INNOVATIONS in pharmacy

Volume 7 | Number 4 Article 11

11-11-2016

# Next Generation Sequencing Claims in Precision Medicine: Questions a Formulary Committee Should Ask

Paul C. Langley
*University of Minnesota*, langley@maimonresearch.com

# Next Generation Sequencing Claims in Precision Medicine: Questions a Formulary Committee Should Ask

*Paul C Langley PhD, Adjunct Professor, University of Minnesota*

**Abstract**

*The enthusiasm with which precision medicine has been embraced over the past 15 years has obscured the fact that the evidence base for biomarker-driven assessments, in particular for next generation sequencing (NGS), is limited. This applies both to the comparative performance of the various assessment tools as well as to the impact of biomarker driven decisions at the patient level. Where a genetic test is being evaluated there are five key questions a formulary committee should ask when assessing whether or not to recommend coverage and reimbursement for the test in target patient populations:*

*(i) has the test met required standards for analytic and clinical validity?*
*(ii) has the test been evaluated against competing tests for analytic and clinical validity?*
*(iii) have the test-based claims met standards for credibility, evaluation and replication?*
*(iv) has the test been accepted as part of the standard of care for patient management in the target disease state?*
*(v) has the introduction of the test improved outcomes, including survivorship, adverse events, quality of life and costs, in the targeted population?*

*The purpose of this commentary is twofold: first, to consider the appropriate evidentiary standards for the evaluation of a test and comparator tests; and, second, to identify questions that a formulary committee should address in submissions made for a test in health care systems. A critical issue is not only comparative claims for the test against the standard of care and comparator tests, but the assessment of test performance for the identified treatment pathways where mutations or variants are linked to recommendations for therapy options. Unless these issues are addressed it is unlikely that the promise of personalized medicine will be realized. The absence of an evidence base will deter both physicians and their patients from adopting NGS based recommendations.*

**Keywords**: Precision medicine, evidentiary standards, modeling, claims protocols, claims evaluation, replication

## INTRODUCTION

A major, if not the principal objection, to precision medicine interventions is the lack of a robust and comprehensive evidence base to support claims that the new paradigm will materially improve patient outcomes and, hopefully, reduce costs [1 2]. Faced with a dearth of evidence that claims for precision medicine interventions are credible, evaluable and replicable, health systems are understandably reluctant to embrace this paradigm.

In many respects, the lack of evidence to support claims in precision medicine and the reluctance to underwrite tests both for older style genomic targeting and the current flavor of the month, next generation sequencing (NGS), is little different from the lack of evidence for claims in traditional pharmaceutical products. More specifically: the lack of claims that are credible, the lack of claims that are evaluable and the lack of claims that can be replicated [3 4]. Formulary committees and other health care decision makers are asked to take on board such claims. These include modeled extrapolations from short-term phase 3 trials as well as the absurd lifetime cost-per-QALY models favored by groups in the US such as the Academy of Managed Care Pharmacy (AMCP) and the Institute for Clinical Effectiveness Research (ICER) and their progenitor the National Institute for Care and Health Excellence (NICE) reference case in the UK [5 6 7]. At the same time, journal editors seem wedded to support the proliferation of lifetime cost-per-QALY models [8 9 10]. Understandably, there is little evidence to suggest that these models actually influence formulary decisions in the US.

Central to the demand for claims that are credible, evaluable and replicable is the issue of clinical utility: the ability of a screening or diagnostic test to support diagnosis, treatment interventions, outcomes, disease management and prevent or ameliorate adverse health outcomes'[11 12]. In precision medicine the test platform is the treatment gatekeeper: the choice of intervention relies upon the accuracy of the test while c claims for clinical utility rest upon test performance. The National Institutes of Health – Department of Energy Task Force on Genetic Testing reporting in 1997 defined the clinical utility of a test as 'the balance of benefits to risks' [13]. While there is obviously a debate as to how broadly the benefits of testing are to be defined, there is agreement that

**Corresponding author**: Paul C Langley, PhD
Adjunct Professor
College of Pharmacy University of Minnesota
Email: langley@maimonresearch.com
Web:www.maimonresearch.net

the primary benefit is the ability of the test to inform clinical practice.

From a public health perspective, the evidence to support health impacts is a critical input to setting public priorities and evaluating clinical utility, with the caveat that these must be evidence based. There is no inherent utility in a test unless patients in targeted groups have access to that test and claims based on that test can be evaluated. The acceptance or otherwise of a test should not rest solely on claims for analytical or clinical validity. Rather the test outcomes should be evaluated against standards that apply to claims for pharmaceuticals: are the claims made for the test in clinical practice credible, evaluable and replicable? In other words, does evidence for the test (or more broadly, the test) meet the standards of normal science [14 15].

If the standards of normal science are applied then we might usefully point to a list of core questions that a formulary committee, health care decision or professional clinical group might ask before approving a particular test for reimbursement or inclusion in a treatment guideline. These are:

(i) has the test met required standards for analytic and clinical validity?
(ii) has the test been evaluated against competing tests for analytic and clinical validity?
(iii) have the test claims for outcomes and costs in target treating populations met standards for credibility, evaluation and replication?
(iv) has the test been accepted as part of the standard of care for patient management in the target disease state?
(v) has the introduction of the test improved outcomes, including survivorship, adverse events, quality of life and costs, in the targeted population compared to the standard of care

There is no reason why a formulary committee of other health decision maker should assume that making a NGS test available, even if restricted to target population groups within disease states, will necessarily improve outcomes. After all, there are questions of the uptake of tests ordered by physicians, how a physician implements the test therapy recommendations, adherence and persistent with therapy choices (which may involve complex polytherapy) and the timeframe over which the outcomes for treated patients are monitored.

Once rates of uptake and implementation, together with adherence and persistence with therapy are factored into the assessment, the result from a well- designed, protocol driven observational study may conclude that the outcomes for the target group are inferior to those with the standard of care. Claims that the outcomes will inevitably be better because they address the issue of tumor heterogeneity may simply evaporate in clinical practice.

The onus is on the test developer or vendor to ensure that there is feedback to health systems, treating physicians and patients that it is worth investing in an NGS test; and, possibly more to the point, that investing in their test is more cost-effective than an investment in any one of a number of competing NGS tests. There are two reasons for this: (i) tests may differ in their platform, process and structure and (ii) there is no necessary expectation that different tests will generate similar recommendations for matching pharmaceutical compounds to genetic markers or that the same genetic markers will be identified. Claims for test performance and outcomes from their application must, therefore, be comparative.

Care must be taken, however, in establishing the basis for NGS test comparison (or a comparison against the standard of care). Irrespective of whether or not the assessment is through a randomized clinical trial (RCT)(e.g., an umbrella trial – see below) or through a prospective, protocol driven observational study, the focus is on the response by patients to the recommended therapies driven by the NGS platform. Matching therapies to mutations may result in a recommendation for a monotherapy matched to a single mutation (or cluster of mutations) or more complex polytherapy choices matched to a different cluster of mutations.

Assessments, therefore, should be in terms of the various therapy choices not just on the expressed mutations. One approach, given the inherent heterogeneity of mutations within tumor types is to focus the assessment on the most frequently reported mutation clusters, defined at the level of the individual patient enrolled in the study. The study may, for example, recruit patients in the top 5 or 10 most frequently reported mutation clusters with their associated therapies. These clusters would define the treatment arms within an RCT design or the selection of patients to be tracked in an observational study. The advantage of the observational study is that the starting point could be physicians who request a test, following through on those physicians who implement (or attempt to implement) the recommendations. Needless to say, both RCTs and observational assessments would have to be disease specific, focusing on target populations within the care pathway. After all, claims for one NGS test platform in one target population is no guarantee that a similar response would be expected in in another disease state.

Where a modeled claim accompanies a formulary submission for a NGS test, the model should attempt to capture

treatment pathways defined in terms of the standard(s) of care and the most frequently reported mutation and hence therapy clusters. Unfortunately, with the limited data yet available on therapy cluster outcomes from either RCTs or observational studies this is difficult to accomplish. We have yet to see the expected wave of extrapolated lifetime cost-per-QALY (quality adjusted life year) precision medicine decision models to support claims from competing NGS platforms. Models that claim that an 'acceptable' accuracy response threshold from a NGS test demonstrates that the sponsor's test will, with appropriate corrections for structural and parameter uncertainty, together with a judicial choice of Markov framework, lead to increased survival, less toxicity, improved quality of life and lower costs in a target population.

Presenting non-evaluable modeled claims that the up-front cost of the genomic test and potentially subsequent tests to check the course of the disease and patient response should be seen as sufficient justification for investing in an NGS test are unacceptable. Attempting to justify the cost-effectiveness of an NGS test through modeling, even if the model generates cost-per-QALY estimates appropriately falling below the magic $50,000 threshold is also not acceptable. Just as in the case of drug products, the claims made will either lack credibility (e.g., cost-per-QALY saved over the lifetime of the patient cohort), will be difficult if not impossible to evaluate with existing data (QALY claims again) and, by extension, be impossible to replicate. In order to forestall such an eventuality, it is important that health care decision makers to set standards for claims credibility, evaluation and replications for NGS tests. The losers in these scenarios will be health systems. They will be asked to support reimbursement for the test while treating physicians will be reluctant to order or implement test claims given the limited evidence for outcomes and costs. The fact that treatment guidelines may recommend genomic assessments at particular disease stages is no guarantee that they will be followed through or their recommendations for drug repositioning acted on. Tests may be ordered and the results put to one side.

The absence of credible, evaluable and replicable claims for drug products in formulary submissions was the genesis behind the recently published proposed guidelines for formulary evaluation developed by the Program in Social and Administrative Pharmacy at the University of Minnesota [16]. The proposed guidelines, while not rejecting modeled claims for clinical comparisons and cost-effectiveness, ask that the claims submitted meet the standards of normal science. To support this, manufacturers making a submission were asked to submit a protocol detailing how the claims were to be assessed and reported back to the formulary committee in a meaningful timeframe. At the same time, even if the manufacturer could claim that it had undertaken such an exercise with another health system and could produce the results as, hopefully, a peer reviewed publication, the option was still open for the health system to request the manufacturer to support a further claims assessment study.

Replication of test claims is, in many ways, the critical issue. Apart from the obvious point that the process of conjecture and refutation underpins our notion of scientific progress (and has done for last least last 350 years – or earlier if we accept the contributions of Francis Bacon [17]), there is a long standing concern with the absence of studies that are directed to replicating claims from clinical trials and, where such efforts have been undertaken, the limited evidence for claims replication. In the case of tests, formulary committee should be insistent that, in approving and reimbursing a test, evidence is presented that the molecular targets identified, their distribution and claims for clinical impact can be replicated in target populations.

**COMMENTARY OUTLINE**

The first part of this commentary addresses the issue of current evidentiary standards for evaluating NGS platforms and, more generally, diagnostic tests. Standards proposed by NICE under the Diagnostic Assessment Programme (DAP) are reviewed. These standards are of interest, not only because they point to the importance of identifying the place of a test in therapy for a target population but they link claims for test accuracy to establishing a modeled cost-per-QALY case for the test. At the same time, it is important to address the critical importance (which NICE emphasizes) of ensuring that evidence presented is robust and meets accepted standards in master protocol trial design, and that the test have been approved by responsible agencies.

As a major theme in this commentary is on the need to evaluate tests that are specific to target populations, the commentary then addresses the importance of identifying the characteristics of the target population and the expected place of a test in the treatment pathway. It is argued that it is unwise to expect blanket approval for an NGS test across target populations in disease areas. Rather, health care decision makers should insist on robust evidence from both clinical trials and observational studies to demonstrate that the test has claims relevant to that group and the stage of disease.

The first step in developing a robust case for any test is to ensure that the test meets standards for analytical and clinical validity. These issues are reviewed from the perspective of both the Center for Medicare and Medicaid Services (CMS) and the Food and Drug Administration (FDA). The issue is whether or not platform NGS tests should be seen as medical devices. If they are (as seems likely) the

question is what standards and possible monitoring will the FDA require to approve these tests?

In addition to possible standards for individual NGS test approval, further issues that are addressed are (i) whether or not NGS test platforms should be compared in terms of both their accuracy and the prognostic claims made for linking mutations to drug products and (ii) the standards that would be expected not only for integrating NGS test recommendations in treatment guidelines. It would be possibly unwise, given issues of efficacy and safety, for treatment guidelines to indicate that any NGS test may be appropriate at a particular point in a treatment pathway. Professional groups may wish to nominate specific NGS platforms with a recognized evidence-based contribution to therapy choices alongside the standard of care. Professional groups should indicate why a particular platform is recommended along with details on the platform structure and standard operating procedures in place for curating and updating the platform.

The commentary details recommended evidentiary standards for both clinical trials and observational studies that support claims for test performance in target populations. These would support modeled claims for the test. Where modeled claims and claims from trials and observational studies are presented these should apply (i) to the overall test performance and (ii) the proposed treatment pathways defined by the recommended monotherapy and polytherapy therapy choices linked to observed mutation clusters.

The commentary concludes by detailing (i) a proposed submission format for an NGS test and (ii) a list of 40 questions that a formulary committee should consider in evaluating the submission. As emphasized above, the claims made should be credible, evaluable and replicable. The submission should be accompanied by a protocol to describe how the prognostic claims from the test might be evaluated. The protocol should take into account mutation heterogeneity within, for example, tumors and whether the test should be re-administered given tumor progression. The protocol should describe how it is proposed to provide feedback to the health system as well as feedback on enhancements or modifications to the NGS platform.

**THE NICE DIAGNOSTIC ASSESSMENT PROGRAMME**

A useful reference point in considering the evidence standards for formulary submissions to support approval and reimbursement of a test is the DAP introduced by NICE in the UK in 2011[18]. Although the program process and conclusions regarding clinical and cost-effectiveness are focused on the assessment and implementation of the test within the National Health Service (NHS) the DAP is relevant for manufacturers in the US. Manufacturers may on the one hand be considering the UK as a potential market will have to meet DAP standards while on the other hand there may be groups in the US (e.g., the ICER) who may take the opportunity to apply their NICE-derived willingness-to-pay methodology to make recommendations for the cost-effectiveness of competing tests.

Key features of the NICE DAP are:

- **Developing the scope of the assessment**
  - Identify the target patient population, capturing the aetiology of the disease, disease stage, grade or severity, factors that may impact the accuracy of the test, as well as these benefits or risks of treatment such as comorbidities and age, gender and ethnicity
  - Describe of the test, its place in therapy and the setting for the test
  - Detail expected outcomes, costs and time horizon for the analysis within the care pathway
- **Evidence assessment and evaluation**
  - Identify and synthesize the evidence for diagnostic test accuracy utilizing systematic reviews, assessments of variability in results and potential biases, meta-analyses for sensitivity and specificity, likelihood ratios and predictive values, odds ratios, summary receiver operating characteristic (ROC) analyses and hierarchical models
  - Identify and synthesize evidence for health outcomes to include evidence for cost-effectiveness focusing on costs and resource use, duration and quality of life
- **Modeling clinical outcomes and cost-effectiveness**
  - Simplified analyses: where a test is superior to comparators in its analytical validity and no worse in direct side effects, possibly qualified by test cost, more complex assessment may be unnecessary
  - Assessment structure: if final data on test outcomes in target population limited it may be necessary to combine evidence from care pathway to link diagnosis, treatment and final outcomes
  - Test accuracy: summarize analytical validity assessments
  - Diagnostic process: test and potential test sequences
  - Reference case: application of modeled cost-utility analysis over relevant time horizon
  - Outcomes: QALYs
  - Evidence on resource use and costs

- Discounting costs and benefits
- Characterize bias and uncertainty in model structure and data inputs
- Application and selection of patient sub-groups

When a submission is requested by NICE, it is typically referred to an external assessment group who develop a diagnostics assessment report for review by stakeholders (defined in the assessment scope). A final review and report is prepared by NICE and a diagnostic guidance drafted. The draft guidance is reviewed, issues resolved and then published by NICE for the NHS.

While the NICE process is a useful benchmark, it is unlikely that a health system in the US (other than a government agency) would have the resources to undertake such an extensive review. Also, given the lack of interest in QALY measures, it is unlikely that a reference case modeling would resonate with health system decision makers. The questions proposed here are considered more in line with what is feasible in the US environment.

**EVIDENTIARY STANDARDS FOR A TEST SUBMISSION**

The evidence to support claims for analytical validity, clinical validity and outcomes must conform to accepted standards. If these standards are not met, the test submission should be returned to the manufacturer. Recommended standards would include reporting:

**(a) Diagnostic Accuracy and Quality**

Diagnostic Accuracy: All studies referenced to support claims for diagnostic accuracy for the target test and comparators should be evaluated against the STARD 2015 (Standards for Reporting of Diagnostic Accuracy Studies) statement [19]. Each study reported should be scored.

Quality Assessment and Bias: all studies referenced to support claims for diagnostic accuracy should also be appraised for quality and bias against the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies) [20]. Each study reported should be scored.

**(b) Systematic Reviews**

Whenever a systematic review is requested to support the test and comparator evidence base it must conform to the PISMA-P 2015 standards[21]. Apart from the usual databases (e.g. PubMed), reference should also be made for reviews focused on the CDC Public Health Genomics Knowledge Base [22].

**(c) Reporting Randomized Trials**

Reporting of results from randomized clinical trials of test performance should conform to the Consolidated Standards of Reporting Trials (CONSORT) [23] . This is a standard format for reporting on trial organization, analysis and interpretation. The CONSORT Statement comprises a 25-item check list and flow diagram to record the progress of patients through the trial.

**(d) Evidence Hierarchy**

Claims for the efficacy or effectiveness of tests in clinical practice must be founded on high quality and bias-free evidence. Where a submission has undertaken a systematic review or relies upon individual studies to support credible, evaluable and replicable claims the evidence presented should be assessed against the standards established within the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working groups. The GRADE framework has superseded earlier proposals for the ranking of evidence (which typically ranks from randomized trials through to observational studies and anecdotal, key opinion leader evidence) to a more flexible evidence hierarchy addressing the quality of evidence for individual outcomes. Specifically: bias, inconsistency, indirectness, imprecision and publication bias [24].

The GRADE framework is intended to apply to meta-analyses from systematic reviews but can be applied to individual studies or non-quantitative syntheses. The essence of the GRADE approach is that, within each hierarchy level, it allows the downgrading or upgrading of evidence. Downgrading, for example in the case of randomized clinical trials, occurs if there is a risk of bias, inconsistency, indirectness, imprecision and publication bias. Upgrading, for example in the case of non-randomized studies can occur if there is a large magnitude of effect, evidence of a dose response effect and if all plausible confounding factors have been taken into account. The application of the GRADE framework a 4-level quality rating hierarchy. This is detailed in the Cochrane Collaboration handbook [25].

1. *High Quality Rating*: Randomized trials; or double-upgraded observational studies
2. *Moderate Quality Rating*: Randomized trials; or upgraded observational studies

3. *Low quality rating*: Double-downgraded randomized trials; or observational studies
4. *Very low quality rating*: Triple-downgraded randomized trials; or downgraded observational studies; or case series/case reports.

The GRADE evidence approach has figured largely in the Agency for Healthcare Research and Quality *Methods Guide for Comparative Effectiveness Research* to support the Evidence-base Practice Center (EPC) Program [26] . The EPC framework grades the strength of evidence from RCTs as well as observational studies in a systematic review through assessing specific domains: study limitations, directness, consistency, precision and reporting bias. Potential additional domains are: dose-response association, plausible confounding for observed effect and strength of association. Scoring these domains yields four strength of evidence grades:

1. *High*: The reviewers are very confident that the estimate of effect lies close to the true effect
2. *Moderate*: The reviewers are moderately confident that the estimate of effect lies close to the true effect
3. *Low*: The reviewers have limited confidence that the estimate of effect lies close to the true effect
4. *Insufficient*: The reviewers have no evidence, they are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome

**(e) NCCN Evidence Hierarchy**

Given the focus on cancer in claims made for NGS platforms, the evidence standards applied by the National Comprehensive Cancer Network (NCCN) to support guidelines recommendations should be noted [27]. Each recommendation is evaluated by a panel and the strength of evidence (potentially for a specific NGS test) is identified using the NCCN Categories of Evidence and Consensus. While the goal is to base recommendations on high quality evidence from controlled clinical trials, meta-analyses, studies for combination therapies, treatment sequencing and head-to-head studies are often not available. Panels may then use lower level evidence. The category recommendations are:

Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate;

Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate;

Category 2B: Based upon lower level evidence, there in NCCN consensus that the intervention is appropriate; and

Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

All recommendations are 2A unless otherwise noted

The importance of evidence standards in precision medicine claims should not be understated. If the application of NGS (and presumably the generations after that) is to be accepted it is crucial that we have publicly accessible, comprehensive, structured, transparent and bias-free evidence that support both claims development from analytical and clinical validations but subsequent clinical utility claims; claims that are credible, evaluable and replicable. Comparative and systematic reviews are central to this process: comparative because the choice of a test and a belief in the validity of the test is critical to its acceptance by treating physicians as an integral part of the process of care, and systematic because we require the accumulation of evidence from well conducted and credible effectiveness studies to support patient involvement in treatment decisions.

If the hope is that a specific NGS platform be recognized and included as an element in a treatment guideline, then the NCCN standards should be built into any research program to support that platform.

It is essential to recognize and accommodate the dynamic nature of NGS test development. Test platforms are typically in a process of continual upgrading as new evidence is presented for linkages, pathways and targeted therapies. To support this process manufacturers should be asked (possibly as part of annual or bi-annual test reviews) to confirm test performance as part of an ongoing process of re-evaluation and replication in target populations.

***Master Protocol Trials***

The standards proposed by GRADE and EPC for the evaluation individual trials and systematic reviews apply not only to the phase 2/3 classical randomized controlled trials in drug development and post market entry phase 4 effectiveness

trials, but apply with equal force to the trials directed to supporting evidence claims in precision medicine. These so-called master protocol trials include basket trials, umbrella trials, hybrid trials and platform trials [28] . These trial designs have been prompted by the heterogeneity that exists between patients within, for example, a given tumor type and also the heterogeneity that exists within an individual with tumor evolution and metastasis.

Master protocols are an essential part of the process of translating molecular evaluation and genomic target matching into clinically relevant outcomes [29]. These targets include both on-label and off-label prescription of approved compounds as well approval of investigational compounds developed as a result of NGS tests. While off-label matching is suggestive, *ad hoc* claims from clinical practice and systematic reviews of the literature are no substitute for trial based assessments.

A credible and replicable evidence base driven by master protocols is also a potentially significant contributor to identifying the appropriate target mutation pathway. A test potentially identifies relevant multiple molecular aberrations by disease stage. This drives the choice faced by the treating physician in options for actionable single drug or multiple drug combinations. Faced with a number of potentially competing therapies that match to a mutation, prior assessment may also suggest an optimal treatment to support prognostic claims. Indeed, in the early stages of NGS platform modelling, the most important contribution may be to suggesting a hierarchy of target drug and drug combinations for master protocol trials.

Understandably, in the absence of a well-documented, high quality and coherent evidence base, physicians and patients will be reluctant implement NGS recommendations. Prognostic claims that match mutation to therapy are no different in principal from classical drug claims. They should not be taken at face value. One of roles of a health system is to act as gatekeeper to ensure that these claims merit attention in clinical practice. This applies not only to the NGS test overall but to the individual prognostic claims. After all, support for one prognostic claim may be obscured by indifferent results for other prognostic claims within the proposed set of mutation and therapy linkages. Capturing heterogeneity is no guarantee of test performance for the individual therapy pathways.

The common element in master protocol trials is to utilize a centralized screening test platform with a common protocol format for biomarker-driven sub-studies within the target population. This allows multiple parallel drug studies that replace the traditional (time consuming and wasteful) independent classical single target phase 1 through 3 structures where a strong sub-group effect can be masked by a weak overall effect. In classical trial implementation *post facto* identification of sub-groups usually generates suggestions rather than hypotheses as the sub-sample is typically underpowered. In master protocol designs the sub-studies can be flexible in their design to include multi-phase designs to match experimental drugs or the standard of care. The ultimate objective is to confirm the predictive role of the biomarker.

***Basket Trial Designs***

A basket trial is a master protocol where patient eligibility is defined by the presence of a particular biomarker or alteration rather than, for example, a particular cancer type. The claim is that response can be predicted from a tumor's molecular characteristics matched to a target therapy (or therapies) independent of the tumor histology. As the standard of care will vary across tumor types, it is unusual to find a standard of care control arm.

The principal objection to the basket trial design, at least from the perspective of a formulary committee, is that it says nothing about the overall benefits of introducing that test into clinical practice within single types or classes of cancer for target populations by stage of disease. From the perspective of overall clinical benefit and the potential cost-effectiveness of the target therapy the evidence is incomplete and may, in fact, be misleading in failing to capture potential interactions between target therapies in patient groups.

***Umbrella Trial Designs***

In contrast to the basket trial, an umbrella trial focuses on a single type or class of cancers. The tumor is centrally screened and patients are assigned to molecularly defined sub-groups where they can be randomized and matched against the standard of care pathway. The presence of a single test, the randomization to target therapies and the standard of care for sub-trials, together with the ability to establish response, potential interactions, the impact of comorbidities and demographic factors within a guideline recommended treatment pathway for a specific cancer makes the umbrella trial attractive to formulary committees. Again, however, the formulary committee will need to be assured that the test is accurate and has been evaluated against comparator tests in agree target populations.

***Platform and Hybrid Trial Designs***

Platform designs are a randomized design with a common control arm and experimental arms defined by molecular type. Management of the platform allows therapies to be added or dropped through adaptive randomization to reduce

overall sample size and improve efficiency. Again, unless the platform design captures the range of biomarkers and manages therapies against those biomarkers, it is difficult to see the attraction of this design to formulary committees. A hybrid trial is one where there is a mix of umbrella and basket designs capturing combinations of histologies and molecular aberrations.

Although standards have yet to be agreed for the reporting and assessment of master protocol trials, from the perspective of test choice, the umbrella trial is of particular interest. However, the results of a single umbrella trial only reflect the ability of the test to assign patients to molecular sub-types. Claims for the performance of target therapies against the standard of care will depend on the performance of the test against comparator tests in the accuracy of patient assignment and the ability of the test to match potentially high response therapies to those molecular markers.

**APPROVAL STATUS OF COMPETING TESTS**

It would be assumed by the formulary committee that the tests submitted (proposed and comparator tests) have been approved, either by the FDA or CMS or, given the present uncertainty over the responsible agency given the moves by the FDA to label tests as medical devices, a successor agency.

The submission should detail whether the test has received regulatory or other approvals in the US or in overseas jurisdictions. Where approvals have been given the manufacturer should indicate whether the test: (i) has been approved only for specific targeted disease states; (ii) whether the material submitted to treating physicians has been approved by the regulatory agency (e.g., a restriction on reporting off-label variant matches); (iii) whether any restrictions or monitoring requirements have been required by a regulatory agency for toxicity or adverse events; and (iv) whether the test has been approved for inclusion in a professionally recognized treatment guideline (e.g., to meet NCCN recommendation standards).

**TARGET POPULATION FOR THE TEST**

As describe below in the context of the FDA reference standard for diagnostic tests, the target population for any NGS test assessment must be a target population that meets commonly accepted clinical and diagnostic criteria. If the target group is represented by the stage of disease, then recruitment to the target group for NGS evaluation and claims evaluation must conform to that definition. This allows both for comparisons between claims in that target population from competing tests as well as replication of claims made for that target population in other health jurisdictions. It is most unlikely (and inadvisable) for a health system to give the developer or vendor of an NGS test an 'open season' remit to market the test across target patient groups (or ad hoc patients) in different disease states.

As will be detailed below, the test has to be considered from the perspective of the anticipated benefits and costs that it confers on the target treating population within a specific disease state. Approval for a test to be reimbursed should be disease state specific. The approval should be linked to a management strategy where the test complements and is integral to initial diagnosis, disease staging and assessing relapse and resistance to the standard of care within that disease state. The criteria for introducing, interpreting and implementing a test and its recommendations should be detailed as part of the management strategy.

To date, the FDA has endorsed a number of companion diagnostics where the target population is well defined. However, in many cases the claims for the test are much wider, arguing for example that test is appropriate across a number of disease states (e.g., small population non-responsive cancers).

The promise of NGS tests is, of course, only a first step to the acceptance of a particular test type as an integral and reimbursed element in the standard of care for specific cancer types and other diseases. Claims, for example, for maximal inhibition of viability and proliferation in tumor endpoints, and increased apoptotic effect need to be translated to claims for survivorship, toxicity and quality of life.

Acceptance will also depend upon cost; the cost of the test, how often the test needs to be repeated over the course of the disease, the costs of the targeted drug therapies (which may change following future tests as a target tumor mutates) and the anticipated direct medical costs of supporting the patient over their lifetime. At the same time, the choice of test should be considered in comparative terms: if there are competing tests which claim a more robust evaluation and validation for target drug options their claims need to be set alongside those for the submitted test.

Finally, to add a further level of complexity, there is the question of comorbidities. Older patients, typically those with cancers, will typically present with one or more comorbidities. The questions then become, first, one of asking whether the test takes into account the presence of comorbidities and, second, whether their management qualifies the claims made for cost and survivorship for the cancer target.

The data set to describe the target population should include:

- Selection Criteria: Clinical and diagnostic criteria for target patient population

- Epidemiology: prevalence and incidence estimates for the target population defined by place in therapy or stage of disease defining initial test application
- Demographics: age, gender, ethnicity
- Genetics: distribution of variants
- Insurance status: Medicare, Medicaid, Veterans Administration, commercial
- Comorbidities: prevalence of comorbidities by type and, if appropriate, by the stage of comorbid disease
- Uptake: evidence for uptake or anticipated uptake of test (and comparator tests) to include tests sold in the previous two calendar years
- Implementation: evidence for implementation of test recommended or proposed drug regimen (or regimens)
- Guidelines: recommended treatment guidelines (e.g., NCCN)

**ANALYTICAL AND CLINICAL EVALUATIONS OF THE NGS TEST**

In presenting evidence for the technical merits of the NGS test it is important that the submission makes the distinction between evidence to support its analytical validity and evidence for its clinical validity. The former refers to the test's ability accurately and reliably to detect variants (e.g., mutations, copy number variations) when they are present while the latter refers to the ability of the test to associate the variant with the presence or absence of the target phenotype. Thus, for DNA tests analytic validity requires establishing the probability that the test will be positive when a particular sequence of analyte is present (analytic sensitivity) or the probability it will be negative when sequence is absent (analytical specificity). The accuracy and reliability of the test, through replication in patient samples, is critical.

Similarly, the assessment of clinical validity rests upon sensitivity and specificity, but with an assessment of positive and negative predictive values, with account taken of prevalence, potential heterogeneity and penetrance. Once again, the issue is one of accuracy: is the genetic variant (or variants) being analyzed related to the presence or risk of the target disease.

Quality standards for laboratory testing performed on specimens from humans for the purpose of diagnosis, prevention, treatment of disease or assessment of health are established under the Clinical Laboratory Improvement Amendments (CLIA) of 1988 (as amended in 1997 and 2012). The CLIA covers the certification and oversight of clinical laboratory testing through federal regulations titled 'Standards and Certification: Laboratory Requirements' (42 CFR 493) issued by the Centers for Medicare and Medicaid Services (CMS). The CMS, under CLIA, oversees laboratory processes, not the tests they develop. There is no requirement for a premarket review of a test or for evidence of clinical validity.

In October 2014 the Food and Drug Administration (FDA) published a draft guidance for the oversight of laboratory developed tests (LDTs). It was proposed that the FDA would oversee the quality of these tests alongside CMS. Given some confusion over the respective roles of the FDA and CMS with potential duplication an interagency taskforce was set up to collaborate on the oversight of LDTs. The position taken by the FDA is that its oversight will assure the tests are analytically and clinically valid. The taskforce has yet to report.

If the proposed FDA framework is accepted CMS through CLIA will ensure quality through its focus on laboratory operations and the testing process while the FDA will contribute to quality standards by enforcing compliance with the agency's quality systems regulation of the design and manufacture of, in this case NGS tests. The FDA view is that NGS tests are best considered as medical devices. Under section 201(h) of the Federal Food Drug and Cosmetic (FD&C) Act a product will be regulated as a medical device and is subject to premarketing and post-marketing regulatory controls if it is

- An instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part or accessory which is:
  - recognized in the official National Formulary, or United States Pharmacopoeia, or any supplement to them
  - intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment or prevention of disease, in man or other animals, or
  - intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes.

It is unclear how the Food and Drug Administration (FDA) will, if given jurisdiction, regulate NGS tests. At the present time there are an estimated 60,000 genetic testing products on the US market. Of these, the current estimate is that some 7,600 would be considered high risk by the FDA. While there is debate (and pushback from the industry) as to the statutory authority of the FDA to regulate laboratory developed tests as medical devices, the balance of the

evidence appears to favor the FDA having jurisdiction. In a presentation to Congress in late 2015, Jeffrey Schulen, Director, Center for Devices and Radiological Health at the FDA, argued that the FDA has jurisdiction over diagnostic tests under the flexible, risk-based framework established under the Medical Devices Amendments of 1976 (MDA) [30]. Under this amendment the FDA assigns *in vitro* diagnostic devices (IVDs) to one of three classes that correspond to the level of risk the IVD presents to patients and the public (Classes I through III). In Class III IVDs (where in vitro NGS tests are likely to be assigned), the IVD presents the highest level of risk and should be subject to premarket approval, post-market and other controls to ensure the tests can be used safely and effectively. The primary risk of is that of an undetected inaccurate test result resulting from both false negatives and false positives. The former could lead to unnecessary or the delay of necessary medical procedures – including a false positive flag for a repositioned or new drug regimen; the latter equally could lead to injury or death from the unchecked progression of disease. Both, as Schulen points out, could lead to unnecessary costs. Among the tests that the FDA considers high risk are those for companion diagnostics. Presumably, the concern for injury and death could apply equally well to the therapy options matched by the NGS platform to mutations or mutation clusters.

To ensure a reasonable assurance of safety and effectiveness, the benefits of a test will have to outweigh the risks it poses and that it will provide clinically significant results. For premarket approval there needs to be an independent demonstration of safety and effectiveness. The test has to achieve (i) a satisfactory analytical performance or validity in detecting pre-specified markers and (ii) a satisfactory clinical performance or validity demonstrating that the marker has a clinical significance in correlating with a disease or condition or with the ability to predict a therapeutic response to a drug.

Most tests under Class III would need 501(k) clearance. The evidence presented to support analytical and clinical validity should conform to standards in place or standards proposed by the Food and Drug Administration (FDA). While these standards (and those presently proposed as guidance) apply to medical devices, those making the submission (or developing an NGS test) should assume, in the absence of agreed standards and processes, that the test will be considered a medical device for the purpose of the health system assessors.

***Molecular Diagnostic Services Program (MolDx)***

In 2011 CMS contracted to establish the Molecular Diagnostic Services Program (MolDx) through Palmetto GBA [31]. The purpose of this program was to:

- establish a unique registration identification of molecular diagnostics tests to facilitate claims processing and track utilization
- to establish clinical utility expectations
- to complete technical assessments of published test data to determine clinical utility and coverage
- establish reimbursement for gene and gene components that meet Medicare covered service criteria

As of October 2016, MolDx had approved 152 tests for gene and gene components that met Medicare criteria for a covered service.

The clinical test evaluation process is in two parts (i) the evaluation of analytical and clinical validity and (ii) the evaluation of clinical utility [32]. To meet Medicare 'necessary and reasonable criteria' the test must meet both evaluation standards. Performance standards have been established by MolDx for analytical validity. These include analytical specifications for comprehensive genomic profiling [33], analytical specifications for qualitative tumor only somatic variant detection using circulating tumor DNA.

The fact that a test has met the standards for the medical necessity review and registration under the Molecular Diagnostic Services Program (MolDx) – and consequently received national coverage by Medicare – does not mean that those making the submission can put evidence for analytical and clinical validity to one side. Although it is not the policy of the MolDx to comment on failed applications, it is important to advise on whether the application for this test has been put to one side.

When evaluating claims for clinical utility The MolDx team reviews each clinical trial presented to support claims for clinical utility. The trial data, to demonstrate the strongest clinical utility trial as opposed to other trials are classified as:

(1) mCTD3A: A randomized prospectively controlled trial that directly demonstrates that a therapeutic intervention based on test results leads to statistically and clinically significant improvement in patient outcomes compared to currently accepted standard of care.

(2) mCTD3B: A prospective-retrospective trial using archived samples from a previously reported prospectively controlled trial to demonstrate that treatment based on a molecular test result in a specified patient population is associated with improved outcomes in a statistically and clinically significant manner versus a currently accepted standard of care.

(3) mCTD2A: A prospective observational study enrolling patients in a registry, treating according to a defined pathway using a molecular test as an integral part of a care plan and demonstrating statistically and clinically significant improvement in healthcare outcomes versus a currently accepted standard of care.
(4) mCTB2B: Retrospective data modelling using large data sets to demonstrate statistically and clinically significant improvement in healthcare outcomes when a given molecular test guides treatment versus a standard of care.
(5) mCTD1; Retrospective observational studies that do not stipulate treatment pathways or follow-up results based on results from the molecular test
(6) mCTD0: Preclinical studies where the data are from animal or in vivo experiments or related studies or trials

If the level of evidence presented is below mCTD2B then the application will be rejected; although it is not clear what happens if the evidence based presented includes a mix of studies or where pre-clinical studies are linked to accepted studies.

***Prospective FDA Standards for NGS Tests as a Medical Device***

In July 2016 the FDA issued a draft (i.e., non- binding) guidance for regulatory oversight of NGS sequencing *in* vitro diagnostics [34]. While draft guidance was directed to whole exome DNA sequencing or targeted DNA sequencing NGS-based tests for diagnosing germline diseases or other conditions arising from inherited or de novo germline variants, the standards required are indicative of those the FDA might propose for somatic diseases. The recommendations cover (i) the design, development and validation of NGS-based tests and (ii) FDA-recognized standards for regulatory oversight of those tests.

It should also be noted that the FDA already has in place standards for evaluating diagnostic tests [35] . This guidance is intended for the submission of premarket notification (510(k)) and premarket approval for diagnostic devices or tests where there two possible outcomes: positive or negative. These standards are important because they provide more detail on benchmarks for assessing diagnostic performance. The benchmarks are: (i) comparison to a reference standard defined as the 'best available method for establishing the presence or absence of the target condition'; and (ii) non-reference standards. The benchmark will determine which performance measures are appropriate.

Reference standards are those agreed and adopted by 'opinion or practice within the medical, laboratory or medical community'. This distinction is important in precision medicine as it points to the importance of a consensus opinion in defining target patient groups for reporting clinical trial results and for the assessment of prognostic claims from NGS tests. The FDA recommends that before proceeding they should be consulted to ensure the reference standard meets the agency needs.

Even so, these standards are a long way from standards for NGS tests. Meeting standards for analytical and clinical validity are only a necessary, not a sufficient condition for approval of an NGS test as a medical device. The issue which is probably most vexing is how to evaluate the merits of competing tests. If NGS platforms generate different outputs, then the FDA should possibly consider minimum response standards at both the aggregate and pathway level (in effect to weed out the 'duds' and tests which have a high proportion of 'duds'). It seems possibly unreasonable to put the burden on health care systems; so perhaps, if the FDA adopts a more *laissez faire* approach, the onus should be (as is proposed here) for test developers and vendors to make the comparative case for their product.

It is worth emphasizing the FDA perspective as it raises the questions that an assessment by health care decision makers should address as an NGS-based test for clinical use will typically include reagents, consumables, instruments and software. The presence of one or more of these will be dictated by the workflow and functioning of the test. It is important, therefore, to detail the steps in the test process. These may include: (i) specimen collection and acceptable types for processing and storage; (ii) DNA extraction; (iii) DNA processing and library preparation; (iv) the regions of the genome, including genes and variants, that are interrogated by the test; (v) generation of sequence reads and base calling; (v) variant classification/interpretation and (vi) preparation of test report.

At the same time, the test may involve manual variant interpretation. If this is the case then the technical description should include standard operating procedures (SOPs), decision algorithms and supplementary test components.

It is important that the submission, in particular for tests supporting next generation sequencing, includes a technical description of the predictive platform. Consider predictive simulation modeling, for example, where an *in silico* platform incorporates integrated networks of signaling and metabolic pathways through aggregating functional relationships between proteins and predicts mechanisms using drug combinations that interact to reduce viability, proliferation and other endpoints through the creation of patient avatars [36]. The platform needs to be described and should include, in

this instance, how the effects of drugs or drug combinations are selected and applied from a library of molecularly targeted drugs to short-list potential repurposed or novel drug candidates.

Information should also be provided on how the library is curated, aggregated and interpreted in evaluating linkages. Although there is no requirement here that the library or, as described by the FDA, a 'genetic variant database' should be publicly available, the standards proposed for recognition by the FDA of a genetic variant database as a source of valid scientific information to support claims for the analytical and clinical performance of a test should be met. Two standard operating procedures are critical:

- procedures for the curating, aggregating and interpretation of high quality valid scientific evidence to support claims for the analytical and clinical performance of the platform
- procedures for updating inputs and maintaining the stability and architecture of the platform content and processes

If the test manufacturer employs third parties to collect samples and undertake assessments, the role of these should be identified

***Possible FDA Requirements in Developing an NGS Test***

Design considerations in developing an NGS test (or reporting on a developed test) focus on the activities that the developer should have performed to identify the intended clinical use of the test and to design the test for that use. Developers should establish and justify minimum acceptable and target values for each performance metric appropriate to the indications for that use (e.g., define clinical need that is driving test development, justify required test features, specify specimen types, identify required metrics and threshold performance standards). All test components and methods should be identified to include sequencing platform and controls.

The developer should identify and report on test analytical performance characteristics for the predefined metrics to ensure the test successfully identifies, within statistical bounds, the presence or absence of a variant. At the same time, once the test is in clinical use it should be continuously monitored. The draft guidance suggests a set of performance metrics for (a) accuracy with thresholds for positive percent agreement, negative percent agreement, technical positive predictive value and the rate of 'no calls' or 'invalid' calls; (b) precision - reproducibility and repeatability – for variant and wild type calls; (c) limit of detection for each variant type; and (d) analytical specificity – interference, cross reactivity and cross-contamination.

The developer should also establish minimum acceptable thresholds for test run quality metrics. To determine whether a test run or variant call may need supplemental procedures to query further a variant call.

From the perspective of a formulary committee or health system evaluating the merits of an approved NGS test or competing tests, the draft guidance also suggests information be posted to the public domain (e.g., a manufacturer's website) to cover:

(i) Test performance
(ii) Test design
(iii) Test reports

***Information on Test Performance***

When reporting on test performance, the draft guidance recommends that the following information be posted in the public domain:

- Indications for use
  - Type(s) of sequence variations detected
  - Any test limitations (erg., variants the test cannot detect)
  - The fraction of an affected population for which the test is likely to provide relevant results
- Regions of the genome in which sequences meeting pre-performance specifications can be identified relevant results
- Types of variants the test will report
- For targeted panels list the gene(s) included in panel
- For whole exome sequencing (WES) based tests to describe clinically relevant regions of the exome and relevant coverage for those regions
- For summary performance
  - Results for test accuracy/precision
  - Results of reproducibility studies
  - Results for targeted panels
  - Results for WES
- Test failure
  - Probability of failure from performance data
  - Scenarios under which the test can fail

***Information on Test Design***

Again, this should be posted to the public domain:

- Specify test components
- Describe steps in test design
- Details on specimen type
- Minimum DNA yield and quality
- Method for sequencing DNA
- Level of multiplexing (if applicable)

- Specify all software components
- Detail databases used for analysis
- Criteria for annotation and filtering of variants

***Information for Test Reports***

Also published to the public domain:

- Relationship between reported variants and the clinical presentation of the patient
- Description of the genomic and chromosomal regions detected by the test
- Summary of test performance studies
- List of pathogenic or actionable variants (including those of unknown significance)
    - Variants to be reported using common nomenclature
    - Description of clinical evidence supporting the interpretation reported variants
    - Summary of genes related to the patient's phenotype (with reference to databases for variant interpretation)
    - Additional information that may be required
- Test limitations
- Risk mitigation procedures (if required

**TREATMENT PATHWAY AND AGGREGATE PERFORMANCE CLAIMS**

As described briefly in the introduction to this commentary, a key distinction in validating claims for the reimbursement and placement of an NGS test is between (i) the overall performance of the test against the standard(s) of care that characterize the proposed intervention point for the test in the treatment pathway for a specific disease and (ii) the contribution of the most frequently recommended therapy combinations that the test administrator reports for the individual platform assessments.

Competing claims for NGS tests will rest, therefore, on both aggregate performance and patient response at the therapy choice level. This distinction has a number of implications: (i) there is no certainty that different platforms will generate similar or the same therapy choice recommendations; (ii) adverse events and toxicity may differ between treatment pathways so that risk mitigation may be driven by pathway therapy choices; (iii) if drug-to-drug interactions are a concern, they are best explored at the pathway level; (iv) response to therapy may vary by therapy choice which may lead to qualifications of indication for use; (vi) resource utilization and cost claims may vary by therapy pathway so that the distribution of patients by therapy pathway may be a major, if not the principal, determinant of therapy costs; (vii) adherence and persistence with therapy may differ by pathway, suggesting possible pathway specific interventions to support therapy utilization; and (viii) the impact of comorbidities and their impact on the overall polytherapy burden may qualify response to therapy through adherence and persistence.

**COMPARATIVE NGS EVALUATIONS**

It is unlikely that a formulary committee or health system will only evaluate a single test. With the potential commercial benefits of test adoption, it is more than likely that decision makers will be faced with a number of competing tests. Given this, it is important that the information requested for each test is to a fixed standard that allows a comprehensive assessment of competing claims for clinical and analytical validity.

In the unlikely event that there has been a head-to-head comparison of the analytical and clinical validity of the proposed test against one or more actual or potential comparator tests, a submission should include:

(i) A summary statement of the clinical and analytical benefits of the proposed test over competing tests in the target population or for each of the populations indicated

(ii) A detailed justification for the place of the proposed test and competing tests in the care pathway for the target disease state population (or populations) together with the clinical criteria for implementing the test

(iii) A profile of the anticipated frequency with which the test and competing tests will be replicated to assess tumor progression and the clinical criteria for re-assessment

(iv) A summary of test performance for the test and competing tests

(v) A summary of test design for the test and competing tests

(vi) Evaluation reports for the test and competing tests

(vii) Procedures for reporting updating and monitoring of the test and competitors (if known) with implications for use and reporting of test results

(viii) Procedures for risk mitigation and reporting for the test and competing tests

Techniques for assessing the analytical validity of NGS tests need not be restricted to those data elements proposed in the FDA guidance. There are a range of possible tests that can supplement claims for sensitivity and specificity and prevalence corrected claims for positive and negative predictive value with the appropriate confidence intervals.

**INTEGRATION IN CARE PATHWAYS**

To date, there are only a few treatment guidelines that have included the option of genomic testing as an element in care pathway decisions. As claims for the introduction of NGS tests as an integral part of the process of care will likely run ahead of their introduction into treatment guideless, it is important that those proposing a role for NGS in disease states (by stage of disease for a target population) present a robust, evidence-based case for the test (e.g., to meet NCCN standards).

The test submission should detail, for each disease state, the potential placement of the test in the process of care together with a consensus definition on the target intervention group. Given the focus of precision medicine on cancer, in particular late stage cancer interventions, submissions should utilize as their template for test placement in oncology the treatment pathways and clinical guidance developed by the NCCN [37] In non-oncological disease states the guidelines should be latest proposed by the respective professional associations.

While it is unlikely that a professional group responsible for treatment guideline development would nominate a particular NGS test, it is possible that they would set the minimum performance standards required for an NGS test to be accepted into clinical practice for target patient populations.

The appropriateness, acceptance or otherwise of claims for the analytical and clinical validity of an NGS test are only relevant in the context of the 'footprint' of the test. Is the test designed to:

(i) apply across the board to a range (or any number) of disease states
(ii) apply to a single disease state
(iii) apply to a target tumor across a range (or any number) of disease states
(iv) apply to a target tumor in a single disease state

Each of these footprints should be supported by an evidence base which meets the standards for the quality of the evidence as well as claims for the test which are credible, evaluable and replicable in the restive molecular targets. This is seen in the case of the footprint (i) where the claim is for the test across either 'any' disease state or for a sub-set of disease states.

Irrespective of the breadth of the claim, a formulary committee is faced with the question of whether limited claims for the test in a sub-set of disease states can be assumed to be sufficient to accept the test as the basis for molecular targeting in disease states where the evidence is lacking. Clearly, it would be unwise, given issues of performance and safety outcomes, to assume that this is the case. Formulary committee, therefore, should request substantive and credible evidence for claims in each disease state. This should apply to claims directed towards the test's target population by stage of disease.

An important caveat is the whether the test utilized in a particular trial design is the test proposed by a vendor for approval by the health care system. Claims for analytical and clinical validity are only acceptable if the claims for analytical and clinical validity between competing tests to identify one or more molecular targets and link these to proposed therapy interventions are compatible for all target treating populations within the test 'footprint'.

No test can claim to have identified all potential molecular markers and the potential interactions between variants with their impact on patient response. Even within a given histologic tumor multiple markers may be present.
In the unlikely event that a manufacturer underwrites a head-to-head comparison with competing tests or at least the most likely comparator for a target population in one or more disease states, an assessment of competing claims will rely upon indirect comparisons. While this is a common situation in drug product evaluations, where there is an established methodological literature to support indirect comparisons, comparing competing tests adds a further level of complexity.

Unless an effective indirect comparison between competing tests is undertaken, payers have no basis for assessing the merits of claims made by a developer/vendor. This situation is made more difficult given the footprints of competing tests and the number of tests that will be competing for the attention of payers. Even with common agreement on analytical and clinical validity standards in the case, for example, where a test is considered a medical device and the FDA takes responsibility for class III 501(k) evaluations, there is no certainty that competing tests will generate similar variant profiles for therapy interventions. This is a situation which will make indirect comparisons more complex if there are options for matching therapies to specific variants given individual NGS test results.

**EVIDENCE BASE FOR NGS TESTS**

At the present time, while there is accumulating evidence for the analytical and clinical validity of NGS tests, although minimum performance standards for accuracy have yet to be agreed, there is little evidence from either randomized clinical trials or well conducted observational studies. Formulary committees and other health decision makers are understandably reluctant to support a specific NGS test not only with the absence of, for example, umbrella trial designs matching test results against the standard of care in target populations but, equally importantly, comparative assessments of competing tests.

***Randomized Clinical Trials***

Required performance criteria for NGS tests in target patient populations are no different from the criteria that would apply to non-NGS treatment options. The test would be judged in comparative terms against the outcomes that are typically applied in therapy impact assessments within that disease group and by stage of disease.

In late stage oncology interventions where there are a range of potential roles for NGS evaluations and their contribution to the process of care, clinical standards for outcomes are well established. These capture both the primary outcomes these outcomes – progression-free survival, overall survival, relapse, toxicity – and secondary endpoints that focus on patient reported outcomes, including functional status, pain and both generic and disease specific quality of life measures.

Given that the adoption of NGS tests is at any early stage and that the evidence base to support disease and target-group specific NGS claims is limited, developers of NGS tests face three hurdles. First, to invest in well-designed randomized clinical trials to compare NGS interventions against the standard of care in target disease states; second, to develop protocols to validate claims in treatment practice; and, third, to develop a robust methodology to compare the effectiveness of competing NGS tests.

Although standards for randomized clinical trials are well established, the NGS trial design is necessarily challenged by heterogeneity within patient populations. It is important to recognize the diversity of trial designs that have been proposed under the overall 'umbrella' designation to account for (i) inter and intra patient tumor molecular heterogeneity and (ii) tumor progression and the capture of intra-patient heterogeneity over time [38]. There is a developing literature and examples of alternative umbrella trial designs (e.g., the PANGEA IMBBP Phase1/2 trial design for gastro-esophageal adenocarcinoma) [39] . Given the complexity of these designs in terms of both treatment pathways, definition of standard of care, target drug combinations and variations in the incidence of biomarkers from exome sequencing in target trial populations, the structure of the various trials needs to be adequately described. When the results of trials are reported they should reference claims for analytical and clinical validity to support the structure of the trial and the individual therapy pathways.

***Reporting randomized trials, required data elements:***

- Title and ClinicalTrials.gov identifier
- Sponsor (and collaborator)
- Study design
- Disease diagnosis
- Treatment guideline
- Target population (including tumor type)
- Inclusion/exclusion criteria
- Assessment of tumor progression
- Test for molecular profiling
    - Test platform
    - Test quality and approval status
    - Analytical and clinical validity
- Patient biomarker (oncogene) distribution in target population
- Tumor suppressor therapy distribution to define therapy pathways
- Criteria for optimal pathway prognostic treatment choice
- Clinically actionable oncogene-tumor suppressor groups distribution
- Patient distribution by actionable oncogene-tumor suppression groups
- Primary endpoint measure(s)
- Secondary endpoint measures
- Trial structure and duration of therapy
    - Active arms (actionable tumor suppressor groups)
    - Control arm (standard of care)
    - Allocation
    - Criteria and/or timing for re-testing
- Aggregate outcomes: primary endpoints
- Aggregate outcomes: secondary endpoints
- Pathway outcomes: primary endpoints
- Pathway outcomes: primary endpoints
- GRADE classification

***Systematic Reviews and Meta-analyses***

Submissions to support an NGS test submission to a health system or formulary committee should not rely upon one or two randomized trials. These should meet the PISMA-P 2015 standards. The complexity of gene expression and the range of platforms utilized to identify recommended therapy options means that the greatest credibility will be afforded the test with the most credible evidence base.

Submissions to formulary committees should report on a systematic review of all randomized trials for the specific test as well as for potentially competing tests. This sets the stage for indirect comparisons of competing NGS claims. Again, the systematic review should be restricted to the evidence for the target population. Given inter and intra patient genomic heterogeneity, it cannot be assumed that claims for the NGS test in one target population can be used to support prospective claims for efficacy or effectiveness in other target populations. Nor can it be assumed that test claims from trials reported for the target treating population, given molecular heterogeneity, need necessarily translate to the health system's target population. At the same time there

needs to be a clear statement of the justification for an expected 'optimal' monotherapy or combination therapy choice to matching individual patients to their respective biomarkers.

***Observational Studies***

Observational studies include (i) prospective studies that are protocol driven and are intended to evaluate claims made from models and (ii) retrospective studies that are intended to evaluate claims from integrated databases. It is unlikely, at this time, that retrospective observational studies will play a substantive role in evaluating claims for specific NGS tests until integrated data sets expand the coverage to include identifiers for specific NGS tests and expand their capture of genomic information to include variant or mutation classifications. The most likely source of data will be from results reported for prospective observational studies that adhere to a protocol submitted in support of a specific NGS test. Indeed, the importance of evidence from protocol-driven observational studies, compared to the costs and time involved in undertaking a randomized trial, will place a premium on developers or vendors of NGS tests to engage with health care systems to partner in prognostic claims validation within target populations.

The purpose of the prospective protocol driven observational study is twofold: (i) to evaluate the claims made for the NGS-directed therapy choices for both the targeted patient population and patients assigned to specific therapy groups as 'individualized' packages and (ii) to evaluate claims for resource utilization and, potentially, cost savings again for the targeted patient population and the therapy groups.

The point to emphasize is that unless claims utilizing NGS tests can be validated by health care systems, there is no basis for reimbursement or an agreement on pricing. Randomized clinical phase 2 and phase 3 trials are a key input, but these need to be followed up with observational tracking studies within target patient groups. Accumulating evidence from tracking studies across similar target populations in health care systems will, hopefully, provide a robust and believable evidence base for the individual NGS platform.

***Reporting observational studies: required data elements:***

- Title
- Sponsor (and collaborator)
- Copy of study protocol
- Target patient population
- Criteria for selecting patient population
- Criteria for selecting treatment pathways
- Basis for claims to be assessed
    - Prospective model
    - Retrospective assessment
- Model summary
- Claims to be evaluated by selected treatment pathway
    - Clinical outcomes
    - Safety and toxicity
    - Remission and relapse
    - Survivorship
    - Quality of life
    - Pain
    - Patient satisfaction
    - Other patient reported outcomes
    - Resource utilization
    - Direct medical costs
- Study implementation
- Inclusion/exclusion criteria
- Test for molecular profiling
    - Test platform
    - Test quality and approval status
    - Analytical and clinical validity
- Patient biomarker (oncogene) distribution
- Tumor suppressor therapy distribution
- Criteria for optimal prognostic treatment choice
- Clinically actionable oncogene-tumor suppressor groups distribution
- Patient distribution by actionable oncogene-tumor suppression groups
- Treatment Pathways
- Timelines
- Adherence and persistence
- Aggregate test outcomes
- Treatment pathway outcomes
- GRADE classification

***Modeled Study Claims***

Claims evaluation involving a protocol assessment will typically rely upon cost-effectiveness or cost-utility models. These would take a short term focus with the results reported in a timeframe relevant to decision makers. The model may be structured around a clinical trial or it may extrapolate from that trial. Claims generated by models are only acceptable if they are credible, evaluable and replicable. The structure of the model would probably replicate and extrapolate from the design of umbrella trials reporting for the particular NGS test. Treatment pathway criteria should be justified (e.g., high frequency therapy choice clusters). Inputs to the model would be based upon the results of the systematic review of the trial literature for that test together with those inputs from a review of the literature where trial primary and secondary outcomes fail to capture the particular parameter (e.g., specific patient reported outcomes such as quality of life).

Standards for model building are well established, to include the treatment of structural and parameter uncertainty [40]. Care should be exercised, however, in adopting standards of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) as these include study designs that are not intended to generate evaluable claims [41] [42] [43].

It is expected that the modeled claims made for the NGS package in the targeted patient groups will be presented as part of the submission package (see below) by the NGS developer or vendor. Claims, both clinical and cost-effective (or cost-utility) would conform to the standards of normal science: they would be credible, evaluable and replicable. Unlike traditional cost-effectiveness claims, the claims proposed by the developer would be expected to take account of the tumor heterogeneity in the target population in the choice of treatment pathways to compare to standard(s) of care pathway(s).

At the same time, however, a model would have to accommodate potential confounding factors such as the presence of co-morbidities, drug-to-drug interactions both between the tumor specific drugs as well as with those assigned for other conditions. Within the timeframe for the model, claims should attempt to accommodate adherence and persistence behavior, non-tumor related morbidity and mortality and anticipated responses involving possible re-testing and drug therapy reallocation to accommodate tumor progression. The lifetime of the modeled claim will have to be consistent with ongoing reporting of outcomes to a formulary committee. Unless the anticipated timeframe for survivorship is 'short', lifetime cost-utility claims would be of no interest as they would lack credibility for evaluation and replication.

Any modeled claim for NGS based-interventions has to be comparative. The model would then have to accommodate the standard of care reference point, together with (again over the lifetime of the model) interventions to accommodate the expected disease progression. In some instances, a developer or vendor may attempt to model competing platforms for the target population. Again, comparative claims would have to be credible and open to evaluation.

**NGS UPTAKE AND IMPLEMENTATION**

Modeling a claims for the anticipated impact of NGS platform indicated treatment pathways should also take account of (i) the uptake of the NGS test by the treating physicians (i.e., ordering the test) and (ii) the implementation of the test (i.e., following the recommended therapy pathways for matched patients). Evidence to date, admittedly anecdotal, would suggest that less than fifty per cent of treating physician order tests with even fewer actually attempting to implement the test results. At the same time, we have no data on patterns of adherence and persistence with a test-directed regimen. This situation is made more opaque by the absence of data on the monitoring of patient responses to a therapy choice and the criteria that physicians may follow to order further tests in response to patient relapse and the likelihood of adverse tumor progression.

Reporting requirements should include:

- Uptake of the NGS test within each target population group
- Proportion of those requesting the NGS test that implement the test recommendations
- Distribution of patients by NGS recommended therapy choice and standard(s) of care
- Adherence and persistence with NGS recommended therapy choice for first six months following index prescription(s)
- Additional NGS tests
- Subsequent six monthly reports covering the preceding six months

***Requesting an NGS Test***

The request for an NGS test is the responsibility of the treating physician in consultation with the patient. The submission from the developer should include a *pro forma* of the information to be provided by the treating physician when the test is requested. As a minimum this should detail (i) patient demographics (age, gender, ethnicity); (ii) clinical assessment and stage of disease; (iii) comorbidities; and (iv) current drug regimens. On receipt of a request, the developer should inform the treating physician if FDA-mandated risk mitigation and reporting requirements are in place for any of the anticipated drug or drug combinations that are likely to be suggested for the individual within the target patient group. The FDA may, of course, put risk assessment requirements in place as a blanket requirement for the NGS test.

***Implementing NGS Test Results***

A question that has to be addressed with the advent of NGS is how the treating physician is to act on recommendations made for a repositioned FDA approved therapy for an individual patient. As it is unlikely that there will be a common repositioned therapy (or therapies) for those failing or non-responsive to the standard of care, one role of a formulary committee or health system in approving an NGS platform is to request developers for their recommendations for repositioning are to be translated into clinical practice decisions and how they would assess and report relative benefits/harms. The NGS developer should agree with the health system the content of reports to be provided to treating physicians to guide therapy options and report safety issues.

One approach that should be considered is the heuristic Oxford Centre for Evidence Based Medicine (OCEBM) levels that cover the entire range of clinical questions that a clinician might ask. The OCEBM system allows clinicians and patients to appraise evidence for prevalence, accuracy of diagnostic tests, prognosis, therapeutic effects, rare harms, common harms, and usefulness of (early) screening[44] .

Another and more recent approach is the GRADE Evidence to Decision (EtD) framework [45]. This should be considered as a possible model for informing physicians and supporting clinical decisions. The EtD approach is to assist physicians to use evidence in a structured and transparent way, linking to practice guidelines and going from evidence to decision.

***Feedback on Test Performance***

The credibility of claims and agreement on the place of a specific NGS test in therapy across a number of disease states will require tracking outcomes to be reported back in a timely fashion to health care decision makers. The protocol that is drafted and agreed with the health system should detail the frequency and content of reports on the performance of the test. Confidence that NGS recommendations have a credible role in treatment decisions must be evidence based. They should be reported by the developer or vendor to the health care system. They should also be made available in the public domain through peer reviewed publications and web sites. It is important, therefore that the claims assessment protocol that is agreed with a health system details which claims are to be reported, how they are to be reported and the timeframe for reporting.

**THE NGS SUBMISSION**

Submissions to support NGS adoption should meet a common evidentiary standard. The following outline is proposed:

**I. Executive Summary**
- Place of the proposed NGS test in therapy
- Anticipated benefits and harms

**II. Test Platform Specification**

**III. Proposed Target Population(s)**
- Epidemiology
- Demographics
- Comorbidities
- Test place in therapy
- Criteria for test implementation
- Guideline recommendations for genomic tests

**IV. Test Accuracy**
- Analytical validity
- Clinical validity

**V. Test Processes and Reporting**
- Requesting the test
- Test inputs
- Processing test inputs
- Reporting test results
- Liaison with treating physician

**VI. Evidence Base for Test Claims**
- Randomized clinical trials
- Observational studies
- Outcomes in clinical practice
- Test uptake and implementation in clinical practice

**VII. Clinical Claims for Test in Target Population(s)**
- Survivorship and relapse
- Safety and toxicity
- Quality of life
- Other patient reported outcomes

**VIII. Modelled Cost-Outcomes Claims**
- Model structure
- Model parameters
- Model inputs
- Uncertainty
- Model claims

**IX. Claims Assessment Protocol**
- Agreement on Protocol and IRB Approvals
- Claims to be evaluated
- Risk mitigation
- Timeframe
- Target population
- Recruitment
- Monitoring and Quality Assessment
- Reporting

**X. Reviewing the Test Platform and Claims**
- Agreement on regular reviews
- Re-assessment of test claims

**APPENDIX A:** Comparator NGS Tests: Approvals and Platforms
**APPENDIX B:** Evidence Base Spreadsheet Summary: Proposed Test
**APPENDIX C**: Evidence Base Spreadsheet Summary: Comparator NGS Test
**APPENDIX D:** Targeted Claims Evaluation for Comparator Tests

**QUESTIONS A FORMULARY COMMITTEE SHOULD ASK**

Once a submission to support a NGS product is received, there are a number of questions that a formulary committee should ask (or included as a checklist as part of the submission). These questions relate to: (i) the proposed NGS test; (ii) the performance of the proposed NGS test in each of the target patient populations; and (iii) evidence for the proposed test and claims made against the standard of care for the target patient group and potential comparator NGS tests. The questions are detailed in Table 1:

**TABLE 1**
**NEXT GENERATION SEQUENCING PRODUCT SUBMISSION: QUESTIONS A FORMULARY COMMITTEE SHOULD ASK**

| Question No. | Question | Submission Section | Response Y/N Comments |
|---|---|---|---|
| 1. | Has the submission provided an executive summary that summarizes the place of the NGS test in therapy for the target populations for the test? | I | |
| 2. | Has the executive summary summarized the anticipated benefits and harms for each of the target populations from the introduction of the NGS test? | I | |
| 3. | Has the submission provided a technical specification of the NGS test to include meeting STARD 2015 and QUADAS-2 standards? | II | |
| 4. | Has the technical specification of the test detailed the steps in the test process: (i) specimen collection, processing and storage; (iii) DNA extraction; (iv) DNA processing and library preparation; (v) regions of genome interrogated; (vi) generation of sequence reads and base calling; (vii) variant classification/interpretation; and (viii) preparation of test report. | II | |
| 5. | Has the technical specification of the test detailed standard operating procedures for curating, aggregating and interpreting high quality valid scientific evidence for the analytical and clinical performance of the platform? | II | |
| 6. | Has the technical specification of the test detailed standard operating procedures for updating inputs and maintaining the stability and architecture of the platform content and processes? | II | |
| 7. | Has the test specification identified the roles of third parties in the collection and assessment of samples? | II | |
| 8. | Has the test specification identified any manual variant interpretation? | II | |
| 9. | Has the submission detailed the intended target populations for the test by disease type and by their place in therapy? | III | |
| 10. | Has the submission provided a definition of each of the target population groups and their criteria for selection that meets national and international standards for clinical and diagnostic identification? | IV | |
| 11. | Has the submission identified for each of the target population groups the place in therapy for the proposed genomic assessment given current treatment guidelines and treatment pathways? | III | |
| 12. | Has the submission provided a profile of the target patient groups: (i) epidemiology; demographics; and (iii) comorbidities | III | |
| 13. | Has the submission provided for each of the target patient populations the results of assessments of analytical validity to recommended FDA standards? | IV | |
| 14. | Has the submission provided for each of the target patient populations the results of assessments of clinical validity to recommended FDA standards? | IV | |
| 15. | Has the submission detailed how the results of the test will be reported to treating physicians? | V | |
| 16. | Has the submission detailed how the results of the tests will be linked to recommendations for single and multiple therapy choices as alternatives to the standard of care? | V | |
| 17. | Has the submission detailed how requests from treating physician requesting clarification on test results will be handled? | V | |
| 18. | Does the submission meet the required evidence standards for reporting clinical trials and observational studies (PISMA- 2015 for systematic reviews, CONSORT and GRADE/EPC)? | VI | |
| 19. | Has the submission provided a systematic review of randomized clinical trials, summary evidence tables and a commentary for efficacy and safety for the proposed NGS test in the respective target patient populations? | VI | |

| | | | |
|---|---|---|---|
| 20. | Has the submission provided a systematic review of observational studies, summary evidence tables and a commentary for efficacy and safety for the proposed NGS test in the respective target patient populations? | VI | |
| 21. | Has the submission included a summary of the outcomes achieved in clinical practice for the individual target patient populations where the NGS test has been utilized? | VI | |
| 22. | Has the submission provided a summary of the historical uptake of the NGS test in the individual target patient populations? | Vi | |
| 23. | Has the submission provided credible, evaluable and replicable modeled claims for the anticipated benefits and risks from introducing the NGS test in the target patient populations? | VII/VIII | |
| 24. | Do the modeled claims made for the NGS test in the target populations include (i) survivorship and relapse; (ii) safety and toxicity; (ii) quality of life; and (iv) other patient reported outcomes? | VII/VIII | |
| 25. | Are the claims provided in the submission presented in comparative terms versus the standard of care? | VII/VIII | |
| 26. | In presenting the modeled claims, are these modeled separately for each of the target populations and therapy pathways? | VIII | |
| 27. | Has the submission provided details on: (i) model structure; (ii) model parameters; (iii) model input sources; (iv) treatment of uncertainty; and (v) incremental benefits for outcomes and costs versus standard of care for the test overall and for the therapy pathways? | VIII | |
| 28. | What assumptions has the model made to accommodate the anticipated distribution of therapy options in the target population and links to therapy options? | VIII | |
| 29. | Has an electronic copy of the model been provided for each of the target patient groups? | VIII | |
| 30. | Has the developer or vendor provided a protocol to detail how the claims for the NGS test are to be evaluated for each of the target population groups, together with reporting requirements? | IX | |
| 31. | Is the protocol as presented acceptable? | IX | |
| 32. | Has the developer or vendor agreed to a regular review of test platform enhancements and processes to support claims? | X | |
| 33. | Comparator tests: Has the submission identified the potential comparator NGS tests? | Appendix A | |
| 34. | Comparator tests: Has the submission summarized the proposed target populations for the comparator tests? | Appendices B and C | |
| 35. | Comparator tests: Has the submission detailed how the comparator tests are to be implemented for the proposed target populations? | Appendix D | |
| 36. | Comparator tests: Has the submission detailed the evidence for analytical and clinical validity for the comparator tests? | Appendix D | |
| 37. | Comparator tests: Has the submission detailed the evidence base and claims for the comparator tests from randomized clinical trials and observational studies? | Appendix D | |
| 38. | Comparator Tests: Has the submission detailed the modeled claims for the comparator tests? | Appendix D | |
| 39. | Comparator tests: Do the claims made for the comparator tests meet the standards for credibility, evaluation and replication? | Appendix D | |
| 40. | Comparator tests: Has the submission reported any direct/indirect comparisons between comparator tests, the proposed test and the standard of care in the target patient populations? | Appendix D | |

**CONCLUSIONS**

The purpose of this commentary has been to raise a number of key questions regarding (i) the evidence base for NGS claims; (ii) the standards required to support NGS claims and (iii) the standards required for their introduction to clinical practice.

The current evidence base is inadequate. There is not only limited evidence for the clinical utility of NGS tests but a limited appreciation of the steps required to ensure that claims made are validated and that risks to patients are adequately protected. While there is little doubt that the various NGS tests will be classified as medical devices, we have no direction from agencies such as the FDA as to their role beyond certifying the test meets standards for analytical and clinical validity. Unlike the NICE DAP evidence standards where modeled claims for costs and outcomes are requested, there is no evidence to date that the FDA will attempt further to regulate test claims and the choice of mutation linked therapies. Even so, concerns with safety and adverse events with what are likely off-label use of a range of pharmaceutical products may force the FDA to demand stronger evidence for clinical utility, to include risk-mitigation and monitoring requirements.

In the absence of standards to evaluate and monitor the claims from NGS platforms, the proposal here is that formulary committees and health systems take the initiative. A range of questions that formulary committees might reasonably ask are detailed, together with the proposal that submissions for NGS tests should not only be target patient specific but should be accompanied by an assessment protocol.

At the same time, NGS developers and vendors will face an uphill struggle to convince health care systems, treating physicians and guideline developers to introduce NGS test options into clinical practice. While the evidence base (as emphasized) is just too limited, this situation will likely be compounded by the number of competing tests and the need, from both clinical and cost-effectiveness perspectives, to report on the validity of test claims for well-defined reference target populations and for the mutation-linked therapy choices. Whatever claims are made for individual NGS tests, to include claims for the superiority of one test over another, the claims made must be credible, evaluable and replicable. Health system decision makers will expect no less.

**Acknowledgements:** I would like to thank Jeffrey Waldron, PM Connective for his comments on an earlier draft of this paper.

**REFERENCES**

[1] Meric-Bernstram F, Johnson A, Holla V et al. A decision support framework for genomically informed investigational cancer therapy. *J Natl Cancer Inst*. 2015;107(7):djv098

[2] Phillips KA, SAakowski JA, Trosman J et al. The veconomic value of personalized medicine tests: what we know and what we need to know*. Genet Med*. 2014;16(3):251-7

[3] Centre for Evidence Based Medicine. COMPare Project. Oxford 2016. http://compare-trials.org/results

[4] Camerer CF, Dreber A, Forsell E et al. Evaluating replicability of laboratory experiments in economics. Science. 10.1126/science.aaf0918 (2016)

[5] Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *Innov Pharm*. 2016;7(2): No. 11.

[6] Langley PC**.** Supporting formulary decisions: The discovery of new facts or constructed evidence? *Innov Pharm* . 2016;7(2): No.15.

[7] Langley PC**.** Sunlit uplands: the genius of the NICE reference case. *Innov Pharm*. 2016;7(2): No.12

[8] Langley PC**.** Imaginary worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *Innov Pharm*. 2016;7(2): No. 9.

[9] Langley PC, Rhee TG. Imaginary worlds: A systematic review of the status of modeled cost-effectiveness claims published in the Journal of Medical Economics from January 2015 to December 2015. *Innov Pharm* . 2016;7(2): No. 16.

[10] Langley PC, Rhee TG. Imaginary worlds: The status of modeled economic evaluation claims published in Value in Health January 2015 to December 2015. *Innov Pharm*. 2016;7(2): No. 18.

[11] Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? *Genetics Med*. 2006;8(7):448-50

[12] Burke W. Genetic Tests: Clinical Validity and Clinical Utility. *Curr Proc Hum Genet*. 2014;60(7):995-1003

[13] Holtzman NA, Watson MS. Final Report of the TASK Force on Genetic Testing. Baltimore: Johns Hopkins University Press; 1999

[14] Langley PC. The status of modeled claims. *J Med Econ*. 2015;18(12):991-2

[15] Popper KR. The logic of scientific discovery. New York: Harper, 1959

[16] Langley PC. Nullius in verba: the University of Minnesota Social and Administrative Pharmacy program Proposed Guidelines for Formulary Evaluations. *Innov Pharm*. 2016;7(2): No. 10.

[17] Wootton D. The Invention of Science. New York: Harper Collins, 2015

[18] National Institute for Health and Care Excellence. Diagnostics Assessment Programme Manual. NICE: London, 2011 (previously National Institute for Health and Clinical Excellence).

[19] Bossuyt PM, Reitsma JB, Bruns DE et al. STARD 2015 – An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. PMID: 26511519; *Radiology*. 2015:151516. PMID: 26509226 *Clinical Chemistry*. 2015. pii: clinchem.2015.246280. PMID: 26510957. See also STARD item list [http://www.equator-network.org/reporting-guidelines/stard/]

[20] Whiting PF, Rutjes A, Westwood M et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Int Med*. 2011;155:529-36. See also [http://www.bristol.ac.uk/social-community-medicine/projects/quadas/]

[21] Moher D, Shamseer L, Clarke E et al. Preferred reporting items for systematic reviews and meta-analyses protocol (PRISMA-P) 2015 statement. *Systematic Rev*. 2015;4:1 DOI: 101186/2046-4053-4-1

[22] Yu W, Gwinn M, Dotson WD et al. A knowledge base for tracking the impact of genomics on population health. *Genet Med*. 2016, June 9 DOI: 10.1038/gim.2016.63. See also [https://phgkb.cdc.gov/GAPPKB/phgHome.do?action=home]

[23] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Ann Int Med*. 2010:152(11). See also: The CONSORT Statement 25-item check list [http://www.consort-statement.org/checklists/view/32-consort/66-title] and flow diagram [http://www.consort-statement.org/consort-statement/flow-diagram] to record the progress of patients through the trial.

[24] Meader N, King K, Llewellyn A et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic Rev*. 2014;3:82

[25] Cochrane Handbook: [http://handbook.cochrane.org/part_2_general_methods_for_cochrane_reviews.htm].

[26] Berkman MD, Lohr KN, Ansari M et al. Grading the strength of a body of evidence when assessing health care interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An update. Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13 (14)-EHC130EF. Rockville MD: Agency for Healthcare Research and Quality. November 2013.

[27] National Comprehensive Cancer Network, NCCN Guidelines and Derivative Information Products: User Guide https://www.nccn.org/professionals/transparency.asp

[28] Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials and other master protocols: a review and examples. *Annals Oncology*. 2016: doi:10.1093/annonc/mdw413

[29] Siu LL, Conley BA, Boerner S et al. Next-generation sequencing to guide clinical trials. *Clin Cancer Res Focus*. 2015;21(20): 4536-4544

[30] Shuren J. Examining the Regulation of Diagnostic Tests and Laboratory Operations. Subcommittee on Health, Committee on Energy and Commerce, US House of Representatives. Nov 17, 2015 http://www.fda.gov/NewsEvents/Testimony/ucm473922.htm

[31] Palmetto GBA: [http://www.palmettogba.com/moldx]

[32] Palmetto GBA. Clinical Test Evaluation Process (CTEP) M00096 (v5). November 2015.

[33] MolDx (Palmetto GBA). Analytical Performance Specifications for Comprehensive Genomic Profiling (M00118, V1).

[34] Food and Drug Administration. Use of standards in FDA regulatory oversight of next-generation sequencing (NGS) - based in vitro diagnostics (IVDs) used for diagnosing germline diseases. July 8, 2016 http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM509837.pdf?source=govdelivery&utm_medium=email&utm_source=govdelivery

[35] Food and Drug Administration. Guidance for Industry and FDA Staff. Statistical Guidance on reporting results from studies evaluating diagnostic tests. March 2007.

[36] Doudican NA, Mazumder A, Kapoor A et al. Predictive simulation for designing cancer therapeutic regimens with novel biological mechanisms. *J Cancer*. 2014;5(6):406-16

[37] National Comprehensive Cancer Centre Network (NCCN): [https://www.nccn.org/professionals/physician_gls/f_guidelines.asp].

[38] Catenacci D. Next generation clinical trials: novel strategies to address the challenge of tumor molecular heterogeneity. *Molecular Oncol*. 2015; 9:967-996

[39] PANGEA-IMBBP: Personalized Antibodies for Gastro-Esophageal Adenocarcinoma - A 1st Pilot Metastatic Trial of Biologics Beyond Progression [NCT02213289] (https://clinicaltrials.gov/ct2/show/NCT02213289)

[40] Drummond MF, Sculpher MJ, Torrance GW et al. Methods for the economic evaluation of health care programmes. 3rd Ed. Oxford: Oxford University Press; 2005

[41] Caro JJ, Briggs A, Siebert U et al. Modeling good research practices – Overview: A report of the ISPOR-SMDM Modeling good practices task force – 1. *Value Health*. 2012;15:796-803

[42] Caro JJ, Eddy DM, Kan H et al. Questionnaire to assess relevance and credibility of modeled studies for informing health care decision making: An ISPOR-AMCP-NPC good practice task force report. *Value Health*. 2014;17:174-182

[43] Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices Task Force – 7. *Value Health*. 2012;15:843-850

[44] Oxford Centre for Evidence Based Medicine: (http://www.cebm.net/explanation-2011-ocebm-levels-evidence/).

[45] Alonso-XCoello P, Schünemann HJ, Moberg J et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1:Introduction. *BMJ*. 2016353;12016