Reviews • KEYNOTE REVIEW

Teaser Setting the companion diagnostic performance by selecting the cut-off score integrates scientific, clinical, ethical, and commercial considerations and determines the value of the stratified medicine combination to developers, payers, and patients.

The clinical benefits, ethics, and economics of stratified medicine and companion diagnostics

Mark R. Trusheim and Ernst R. Berndt

MIT Sloan School of Management, 100 Main St. E62-518, Cambridge, MA 01239-4307, USA

The stratified medicine companion diagnostic (CDx) cut-off decision integrates scientific, clinical, ethical, and commercial considerations, and determines its value to developers, providers, payers, and patients. Competition already sharpens these issues in oncology, and might soon do the same for emerging stratified medicines in autoimmune, cardiovascular, neurodegenerative, respiratory, and other conditions. Of 53 oncology targets with a launched therapeutic, 44 have competing therapeutics. Only 12 of 141 Phase III candidates addressing new targets face no competition. CDx choices might alter competitive positions and reimbursement. Under current diagnostic incentives, payers see novel stratified medicines that improve public health and increase costs, but do not observe companion diagnostics for legacy treatments that would reduce costs. It would be in the interests of payers to rediscover their heritage of direct investment in diagnostic development.

Introduction

By using CDx, stratified medicines target those patients who are most likely to benefit from them. However, no one CDx performs perfectly. Selecting the diagnostic score (i.e., the cut-off value) that separates those who then qualify for treatment from those who do not is usually an ambiguous choice. The science often provides a range of possibilities that trade including more who might benefit (true positives) at the expense of also including more who will not (true negatives who thus become false positives). Here, we focus on the clinical, economic, and ethical ramifications of choosing that companion diagnostic cut-off value. That choice fundamentally impacts the performance of the therapeutic in the treated population and, therefore, also the incentives of regulators to approve the therapeutic, of patients and providers to use it, of payers to cover and reimburse the therapy, and of innovators and investors to develop the therapeutic in the first place.

Stratified medicine tightens the links among science, the clinic, and the marketplace. Setting the CDx cut-off value is perhaps the crucial connection among all three, with no easy rule of thumb to guide the choice. Each stratified medicine opportunity faces unique facts and circumstances that require balancing ethical, scientific, and financial concerns. Beyond this selection,

Mark Trusheim is a Visiting Scientist at the MIT Sloan School of Management. He has also been a Special Government Employee for the FDA's Office of the Commissioner and is the Founder of Co-Bio Consulting, LLC. Mark's research focuses on modeling the introduction of new innovations in healthcare, such as stratified medicines, to inform public policy, corporate strategy and product development programs. He is a former President and Board member of the Massachusetts Biotechnology Council. Earlier, he held multiple private sector roles including, founder of Cantata Laboratories a biochemical profiling biomarker discovery and diagnostics firm; Monsanto, culminating his career there as Co-President and Chief Operating Officer of Cereon Genomics, LLC.; Kenan Systems Corporation which focused on developing quantitative models and artificial intelligence based applications for large corporations and government agencies; and Wang Laboratories. He holds degrees in Chemistry from Stanford University and Management from MIT.

Ernst Berndt is the Louis E. Seley Professor in Applied Economics at the Massachusetts Institute of Technology (MIT). His research has examined how stratified medicine is changing pharmaceutical economics, how innovations have affected the costs of treating selected diseases over time, how industry funding of the FDA has affected review times and safety withdrawal rates, and factors affecting the globalization of clinical trials into emerging economies. In other research, Berndt has assessed how illness and its treatment affect individuals' ability to function at work. More generally, Berndt's research deals with assessing the sources of productivity growth and how productivity is measured. Berndt has also implemented methods for adjusting prices for changes in quality in the pharmaceutical, health care, personal computer hardware, prepackaged software and personal digital assistant industries. He has also published research on the impact of direct-to-consumer marketing of prescription pharmaceuticals on drug utilization in areas such as depression and schizophrenia. Currently he is on the editorial board of Health Affairs. He holds economics degrees from Valparaiso University and the University of Wisconsin and honorary doctorates from Uppsala University and University of Basel.

Corresponding author:. Trusheim, M.R. (trusheim@mit.edu), (mtrusheim@cobioconsulting.com)

GLOSSARY

510(k) a US FDA medical device and/or diagnostic approval pathway emphasizing the comparison of the performance of the new device seeking approval with that of a predicate device already FDA approved. 510(k) refers to the section of the Food and Drug Cosmetic Act instituting the pathway (http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/510kClearances/).

CE marking a European Union (EU) diagnostic authorization pathway emphasizing manufacturer self-declaration that the diagnostic complies with the essential requirements of the EU Product Directives (https://www.gov.uk/guidance/ce-marking).

Diagnostic negative predictive value (NPV) the percentage of patients who test negative who are true negatives.

Diagnostic positive predictive value (PPV) the percentage of patients who test positive who are true positives.

Diagnostic sensitivity the percentage of true positive patients who test positive.

Diagnostic specificity the percentage of true negative patients who test negative.

Incremental cost effectiveness ratio (ICER) the difference in cost between two health interventions, often a new product versus a standard of care, divided by the difference in their health effect [44].

Laboratory-developed tests (LDT) according to the FDA, a diagnostic test that is designed, manufactured, and used within a single laboratory (http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm407296.htm).

Pre-market approval (PMA) a US FDA medical device and/or diagnostic approval pathway that uses information concerning the performance of the new device seeking approval. This pathway usually requires more clinical data than the 510(k) pathway (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/ucm2007514.htm).

Quality adjust life year (QALY) the health benefit of an intervention taking into account both the duration and effectiveness in restoring full health (https://www.nice.org.uk/glossary?letter=q).

the underlying health economics impact dynamics among competing developers, the current bias of stratified medicines for only novel therapies, and the resultant consequences for payer costs and patient benefits. Diagnostic and biopharmaceutical firms, and even basic researchers, are also affected.

Most therapeutics provide benefit to only a fraction of those who take them. For example, clinical remission rates for tumor necrosis factor alpha (TNFα) inhibitors in autoimmune diseases, such as Crohn's disease and ulcerative colitis, are approximately 25–40% [1]. For many oncology therapeutics, perhaps only 20–30% of patients respond and have their life expectancy (months of overall survival) increased [2]. In a large population, treatment with therapeutics such as statins, and others, will result in a much smaller proportion of the population receiving a health outcome benefit, such as an avoided heart attack or death. According to a review by David Newman, the number needed to treat (NNT) with a statin for 5 years to save a life is 83 and the NNT to avoid a single heart attack is approximately 39; that is, for every 39 patients with known heart disease treated continuously with a statin for 5 years, one nonfatal heart attack is avoided [3,4]. If we knew which single person out of the 39 would avoid the event, the other 38 would not

need to take their medicine, eliminating the drug costs and adverse effects in those 38 patients. Unfortunately, we cannot currently predict which individual would avoid an event if they lowered their cholesterol, so we encourage all to take their medicine.

At a cost of US$120 per year per patient for a generic statin (https://www.nice.org.uk/glossary?letter=q), it costs a total of US$50 000 over 5 years to save a single life from a cardiovascular event, not counting the physician office fees or other costs. As recently as 2013 for branded statins, the cost was four times higher at US$208 000 [5]. By comparison, the TNFα inhibitors NNT for clinical remission response is approximately three (costing approximately US$45 000–US$75 000 per clinical remission per patient per year [6]). The NNTs for oncology drugs are approximately three to five based on the response rates cited above. Stratified medicine approaches promise to target responders better and so to reduce the NNT for a therapeutic, with financial implications for developers and payers as well as benefits to patient and public health.

In this review, we focus on the post-launch effects of stratified medicines rather than on the impacts on clinical development, which has been a focus of other research [7–10].

The value of an ideal companion diagnostic

Most drugs are prescribed empirically to 'all-comers' even though some, perhaps many, will not respond. With a stratified medicine, a companion diagnostic is used to identify a patient subpopulation having a differential expected clinical response. Companion diagnostics have been used in oncology since at least the 1990s, when trastuzumab (Herceptin) was launched with a companion diagnostic for human epidermal growth factor 2 (HER2) overexpression [24]. Since then, many stratified medicines been introduced not only into oncology, but also into fields such as infectious disease [sofosbuvir Sovaldi for hepatitis C (HCV) genotypes 1–4], respiratory disease [omalizumab (Xolair) and immunoglobulin E (IgE) levels for both patient selection and dosing], and neurodegenerative disease [natalizumab (Tysabri) and John Cunningham virus (JCV) testing] (http://labels.fda.gov/).

An ideal companion diagnostic perfectly identifies treatment responders. In Fig. 1a, the two curves represent all the patients with the disease who might be treated with the drug. The yellow curve to the left represents the patients who will not respond to the therapeutic, with the orange flat bar indicating this zero response. The blue curve to the right represents those patients who will respond, with the high orange bar denoting these patients' consistently high response. The companion diagnostic test score along the x-axis perfectly separates the two patient populations. Any patient with a companion diagnostic score larger than (to the right of) that indicated by the vertical dashed line will respond to the drug, whereas those with a lower score are nonresponders. By eliminating the nonresponders in a clinical trial, the observed therapeutic effect will increase. For example, imagine an all-comers cancer trial in which all possible patients are treated, and the average outcome is 6 months overall survival. By contrast, in the idealized example in which one half (1/2) of patients respond uniformly well (gaining an additional 12 months overall survival) and the remainder do not respond at all, the average overall survival in the companion diagnostic

A stratified medicine with an ideal companion diagnostic
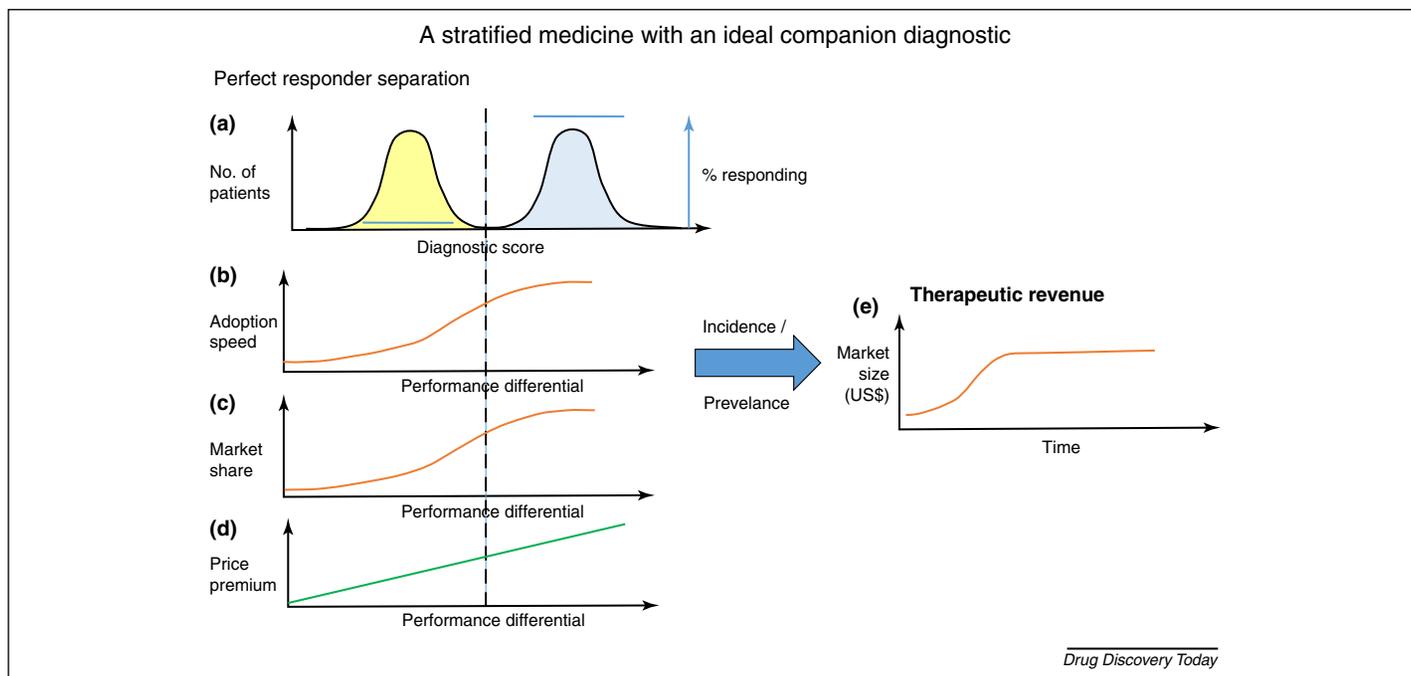
FIGURE 1

An ideal companion diagnostic perfectly separates therapeutic responders from nonresponders resulting in a positive clinical performance differential compared with an all-comers approach, which in turn could lead to faster clinical adoption, greater market share, and a price premium. See main text for more details.

selected population (the blue curve in Fig. 1a) will be 12 months, double that in the all-comers clinical trial design (the combined yellow and blue populations). Fig. 1b–d suggests that this improved therapeutic performance translates into faser clinical adoption greater market share, and a higher therapeutic price. Although not strict mathematical relations as the charts imply, it is plausible within a therapeutic class that clinical and market enthusiasm for a drug corresponds at least approximately with the therapeutic net benefit profile [11]. Combining these clinical and commercialization characteristics with the incidence and prevalence of the conditions creates a therapeutic market forecast shown Fig. 1e, in which revenues grow over time and then plateau as peak sales are achieved. In this simplistic example, price increases or decreases over time are ignored and the postexclusivity period is not shown.

Similarly, a companion diagnostic revenue line could be drawn. Given the generally far lower reimbursement and near-immediate competition from similar diagnostics because of limited patent, regulatory, or other market protections, the companion diagnostic revenues are usually lower than the therapeutic revenues. The sponsoring diagnostic firm receives little of the health value created by the drug–diagnostic combination, as discussed below.

The value of a realistic companion diagnostic

Unfortunately, in practice, no diagnostic performs ideally. All diagnostics experience some level of error. Biological relations are often inherently approximate. Any instrument has some measurement variability from day to day and sample to sample. Some phenomena are difficult to measure because of the molecular properties of the analyte or because the phenomena themselves are more qualitative or even subjective, such as pain levels,

walking gate, or activities of daily living. Even when analytes can be measured precisely, patients can vary significantly in the pharmacological relation of an analyte to a gross clinical phenotype of interest. Beyond the biological, behavioral factors can also affect the performance of a diagnostic in predicting a patient's response. For example, patients might be inconsistent in their dosing or in adhering to dietary restrictions. The result is that, for any companion diagnostic, some patients will receive false positive results; that is, scores indicating that they will respond, but when treated, they do not respond. Other patients will receive false negative results; that is, scores indicating that they will not respond, but if treated they would respond. Choosing the cut-off based on a highly controlled clinical trial setting might not fully reflect the performance in a more heterogeneous community practice setting. Those differences might need to be considered as physicians develop guidelines and sponsors negotiate coverage and reimbursement with payers.

Fig. 2 illustrates this more realistic scenario. Once again, the yellow curve in Fig. 2a represents those patients who will not respond. It is now larger to reflect what, unfortunately, seems the more common state in practice. The blue curve represents those fewer patients who will respond. In this case, however, the companion diagnostic score does not perfectly separate the two patient populations, which instead overlap. This overlap leads to false positive and false negative results. Depending on the cut-off value chosen, the number of false positive and false negative varies as suggested by the two vertical dashed lines in Fig. 2. False positives are those patients represented by the portion of the yellow curve to right of the vertical dashed line. False negative patients are those patients represented by the portion of the blue curve to the left of the dashed line. The blue curve in Fig. 2a indicates that the
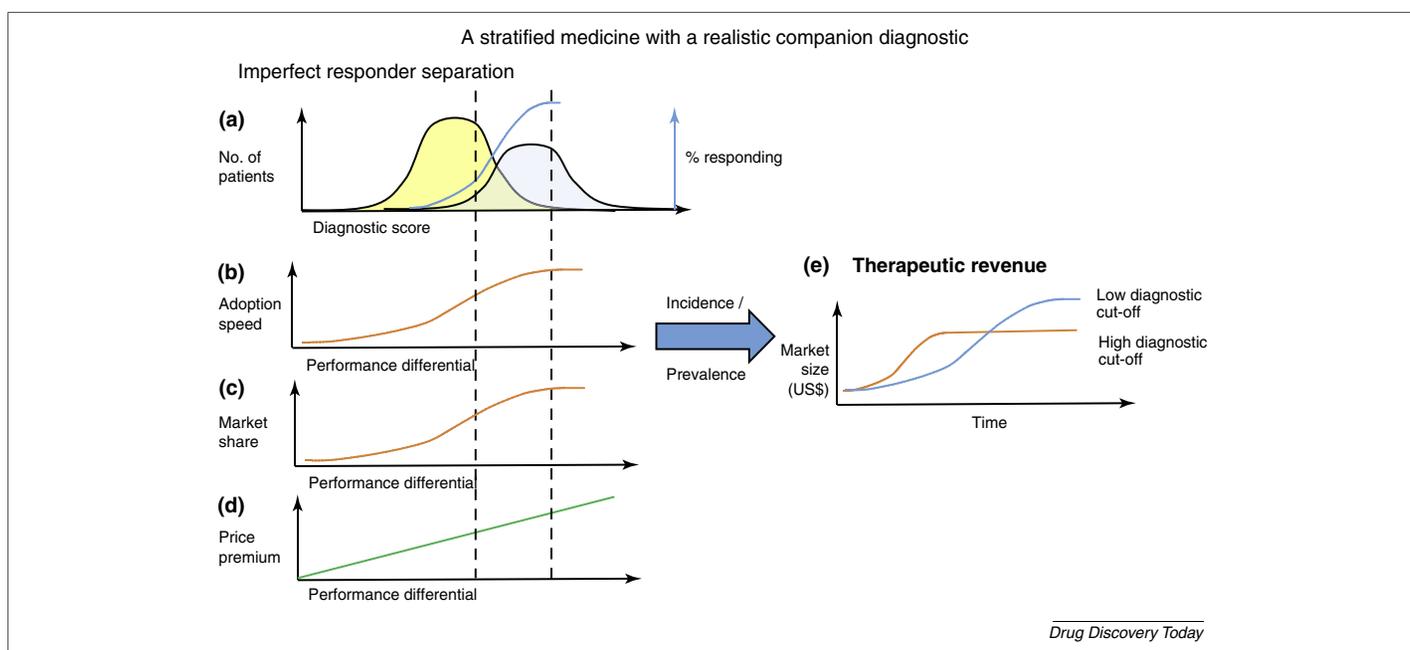
FIGURE 2

A realistic companion diagnostic imperfectly separates responders from non responders, presenting a range of possible cut-off values. The resulting range of potential performance differentials leads to similarly varying revenue results depending on the resulting changes to adoption speed, market share, and price as well as the prevalence of therapeutic responders. See main text for more details.

therapeutic response increases from low to high as the companion diagnostic score value and the relative fraction of responders increase.

This imperfect, overlapping situation occurs with genetic markers as well as with more continuous molecular analytes, such as protein levels, cell counts, metabolite levels, blood pressure, low-density and high-density lipoprotein (LDL and HDL) cholesterol, and so on. Although individual genetic differences (alleles) are binary, most genes have many such allelic changes. In addition, a given genetic change might exist on one or both copies of the gene on a chromosome pair (heterozygous or homozygous variants). Selecting which genetic variants to include, and whether to include heterozygous as well as homozygous variants, can lead to a range of companion diagnostic results similar to the continuous companion diagnostic scale scores shown in Fig. 2.

Setting the cut-off value (vertical dashed lines in Fig. 2) for the imperfectly performing companion diagnostic presents multiple challenges to the scientist, regulator, ethicist, marketer, clinician, and payer. Scientists might seek natural break points connected to a biological mechanistic rationale, or struggle to define the proper balance between diagnostic sensitivity and specificity (see Glossary). Regulators might seek a division that maximizes the benefit:risk ratio with the greatest certainty. Ethicists might be concerned with issues of denying care to some or knowingly causing harm to some (statistically) to benefit others. Marketers might seek to optimize revenues by balancing efficacy improvements, and the correlated pricing and market share, with the number of eligible patients in the market. Clinicians might seek to know the likelihood that their individual patient will respond to treatment or will incur an adverse event. Payers might focus on the net clinical benefit to their specifically covered population and the overall affordability of the resulting net total outlays for the

actually treated population. Although clearly having overlapping perspectives, when selecting the CDx cut-off each stakeholder brings its own unique view of the issues to emphasize and the proper metrics to optimize.

Setting a high cut-off value towards the far right of the yellow curve in Fig. 2a excludes nearly all nonresponding patient population scores, ensuring that nearly all of the selected and then treated patients will respond. This also results in few nonresponding patients being exposed to the adverse effect risks of the drug or the treatment time opportunity cost of pursuing an ineffective treatment while the disease worsens or death occurs. In technical terms, a cut-off value has been chosen to create a companion diagnostic with a high clinical specificity; that is, few false positives will occur.

This of course assumes that patients are tested and the treatment action corresponding to the test result is undertaken. In practice, this does not always occur. A study of patients with breast cancer in the US found that one-third of them were not tested for HER2 overexpression and that 20% of those treated with trastuzumab (Herceptin; whose label requires high HER2 expression to qualify for treatment) were treated despite having no record of being tested [12]. A similar study found the same testing rate for Canada [13]; thus, this is not necessarily a feature unique to the US healthcare delivery system.

Implications of a high cut-off

Setting a high cut-off ensures the best possible clinical trial efficacy results, again for the reason that few false positive, nonresponding patients are selected and treated. When one follows the right vertical dashed line downward in Fig. 2b–d, one observes that, with the resulting performance differential, the therapeutic could achieve the high end of its potential adoption speed, price, and market share in the selected subpopulation.

Ethically, a high test value cut-off has the inherent negative characteristic of denying treatment to some patients; that is, those who would respond to treatment but received a low test result from the imperfect companion diagnostic (the false negatives). For a severe condition with few treatment options, this might be unacceptable. For a condition with many and similarly efficacious treatment options available, or perhaps a condition with low morbidity and mortality, this might be acceptable.

For the innovative manufacturer, beyond the ethical concerns, a high cut-off value risks capping revenues below what they might be with a lower cut-off value. As Fig. 2e illustrates, while using a high cut-off might achieve more rapid uptake, the peak revenues could be less than using a lower cut-off value. Given the limited number of patients selected with a high cut-off, even higher pricing and greater penetration might not offset the foregone larger number of patients eligible to receive the drug if a lower cut-off were used.

Under most current clinical trial designs and resulting regulatory approvals, providing treatment to patients below the cut-off value would likely be classified 'off-label'. Product manufacturer employees may not suggest such treatment under penalty of fine, imprisonment, or both while payers will likely not reimburse. However, payers do make exceptions, particularly when a respected professional society guideline recommends an off-label use of a therapeutic. Although the current system does not perfectly forbid off-label use, using step-edit and prior authorization utilization tools, it does make it difficult, and increasingly so for expensive drugs, a common feature of stratified medicines.

Implications of a low cut-off

Setting a low cut-off value for a companion diagnostic does not resolve the issues raised by selecting a high cut-off, but merely presents the converse of the high-cut off issues, as shown in Fig. 2 by the left dashed line. Instead of excluding nearly all patients who would not respond, setting a low cut-off value towards the far left of the blue curve of responding patient population scores, includes nearly all patients who will respond; in this case, the companion diagnostic has high clinical sensitivity.

Although few patients who might benefit are denied treatment, the number of nonresponding patients classified as test positive, and so receiving ineffective treatment, increases. The companion diagnostic clinical specificity falls to reflect this compromise.

Ethically, a low cut-off implies knowingly exposing more patients to the therapeutic who will not benefit but will incur its adverse effect risks and delays in seeking other treatment. For a therapeutic with significant, irreversible adverse effects, this might be unacceptable. For a therapeutic with few adverse effects or for a condition with few treatment alternatives, this greater exposure to potential harm, particularly if well communicated to the patient, might be appropriate.

When one follows the left vertical dashed line downward in Fig. 2b–d, one observes that the more modest performance differential (compared with the high cut-off case) leads to more modest improvements in adoption speed, price, and market share. However, the number of treated patients (the number of patients with scores greater than the cut-off) is potentially considerably larger than with the high cut-off case and the performance is possibly

superior to the all-comers (unselected, no companion diagnostic) case.

However, this theoretical larger number of patients might not occur in practice if the efficacy of the drug at the lower cut-off fails to inspire patient interest and use, or if the large number of false positive patients so lowers an individual patient's expectation of response that few patients and the providers choose to pursue that treatment. Paradoxically, a low cut-off in such cases might result in fewer patients receiving the therapeutic than if a higher cut-off had been chosen. In such cases, all stakeholders would seem to lose out, with the possible exceptions of those payers solely focused on net total outlays regardless of net population clinical benefit.

For the innovative manufacturer, setting a low cut-off value with the corresponding larger number of test qualified patients, eventual peak revenues might grow higher but take longer to achieve, and so fail to meet original revenue projections and disappoint investors. If remaining patent life is short, peak revenues might never be attained. A low cut-off value might also run the risk of demonstrating insufficient benefit to gain regulatory approval, particularly if another competitor demonstrates a higher observed efficacy from greater responder enrichment owing to a high cut-off value approach.

Each candidate therapeutic faces unique circumstances of unmet medical need, therapeutic performance, companion diagnostic performance, and market adoption dynamics. General rules of thumb for preferring high or low cut-offs are not obvious either ethically or financially. Each case requires individual analysis to balance its specific factors. Multiple tools and frameworks for evaluating these trade-offs have been presented elsewhere in the literature [14–20].

Behavioral change benefits

Additional potential benefits not shown in Fig. 2 might result from potential behavior changes induced by a stratified medicine: namely, willingness to seek, initiate, and adhere to the treatment regimen. Patients might be encouraged to seek treatment for their condition if a test exists to recommend a therapy. This effect expands the absolute number of patients and so increases the overall market size. Perhaps more importantly, a companion diagnostic prospectively indicating probable response to a therapy might make physicians more inclined to consider and recommend the therapy. By providing higher, but not complete, assurance that the therapy will specifically work for them, the test shifts an individual patient's benefit odds and so helps overcome any barriers faced, from fear to inconvenience. This effect improves therapeutic market share, assuming more than one treatment exists. Even if the therapy is the only available treatment, a CDx might encourage providers and patients to initiate care and so expand the actual penetration of the label-qualified intended-to-treat population.

If the companion diagnostic inspires greater confidence that the therapeutic is the best course for the patient, stratified medicines might also benefit from improved patient adherence. If, in addition, the companion diagnostic also proves useful in monitoring disease progression, treatment effects or both, adherence might improve even more. For example, before the introduction of viral load testing, patients with AIDS were generally poorly adherent to antiretroviral treatments. Effective treatment caused flu-like

symptoms as the body cleared the virus. Believing the drug made them worse, many halted treatment. With the advent of the viral load test, both physician and patient had an independent, objective measure of treatment success, and patient adherence rates soared [21]. Physicians, patients, and payers generally respond well to diagnostics that also monitor therapeutic impact, such as hemoglobin A_{1C} (HBA$_{1C}$) for diabetes [22], or cholesterol levels for cardiovascular risk. Although not yet studied or demonstrated, stratified medicines in many therapeutic areas, not only oncology (e.g., HCV), might similarly experience increased adherence with commensurate benefits for patients, public health, and manufacturer revenues.

Stratified medicine competition

Our discussion thus far has considered the case of a single candidate therapy for a target. Additional dynamics arise if multiple stratified medicines compete for the same target and indication. Fig. 3 shows that, in oncology, many targets have multiple products launched, in clinical development, or both. Based on September 2014 PharmaProjects data, Fig. 3 shows for each of the 298 then active targets the number of unique therapeutic entities launched or being investigated in a clinical trial. Therapeutics whose targets were classified by PharmaProjects as unspecified, not applicable, or simply blank have been excluded. If a therapeutic engaged multiple targets, only the target classified by PharmaProjects as the primary target was plotted.

In total, 181 launched products were approved for 53 targets. Of those 53 targets, only nine had a single therapeutic facing no competitors in the market or under clinical development; 23 targets already had two or more launched products. Another 21 targets with only a single launched product faced at least one candidate therapeutic in development, and 16 of those 21 faced two or more competitors in Phase I, II, or III clinical development.

In addition, 245 targets had one or more candidate therapies in clinical development, but not a launched product. Of those 245, 82 (33%) already had two or more candidate therapeutics in development; and 31 of the 245 targets (12.6%) had four or more candidates in development.

As targets approach the market, the competition increases: 35 targets had no launched product but at least one Phase III development candidate therapy, 21 of those (60%) had three or more candidate therapies in Phase I, II, or III clinical trials. From the candidate therapeutic perspective, of the 141 therapeutics in clinical development for those 35 targets, only 12 candidate therapeutics (8.5%) had a unique niche facing no direct competition for their target.

Therefore, at least in oncology, developers appear likely to face competition not only after they reach market, but also in their quest to be first-in-class. As noted above, companion diagnostics use is also now observed in other areas, such as cardiovascular disease, infectious disease, neurodegenerative disorders, and respiratory ailments. If those trends continue, these competitive dynamics will also likely replay in those indications. With such frequent and pervasive competition, understanding and developing a perspective on how stratified medicine approaches could structure the market would seem important for most development programs.

Under competition, three essentially identical drugs might receive dramatically different labels, incremental cost-effectiveness ratio (ICER)-justified pricing, and market positioning depending on their stratification approach (discussed more fully below). It appears superior to use an imperfect biomarker to none at all. It is less obvious whether patients, payers, and firms prefer the same cut-off values for the companion diagnostic, or even whether each stakeholder a priori prefers the high, low, or perhaps some other CDx cut-off value.
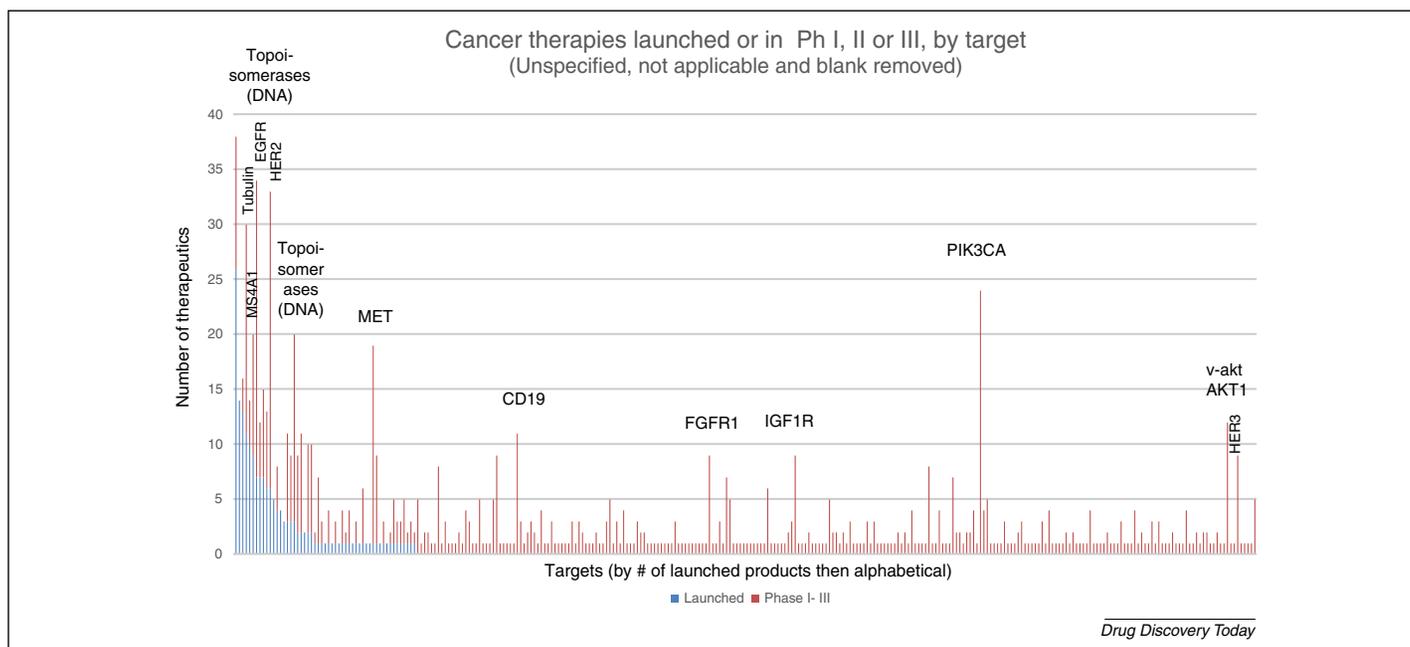

FIGURE 3

Oncology competition by target is common, with over 80% (44 of 53) therapeutic known targets having two or more competing medicines. Many targets with only candidate therapeutics in clinical development also see competition. Based on an analysis of the PharmaProjects pipeline database as of September 2014.

The competing development teams might face a version of the game theory 'prisoner's dilemma' in which the optimal result for patients and all firms would be to select a low or mid companion diagnostic cut-off value but the advantages of a potentially differentiating high efficacy claim might drive developers to select a high cut-off value. If all choose this approach, the overall value might be reduced, with many patients excluded from treatment, but those being treated paying very high prices. However, the potential advantage of a higher cut-off value might prove too alluring, or the fear of a competitor selecting one might drive all to do so. Each situation will depend on the specific facts of the indication, therapeutic, companion diagnostic, and actions by competitors.

Development teams and their firms do not entirely control the selection and use of a CDx. In a 2012 payer survey, over one half of the respondents (five of nine) required evidence of HER2 testing before reimbursing for trastuzumab (Herceptin) and 60% of those (three of five) required reporting the test result, presumably to discourage treatment and reimbursement of CDx-negative patients, even though the HER2 CDx is not a perfect treatment response predictor [23]. In the case of cetuximab (Erbitux), physicians and their professional societies [American Society of Clinical Oncology (ASCO) and National Comprehensive Cancer Network (NCCN)] changed the CDx entirely in 2009 from the epidermal growth factor receptor (EGFR) overexpression CDx on the US Food and Drug Administration (FDA) cetuximab label to KRAS genetic variants (with those not containing KRAS mutations in codon 12 or 13 in their colorectal cancer being suitable for treatment) based on a combination of randomized controlled studies conducted by academics, not the sponsors [24]. Three years later, in 2012, the FDA and sponsors updated the cetuximab label (http://labels.fda.gov/). In response to generic drug companies not being held liable by the US Supreme Court for reporting adverse events, in 2013 the FDA suggested that it should be possible to amend therapeutic labels independently of the original sponsor's involvement, and reopened the comment period for the proposed change again in 2015 [25,26]. If broadly implemented, this might also allow diagnostic companies, academics, or others to present CDx data to amend therapeutic labels. Once launched, therapeutic sponsors might not retain complete control of the qualifying populations or even the CDx used.

A competitive example

To illustrate implications of stratified medicine competition, consider the hypothetical, but plausible situation of three oncology candidate therapies in a race to be first- and best-in-class to treat the same novel target for which a candidate companion diagnostic for likely drug responders has been identified. For this example, we assume that the three candidate drugs are essentially similar in their chemical structure, pharmacology, formulation, therapeutic index, and other relevant properties. This enables us to focus on the decisions and implications regarding whether and how to use the candidate companion diagnostic.

Fig. 4 demonstrates three possible choices facing the firms based on these assumptions. Firm A chooses an all-comers approach for
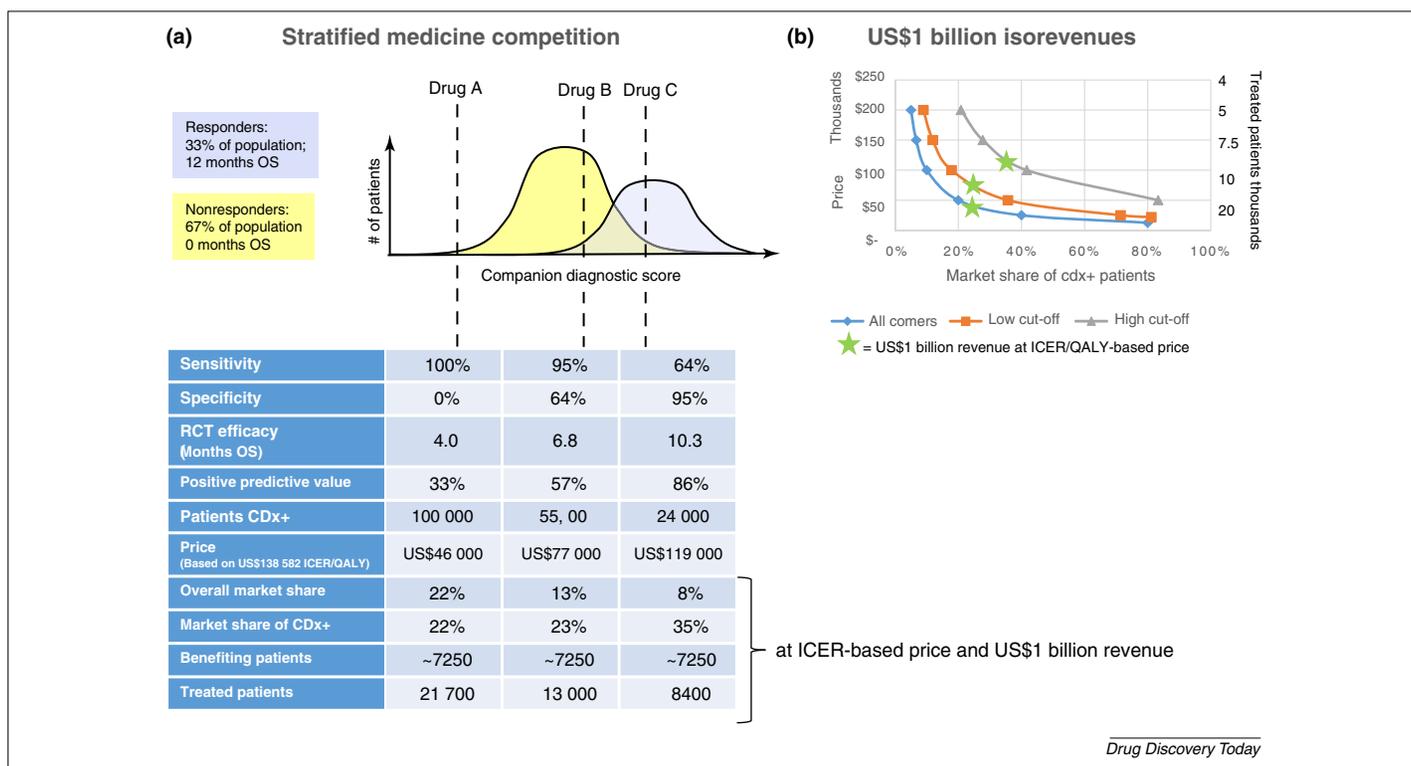

(a) Stratified medicine competition

	Drug A	Drug B	Drug C
Sensitivity	100%	95%	64%
Specificity	0%	64%	95%
RCT efficacy (Months OS)	4.0	6.8	10.3
Positive predictive value	33%	57%	86%
Patients CDx+	100 000	55, 00	24 000
Price (Based on US$138 582 ICER/QALY)	US$46 000	US$77 000	US$119 000
Overall market share	22%	13%	8%
Market share of CDx+	22%	23%	35%
Benefiting patients	~7250	~7250	~7250
Treated patients	21 700	13 000	8400

Responders: 33% of population; 12 months OS

Nonresponders: 67% of population 0 months OS

at ICER-based price and US$1 billion revenue

(b) US$1 billion isorevenues

— All comers — Low cut-off — High cut-off
★ = US$1 billion revenue at ICER/QALY-based price

Drug Discovery Today

FIGURE 4

Stratified medicine competition. **(a)** Pharmacologically similar therapeutics addressing the same target but applying companion diagnostics differently, if at all, might appear distinct in their randomized clinical trial (RCT)-reported efficacies and response rates, as well as incremental cost effectiveness ratio (ICER)-justified prices and commercially reported market sizes. **(b)** Market shares of the companion diagnostic selected population must increase compared with the all-comers therapeutic to achieve US$1 billion revenue. In this oncology disease example, achieving the same number of treated patients at a given price requires as much as four times the eligible patient market share for the stratified medicine as for the all-comers medicine. *Abbreviations*: CDx+, companion diagnostic test positive; OS, overall survival.

Reviews • KEYNOTE REVIEW

its Drug A, which does not use the companion diagnostic. Firms B and C both choose to pursue a companion diagnostic approach, but set different cut-off values. Firm B chooses a low diagnostic cut-off value for its Drug B, whereas Firm C sets a high diagnostic cut-off for its Drug C.

A stratified medicine approach holds the potential for smaller, faster, and less expensive clinical development because of the higher anticipated therapeutic effect owing to companion diagnostic use. However, the approach also requires the development of the diagnostic and its associated risk of failure, more complex patient recruitment, and possibly no savings in trial size because of the potential need to examine patients with negative test results and the continuing need to develop an acceptably large patient safety database. For this example, we consider that the factors approximately off set one another.

To keep the mathematics simple, we further assume that early translational medicine evidence suggests that 33% of the 100 000 patients with this condition respond to treatment addressing the target. Assume that each responder gains 12 months overall survival compared with the standard of care, and that the remaining patients receive zero incremental benefit. The Drug A clinical trial design enrolling all-comers meeting the disease and staging criteria would then be expected to obtain an average clinical benefit of 4 months overall survival (the weighted average of the one-third of patients who respond with the two-thirds of patients who do not).

Even though no companion diagnostic was used, one can consider that an all-comers trial has 100% sensitivity (it selects all patients who might respond), and has 0% specificity (it excludes none who will not benefit). Another diagnostic metric, Positive Predictive Value (PPV), states the fraction of CDx+ patients that do respond. More technically, PPV measures the number of true positives as a portion of all those who test positive. The all-comers PPV can be said to be 33% (the responder prevalence rate).

As Firm A anticipates launching Drug A, it faces the blue isorevenue in Fig. 4b. The all-comers label supported by its clinical trial will allow its marketing to suggest that all 100 000 patients with the condition are eligible for treatment. To achieve US$1 billion blockbuster-level sales, Drug A must achieve 20% market share (be used by 20 000 patients) at a US$50 000 1-year drug regimen price with its 4 months' overall survival improvement.

Firm A could choose any other price, and the blue isorevenue indicates what market share Drug A must achieve for US$1 billion in sales. For instance, at a price of US$200 000, Drug A must be used, and paid for, by 5000 patients, which conveniently equals 5% market share in this example. At a price of US$12 500, 80% market share must be achieved (80 000 treated and paid patients) to reach US$1 billion in sales. A recent literature review suggests that, for oncology therapeutics in the USA, the mean ICER using the quality-adjusted life year (QALY) for the health benefit metric is US$138 582/QALY [27]. If the ICER guides payer reimbursement, Firm A might expect a Drug A price of approximately US$46 000 (one-third of the mean ICER based on an expected average 4 months' overall survival improvement). Per Fig. 4b, at that price, Drug A would need to achieve 22% market share to generate US$1 billion in annual revenue, as indicated by the green star.

Firm B chooses to use a companion diagnostic approach that selects nearly all patients who will respond by setting a low CDx cut-off. In this hypothetical case, the cut-off is set to generate 95%

sensitivity (95% of responders will receive a positive test score; approximately 31 500 of 33 000). The hypothesized test is assumed to be good, but not perfect. The low cut-off value results in a 64% specificity (64% of nonresponders will test negative; approximately 43 000 of 67 000). This means that 36% of the nonresponders will test positive (approximately 24 000 of 67 000). For an oncology companion diagnostic, this is a superior performance, especially with such a high negative predictive value of 97% (43,000 of 44,500 CDx negative patients are true negatives). One of the more powerful companion diagnostics known, the KRAS test for detecting likely responders and nonresponders to cetuximab (Erbitux) in colorectal cancer, has an estimated 75% sensitivity and 35% specificity [28].

Using the same patient response assumptions as for Drug A, the Drug B clinical trial would be expected to show a mean treatment benefit of 6.8 months' overall survival (31 500 with 12 months' additional survival and 24 000 with 0 months' additional benefit), which is 70% longer than that for Drug A (2.8 months longer than 4.0 months). Of the patients testing positive (CDx+), 57% would be expected to respond. This is 24 percentage points higher than the 33% of treated patients responding to Drug A with its unenriched population.

However, the label for Drug B will specify that it should only be used for those who test positive. Given the postulated 95% sensitivity, 64% specificity, and 33% responder prevalence, nearly 56% of patients with the disease will test positive (coincidentally similar to the 57% PPV). Instead of the 100 000 eligible patients for Drug A, only the nearly 56 000 CDx+ patients will be eligible for Drug B. To achieve US$1 billion in revenues, Firm B faces the middle isorevenue in Fig. 4b. Firm B will need a higher price for Drug B, a higher share of the patients testing positive, or both. Using the US$138 582 QALY-based oncology ICER, Firm B might expect a price of US$77 000. This is higher than the Drug A price because of the higher overall survival benefit in the intended-to-treat population. Drug B would need to be used, and paid for, by 13 000 patients (23% market share of the 56 000 eligible patients) to achieve US$1 billion in annual revenues. Note that the Drug B share of all patients with the disease is only 13% compared with the 22% share of all patients with the disease Drug A required to reach US$1 billion.

Recall that Drug B does not perform any better than Drug A in those who respond (12 months' overall survival) and that both drugs are competing in the same disease indication of 100 000 oncology patients of whom only 33 000 will respond to treating this target. By using a companion diagnostic with a low cut-off, Firm B sees its value based, ICER/QALY justified price increase by over US$30 000 per patient. Payers pay the same amount, US$1 billion for about the same number of patients benefiting: approximately 7200 for both drugs. However, with Drug B, nearly 9000 fewer patients receive treatment that does not benefit them, assuming that both firms achieve US$1 billion in revenues and an ICER-based reimbursement.

In clinical practice, Firm A might find it difficult to market and compete versus Firm B because of the difference in expected patient overall survival from the use of the companion diagnostic. From a public health perspective, an ordered market with selection bias in which more patients choose Drug B with the companion diagnostic could make the realized benefits from Drug A even

lower. Given that both work on the same biological target, Drug A might see a patient population in practice that is responder depleted by Drug B. Thus, rather than seeing an all-comers population with 33% responders, Drug A might see a remaining all-comers population with only 22%, 15%, or even lower levels of responders if Drug B achieves 60%, 80%, or even higher levels of market share among its companion diagnostic positive population. If therapeutic ordering occurred, real-world payer studies might therefore show Drug A value substantially below the already lower average benefit seen in its clinical trials, whereas Drug B benefit would align with the clinical trial observation.

Firm C chooses to use a companion diagnostic approach that excludes nearly all patients who will not respond by setting a high diagnostic cut-off. In this hypothetical case, the cut-off is set to generate 95% specificity [95% of nonresponders (approximately 63 500 of 67 000) will receive a negative test score (CDx–)]. As shown by the far right vertical line in Fig. 4a, the high cut-off also excludes some patients who would benefit from treatment. In this hypothetical case, the corresponding sensitivity is 64% (approximately 21 000 of approximately 33 000 patients who would respond will score CDx+); 36% of patients who might benefit (approximately 12 000 of those approximately 33 000) will receive a negative companion diagnostic test result.

Using the same 33% responder prevalence and 12 month overall survival benefit assumptions as for the other two postulated drugs, the Drug C clinical trial would be expected to show a mean treatment survival benefit 2.5 times greater than Drug A: 10.3 months' increased overall survival (21 000 with 12 months additional survival and 3500 with 0 months' additional benefit) versus 4.0 months. The power of the high cut-off is demonstrated by its high PPV: 86% of patients testing positive would be expected to respond. This is 53 percentage points higher than the 33% response rate expected for Drug A. The Drug C reported overall survival benefit also will probably be 51% longer (3.5 months longer than 6.8 months) than that for Drug B.

However, the fraction of the disease patient population testing positive (the selected population) is much smaller than for Drug B: only 24 000 of the 100 000 patients. The smaller selected population results both from better excluding nonresponders (reducing false positives) and from excluding many patients who would have benefited from Drug C (increasing false negatives). Firm C has chosen a highly enriched, but smaller population strategy.

Competitively, Drug C will have superior efficacy outcome evidence and label claims, even though its underlying target and drug properties are essentially identical to the other two drugs. Based upon the US$138 582 ICER oncology rate and the 10.3 month overall survival benefit, Firm C might expect US$119 000 Drug C reimbursement; over 150% more than Drug A and 50% more than Drug B. At that price, Drug C needs to achieve 35% market share of the selected population to generate US$1 billion in annual revenue. As with the other two drugs, approximately 7200 patients will respond, but because of the high PPV, only 8400 patients in total will receive the therapeutic. Drug B needs to treat 13 000 and Drug A nearly 22 000 to reach the same number of responding patients at a payer cost of US$1 billion.

However, because of the high CDx cut-off, Drug C could never reach approximately 12 000 patients who could benefit, even if Drug C achieved 100% market share of its CDx-selected population.

Recall that the 64% CDx sensitivity indicates that 36% of the 33 000 potential responders will receive a negative score and be denied Drug C, resulting in approximately 12 000 lost person years of benefit. Drug B would reach 95% of all potential responders if it achieved 100% market share of the CDx-selected population, missing only approximately 1500 patients and so 1500 QALYs. Given that Drug A does not use a CDx, it could reach all potential responders. Not knowing what alternative treatments might exist for the nonresponders, we do not estimate the off-setting QALY losses from nonresponders being given ineffective care and forgoing other more effective treatment.

Recognizing that all three drugs are in fact nearly identical, a payer faces the following costs and benefits (not off-set for nonresponder harm) for providing ubiquitous access to all qualifying patients via a single drug formulary with no patient co-payment. Choosing Drug A would cost US$4.6 billion per year and provide 33 000 QALYs per year. Drug B would cost US$4.3 billion per year and provide approximately 32 000 QALYs per year. Drug C would cost US$2.9B billion per year, but only provide approximately 21 000 QALYs. Given the assumed ICER-based pricing, the payer cost per QALY is identical at approximately US$138 000/QALY.

Drug B and Drug C also require the use, and cost, of the CDx. The CDx cost is not considered here because of the usually relatively low cost of diagnostics, but a payer should technically reduce the ICER-based drug reimbursement to account for the CDx reimbursement.

A savvy payer or integrated provider might recognize that the drugs are essentially identical and, therefore, use the CDx with Drug A (or negotiate discounts with Firms B and C to match Drug A pricing). Using the low cut-off to reach nearly all responders but with Drug A pricing would lower the ICER-based price to approximately US$81 000 and total cost to US$2.6 billion to achieve the nearly perfect health benefit of approximately 32 000 QALY/year and save the payer US$1.7 billion, over 35% compared with the ICER-justified Drug B price. Such actions would of course reduce incentives for future developers to develop stratified medicines if in the end they still only receive the all comers, nonstratified value.

The companion diagnostic developer faces lower revenue prospects. Even assuming a high reimbursement to the clinical laboratory of US$400 of which the CDx developer receives 50% for the test kit, the entire CDx testing market is only US$20 million (for a selection, but not monitoring CDx) compared with a market measured in billions of dollars for the therapeutics. If the CDx uses a standard technology, such as an immunoassay whose kit prices are often effectively limited to US$25 or less per test, the total market falls to merely US$2.5 million. Given rapid competition, most CDx developers will receive half or less of the potential testing market, and not all patients will be tested, making these already comparatively small amounts even smaller. Even a manufacturer test price of US$2000 per patient with 100% testing captured by the firm only produces US$200 million per year.

In summary, under plausible market competition characteristics, three essentially identical drugs receive dramatically different labels, ICER-justified pricing, and market positioning, whereas under all circumstances, the CDx developer likely receives 1% or less of the revenue flowing to the therapeutic.

Valuable diagnostics poorly valued

Companion diagnostics create the difference in value between an all-comers therapeutic and the stratified medicine. As part of the stratified medicine combination, companion diagnostics select patients who are more likely to benefit, encourage patient confidence to initiate and adhere to treatment, reduce patient nonresponse treatment opportunity cost, speed clinical adoption, and perhaps increase reimbursement for their associated therapeutics while better focusing payer spending.

Diagnostics have a history of relatively low reimbursement in which the exceptions, such as high end imaging (CAT, MRI, PET) or a few genetic tests such as Oncotype DX®, prove the rule. Even so, the high-priced tests are relatively inexpensive at a few thousand dollars compared with the associated surgeon fees, hospital bed day costs, or branded specialty medicines. Although companion diagnostics create the clinical value, in general the financial value is captured by therapeutic developers and perhaps payers.

Multiple factors drive low diagnostic pricing and margins. Many diagnostic firms pursue business models that employ non-exclusive use of specific tests. Diagnostic instrument companies, test aggregators (especially in next-generation sequencing, such as Caris Life Sciences, Foundation Medicine, and Pathway Genomics), and academic medical centers mostly use nonproprietary markers, for which they pay low or no royalties or other compensation to those who discovered the biomarker or developed it for clinical insight.

Biopharmaceutical companies often consider biomarkers and diagnostics development precompetitive activities that they then perform through consortia and make broadly available. Although precompetitive to drug developers, such consortia of course directly compete with diagnostics developers.

The US Supreme Court has recently ruled in the AMP v. Myriad Genetics [29], and Mayo Collaborative Services v. Prometheus Laboratories [30] cases that some molecular diagnostics are ineligible for patent protection because, admittedly oversimplifying the rulings, the molecular phenomena underlying them are laws of nature rather than inventions.

Government regulators, such as the FDA, have used their regulatory discretion to require first-to-market companion diagnostics to obtain FDA approval, often via the premarket application (PMA) process, which requires significant development time and costs. Later entrants might also require PMAs or, perhaps, be allowed the swifter 510(k) process. Laboratory-developed tests (LDTs) offered as a service rather than a test kit can also be used as a first-to-market CDx, provided that the LDT has developed the evidence required to obtain FDA approval. In Europe, CDx might only require CE Marking. Thus, innovators face a variety of possible technical and regulatory paths as they develop products for global distribution.

To date, the FDA has also used its regulatory discretion to not remove competing nonapproved diagnostics from the market. The resulting product variety in practice can create challenges for the associated biopharmaceutical company to aid in diagnostic market promotion for fear of indirectly promoting unapproved diagnostics and so technically off-label drug use. Payers such as Medicare have long established technology platform reimbursement approaches that do not consider the clinical value of a test. In combination, these factors lead to nearly immediate follow-on competition and low reimbursement for diagnostic tests. This greatly reduces incentives for their development [31]. Moreover, any movement towards value-based reimbursement to encourage diagnostic development will likely need to avoid generating financial windfalls for existing diagnostics to gain payer support.

Industry participants and academics have suggested alternative diagnostic innovation incentives, including a move towards value-based rather than technology-based reimbursement for select diagnostics as well as market exclusivity and development support subsidies, among others [32–36]. Legislation was introduced in the US House of Representatives in 2013 to address concerns about diagnostic innovation, but was not enacted (https://www.congress.gov/bill/113th-congress/house-bill/2085). In January 2015, the Energy and Commerce Committee issued draft legislation for 21st-Century Cures that, among other goals, seeks to streamline diagnostic regulation, but proposes little to increase their reimbursement or intellectual property protection.

While waiting for more systemic change, innovative diagnostic organizations are finding opportunities in aggregating genetic diagnostic tests into panels for oncology and other disease areas (e.g., Foundation Medicine, Quest, LabCorp, and academic medical centers). Biopharmaceutical companies sponsor new companion diagnostic development associated with their new candidate therapeutics. Federally sponsored research continues to discover putative biomarkers, but usually lacks the resources for large-scale confirmatory clinical trials. Consortia, such as The Biomarkers Consortium, pool resources to develop biomarkers sufficiently for regulatory and clinical use in fields such as cancer, immunology, metabolic disorders, and neuroscience (http://www.biomarkersconsortium.org/whatwedo.php). These admirable efforts usually focus on developing the science for biomarkers for new precision medicine therapeutics or to improve drug development generally. However, they typically do not address precision medicine needs for legacy therapies that comprise most patient treatments and medical costs.

Payer perspective

Payers have expressed concern regarding the affordability of stratified medicine because of increases in the number of high-priced drugs as diseases fragment into orphan and ultra-orphan indication population sizes [37]. However, as the NNT analysis of cardiovascular disease demonstrated above, the cost of a single avoided event from a broadly empiric treatment, despite low per patient per year costs, might not be much less than the cost of new stratified medicines. The discussion in the 'Competitive example' section above demonstrates how improved selection of patients who are most likely to benefit from a therapeutic can dramatically increase the value per treated patient and so the ICER-based price per dose.

Stratified medicines tend to focus on significant unmet needs, as indicated by their disproportionate priority review designation and qualification for accelerated or breakthrough medicine approval [38]. Payers, and the governments and employers who fund them, now face the challenges of paying for the emerging successes in meeting those unmet medical needs. Unsurprisingly, incremental improved public health is likely to have incremental healthcare costs, even if a stratified medicine approach proves reasonably efficient at identifying those who will benefit.

Israel provides an example that explicitly struggles with this issue through a Basket Committee that selects which new drugs, diagnostics, and devices will be made available through its annual update of the standard health services basket funded by special budget allocation approved by the Israeli Treasury, with increases usually limited to the rate of healthcare inflation, GDP growth, or some combination thereof [39,40]. Explicit refusal to include new health technologies in the basket occurs routinely.

Delivering existing public health status at lower cost is also a social good that stratified medicine could help achieve. Both patients and payers would benefit from companion diagnostics for 'met medical needs' whose only partially effective treatments could be stratified to reduce costs for similar public health benefit. Drug manufacturers have minimal if any incentives to stratify generic or soon-to-be generic drugs. Diagnostic companies have neither resources nor incentives to invest in the required clinical evidence development.

Payers, or the governments that often fund them, would seem to have the most obvious incentive to invest in companion diagnostic development that would lower their medical loss ratios for legacy treatments. Payers once made such investments in healthcare technologies, but have abandoned that heritage. In 1963, the Health Insurance Plan of Greater New York, now an EmblemHealth company, sponsored the first mammogram screening trial (HIP study) [41–43]. Payers today generally do not invest in creating technologies to improve public health. However, it is clearly in government payers' interest to do so. Private payers, whether health insurers or the more specialized pharmacy benefit managers (PBMs), might fear free rider or scale issues. Free rider issues could be addressed either through precompetitive consortia efforts or by using these technologies to help differentiate their insurance product offerings, just as their data analytics and disease management capabilities already attempt to do. Regarding scale, the number of lives covered by the largest US insurers is comparable to the populations of European nations.[1]

The contrast in government behavior between military (and space) technology development and health technology development is striking. For military technology, major governments coordinate integrated requirements specification through development and deployment supply chains. They specify their desires for weapons systems, fund companies to develop prototypes, competitively test those prototypes, request bids for supply, and coordinate the global dissemination of the resulting weapons to their armed services and other nations. The private sector has crucial, and profitable, roles throughout that chain.

In healthcare, even national government behavior is more fragmented and uncoordinated, despite governments funding basic research at the beginning, paying for the technology at the end, and regulating its development throughout. Even in the more privatized healthcare of the USA, the Federal Government now provides healthcare directly for 34% of the population through Medicare, Medicaid, active military healthcare, and veterans' healthcare (http://www.census.gov/hhes/www/hlthins/data/incpovhlth/2013/tables.html). Not included in those 2013 Census figures are lives covered under federal employee health benefits (2.7 million employees plus their families) (https://www.opm.gov/policy-data-oversight/data-analysis-documentation/federal-employment-reports/historical-tables/total-government-employment-since-1962/), prisoner healthcare (210 000 inmates), and now the Affordable Care Act insurance subsidies. Adding those plus the tax subsidies (corporate and individual) for employer-provided healthcare insurance and health savings accounts increases the level of US Government funding for healthcare even further. With such a large interest, more directed, substantial government investments in companion diagnostics development to enable better stratification of legacy treatments would appear justified.

Concluding remarks

Stratified medicine tightens the links among science, the clinic, and the marketplace. Setting the companion diagnostic cut-off value is a crucial shared connection among all three, with no easy rule of thumb to guide the choice. Each stratified medicine opportunity faces unique facts and circumstances that require balancing ethical, scientific, and financial concerns.

Today, stratified medicine economic incentives favor new medicine developers and the patients they serve. Payers benefit from more efficient new treatment for unmet medical needs, but likely face increased total costs for the resultant increase in overall public health, with little or no offsetting cost savings from companion diagnostics better stratifying legacy treatments. Diagnostic companies are generally paid for their services, but not sufficiently to invest independently in companion diagnostic development. Current economics do not reliably signal true healthcare needs to therapeutic and diagnostic developers, and even less so to the discovery scientists at the beginning of the innovation value chain.

Improving the stratified medicine innovation chain through better economics requires incremental, but significant changes. Greater direct payer sponsorship of medical technology development has precedent in both civilian and military contexts, but seems unlikely for stratified medicine in the near term. Proponents of changes already occurring, such as alternative payment methods and accountable care organizations, hope that they will better connect healthcare decision making with healthcare technology providers. The introduction of improved information technologies from electronic health records, big data analytics, patient wearable devices, and improved data sharing in the sciences also promise improvements.

Companion diagnostics and the stratified medicines that they enable are a growing category of new and legacy therapies in oncology and other disease areas. Their ultimate success depends upon more than scientific discovery. They unite clinical benefits,

---

[1] UnitedHealth Group and Express Scripts International each serve 85 million lives, about the same as the population of Germany (http://www.unitedhealthgroup.com/About/Default.aspx; http://blogs.wsj.com/pharmalot/2015/01/06/the-big-issue-has-not-been-choice-but-access-express-scripts-miller-explains/). Anthem serves 37 million people (http://www.antheminc.com/prodcontrib/groups/wellpoint/@wp_news_main/documents/wlp_assets/pw_e226844.pdf) about the size of Poland and somewhat smaller than Spain. Humana has 13 million members, about the size of Belgium. Kaiser Permanente serves 9.1 million (http://share.kaiserpermanente.org/static/kp_annualreport_2013/index.html#by-the-numbers), which is slightly more than the populations of Switzerland or Austria.

ethical choices, and economic incentives in ways that significantly accelerate decision timing, decrease therapeutic outcome uncertainty, shift competition, and might increase ICER-justified product prices. Mechanisms to create, determine, and share value among all stakeholders from patients, providers, and payers to regulators, developers, and discovery scientists must also advance.

References

1 US Food and Drug Administration (2015) *Remicade (infliximab) Label, Revised: January 2015*. US Food and Drug Administration

2 Helwick, C. (2014) Pathologic complete response: understanding the subtleties. *ASCO Post 5* October 15

3 Newman, D. (2013) Statin drugs given for 5 years for heart disease prevention (with known heart disease). *TheNNT.com* November 2

4 Ridker, P.M. (2009) The JUPITER Trial: results, controversies and implications for prevention. *Circ. Cardiovasc. Quality Outcomes* 2, 279–285

5 Consumer Reports (2014) *Evaluating Statin Drugs to Treat: High Cholesterol and Heart Disease, Comparing Effectiveness, Safety and Price, Consumer Reports*.

6 Schabert, V.F. *et al.* (2012) Annual costs of tumor necrosis factor inhibitors using real-world data in a commercially insured population in the United States. *J. Med. Econ.* 15, 264–275

7 Simon, R. (2008) Designs and adaptive analysis plans for pivotal clinical trials of therapeutics and companion diagnostics. *Expert Opin. Med. Diagn.* 2, 721–729

8 Taube, S.E. *et al.* (2009) A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J. Natl. Cancer Inst.* 101, 1453–1463

9 Phillips, K.A. (2006) Diagnostics and biomarker development: priming the pipeline. *Nat. Rev. Drug Discov.* 5, 463–469

10 Trusheim, M.R. and Berndt, E.R. (2012) Economic challenges and possible policy actions to advance stratified medicine. *Personal. Med.* 9, 413–427

11 Howard, D.H. *et al.* (2015) Pricing in the market for anticancer drugs. *J. Econ. Perspect.* 29, 139–162

12 Phillips, K.A. *et al.* (2009) Clinical practice patterns and cost-effectiveness of HER2 testing strategies in breast cancer patients. *Cancer* 115, 5166–5174

13 Ferrusi, I.L. *et al.* (2013) Closing the personalized medicine information gap: HER2 Test documentation practice. *Am. J. Managed Care* 19, 838–844

14 Trusheim, M.R. *et al.* (2011) Quantifying factors for the success of stratified medicine. *Nat. Rev. Drug Discov.* 10, 817–833

15 Trusheim, M.R. *et al.* (2014) The Janus initiative: a multi-stakeholder process and tool set for facilitating and quantifying adaptive licensing discussions. *Health Policy Technol.* 3, 241–247

16 NICE. (2013) *Gene Expression Profiling and Expanded Immunohistochemistry Tests to Guide the Use of Adjuvant Chemotherapy in Early Breast Cancer Management: MammaPrint, Oncotype DX, IHC4 and Mammostrat.* National Institute for Health and Care Excellence

17 Faulkner, E. (2012) Challenges in the development and reimbursement of personalized medicine-Payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR personalized medicine special interest group. *Value Health* 15, 1162–1171

18 Annemans, L. *et al.* (2013) Current methodological issues in the economic assessment of personalized medicine. *Value Health* 16, S20–S26

19 Jeter, E. *et al.* (2014) *MolDx clinical test evaluation process (CTE)P M00096 Version 2.0.* Palmetto GBA

20 Deverka, P. *et al.* (2013) *Evaluation of Clinical Validity and Clinical Utility of Actionable Molecular Diagnostic Tests in Adult Oncology.* Center for Medical Technology Policy

21 Bonner, K. *et al.* (2013) Viral load monitoring as a tool to reinforce adherence: a systematic review. *J. Acquir. Immune Defic. Syndr.* 64, 74–78

22 Cagliero, E. *et al.* (1999) Immediate feedback of HbA1c levels improves glycemic control in type 1 and insulin-treated type 2 diabetic patients. *Diabetes Care* 22, 1785–1789

23 Cohen, J. *et al.* (2013) Clinical and economic challenges facing pharmacogenomics. *Pharmacoecon. J.* 13, 378–388

24 Carson, B. (2009) KRAS testing: optimizing cancer therapy. *Biotechnol. Healthcare* 6, 7–9

25 Federal Register (2013) *Supplemental Applications Proposing Labeling Changes for Approved Drugs and Biological Products.* Federal Register

26 Federal Register (2015) *Supplemental Applications Proposing Labeling Changes for Approved Drugs and Biological Products; Public Meeting; Request for Comments; Reopening of the Comment Period.* Federal Register

27 Bae, Y.H. and Mullins, C.D. (2014) Do value thresholds for oncology drugs differ from nononcology drugs? *J. Manag. Care Pharm.* 20, 1086–1092

28 Westwood, M. *et al.* (2014) *KRAS Mutation Testing of Tumours in Adults with Metastatic Colorectal Cancer: A Systematic Review and Cost-Effectiveness Analysis.* NIHR Journals Library

29 Supreme Court of the United States. Ruling No. 12-398. Association for Molecular Pathology *et al. v.* Myriad Genetics Inc. Decided June 13, 2013

30 Supreme Court of the United States. Ruling No. 10-1150. Mayo Medical Laboratories, *et al. v.* Prometheus Laboratories, Inc. Decided March 20, 2012

31 Trusheim, M.R. *et al.* (2013) Uncertain prognosis for high-quality diagnostics: clinical challenges, economic barriers and needed reforms. *Pharmacogenomics* 14, 325–334

32 Aspinall, M.A. and Hamermesh, R.G. (2007) Realizing the promise of personalized medicine. *Harvard Business Rev.* 85, 108–117

33 Cohen, J.P. and Felix, A.E. (2014) Personalized medicines's bottleneck: diagnostic test evidence and reimbursement. *J. Personal. Med.* 4, 163–175

34 Personalized Medicine Coalition (2014 July) *The Future of Coverage and Payment for Personalized Medicine Diagnostics*, Personalized Medicine Coalition.

35 Miller, I. *et al.* (2011) Market access challenges in the EU for high medical value diagnostic tests. *Personal. Med.* 8, 137–148

36 Garau, M. *et al.* (2013) Can and should value-based pricing be applied to molecular diagnostics? *Personal. Med.* 10, 61–72

37 Aronson, N. (2015) Making personalized medicine more affordable. *Ann. N.Y. Acad. Sci.* Published online February 27, 2015, http://dx.doi.org/10.1111/nyas.12614

38 Hu, S.X. *et al.* (2013) Defining and quantifying the use of personalized medicines. *Nat. Rev. Drug Discov.* 12, 896–897

39 Sax, P. (2014) The shaping of pharmaceutical governance: the Israeli case. *Isr. J. Health Policy Res.* 3, 16

40 Golan, O. and Hansen, P. (2012) Which health technologies should be funded? A prioritization framework based explicitly on value for money. *Isr. J. Health Policy Res.* 1, 44

41 Randal, J. (2000) After 40 years, mammography remains as much emotion as science. *J. Natl. Cancer Inst.* 92, 1630–1632

42 Gold, R.H. *et al.* (1990) Radiologic History Exhibit: highlights from the history of mammography. *RadioGraphics* 10, 1111–1131

43 Shapiro, S. (1997) Periodic screening for breast cancer: the HIP randomized controlled trial. *J. Natl. Cancer Inst. Monogr.* 22, 27–30

44 Phillips, C. (2009) *What is Cost-Effectiveness. Health Economics* (2nd edn), Hayward Medical Communications