Challenges in health technology assessments of genetic tests

Xuanqian Xie[1], Olga Gajic-Veljanoski[1], Lindsey Falk[1], Alexis K. Schaink[1], Anna Lambrinos[1], Myra Wang[1], Vivian Ng[1], Wendy J. Ungar[2], Nancy Sikich[1]

[1]Ontario Health, Toronto, Canada; [2]Program of Child Health Evaluative Sciences, the Hospital for Sick Children Research Institute, Toronto, ON, Canada, and the Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada

Contributions: (I) Conception and design: X Xie, O Gajic-Veljanoski, M Wang, WJ Ungar, N Sikich; (II) Administrative support: X Xie, L Falk; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: X Xie, O Gajic-Veljanoski; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xuanqian Xie, Senior Health Economist. Ontario Health, 130 Bloor Street West, 10th floor, Toronto, ON, M5S 1N5, Canada. Email: shawn.xie@ontariohealth.ca.

Abstract: In the past decade, the use of genetic tests has grown rapidly worldwide. These technologies are often used in prenatal screening, carrier testing, prognostic testing, and diagnostic testing. Standard methods for meta-analysis of diagnostic test accuracy and cost-effectiveness analysis may need adaptation for health technology assessments (HTAs) of genetic tests. We provide some considerations relevant to these evaluations. We briefly address the following challenges for HTAs of genetic tests: (I) performing Bayesian meta-analysis of diagnostic test accuracy for rare genetic conditions; (II) performing Bayesian meta-analysis of diagnostic test accuracy in the absence of a perfect reference standard; (III) constructing economic models that account for conditional dependence between tests in the absence of a perfect reference standard; (IV) ascertaining the true prevalence of genetic conditions for economic analyses; (V) defining the time horizon and specifying health outcomes for economic modelling; and (VI) cost item measurement and valuation.

Keywords: Health technology assessment (HTA); genetic test; meta-analysis; economic modelling; time horizon; costing

Introduction

Genetic tests (including genomic tests) detect changes in chromosomes, genes, or proteins (1). With the current increasing availability of laboratory technology and techniques, the use of and total expenditures for genetic tests have grown rapidly worldwide. It was reported that there are approximately 75,000 genetic tests on the market and about 10 new tests enter the market each day in the United States (2). These tests are often used in prenatal screening, carrier testing, prognostic testing, and diagnostic testing (1-3). Compared with traditional diagnostic tests, genetic tests have new features that bring additional challenges to health technology assessments (HTAs). Consequently, standard methods for meta-analysis of diagnostic test accuracy and economic evaluations may need

adaptation for HTAs of genetic tests. This review provides some considerations relevant to the evaluation of genetic tests. While these issues are especially relevant to HTAs of genetic tests, they may also apply to non-genetic health technologies.

In our review, first, we briefly describe the Bayesian approach for a meta-analysis of diagnostic test accuracy and present an example related to the synthesis of published data for a very rare genetic condition. Next, we present a meta-analysis and suggest an economic model for combined tests in the absence of a perfect reference standard. Then we share some considerations for defining the prevalence of genetic conditions for economic models, given that the true prevalence is often unknown. We also discuss issues related to choosing appropriate health outcomes and time horizons in economic modelling of genetic tests. Finally, we outline some

Table 1 Non-invasive prenatal testing results for trisomy 13

Author, year	True positive	False positive	False negative	True negative
Bianchi *et al.*, 2014, (6)	1	1	0	897
del Mar Gil, 2014, (7)	1	0	0	191
Langlois *et al.*, 2017, (8)	0	1	0	1,151
Norton *et al.*, 2015, (9)	2	2	0	11,181
Palomaki *et al.*, 2017, (10)	2	2	0	2,527
Quezada *et al.*, 2015, (11)	2	2	3	2,778
Song *et al.*, 2013, (12)	1	0	0	1,740

important issues associated with the costing of genetic tests.

Bayesian meta-analysis of diagnostic test accuracy for a rare genetic condition: a case of trisomy 13 in prenatal screening

Any pregnancy has a small chance of having a baby with a chromosomal anomaly (4). Anomalies can include an incorrect number of chromosome copies (called chromosomal aneuploidies, and includes trisomies 13, 18, and 21) and small missing pieces from chromosomes (called microdeletions). Prenatal screening can detect some chromosomal anomalies in a pregnancy. Traditionally, either first-trimester screening or maternal serum screening is used to screen for trisomy 21 and 18. More recently, noninvasive prenatal testing (NIPT) has been introduced to prenatal screening. NIPT is a maternal blood test that examines cell-free fetal DNA to screen for common chromosomal anomalies (4). NIPT is more accurate than traditional prenatal screening and can also detect additional chromosomal anomalies such as trisomy 13, sex chromosome aneuploidies, and microdeletion syndromes (5).

We previously conducted an HTA on NIPT for select chromosomal anomalies (4). We identified 7 studies reporting the accuracy (sensitivity and specificity) of trisomy 13 (6-12). The results of these studies were presented as counts of true positives, false positives, false negatives, and true negatives in *Table 1*. Six of the 7 included studies had at least one zero cell in their 2×2 table and three studies had two zero cells. To overcome the computational problems of zero counts, some software packages add a small fixed value to all cells (i.e., continuity corrections). However, this approach is not well accepted. When multiple cells in a 2×2 table are zero or small in number, adding a fixed value (e.g., 0.5) may impact parameter estimates. More recently, software packages [e.g., SAS macro: MetaDAS; and R package: lme4 from Cochrane (13,14)] have been developed to estimate the pooled estimates of sensitivity and specificity. The statistical models (e.g., generalized linear mixed models) in these software packages allow for zero cells and therefore do not apply continuity corrections (14). The concept of these models can be simply understood in the following way: if a random variable X (e.g., the number of people with false negative results) follows a binomial distribution, X ~ Binomial (n, p) (e.g., n: the total number of true disease cases; and p: 1 – sensitivity), X has a chance to be 0, especially when n is small (i.e., low disease prevalence) or p is small (e.g., high test sensitivity).

Frequentist methods generally use the normal approximation approach but may have issues with convergence due to sparse data. In our HTA on NIPT, we were unable to obtain summary estimates of the pooled sensitivity and specificity using the SAS Macro "MetaDAS" and R package "lme4" because the models did not converge. We instead applied a Bayesian bivariate meta-analysis model for diagnostic test accuracy that uses a beta distribution for the prior probability distribution of disease prevalence, which is appropriate for cases of low disease prevalence. The Bayesian model was computed using WinBUGS 1.4.3 to avoid issues of non-convergence (15,16) and we found that the median of the pooled sensitivity and specificity for trisomy 13 were 76.0% [95% credible interval (CrI), 46.2% to 94.1%] and 99.9% (95% CrI, 99.9% to 100.0%), respectively. The WinBUGS code used in this section and the section below is available by request from the corresponding author.

Meta-analysis of diagnostic test accuracy and economic models in the absence of a perfect reference test: the case of non-small cell lung cancer

Genetic diagnostic tests commonly lack a perfect standard (i.e., the inability to definitively determine disease status), for example when the comparator quantifies a biomarker such as enzyme activity where the cut-point varies, or if the comparator is a genetic test that includes different variants (17). We illustrate this challenge in the context of a genetic test for non-small cell lung cancer (NSCLC).

Lung cancer is characterized by cancer cells forming in the tissue of one or both lungs (18). NSCLC includes several types of lung cancer except for small-cell lung cancer and accounts for 75% to 85% of all lung cancers. Some lung cancers will progress and the tumour cells develop a DNA resistance mutation in the epidermal growth factor receptor (*EGFR*) T790M gene. Identifying this resistance mutation can help physicians choose appropriate treatment (i.e., osimertinib if positive and alternate chemotherapy if negative). Traditionally, at the stage of disease progression, *EGFR* resistance mutation testing is done on DNA extracted from a tumour sample obtained by tissue biopsy; however, this is an invasive test for people with advanced NSCLC and is also costly. Consequently, cell-free circulating tumour DNA (ctDNA) blood testing (i.e., "liquid biopsy") was developed to enable detection of the resistance mutation *EGFR* T790M in people with advanced NSCLC noninvasively.

Tissue biopsy is an imperfect reference standard (i.e., the sensitivity and specificity are not 100%) because it is a sample of tumour cells and the mutation is not necessarily found in all sections of the tumour. Lack of a perfect reference standard complicates the evaluation of diagnostic test accuracy and the cost-effectiveness of tests like liquid biopsy for NSCLC. If the test being evaluated has higher sensitivity and specificity than the reference standard, and the reference standard is assumed to be perfect, then the additional patients correctly classified by the new test would be erroneously treated as false positives or false negatives. Dendukuri *et al*. (19) extended the Bayesian hierarchical summary receiver operating characteristic (HSROC) model (20) for meta-analysis of diagnostic test accuracy to the case where one or more imperfect reference standards are used in individual studies. Within each study, both the index test (liquid biopsy) and reference standard (e.g., tissue biopsy) are assumed to be imperfect measures of a common

underlying dichotomous latent variable D, the true disease status. The model provides estimates of the pooled sensitivity and specificity of the index test across studies, and the sensitivity and specificity for the reference standard. The model can adjust for possible conditional dependence between the index test and the reference standard within each latent class through the covariance of the sensitivities and specificities. Additionally, historical information or subjective knowledge about some test parameters (e.g., the sensitivity and specificity of the reference standard) may be incorporated though informative prior distributions.

We conducted Bayesian meta-analysis of diagnostic test accuracy for liquid biopsy for NSCLC using three different models: (I) perfect reference standard model; (II) imperfect reference standard model, assuming conditional independence; and (III) imperfect reference standard model, adjusting for conditional dependence (18). We used the deviance information criterion (DIC) to compare the three models. The DIC incorporates the goodness-of-fit and complexity of a model and a lower DIC is indicative of a better model. The imperfect reference standard model that adjusted for conditional dependence had the lowest DIC. The pooled sensitivity of liquid biopsy was 0.68 (95% CrI, 0.46 to 0.88) and the pooled specificity was 0.86 (95% CrI, 0.62 to 0.98). The sensitivity and specificity of tissue biopsy for NSCLC (reference standard) in this model were 0.86 (95% CrI, 0.75 to 0.98) and 0.93 (95% CrI, 0.85 to 0.99), respectively.

In clinical practice, patients may also receive a combination of tests. In the above NSCLC example, patients who have a negative result after liquid biopsy may go on to receive a tissue biopsy to confirm the results. When modelling two sequential or simultaneous tests, it is important to consider the potential correlation between tests. We summarize the probability of obtaining each possible test result from two combined tests in *Table 2*. Estimates can be derived from the meta-analysis model by sampling the posterior distribution of these parameters: the pooled estimates of sensitivity (S2) and specificity (C2) of the reference standard (T2; e.g., tissue biopsy), the sensitivity (S1) and specificity (C1) of the index test (T1; e.g., liquid biopsy), and the disease prevalence (Pi). Let Cov_S and Cov_C denote the covariance of the sensitivities and specificities for the two tests. Therefore, the maximum Cov_S is [min (S1, S2) – S1 × S2] and the maximum Cov_C is [min (C1, C2) – C1 × C2], based on a previous proof (21). The meta-analysis models above can be used to determine the magnitude of conditional dependence (i.e., a proportion

Table 2 Probabilities of diagnostic test results by true disease status

Test result	Disease-positive group (D+)	Disease-negative group (D−)
T1+ and T2+	Pi × (S1 × S2 + Cov_S)	(1 − Pi) × ((1 − C1) × (1-C2) + Cov_C)
T1+ and T2−	Pi × (S1 × (1 − S2) − Cov_S)	(1 − Pi) × ((1 − C1) × C2 − Cov_C)
T1− and T2+	Pi × ((1 − S1) × S2 − Cov_S)	(1 − Pi) × (C1 × (1 − C2) − Cov_C)
T1− and T2−	Pi × ((1 − S1) × (1 − S2) + Cov_S)	(1 − Pi) × (C1 × C2 + Cov_C)

D, dichotomous latent variable of the true disease status; D+, people with disease; D−, people without disease; T1, index test; T2, reference standard(s); S1, sensitivity of the index test; S2, sensitivity of the reference standard; C1: specificity of the index test; C2, specificity of the reference standard; Cov_S, covariance between sensitivities; Cov_C, covariance between specificities; Pi, disease prevalence. Note: Setting Cov_S = 0 and Cov_C = 0 will imply the tests are conditionally independent.

of the maximum Cov_S and Cov_C) based on the minimal DIC.

In an economic analysis, we may compare combined tests with a single test, such as the reference standard alone. When the Cov_S and Cov_C are ignored, the cost-effectiveness results of combined tests may be overestimated (22). Two commonly used criteria to define a composite decision rule based on two tests are the conjunctive positivity criterion (composite test result is positive only if both tests are positive) or the disjunctive positivity criterion (composite test result is positive when either test is positive) (23). If we apply the conjunctive positivity criterion, the composite test gains specificity but loses sensitivity compared with either test alone, whereas if we apply the disjunctive positivity criterion, the composite test gains sensitivity but loses specificity. The composite test strategy may be advantageous in specific situations, such as when high test sensitivity is preferred to maximize case detection (23). Based on how patients are managed after test results, we can further model health outcomes that reflect the clinical utility of the test.

Defining the true disease prevalence: cases of sex chromosome aneuploidy and trisomy 21

Disease prevalence is often one of the key parameters to determine the cost-effectiveness of a test (24). However, in many cases the true prevalence of a genetic condition is unknown. For example, patients with negative screening test results (including false negatives) are often not investigated further. When the observed prevalence may not reflect the true prevalence, researchers should estimate the expected true prevalence under some assumptions. For example, the expected prevalence of the sex chromosome aneuploidy XXY syndrome (47,XXY) from the literature is as high as 15.3 per 10,000 male fetuses (25), while the

observed prevalence is 1.94 per 10,000 male fetuses from registry data (26). Because phenotypes may vary widely for sex chromosome aneuploidies, cases are likely to be underdiagnosed or underreported. In addition, people with sex chromosome aneuploidies may be identified gradually as they age and not immediately at the time of birth, so additional cases may not be fully captured in registry data. The expected prevalence can be thought of as the "true" prevalence based on screening and subsequent confirmatory diagnostic testing (25).

We present an example of estimating the true prevalence of trisomy 21 (Down syndrome) among fetuses at 12 weeks of pregnancy (the typical time that prenatal screening for trisomy 21 occurs in Canada). The prevalence of trisomy 21 from a population of live births may differ from that at 12 weeks of pregnancy because of the spontaneous loss of pregnancies affected with trisomy 21 (27) or voluntary termination of pregnancy (28). Thus, when estimating the prevalence of trisomy 21 in viable fetuses at 12 weeks of pregnancy (first trimester) to evaluate the cost-effectiveness of prenatal screening, researchers need to adjust the observed live birth prevalence for spontaneous pregnancy loss (29). Live birth prevalence (excluding voluntary termination of pregnancy due to prenatal detection) is approximately equal to the prevalence of viable fetuses at a given time minus the spontaneous pregnancy loss at the same time point (4):

$$P_{Birth} = P_{12w} \times \left(1 - L_{Spon}\right) \tag{1}$$

where P_{Birth} is the live birth prevalence of a chromosomal anomaly in the absence of a prenatal diagnosis and voluntary termination of pregnancy, P_{12W} is the prevalence of a chromosomal anomaly in viable fetuses at 12 weeks of pregnancy, and L_{Spon} is the spontaneous pregnancy loss

rate from 12 weeks (first trimester) to term for a given chromosomal anomaly.

$$\text{Then, } P_{12w} = P_{Birth} \div \left(1 - L_{Spon}\right) \qquad [2]$$

The live birth prevalence of trisomy 21 in the absence of prenatal screening and voluntary termination of pregnancy was 8.8 per 10,000 births for women aged 30 years old in the UK (P_{Birth}=0.00088) and the spontaneous pregnancy loss between 12 weeks (first trimester) and term was 43% (L_{Spon}=0.43) (27,29). Then, the prevalence of trisomy 21 at 12 weeks of pregnancy (i.e., the time of prenatal screening) is approximately 0.0015 (P_{12w}=0.0015), assuming that the risk of spontaneous pregnancy loss is low for fetuses without chromosomal anomalies. It is important to accurately estimate prevalence in cost-effectiveness analyses, and if possible, approaches like the one above should be used. In situations where prevalence remains uncertain, sensitivity analyses should be used to explore the impact of this uncertainty on results.

Health outcomes and time horizons used in the economic modelling of genetic tests

The effectiveness of a health technology and the time horizon over which its benefits and adverse effects are assessed are two closely related elements. Often evidence on the long-term effectiveness of a health technology is scarce. Consequently, researchers must make assumptions about the long-term health outcomes (often not studied) to model lifetime cost-effectiveness. Economic evaluation guidelines often recommend using quality-adjusted life-years (QALYs) as the primary measure of effectiveness along with long-term time horizons (30-33). According to US guidelines, the time horizon of a cost-effectiveness analysis should be as long as possible to incorporate the differences in the intended and unintended consequences of the compared alternatives (32). However, long-term time horizons for genetic tests may not always be appropriate or feasible. For instance, Edlin *et al.* indicate that a shorter time horizon for health technologies is acceptable for decision-makers when the costs and benefits are incurred only over a short-term period, or when the long-term evidence is limited such that extrapolation would lead to an unreliable decision (33). A recent simulation study demonstrated that when longer time horizons were applied in cases of weak evidence, there was a substantial increase in bias that led to an overestimation of beneficial health outcomes (e.g., life-years or QALYs) (34).

Challenges associated with the long-term economic modelling of genetic tests are typically related to the values assigned to modelled health states, which includes assumptions regarding utilities (preference weights), other outcomes of people affected by the genetic test, and the potential positive and negative short- and long-term consequences of secondary findings for the person or their family (32). Thus, the following potential methodological issues require careful consideration:

❖ *Prognostic value of the genetic test:* if the genetic test has prognostic value, a lifetime cost-effectiveness model may be warranted for the reference case analysis. However, since long-term economic models need to include all relevant outcomes, issues may arise with the availability and quality of model inputs related to the short- and long-term effectiveness of the genetic test and assumptions regarding the treatment and management choices prompted by the test results. All relevant assumptions should be transparent, justified, and tested in sensitivity analyses. Guidelines also recommend exploring the influence of the time horizon in sensitivity analysis (32). Noninvasive fetal RhD blood group genotyping for identifying RhD-blood type incompatibility in managing RhD-negative pregnancies without existing antibodies (i.e., nonalloimmunized RhD-negative pregnancies) is an example of a genetic test that may change only the short-term course of clinical care for this specific population (35). Incompatibility occurs when the fetus's blood type is RhD-positive and the mother's is RhD-negative. This genetic test is a type of cell-free fetal DNA test and is performed at an early stage of pregnancy to determine incompatibility in the RhD blood group between the fetus and the mother. If incompatibility is present, the mother receives a treatment (Rh immunoglobulin injection) during the pregnancy to prevent anti-D antibodies from developing and attacking future incompatible fetuses (35). This technology is less likely to impact clinical outcomes and QALYs substantially after the pregnancy; thus, a long-term time horizon for this subpopulation of nonalloimmunized RhD-negative women may result in overestimation of the benefit and a biased (favourable) estimate of the incremental cost-effectiveness ratio (ICER). Lastly, short- or long-term cost-effectiveness modelling of all relevant

health outcomes (e.g., probability of having a live baby) rather than QALYs alone may be of value to inform decision-makers.

❖ *Secondary findings of genetic testing*: secondary findings that are unrelated to the primary purpose of a genetic test may arise. Secondary findings are not associated with symptoms, but may have an impact on present or future health, as they indicate people with a specified genetic condition may be at higher risk of developing a disease in the future that is unrelated to the original indication for testing (1). The American College of Medical Genetics and Genomics recommends that in addition to the primary findings of a test, laboratories performing whole exome and whole genome sequencing include reports of certain secondary variants (1,36) for which earlier diagnosis (or risk detection) and earlier intervention can improve health outcomes. However, there is uncertainty around the economic consequences of secondary findings and their appropriate incorporation in economic analyses is challenging (37,38). This increases the complexity of economic evaluations of genetic tests. The interpretation of health benefits is challenging and weighing and joint modelling of the benefits associated with both primary and secondary findings is necessary. Furthermore, because secondary findings are not the primary outcome of interest, it may be challenging to define a proper comparator (e.g., traditional tests may be unable to capture secondary findings) (38). Modelling secondary findings of genetic tests is rarely considered due to limited evidence and the inability to propose reasonable assumptions about relevant health effects and costs accrued over the long term.

❖ *Availability and validity of utility data (valuing health preferences)*: for the most reliable estimate of incremental cost per QALY of a technology, health preferences measured by recommended elicitation methods need to accurately reflect the decision problem and the health states modelled (32). In economic evaluations of genetic tests, especially in children, there is often no source of utilities that fully matches the target population and the use of a proxy measure or adult utility is necessary. Furthermore, only a few multi-attribute generic preference measures [e.g., Health Utilities Index (HUI)] have been validated. As HTA researchers,

we need to understand and interpret if a change (difference) in the utility score over time is minimally important, clinically important, and important to patients. More importantly, the magnitude of change in QALYs in a model-based cost-effectiveness analysis should be carefully analyzed, given that a number of health states are combined to estimate a change in QALYs over the specified time horizon (32). Another issue is whose health preferences (utilities) should be used in cost-effectiveness analyses of genetic tests. Although most economic guidelines suggest the use of community preferences, target populations of these analyses often include children for whom patient preferences rather than community preferences would be more appropriate, but they are difficult to obtain due to age limitations (32,39). Also, in economic evaluations of genetic tests, it may be important to consider utilities associated with diagnostic testing; if possible, we need to account for preferences for false positive and false negative test results. For example, a RhD-negative mother with developed anti-D antibodies carrying a RhD-negative fetus (i.e., lack of RhD incompatibility) may be negatively impacted by concerns for fetal health and the need for intensive monitoring during pregnancy due to a false positive RhD result. This disutility could be jointly modelled with other disutilities associated with the adverse effects of various monitoring procedures over this misclassified "at risk" pregnancy. Lastly, methodological issues could arise if health preferences of both the mother and baby are considered throughout the pregnancy and followed into childhood and throughout life. Joint modelling of multiple effectiveness outcomes including the utilities of both the mother and child are challenging because of the inherent features of an ordinary Markov (state transition) approach as well as the specific challenges of ascertaining a child's (let alone an infant's) health state preferences. If a Markov model is used, we need to decide which or whose utilities (health preferences) and outcomes are the most important for the cost-effectiveness analysis. In this case, limitations related to the simplification of the model structure and the uncertainty in estimation of QALYs should be recognized. Otherwise, more complex agent-based

or dynamic cohort models could be considered to adequately address disease management at the time of diagnosis and any long-term health consequences for both the mother and the affected baby.

❖ *Inclusion of caregiver and familial spill-over effects of the genetic test:* many genetic conditions are associated with negative impacts on the quality of life and the physical health of family members or caregivers of an affected person, but data may be limited on how the genetic test affects the outcomes of individuals other than the patient in economic evaluation. Recognition of spill-over effects is suggested (32) and their inclusion in the analysis requires reliably measured data inputs and joint modelling of outcomes (e.g., the loss of utility for patients and caregivers affected by the disease). Also, we ought to consider time-dependent changes in the outcomes (e.g., utilities) due to dilution of the spill-over effects over time resulting from possible adaptations to the disease and circumstances. Furthermore, there are potential family spill-over effects arising from knowledge of a genetic risk that may introduce some benefit or harm to family members. When a person is positive for a pathogenic variant in an autosomal dominant inherited condition, their first-degree relatives (parents, siblings, children) and extended family members may be eligible for testing of the familial pathogenic variant (termed "cascade testing") (40). Cascade testing is an efficient way to track familial pathogenic variants, such as familial hypercholesterolemia (40). For relatives who have an inherited risk factor and are currently disease-free, disease prevention, close follow-up and monitoring may lead to health benefits and reduced health care use if the disease is prevented and costly to treat (41). At the same time, knowledge of an inherited risk factor may lead to anxiety. Researchers ought to consider incorporating these spill-over effects in economic analyses to gain a full understanding of a test's impact on health. However, when evaluating cascade testing, researchers also need to consider contextual factors such as legal issues (e.g., privacy regulations in the US do not allow patients' doctors to directly contact their relatives) and the accessibility of cascade testing (e.g., preventive genetic testing in family members may not be covered by some payers) (41).

Costing of genetic tests and the analytic perspectives used in economic modelling

Costing of genetic tests can be complex. In addition to the cost of the genetic test itself, there are often other associated cost components, such as sampling, laboratory preparation, bioinformatic analysis, data management and storage (e.g., linking genetic test results with other data from reference libraries of genetic results and administrative health datasets), and interpretation and reporting of genetic test results (42-44). Also, the teams that manage data from genetic tests may vary and can consist of a clinician, a molecular biologist, a genetic counsellor, and data scientists (2). Compared with other types of tests, clinical interpretation of genetic results can be much longer and more costly (42). For example, in addition to regular physician visits for testing, some genetic tests (e.g., NIPT for chromosomal anomalies) may also require pre-test and/ or post-test genetic counselling to discuss the detectable genetic conditions of interest, the test's detection limitations and its role in detecting other conditions, family or personal history of the conditions tested, or appropriate further testing options (4). Genetic counselling also incorporates ethical and legal components. In summary, the full cost associated with genetic tests are generally greater than the cost of the test alone.

Adding to the complexity, genetic tests may be performed in commercial, hospital, or community laboratories. For commercial tests, industry often sets a list price based on the cost of labour, infrastructure, equipment, predicted test volume, equipment maintenance, potential test transportation, and validation and certification of testing. However, price negotiations may be possible between public payers and manufacturers for commercial tests. For hospital-based tests, it can be challenging to estimate the cost incurred by the hospital's laboratory. Costing of hospital-based genetic tests must include the costs of disposables and laboratory operating costs. When we consider whether the infrastructure, equipment, and setup costs should be incorporated, we need to understand whether existing facilities will be used or new facilities need to be developed. In economic evaluations, we are generally only interested in opportunity costs incurred in the future. Thus, we may assume the costs of existing infrastructure and equipment have already been incurred and cannot be recovered.

These costs are sunk costs (i.e., retrospective costs) and are generally not considered by decision-makers. If an existing facility requires upgrades or a new facility needs to be developed, these new costs should be included. We also need to understand whether the hospital needs to hire additional employees for the genetic test and appropriately allocate these new staffing costs to the genetic test.

Lastly, guidelines suggest that an economic evaluation explores, assesses, and presents the outcomes from both the health care sector and societal perspectives as two separate reference case analyses (32). The societal perspective is particularly relevant for genetic tests since spill-over effects may often occur.

Conclusions

When conducting HTAs of genetic tests, researchers need to understand the features of the genetic test and select the appropriate methods. Standard literature review and economic HTA methods may require adaptation. The following considerations may be relevant when embarking on an HTA for genetic tests:

- ❖ Bayesian meta-analysis of diagnostic test accuracy is particularly useful for rare genetic conditions and in the absence of a perfect reference standard;
- ❖ When developing economic models of combined tests, researchers need to consider adjusting for conditional dependence between tests;
- ❖ When evaluating the cost-effectiveness of genetic tests, researchers should pursue the unobserved "true" prevalence of the genetic condition of interest;
- ❖ Long-term time horizons and QALYs may not always be suitable in cost-effectiveness analyses of genetic tests;
- ❖ In addition to the cost of the genetic test itself, there are often other cost components associated with testing that should be explored.

Acknowledgments

Footnote

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Disclaimer: The opinions expressed in this publication do not necessarily represent the opinions of Ontario Health. No endorsement is intended or should be inferred.

References

1. U.S. National Library of Medicine. What is genetic testing? [published 23 June 2020, accessed 25 June 2020]. Available online: https://ghr.nlm.nih.gov/primer/testing/genetictesting
2. Phillips KA, Deverka PA, Hooker GW, et al. Genetic Test Availability And Spending: Where Are We Now? Where Are We Going? Health Aff (Millwood) 2018;37:710-6.
3. Djalalov S, Musa Z, Mendelson M, et al. A review of economic evaluations of genetic testing services and interventions (2004-2009). Genet Med 2011;13:89-94.
4. Health Quality Ontario. Noninvasive prenatal testing for trisomies 21, 18, and 13, sex chromosome aneuploidies, and microdeletions: a health technology assessment. Ont Health Technol Assess Ser 2019;19:1-166.
5. Xie X, Wang M, Goh ES, et al. Noninvasive Prenatal Testing for Trisomies 21, 18, and 13, Sex Chromosome Aneuploidies, and Microdeletions in Average-Risk Pregnancies: A Cost-Effectiveness Analysis. J Obstet Gynaecol Can 2020;42:740-749.e12.
6. Bianchi DW, Parker RL, Wentworth J, et al. DNA sequencing versus standard prenatal aneuploidy screening.

N Engl J Med 2014;370:799-808.

7. del Mar Gil M, Quezada MS, Bregant B, et al. Cell-free DNA analysis for trisomy risk assessment in first-trimester twin pregnancies. Fetal Diagn Ther 2014;35:204-11.

8. Langlois S, Johnson J, Audibert F, et al. Comparison of first-tier cell-free DNA screening for common aneuploidies with conventional publically funded screening. Prenat Diagn 2017;37:1238-44.

9. Norton ME, Jacobsson B, Swamy GK, et al. Cell-free DNA analysis for noninvasive examination of trisomy. N Engl J Med 2015;372:1589-97.

10. Palomaki GE, Kloza EM, O'Brien BM, et al. The clinical utility of DNA-based screening for fetal aneuploidy by primary obstetrical care providers in the general pregnancy population. Genet Med 2017;19:778-86.

11. Quezada MS, Gil MM, Francisco C, et al. Screening for trisomies 21, 18 and 13 by cell-free DNA analysis of maternal blood at 10-11 weeks' gestation and the combined test at 11-13 weeks. Ultrasound Obstet Gynecol 2015;45:36-41.

12. Song Y, Liu C, Qi H, et al. Noninvasive prenatal testing of fetal aneuploidies by massively parallel sequencing in a prospective Chinese population. Prenat Diagn 2013;33:700-6.

13. Cochrane Methods Groups. Software for meta-analysis of DTA studies. [accessed 25 June 2020]. Available online: https://methods.cochrane.org/sdt/software-meta-analysis-dta-studies.

14. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol 2006;59:1331-2.

15. Verde PE. Meta-analysis of diagnostic test data: a bivariate Bayesian modeling approach. Stat Med 2010;29:3088-102.

16. Lunn DJ, Thomas A, Best N, et al. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000;10:325-37.

17. Zur RM, Roy LM, Ito S, et al. Thiopurine S-methyltransferase testing for averting drug toxicity: a meta-analysis of diagnostic test accuracy. Pharmacogenomics J 2016;16:305-11.

18. Ontario Health. Cell-Free Circulating Tumour DNA Blood Testing to Detect EGFR T790M Mutation in People With Advanced Non-Small Cell Lung Cancer: A Health Technology Assessment. Ont Health Technol Assess Ser 2020;20:1-176.

19. Dendukuri N, Schiller I, Joseph L, et al. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics

2012;68:1285-93.

20. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001;20:2865-84.

21. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 2001;57:158-67.

22. Novielli N, Cooper NJ, Sutton AJ. Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency? Value Health 2013;16:536-41.

23. Xie X, Sinclair A, Dendukuri N. Evaluating the accuracy and economic value of a new test in the absence of a perfect reference test. Res Synth Methods 2017;8:321-32.

24. Kasztura M, Richard A, Bempong NE, et al. Cost-effectiveness of precision medicine: a scoping review. Int J Public Health 2019;64:1261-71.

25. Bojesen A, Juul S, Gravholt CH. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. J Clin Endocrinol Metab 2003;88:622-6.

26. European Registry of Congenital Anomalies and Twins. Cases and prevalence (per 10,000 births) of anomalies in 2015 in the United Kingdom. [accessed 27 May 2018]. Available online: http://www.eurocat-network.eu/accessprevalencedata/prevalencetables

27. Morris JK, Wald NJ, Watt HC. Fetal loss in Down syndrome pregnancies. Prenat Diagn 1999;19:142-5.

28. Natoli JL, Ackerman DL, McDermott S, et al. Prenatal diagnosis of Down syndrome: a systematic review of termination rates (1995-2011). Prenat Diagn 2012;32:142-53.

29. Savva GM, Walker K, Morris JK. The maternal age-specific live birth prevalence of trisomies 13 and 18 compared to trisomy 21 (Down syndrome). Prenat Diagn 2010;30:57-64.

30. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies. [published March 2017, accessed 25 June 2020]. Available online: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf

31. National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance. 3rd ed. [Published 26 September 2012, accessed 25 June 2020]. Available online: https://www.nice.org.uk/process/pmg4/resources/methods-for-the-development-of-nice-public-health-guidance-third-edition-pdf-2007967445701

32. Neumann P, Sanders G, Russell L, et al. editors. Cost-Effectiveness in Health and Medicine 2nd Edition. New York: Oxford University Press, 2017.

33. Edlin R, McCabe C, Hulme C, et al. Cost Effectiveness Modelling for Health Technology Assessment: The Australasian Drug Information Service (ADIS), 2015.

34. Xie X, Yeung MW, Wang Z, et al. Comparison of the expected rewards between probabilistic and deterministic analyses in a Markov model. Expert Rev Pharmacoecon Outcomes Res 2020;20:169-75.

35. Ontario Health. Noninvasive Fetal RhD Blood Group Genotyping: A Health Technology Assessment (Draft). [published January 2020, accessed 25 June 2020]. Available online: https://www.hqontario.ca/Portals/0/Documents/evidence/open-comment/hta-noninvasive-fetal-rhd-blood-group-genotyping-2001.pdf

36. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med 2017;19:249-55.

37. Regier DA, Weymann D, Buchanan J, et al. Valuation of Health and Nonhealth Outcomes from Next-Generation Sequencing: Approaches, Challenges, and Solutions. Value Health 2018;21:1043-7.

38. Christensen KD, Phillips KA, Green RC, et al. Cost Analyses of Genomic Sequencing: Lessons Learned from the MedSeq Project. Value Health 2018;21:1054-61.

39. Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? Pharmacoeconomics 2011;29:641-52.

40. Knowles JW, Rader DJ, Khoury MJ. Cascade Screening for Familial Hypercholesterolemia and the Use of Genetic Testing. JAMA 2017;318:381-2.

41. Caswell-Jin JL, Zimmer AD, Stedden W, et al. Cascade Genetic Testing of Relatives for Hereditary Cancer Risk: Results of an Online Initiative. J Natl Cancer Inst 2019;111:95-8.

42. Schwarze K, Buchanan J, Fermont JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. Genet Med 2020;22:85-94.

43. Wordsworth S, Doble B, Payne K, et al. Using "Big Data" in the Cost-Effectiveness Analysis of Next-Generation Sequencing Technologies: Challenges and Potential Solutions. Value Health 2018;21:1048-53.

44. Jegathisawaran J, Tsiplova K, Hayeems R, et al. Determining accurate costs for genomic sequencing technologies-a necessary prerequisite. J Community Genet 2020;11:235-8.