of study design is not ideal for diagnostic tests and is rarely used in practice for several reasons. They are harder to design due to the complex interaction of factors during the diagnosis, treatment and outcome stages of the patient care pathway. That is, diagnostic RCTs convert the research study into a hybrid design that evaluates a combined test–drug intervention. In addition, RCTs are limited by short follow-up periods, large sample sizes, selected patient populations and highly controlled settings when evaluating the impact on patient outcomes [25]. Decision analytic modeling is an alternative method that can reduce, although not eliminate, the risks of investing in research for technologies that may not be clinically or cost effective (see Figure 1B).

This modeling technique can provide reasonable judgments on clinical utility in consultation with patients, providers and payers. As proposed for molecular diagnostic tests in oncology, decision analysis can model 'what if' scenarios, including the baseline and best or worst case scenarios, to provide plausible evidence of clinical utility [25]. Relevant parameters can then be varied in a decision analytic model to represent uncertainty in a sensitivity analysis. Furthermore, decision analysis provides a less expensive solution compared with conducting RCTs. When the 13C UBT was being evaluated, one small RCT was identified in which 43 patients were randomized to one of four management strategies, including 13C UBT and serology [26]. While symptom resolution, medication use and number of physician visits were similar across all arms, the average medical costs per patient were twice as much in the 13C UBT arm compared with serology (\$1726 vs \$898). Findings for health-related quality of life were variable, as patients tested with 13C UBT reported lower mental health scores but higher dyspepsia-specific quality of life scores compared with those randomized to serology. While everything appears 20/20 in hindsight, it is plausible that the RCT did not have to be conducted at all. A decision analytic model may well have predicted that ultimately the 13C UBT would not be cost–effective.

Decision analysis & cost–effectiveness

Each component of the evidence base, including diagnostic accuracy, clinical utility and cost–effectiveness, are necessary but not sufficient on their own to provide decision makers with the information to allocate scarce public resources. Decision analytic models attempt to combine multiple sources of evidence (e.g., clinical, economic, expert opinion) into a single framework to assess the potential costs and benefits of adopting a new test [27]. The model connects intermediate end points such as true or false positive and negative results with the test's impact on health outcomes and health-related quality of life (Figure 2). Moreover, it is important to consider the entire care pathway and measure downstream benefits, which may help offset some of the technology's upfront costs [8]. In formal economic evaluations, at least two strategies are compared by estimating the incremental costs and incremental benefits, often measured in quality-adjusted life years (QALYs). If one strategy is both more costly and more effective, then it is necessary to calculate the incremental cost–effectiveness ratio (ICER), which is the additional cost of the new intervention above the comparative intervention divided by the difference in QALYs [28].

A major advantage of building decision analytic models is the ability to vary assumptions (e.g., test cost and accuracy) to explore the key determinants that affect the overall cost effectiveness. There is the potential for threshold analysis to determine in advance how good (or how cheap) a test needs to be for it to be a viable option moving forward in the development process. This approach can be seen as a univariate sensitivity analysis in which one parameter is varied over a plausible range while all others are held constant to measure its independent effect on the resulting incremental costs and QALYs. Using existing epidemiologic data on incidence, prevalence, or mortality rates, plausible ranges of a test's sensitivity and specificity can be varied to determine the diagnostic performance that is necessary and at what cost for it to be cost effective (Figure 3).

One such study of a molecular test to improve the diagnosis of thyroid cancer in patients with indeterminate biopsy results found that the test's sensitivity had a larger effect on the outcomes compared with specificity [15]. This same approach can be used to assess threshold values for other model parameters, such as the cost of the test (i.e., headroom). In the thyroid cancer example, the molecular testing strategy was cost saving at \$1087 per QALY gained, excluding the direct cost of the test itself and assuming a sensitivity and specificity of 95%. Thus, this value is the threshold for the cost of the test at which this strategy would be considered cost neutral. In fact, if the cost of the molecular test was less than \$500, then there was 100% certainty that this strategy would be dominant (more effective and less costly) at higher levels of both sensitivity and specificity. When the sensitivity decreased to 87.5% or less, however, the incremental QALYs were negative, indicating that it would not be beneficial to use the molecular test.

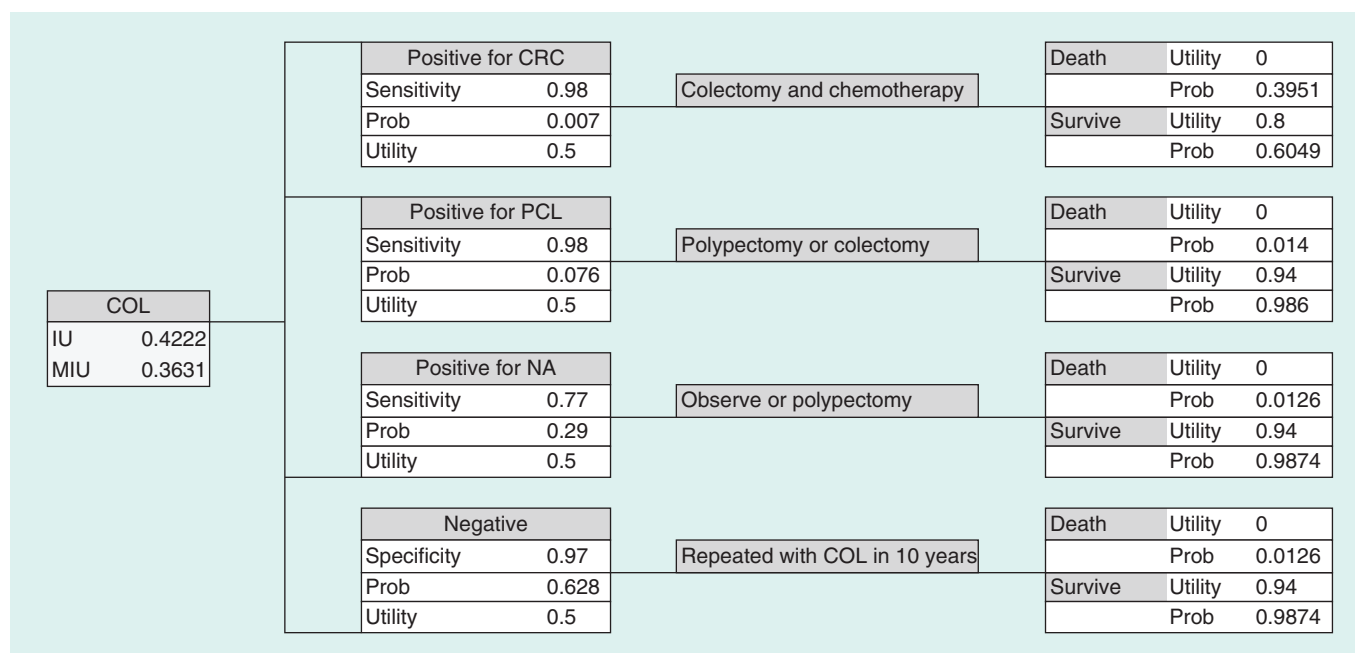


Figure 2. Example of a decision analysis tree to evaluate the cost-effectiveness of a screening strategy with colonoscopy for diagnosis of colorectal cancer.

The expected benefits (or utilities) are shown, while the same model can be used to calculate the expected costs using the probabilities of each outcome.

COL: Colonoscopy; CRC: Colorectal cancer; IU: Integrated utility; MIU: Modified integrated utility; NA: Non-advanced adenoma; PCL: Precancerous lesions.

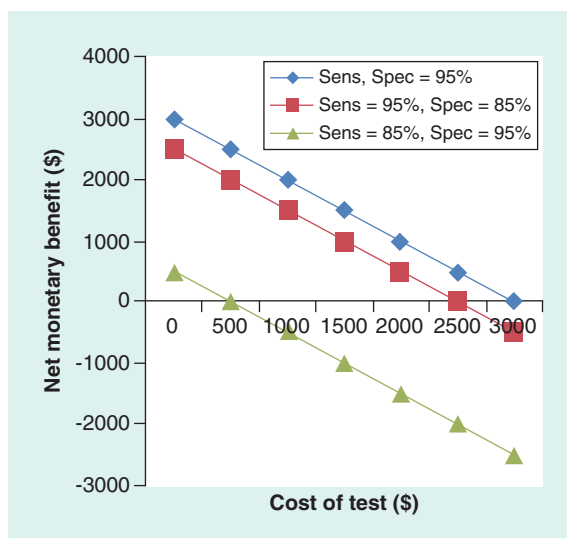


Figure 3. Net monetary benefit of a hypothetical molecular test to diagnose thyroid cancer at varying levels of sensitivity/specificity and different costs of the test.

The net monetary benefit is defined as (willingness-to-pay threshold x incremental QALY – incremental cost).

Adapted with permission from [15].

In another example, challenge ROC curves were used to assess the cost effectiveness of noninvasive diagnostic tests, such as magnetic resonance imaging (MRI) and computed tomography (CT), to diagnose coronary heart disease [29]. Threshold pairs of sensitivity and specificity that are on the challenge ROC curve or to its upper left represent the diagnostic performance required for a test to be cost effective at a given willingness to pay threshold. The study found that a new test had to be relatively inexpensive (<\$1000) and have near-perfect sensitivity and specificity (>90%) to warrant further development. In another use of the threshold approach, various parameters were evaluated for their effect on changing the threshold probability of using MRI to diagnose multiple sclerosis in patients with mild neurological symptoms [30]. The thresholds indicated the lowest probability for treating

the disease in which immediate MRI is more beneficial than waiting for further symptoms to develop. The authors found that reducing the years of anxiety from 20 to 10 years resulted in lowering the treatment threshold substantially enough such that use of MRI had value in ruling out or confirming disease. Thus, quality of life was a major driving force on the clinical utility of MRI.

Value of information analysis

While threshold analysis is a type of one-way sensitivity analysis, there is always a degree of uncertainty in the parameters used to populate a decision analytic model. In a probabilistic sensitivity analysis (PSA), probability distributions are assigned to all parameters simultaneously to facilitate the evaluation of their joint uncertainty around the incremental costs and QALYs. Using PSA allows for subsequent use of value of information (VOI) analysis, which can identify the potential value of conducting future research for key parameters [31,32]. VOI analysis has the advantage over conventional sensitivity analysis in that it provides a quantifiable estimate of the potential economic gain of obtaining additional information [5]. Estimates from this analysis, such as the expected value of perfect information (EVPI) and expected value of partial perfect information (EVPPI), indicate the upper bound on the value of conducting future research to reduce the uncertainty in specific model parameters.

To estimate the EVPI, the cost of uncertainty for each simulated cost-QALY pairing is calculated by taking the difference between the maximum net benefit of a new technology for that simulation and the expected maximum net benefit across all the simulations [33]. Taking the average of this net benefit distribution across all simulations and multiplying by the expected number of patients who will benefit from additional information gives the EVPI. To estimate the EVPPI, two simulation loops are needed: an outer loop that involves random sampling of the parameter of interest and an inner loop that involves random sampling from the other parameters [34]. First, for each outer loop simulation the net benefit of the technology that gives the maximum net benefit across the inner loop simulations is calculated. Second, averaging the net benefit distribution across all outer loop simulations gives the net benefit of perfect information for the parameter of interest. Third, the difference between this value of additional information and the expected net benefit of a new technology based on current information is calculated. Finally, multiplying this value by the expected number of patients who will benefit from additional information gives the EVPPI [35].

While most studies report the EVPI or EVPPI, it must be noted that these estimates do not allow decision makers to understand the costs of reducing partial uncertainty and whether future research is worthwhile. The expected value of sample information (EVSI) and expected net gain of sampling (ENSG) may be more appropriate VOI measures to fully inform research investment decisions. The EVSI measures the gain of reducing uncertainty through obtaining new data in a study. A large study will give a larger EVSI than a small one. However, since a large study costs more to conduct, the difference between the EVSI and the new study cost is the ENSG. The ideal sample size for the study is the one that maximizes the ENSG and would give the highest return on investment [36].

It is expected that decisions on adopting a new technology come with uncertainty, which may be reduced with additional evidence. In the VOI analysis, the cost of this uncertainty is multiplied by the number of people affected to get a population estimate that represents a threshold value (i.e., EVSI) on whether it is better to make a decision based on the existing evidence or wait to conduct more research [37]. In theory, this value is the maximum amount that a policy maker should be willing to pay to gather more information before having to make a decision [5]. The threshold value is then compared with the cost of conducting a study to determine whether there is a positive net gain (i.e., ENSG). If the cost of the proposed research exceeds this value, then it is not worthwhile to delay the decision to obtain further evidence.

Figure 4 shows the results from a VOI analysis for the benefits of statin therapy in rheumatoid arthritis patients. The most uncertain parameters that warranted further research were the disease-activity benefits and health-utility changes produced by statins over the long term. Given that the maximum values on these parameters were in the hundreds of millions of dollars, it is almost guaranteed that conducting further studies would be appropriate before making a policy decision. According to Figure 4, if a study that lasts longer than 12 months can be conducted for less than \$250 million, which is very likely, then it is worthwhile to delay the funding decision and invest in a study to obtain more data.

From the perspective of the research funder, VOI analysis can also be used to inform funding decisions when many grant proposals are competing for the same pool of funds. When results from a VOI analysis are included as part of these proposals, the applications can then be compared objectively and prioritized. This process also offers a sense of transparency on decisions for research funding. In a retrospective analysis of the Corticosteroid

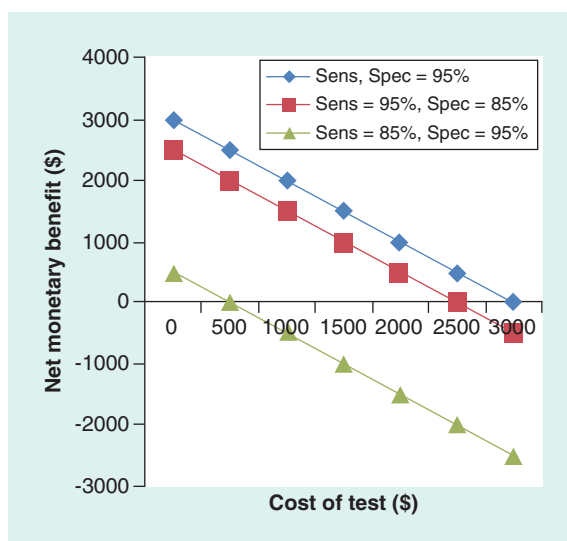


Figure 4. Population value of information of relevant parameters for future research on the use of statins for rheumatoid arthritis.

The population value of information is calculated by multiplying the cost of uncertainty for each parameter by the number of rheumatoid arthritis patients. A value of 0 for coronary heart disease events indicates that there is no value in conducting a future study for this parameter.

CHD: Coronary heart disease; CRP: C-reactive protein; DAS: Disease activity score; HAQ: Health assessment questionnaire.

Adapted with permission from [37].

Randomisation After Significant Head Injury (CRASH) trial, the VOI analysis indicated that the UK National Health Service would have to spend £205 million on other programs in order to obtain the same health benefits conferred by the CRASH trial [38]. This was almost a 100-fold increase from the actual cost of the study (£2.2 million). Clearly, the successful proposal for the CRASH trial was justified and a very efficient use of resources. However, future studies assessing the same type of outcomes would not be worthwhile. The definitive CRASH trial would render the VOI for subsequent research into these outcomes to be near zero.

Applications to personalized medicine

Until recently, VOI analysis has focused on the average costs and benefits at the population level (EVPI). However, the advent of personalized medicine makes individualized care and patient preferences important considerations for evaluating cost effectiveness [39]. Basu and Meltzer introduced the expected value of individualized care (EVIC) framework, in which the potential value of providing physicians with information on patient preferences for making treatment decisions is calculated. While the EVPI captures population parameter uncertainty, the EVIC captures patient heterogeneity and (unknown) variability [40]. That is, the optimal treatment decision depends on many values/weights of patient-level attributes, as opposed to a single value for one parameter in the population. It is important to note that the values of EVPI and EVIC are independent from one another [40].

The EVIC provides guidance on when the value of providing individualized care exceeds the value of cost-effective decisions at the population level, and the EVIC can also be calculated for specific parameters (analogous to the EVPPI) to rank those that are most important to the individualized treatment decision [39]. While the EVIC can vary with different insurance structures, the implication of this framework is that treatment coverage decisions (e.g., radiation for prostate cancer) should not be based on average cost-effectiveness alone if most patients will not benefit or do not prefer that treatment. Similarly, the EVIC, which can also be viewed as the costs of ignoring patient preferences, can be used to avoid limiting coverage for a treatment option if certain patients will benefit [39].

Compared with the EVPI, calculating the EVIC presents additional challenges, including the high costs of eliciting patient preferences [39,41] and the availability of data on each treatment outcome for each patient [40]. Thus, special study designs such as individualized comparative effectiveness research [42] or advanced analysis techniques such as patient-level simulations are needed [15,43]. In an applied example of using either high or low intensity treatment for glaucoma, the EVPI from a discrete event simulation model of individual patients was €0, indicating no value of conducting further research at the population level [40]. However, the EVIC was €580 per patient, indicating that there is value in implementing individualized care. The parameter-specific EVIC for disease progression rate, the patient-level attribute that explained much of the variance, was calculated to be €130 per patient. This value becomes the threshold for which investment of a hypothetical predictive test for individual glaucoma progression becomes worthwhile (with 100% sensitivity and specificity) [40]. In addition, the parameter-specific EVIC calculation can give rise to the test's cut-off value for when high-intensity treatment is preferred (incremental net benefit > 0) in order to individualize care in clinical practice.

Discussion & implications

We have described an overall process that can be used by research funders to inform decisions on investments for new technologies. More generally, it can inform which technologies may be viable and should be pursued. According to diffusion theory, reimbursement for a certain test will lead to its increased use [44]. Spending on medical tests that provide little or no benefit will lead to waste and inefficiencies in clinical practice. Furthermore, once a test is embedded into the healthcare system, it is hard to decommission its use, even if new evidence suggests that it is not cost effective or has no clinical utility. Decision analytic models allow for the early evaluation of new technologies along the patient care pathway.

Early evaluations of cost effectiveness can inform the entire life cycle of a new technology, even before development has begun. That is, economic evaluations offer a go/no go checkpoint to proceed to further stages of development. If a new test can meet both the diagnostic accuracy and willingness-to-pay thresholds for adoption by the healthcare system, then research investment is warranted. In addition, the cost of the test has to be set at a reasonable price so that it can be implemented. On the other hand, a proposed technology should not be funded if it is unlikely to be cost effective, and efforts to develop it further should be abandoned. However, if more data are needed to determine its cost effectiveness, then VOI analysis can be used to assess whether the cost of further research is in itself cost-effective. Phelps and Mushlin have proposed a two-hurdle process for diagnostic technologies [45]. They describe the first hurdle as the EVPI beyond a ‘fallback’ treatment strategy in the absence of the diagnostic information provided by the technology. This initial hurdle serves as a preliminary screen using published data, as only technologies with a positive EVPI would proceed to clinical studies of diagnostic accuracy for informing cost effectiveness, which is the second hurdle.

This type of analysis also presents criteria to judge which studies to fund and where research efforts should be directed when there are many competing proposals. It can also identify priority areas for targeted research calls, which should be linked to the areas of greatest potential uncertainty or benefit. In practice, however, use of VOI analysis has been limited due to many reasons. The results are highly dependent on the model structure and time horizon [36], and there are assumptions on the distribution of parameters as part of the PSA. Estimates on the treatment costs, study costs, patients’ QALYs and population size may not be available [46]. For personalized medicine studies in particular, the cost of future trials for subgroup analyses may be unknown [41], as are the patient-level attributes that may be unknown and/or unmeasurable at the time the treatment decision is made [40]. Furthermore, VOI analysis is a Bayesian method that is computationally intensive and hard to implement. To that end, recent developments have been made to improve the efficiency of economic models. A nonparametric regression approach and Bayesian Laplace approximations have been shown to have similar accuracy compared with conventional Monte-Carlo simulation when calculating EVPPI and EVSI [47–49]. Finally, the methodology may not be understandable to most decision makers. As a result, estimates from VOI analysis may not be used in funding decisions [46].

Economic modeling can also be adapted to evaluate future interventions in specific settings. One study has modeled the economic impact of three hypothetical interventions for COPD: decreasing smoking rates (using a molecular test that measures predisposition to COPD among early smokers), developing new drugs to prevent disease progression and using a molecular test to predict exacerbations [50]. Contrary to the notion that smoking cessation would be the most effective way to reduce COPD burden, the study found that using a predictive test to reduce the frequency of exacerbations would have the greatest impact in terms of monetary benefits. Similarly, having an explicit framework can help predict the potential use of resources at the societal level through a budget impact analysis. Decision makers are increasingly demanding evidence of the global costs of a new technology and its financial impact within their particular context. By applying the costs and health outcomes of an economic model to population-based cohorts, one can estimate the real-world costs of implementing a new test. In this manner, the hypothetical modeling approach is translated into values for an actual patient population.

Finally, early economic evaluations will have more global implications as the promise of personalized medicine starts to bear fruit. For research funders and the healthcare system, both of which are publicly funded in Canada, cost–effectiveness studies will result in better use of limited resources and potentially less waste. With increasing development of genomic tests and molecular biomarkers, it becomes imperative to avoid flooding the health system with technologies that do not provide sufficient value for money, leading to ‘new test fatigue’ [51] and the risk of undermining the overall benefits of the personalized medicine movement. Horizon scanning, in which a systematic search of the literature is performed to identify emerging technologies, has been used by regulatory

agencies to evaluate forthcoming changes in healthcare delivery [52]. Going further, Bryan *et al.* have called for shifting the focus away from technology adoption and moving toward technology management, which includes greater emphasis on disinvestment of technologies that provide marginal to no benefits at high costs [53]. Removing low-value technologies will only help make room in already constrained budgets for novel technologies that can provide better clinical and economic outcomes.

Conclusion

We have described an approach to consider premarket uncertainties for personalized medicine technologies from the perspective of research funders. The steps involved will require close collaboration with health economists in the preparation of proposals for funding. Considerations of cost effectiveness early in the stages of development will reduce the number of new tests that are either not adopted by the healthcare system or that have no utility in clinical practice. While the desire for innovative technologies that can deliver major benefits to patients requires a certain 'leap of faith', both decision modeling and the subsequent VOI analysis can help mitigate the potential risks of investment.

Future perspective

There is a need for model-based, cost-effectiveness analysis to be an iterative process [6,54]. That is, early models are developed and revised as more evidence becomes available. With the complexities of personalized medicine tests and its use in practice, the intervention itself and the relevant comparators in economic evaluations may change over time. Structural uncertainty, in addition to the omnipresent parameter uncertainty, becomes important [55,56]. To date, most of the focus on VOI methods have been on parameter uncertainty [57], but the structure of the model is also relevant in the context of personalized medicine. Scenario analysis, in which select factors are used to represent specific scenarios, may be a possible solution to evaluate complex test-and-treat interventions with multiple components [55,56].

Executive summary

Background

- The pipeline for personalized medicine is rapidly growing; diagnostic tests and biomarkers are crucial to this area.
- Only a fraction of these tests will be cost effective.
- Public research funders are aware that they are investing in technologies that ultimately will not be adopted by the healthcare system.

A stepwise approach

- Cost-effectiveness analyses that are performed early in the technology's development process can help mitigate the potential risks of investment.
- Decision analytic models combine multiple sources of evidence into a single framework to assess the potential costs and benefits of adopting a technology.

Discussion & implications

- We describe an approach that can help guide which technologies make better research investments based on early evaluations of its likelihood of being adopted into routine practice.
- With increasing development of genomic tests and molecular biomarkers, it becomes imperative to avoid flooding the health system with technologies that provide few benefits.

Future perspective

- Personalized medicine technologies are complex interventions with multiple components along the patient care pathway.
- Building the model structure requires an iterative approach that integrates new evidence and the need to address structural uncertainty through scenario analysis.

Conclusion

- Our approach will help research funders to better manage their risks of investment and maximize the impact of public funds to achieve health and economic benefits for society.

Acknowledgments

The authors would like to thank Stirling Bryan and Bruce McManus for providing helpful insights to the ideas in this manuscript. In addition, we acknowledge Alex Jiang for providing the example of a decision analysis model.

Financial & competing interests disclosure

DI Ling is supported by a Postdoctoral Fellowship Award from the Michael Smith Foundation for Health Research. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1. Ling D. Carbon-13 urea breath test for *Helicobacter pylori* infection in patients with uninvestigated ulcer-like dyspepsia: an evidence-based analysis. *Ont. Health Technol. Assess. Ser.* 13(19), 1–30 (2013).
2. Masucci L, Blackhouse G, Goeree R. Cost-effectiveness of the carbon-13 urea breath test for the detection of *Helicobacter pylori*: an economic analysis. *Ont. Health Technol. Assess. Ser.* 13(20), 1–28 (2013).
3. Carbon-13 urea breath test for *Helicobacter pylori* infection in patients with uninvestigated ulcer-like dyspepsia: OHTAC recommendation. www.hqontario.ca/Portals/0/Documents/evidence/reports/recommendation-urea-breath-test-131003-en.pdf (2013).
4. Blumenthal-Barby JS, McGuire AL, Green RC, Ubel PA. How behavioral economics can help to avoid the last mile problem in whole genome sequencing. *Genome Med.* 7(1), 3 (2015).
5. Hall PS, McCabe C, Brown JM, Cameron DA. Health economics in drug development: efficient research to inform healthcare funding decisions. *Eur. J. Cancer* 46(15), 2674–2680 (2010).
6. Buisman LR, Rutten-Van Molken MP, Postmus D, Luime JJ, Uyl-De Groot CA, Redekop WK. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. *Int. J. Technol. Assess. Health Care* DOI:10.1017/S0266462316000064 1–8 (2016).
7. *Evaluation of Translational Omics: Lessons Learned and the Path Forward*. Micheel CM, Nass SJ, Omenn GS (Eds). Institute of Medicine, The National Academies Press, DC, USA (2012).
8. Phillips KA, Ann Sakowski J, Trosman J, Douglas MP, Liang SY, Neumann P. The economic value of personalized medicine tests: what we know and what we need to know. *Genet. Med.* 16(3), 251–257 (2014).
9. Guidelines for the Pharmaceutical Industry on preparation of submissions to the pharmaceutical Benefits Advisory Committee. <https://pbac.pbs.gov.au/> (2006).
10. Hutton J, Maynard A. A nice challenge for health economics. *Health Econ.* 9(2), 89–93 (2000).
11. Schmidt C. Researchers consider value-of-information theory for selecting trials. *J. Natl Cancer Inst.* 102(3), 144–146 (2010).
12. Ranganathan P. Pharmacogenomics of tumor necrosis factor antagonists in rheumatoid arthritis. *Pharmacogenomics* 6(5), 481–490 (2005).
13. Sin DD, Hollander Z, Demarco ML, Mcmanus BM, Ng RT. Biomarker development for chronic obstructive pulmonary disease. From discovery to clinical implementation. *Am. J. Respir. Crit. Care Med.* 192(10), 1162–1170 (2015).
14. Proof. The High Cost of COPD. (2015).
15. Najafzadeh M, Marra CA, Lynd LD, Wiseman SM. Cost-effectiveness of using a molecular diagnostic test to improve preoperative diagnosis of thyroid cancer. *Value Health* 15(8), 1005–1013 (2012).
16. Price CP, Christenson RH. Evaluating new diagnostic technologies: perspectives in the UK and US. *Clin. Chem.* 54(9), 1421–1423 (2008).
17. Walley T. Evaluating laboratory diagnostic tests. *BMJ* 336(7644), 569–570 (2008).
18. Payne K, Mcallister M, Davies LM. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Econ.* 22(3), 258–271 (2013).
19. Ferrante Di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ* 344, e686 (2012).
20. Fineberg HV. Evaluation of computed tomography: achievement and challenge. *Am. J. Roentgenol.* 131(1), 1–4 (1978).
21. Ling DI, Crepeau CA, Dufresne M *et al.* Evaluation of the impact of interferon-gamma release assays on the management of childhood tuberculosis. *Ped. Infect. Dis. J.* 31(12), 1258–1262 (2012).
22. Thompson AJ, Newman WG, Elliott RA, Roberts SA, Tricker K, Payne K. The cost-effectiveness of a pharmacogenetic test: a trial-based evaluation of TPMT genotyping for azathioprine. *Value Health* 17(1), 22–33 (2014).
23. Alldred SK, Takwoingi Y, Guo B *et al.* First trimester serum tests for Down's syndrome screening. *Cochrane Database Syst. Rev.* 11, CD011975 (2015).
24. Norton ME, Jacobsson B, Swamy GK *et al.* Cell-free DNA analysis for noninvasive examination of trisomy. *New Engl. J. Med.* 372(17), 1589–1597 (2015).
25. Deverka P, Messner DA, McCormack R *et al.* Generating and evaluating evidence of the clinical utility of molecular diagnostic tests in oncology. *Genet. Med.* doi:10.1038/gim.2015.162 (2015).

26. Cuddihy MT, Locke GR 3rd, Wahner-Roedler D *et al.* Dyspepsia management in primary care: a management trial. *Int. J. Clin. Pract.* 59(2), 194–201 (2005).
27. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. *Med. Decis. Making* 29(5), E22–E29 (2009).
28. Detsky AS, Laupacis A. Relevance of cost–effectiveness analysis to clinicians and policy makers. *JAMA* 298(2), 221–224 (2007).
29. Hunink MG, Kuntz KM, Fleischmann KE, Brady TJ. Noninvasive imaging for the diagnosis of coronary artery disease: focusing the development of new diagnostic technology. *Ann. Intern. Med.* 131(9), 673–680 (1999).
30. Mooney C, Mushlin AI, Phelps CE. Targeting assessments of magnetic resonance imaging in suspected multiple sclerosis. *Med. Decis. Making* 10(2), 77–94 (1990).
31. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. *Health Technol. Assess.* 8(31), 1–103, iii (2004).
32. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ.* 5(6), 513–524 (1996).
33. Philips Z, Claxton KP, Palmer S, Bojke L, Sculpher MJ. Priority setting for research in health care: an application of value of information analysis to glycoprotein IIb/IIIa antagonists in non-ST elevation acute coronary syndrome. *Int. J. Technol. Assess. Health Care* 22(3), 379–387 (2006).
34. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med. Decis. Making* 27(4), 448–470 (2007).
35. Woods B, Veenstra D, Hawkins N. Prioritizing pharmacogenetic research: a value of information analysis of CYP2D6 testing to guide breast cancer treatment. *Value Health* 14(8), 989–1001 (2011).
36. Wilson EC. A practical guide to value of information analysis. *Pharmacoeconomics* 33(2), 105–121 (2015).
37. Bansback N, Ara R, Ward S, Anis A, Choi HK. Statin therapy in rheumatoid arthritis: a cost–effectiveness and value-of-information analysis. *Pharmacoeconomics* 27(1), 25–37 (2009).
38. McKenna C, Griffin S, Koffijberg H, Claxton K. Methods to place a value on additional evidence are illustrated using a case study of corticosteroids after traumatic brain injury. *J. Clin. Epidemiol.* doi:10.1016/j.jclinepi.2015.09.011 (2015).
39. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. *Med. Decis. Making* 27(2), 112–127 (2007).
40. Van Gestel A, Grutters J, Schouten J *et al.* The role of the expected value of individualized care in cost–effectiveness analyses and decision making. *Value Health* 15(1), 13–21 (2012).
41. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The value of heterogeneity for cost–effectiveness subgroup analysis: conceptual framework and application. *Med. Decis. Making* 34(8), 951–964 (2014).
42. Basu A. Individualization at the heart of comparative effectiveness research: the time for i-CER has come. *Med. Decis. Making* 29(6), NP9–NP11 (2009).
43. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 26(2), 131–148 (2008).
44. Green LW, Ottoson JM, Garcia C, Hiatt RA. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Ann. Rev. Public Health* 30, 151–174 (2009).
45. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Med. Decis. Making* 8(4), 279–289 (1988).
46. Conti R, Veenstra DL, Armstrong K, Lesko LJ, Grosse SD. Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost–effectiveness, and future research priorities. *Med. Decis. Making* 30(3), 328–340 (2010).
47. Tuffaha HW, Strong M, Gordon LG, Scuffham PA. Efficient value of information calculation using a nonparametric regression approach: an applied perspective. *Value Health* 19(4), 505–509 (2016).
48. Brennan A, Kharroubi SA. Efficient computation of partial expected value of sample information using Bayesian approximation. *J. Health Econ.* 26(1), 122–148 (2007).
49. Heath A, Manolopoulou I, Baio G. Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation. *Stat. Med.* 35(23), 4264–4280 (2016).
50. Najafzadeh M, Marra CA, Lynd LD *et al.* Future impact of various interventions on the burden of COPD in Canada: a dynamic population model. *PLoS ONE* 7(10), e46746 (2012).
51. Cobelens F, Van Den Hof S, Pai M *et al.* Which new diagnostics for tuberculosis, and when? *J. Infect. Dis.* 205(Suppl. 2), S191–S198 (2012).
52. CADTH. Horizon scan roundup: a compilation of new and emerging health technologies from around the world. 18 (2015).
53. Bryan S, Mitton C, Donaldson C. Breaking the addiction to technology adoption. *Health Econ.* 23(4), 379–383 (2014).
54. Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. *J. Health Serv. Res. Policy* 2(1), 26–30 (1997).

55. Annemans L, Redekop K, Payne K. Current methodological issues in the economic assessment of personalized medicine. *Value Health* 16(6 Suppl.), S20–S26 (2013).
56. Fugel HJ, Nuijten M, Postma M, Redekop K. Economic evaluation in stratified medicine: methodological issues and challenges. *Front. Pharmacol.* 7, 113 (2016).
57. Rogowski W, Payne K, Schnell-Inderst P *et al.* Concepts of personalization in personalized medicine: implications for economic evaluation. *PharmacoEconomics* 33(1), 49–59 (2015).
58. Husereau D, Arshoff L, Bhimani S, Allen N. Medical device and diagnostic pricing and reimbursement in Canada. Institute of Health Economics, Canada (2015).

