NBER WORKING PAPER SERIES

AN OVERVIEW OF THE STRATIFIED ECONOMICS OF STRATIFIED MEDICINE

Mark R. Trusheim
Ernst R. Berndt

Working Paper 21233
http://www.nber.org/papers/w21233

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2015

This research was not sponsored. The opinions expressed herein are those of the authors, and are not necessarily those of the Massachusetts Institute of Technology. This manuscript is forthcoming in Drug Discovery Today (http://www.drugdiscoverytoday.com), with publication tentatively scheduled for late 2015. Additional information regarding the journal and access to the final publication when released can be obtained through Science Direct at http://www.sciencedirect.com/science/journal/13596446. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

An Overview of the Stratified Economics of Stratified Medicine
Mark R. Trusheim and Ernst R. Berndt
NBER Working Paper No. 21233
June 2015
JEL No. D04,D21,I11,I18,L11,L65

ABSTRACT

The economics of stratified medicine depend critically on setting the cut-off score of the companion diagnostic (CDx). This action integrates scientific, clinical, ethical and commercial considerations, and simultaneously determines the value of the stratified medicine to developers, providers, payers and patient. Setting a high cut-off ensures a larger response by excluding more non-responders but also denies treatment to patients who would respond. This creates ethical and clinical concerns, and limits market size. Setting a low cut-off includes more patients who can benefit but includes more non-responders with commensurate costs, side effects and lost time. CDx's capture little value under current reimbursement and exclusivity protections. Combined with low CDx investment incentives for generic drug manufacturers, little CDx development occurs for older legacy drugs. Therefore payers face an asymmetric situation of novel stratified medicines raising public health and payers' costs, but no CDx's for legacy treatments to reduce costs. It would be in payers' interests to rediscover their heritage of direct investment in diagnostic development.

Mark R. Trusheim
Sloan School of Management
MIT
50 Memorial Drive
Cambridge, MA 02142
trusheim@mit.edu

Ernst R. Berndt
MIT Sloan School of Management
100 Main Street, E62-518
Cambridge, MA 02142
and NBER
eberndt@mit.edu

Introduction and Background

Stratified medicine tightens the links among science, the clinic and the marketplace. Setting the companion diagnostic cut-off value is a critical connection among all three with no easy rule of thumb to guide the choice. Each stratified medicine opportunity faces unique facts and circumstances that requires balancing ethical, scientific and financial concerns. Beyond this selection, the underlying health economics impact dynamics among competing developers, the current bias of stratified medicines for only new 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 auto-immune diseases such as Crohn's disease and ulcerative colitis are approximately 25-40% [1]. For many oncology therapeutics perhaps only 20-30% respond and have their lives extended (as measured by months of overall survival from treatment initiation) [2]. For large population therapeutics such as statins, and others, the proportion that receive a health outcome benefit such as an avoided heart attack or death is far lower. According to a review performed by David Newman, the Number Needed to Treat (NNT) by a statin for five years to save a life is 83 and the NNT to avoid a single heart attack is approximately 39 – for every 39 patients with known heart disease treated continuously with a statin for five years, one non-fatal 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 side effects in those 38. Unfortunately currently we cannot predict which individual would avoid an event if s/he lowered their cholesterol, so we encourage all to take their medicine.

At $120/yr/patient for a generic statin [5], it costs a total of $50,000 over five years to save a single life from a cardiovascular event, not counting the physician office fees or other costs. As recently as 2013 with branded statins, the cost was four times higher at $208,000 [6]. In comparison, the TNFα inhibitors NNT for clinical remission response is about three (costing about $45,000-$75,000 per clinical remission patient year [7]). The NNTs for oncology drugs are about three to five based on the response rates cited above.

Stratified medicine approaches promise to better target responders and so reduce the NNT for a therapeutic. This has financial implications for developers and payers as well as health benefits to patients and the public.

This review focuses on the post-launch effects of stratified medicines rather than the impacts on clinical development – a focus of other research [8,9,10,11].

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 (CDx) is used to identify a patient sub-population having a differential expected clinical response.

An ideal companion diagnostic perfectly identifies and distinguishes treatment responders from those who will not (Figure 1). In the top left portion, the two curves represent all the patients with the disease who might be treated with the drug. The left curve represents the patients who will not respond to the therapeutic, with the low horizontal line indicating this zero response. The right curve represents those patients who will respond, with the high horizontal line denoting these patients' consistently high response.

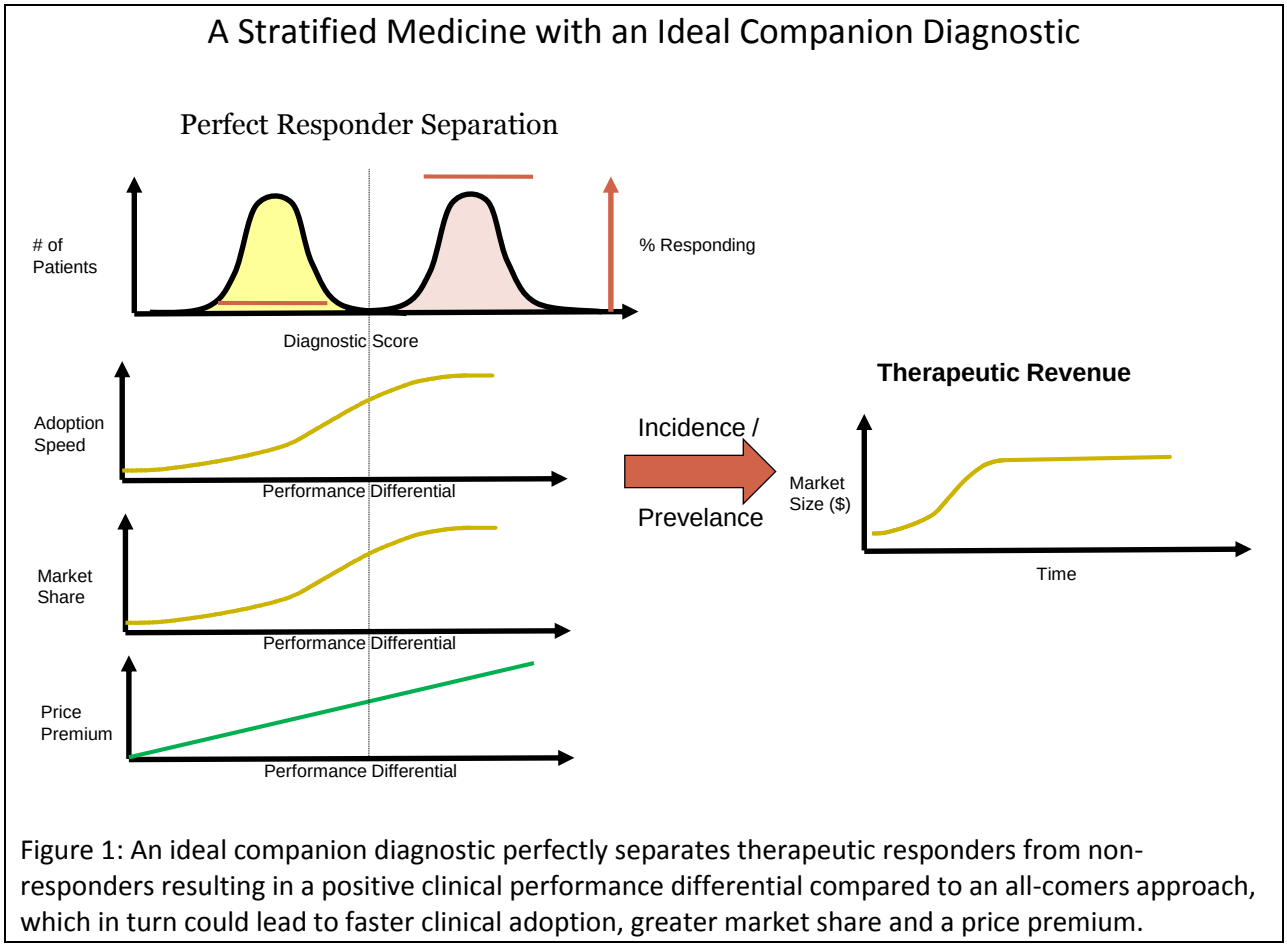


Figure 1: An ideal companion diagnostic perfectly separates therapeutic responders from non-responders resulting in a positive clinical performance differential compared to an all-comers approach, which in turn could lead to faster clinical adoption, greater market share and a price premium.

In Figure 1 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.

By eliminating the non-responders in a clinical trial, the observed therapeutic effect will be enhanced. For example imagine an all-comers cancer trial in which all possible patients are treated, and the average outcome is six 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 selected population (the right curve) will be 12 months, double that in the all-comers clinical trial design (the combined left and right populations).

The three lower panels in Figure 1 suggest that this improved therapeutic performance translates into increased clinical adoption speed, greater market share and a higher therapeutic price. While not necessarily following strict mathematical relationships as the charts imply, it is plausible at least within a therapeutic class that clinical and market enthusiasm for a drug corresponds at least roughly with the therapeutic net benefit profile [12].

Combining these characteristics with the incidence and prevalence of the conditions creates a therapeutic market forecast shown on the right side chart 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 post-exclusivity period is not shown.

An analogous companion diagnostic revenue line could also be drawn. Due to generally far lower reimbursement and near immediate competition from similar diagnostics due to limited patent, regulatory or other market protections, the companion diagnostic revenues are usually far lower than the therapeutic revenues. The sponsoring diagnostic firm receives little of the health value created by the drug-diagnostic combination.

The value of a realistic companion diagnostic

Unfortunately, in practice no diagnostic performs ideally. All diagnostics experience some level of error. In the case of a companion diagnostic, some patients will receive false positive results, scores indicating they will respond, but will not when treated. Other patients will receive false negative results, scores indicating they will not respond, but if treated they would.

Figure 2 illustrates this more realistic scenario. Once again the left curve represents those patients who will not respond. It is now larger to reflect what, unfortunately, seems the more common state in practice. The right curve represents those who will respond -- a minority of the patients. 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. False positives are those patients represented by the portion of the left curve to the right of the vertical dashed line. False negative patients are those patients represented by the portion of the right curve to the left of the dashed line.

The S-shaped curve in the top panel of Figure 2 indicates that the therapeutic response increases from low to high as the companion diagnostic value increases and the relative fraction of responders increases.

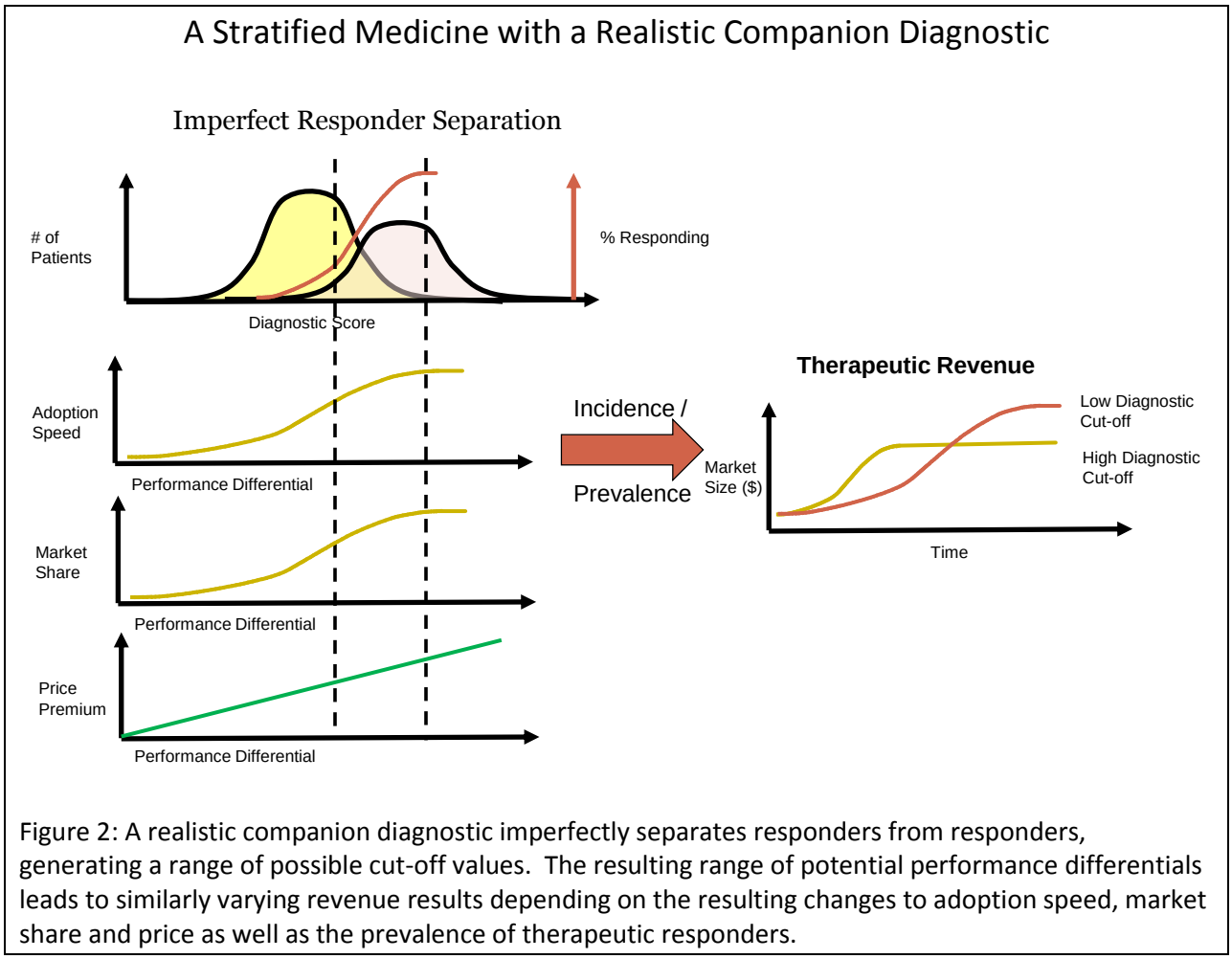


Figure 2: A realistic companion diagnostic imperfectly separates responders from responders, generating a range of possible cut-off values. The resulting range of potential performance differentials leads to similarly varying revenue results depending on the resulting changes to adoption speed, market share and price as well as the prevalence of therapeutic responders.

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 and the like. While individual genetic differences (alleles) are binary, most genes possess many such allelic changes. In addition, a given genetic change may exist on one or both copies of the gene on a chromosome pair (heterozygous or homozygous mutations). Selecting which mutations to include, and whether to include heterozygous as well as homozygous mutations can lead to a range of companion diagnostic score results quite similar to the continuous companion diagnostic scale shown in Figure 2.

Setting the cut-off value (vertical dashed line) for the imperfectly performing companion diagnostic presents multiple challenges to the scientist, regulator, ethicist, marketer, clinician and payer. Setting a high cut-off value towards the far right of the left curve excludes nearly all non-responding patient population scores, ensuring that nearly all the selected and then treated patients will respond. This also results in few non-responding patients being exposed to the side 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-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 U. S. breast cancer patients found that one-third were not tested for HER2 gene over-expression 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 [13]. A similar study found the same testing rate in Canada [14], so this is not necessarily a feature of the U.S. 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, non-responding patients are selected and treated. When one follows the right vertical dashed line downward in Figure 2 to the lower three panels, 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 sub-population.

Ethically, a high test value cut-off possesses the inherent negative characteristic of denying treatment to some patients – 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 may be unacceptable. For a condition with many and similarly efficacious treatment options or perhaps a condition with low morbidity and mortality, this may be quite 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 the right chart in Figure 2 illustrates, while employing a high cut-off may achieve faster uptake, the peak revenues could prove lower than using a lower cut-off value. Due to the limited number of patients selected with a high cut-off, even higher pricing and greater penetration may not offset the 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 off-label treatment under penalty of fine, imprisonment or both while payers will likely not reimburse. Payers do however, make exceptions, particularly when a respected professional society guideline recommends an off-label use of a therapeutic. While the current system does not perfectly forbid off-label use, it does make it more 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 Figure 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 right responding patient population scores, includes nearly all patients who **will** respond. In this case the companion diagnostic possesses high clinical sensitivity.

While few patients who might benefit are denied treatment, the number of non-responding 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 incur its side effect risks and delays in seeking other treatment. For a therapeutic with significant, irreversible side effects this may be unacceptable. For a therapeutic with few side effects or for a condition with few treatment alternatives this greater exposure to potential harm, particularly if well communicated to the patient, may be entirely appropriate.

When one follows the left vertical dashed line downward in Figure 2 to the lower three panels, one observes that the more modest performance differential (compared to 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 CDx scores greater than the cut-off) is potentially considerably larger than with the high cut-off case and the revenue outcome is greater than that in the all-comers (unselected, no companion diagnostic) case.

For the innovative manufacturer, setting a low cut-off value with the corresponding larger number of test-qualified patients, eventual peak revenues may grow larger but take longer to achieve which may disappoint markets. If remaining patent life is short, potential peak revenues may never be attained. A low cut-off value may 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 from 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. The outcomes of balancing the factors therefore require distinct analyses with general rules of thumb for preferring high or low cut-offs not obvious either ethically or financially.

Multiple tools and frameworks for analyzing these trade-offs have been presented in the literature by the current authors (the MIT Stratified Medicine Model and the Janus Program), other academics and health technology assessment organizations [15,16,17,18,19,20,21].

Behavioral change benefits

Additional potential benefits not shown in Figure 2 may result from potential behavior changes induced by a stratified medicine, e.g., willingness to seek, initiate and adhere to the treatment regimen. Patients may 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 likely response to a therapy may 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.

Stratified medicines may also benefit from improved patient adherence if the companion diagnostic inspires greater patient confidence that the therapeutic is the best course for them. If, in addition, the companion diagnostic also proves useful in monitoring disease progression, treatment effects or both, adherence may improve even more. For example, prior to the introduction of viral load tests, AIDS patients were generally poorly adherent to anti-retroviral 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. Patient adherence rates soared [22]. While not yet studied or demonstrated, stratified medicines may 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 can emerge when multiple stratified medicines compete for the same target and indication.

Figure 3 shows that in oncology, many targets have multiple products launched, in clinical development or both. Based on September 2014 PharmaProjects data, the chart 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 engages multiple targets, only the target classified by PharmaProjects as the primary target is plotted.

181 launched products are approved and launched for 53 targets (dashed line). Of those 53 targets, only nine have a single therapeutic facing no competitors in the market or under clinical development. 23 targets already have two or more launched products. Another 21 targets with only a single launched product face at least one candidate therapeutic in development, 16 of those 21 face two or more competitors in Phase I, II or III clinical development.

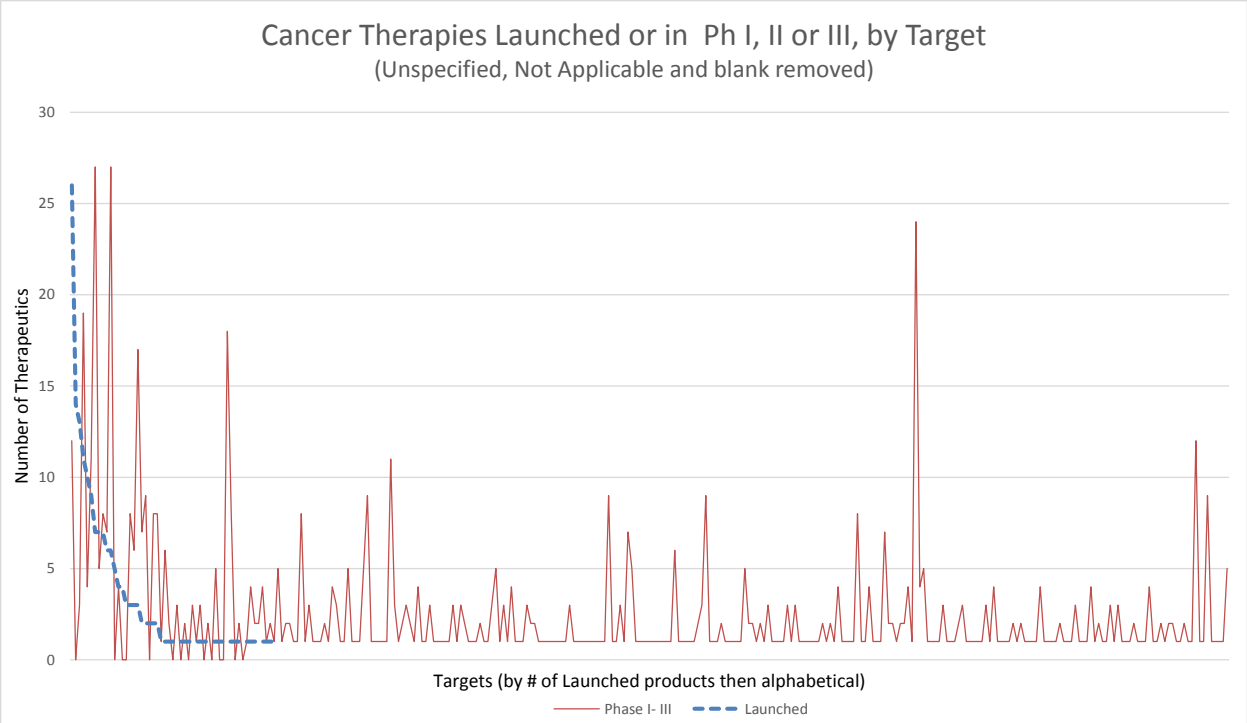


Figure 3: Oncology competition by target is common with over 80% (44 of 53) therapeutic known targets having two or more competing medicines. Many targets with only candidate therapeutics in clinical development also see competition. (Source: Authors' analysis of PharmaProjects pipeline database as of September 2014)

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

As targets proceed through the development process the competition increases. 35 targets have no launched product but at least one Phase III development candidate therapy, 21 of those (60%) have 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%) have 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 the market but also in their quest to be first-in-class. With such highly prevalent competition, understanding and developing a perspective on how stratified medicine approaches could structure the market would seem important for most development programs.

As discussed in Box 1, under competition, three essentially identical drugs may receive dramatically different labels, incremental cost-effectiveness ratio (ICER) justified pricing, and market positioning depending on their stratification approach. 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.

The competing development teams may 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 may drive developers to select a high cut-off value. If all choose this approach, overall value may be reduced with many patients excluded from treatment. But the potential advantage of a higher cut-off value may prove too alluring, or the fear of a competitor selecting one, may drive all to do so. Each situation will depend on the specific facts of the indication, therapeutic, companion diagnostic and competitors.

Box 1: 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-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 allows us to focus on the decisions and implications regarding whether and how to use the candidate companion diagnostic.

Figure B1 demonstrates three possible choices facing the firms based on these assumptions. Firm A chooses an all-comers approach for its Drug A which does not employ the companion diagnostic at all. 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 while 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 due to 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 due to the potential need to

examine negative test result patients and the continuing need to develop an acceptably large patient safety database. For this work we consider that these two cost factors approximately offset 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 to standard of care, and that the remaining patients receive zero 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 four months overall survival - the weighted average of the one-third of patients who respond with the two-thirds of patients who do not.

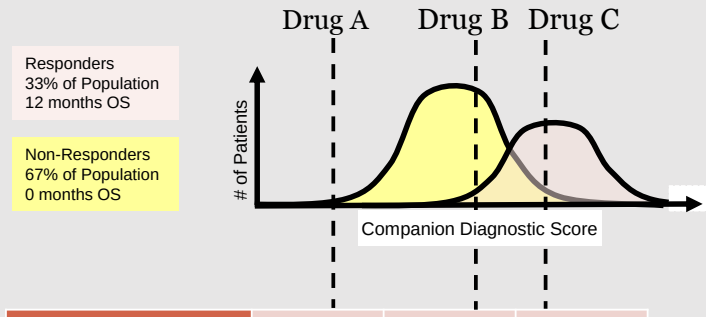
Even though no companion diagnostic was employed, one can consider that an all-comers trial has 100% sensitivity, i.e., it selects all patients who might respond, and 0% specificity, i.e., it excludes none who will not benefit. Another diagnostic metric, Positive Predictive Value (PPV) states the fraction of companion diagnostic positive (CDx+) patients that actually 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 lowest isorevenue curve (diamond marker, solid line) in Figure B1B. 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 $1 billion blockbuster level annual sales, Drug A must achieve 20% market share (be used by 20,000 patients) at a $50,000 one year drug regimen price with its four months overall survival improvement.

Firm A could choose any other price, and the lowest isorevenue curve (diamond marker, solid line) indicates what market share Drug A must achieve for $1 billion in annual sales. For instance, at a price of $200,000 Drug A must be used, and paid for, by 5,000 patients which conveniently equals 5% market share in this example. At a price of $12,500, 80% market share must be achieved (80,000 treated and paid patients) to reach $1B in sales.

Stratified Medicine Competition

A)



Sensitivity	100%	95%	64%
Specificity	0%	64%	95%
RCT Efficacy (Months OS)	4.0	6.7	10.3
Positive Predictive Value	33%	56%	86%
Patients CDx+	100,000	56,000	24,000
Price (Based on $138,582 ICER/QALY)	$46,000	$77,000	$119,000
Overall Market Share	22%	13%	8%
Market Share of CDx+	22%	23%	35%
Benefiting Patients	~7,250	~7,250	~7,250
Treated Patients	21,700	13,000	8,400

at ICER based price and $1B Revenue

B)

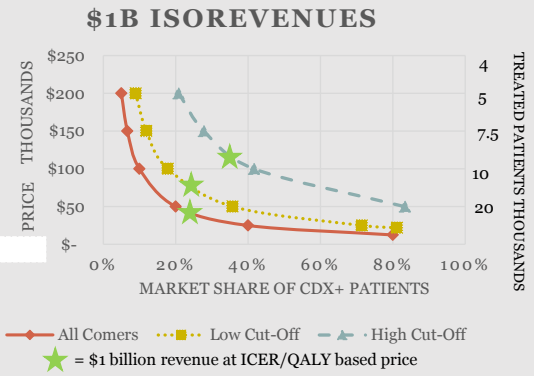


Figure B1: Stratified medicine competition**.** A) Pharmacologically similar therapeutics addressing the same target but applying companion diagnostics differently, if at all, may appear quite distinct in their RCT reported efficacies and response rates as well as commercially reported market sizes. B) Market shares of the companion diagnostic selected population must increase compared to the all comers therapeutic to achieve $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. CDx+ = Companion diagnostic test positive. OS = Overall Survival. # = Number

A recent literature review suggests that for oncology therapeutics in the U.S., the mean incremental cost-effectiveness ratio (ICER) per quality adjusted life year (QALY) is $138,582 [23]. If the ICER/QALY guides payer reimbursement, Firm A might expect a Drug A price of about $46,000 (one-third of the mean ICER/QALY based on an expected average four months overall survival improvement). Per Figure B1B, at that price Drug A would need to achieve 22% market share to generate $1 billion in annual revenue, as indicated by the star on the All Comers isorevenue curve.

Firm B chooses to employ a companion diagnostic approach which 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. The hypothesized test is assumed to be quite good, but not perfect. The low cut-off value results in a 64% specificity – 64% of non-responders will test negative. This means that 36% of the non-responders will test positive. For an oncology companion diagnostic this is quite superior performance. One of the more powerful companion diagnostics known, the KRAS test for detecting likely responders and non-responders to cetuximab (Erbitux) in colorectal cancer, has an estimated 75% sensitivity and 35% specificity [24].

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 6.7 months overall survival – 68% longer than that for Drug A (2.7 months longer than 4.0 months). 56% of the patients testing positive (CDx+) would be expected to respond. This is 23 percentage points higher than the 33% of treated patients responding to Drug A with its unenriched population.

The label for Drug B, however, will specify that it should only be used for those who test positive. Given the postulated 95% sensitivity, 64% specificity and 33% responder prevalence, 56% of patients with the disease will test positive (coincidentally identical with the 56% PPV). Instead of the 100,000 eligible patients for Drug A, only the 56,000 CDx+ patients will be eligible for Drug B. To achieve $1 billion in revenues, Firm B faces the middle isorevenue (square marker, dotted line) in Figure B1B. Firm B will need a higher price for Drug B, a higher share of the patients testing positive, or both, than Firm A. Using the $138,582 oncology ICER/QALY, Firm B might expect a price of $77,000. This is higher than the Drug A price due to 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 $1 billion in annual revenues. Note that the Drug B share of all patients with the disease is only 13% compared to 22% share of all patients with the disease Drug A required to reach $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 $30,000 per patient. Payers pay the same amount, $1 billion for about the same number of patients benefiting: approximately 7,200 for both drugs. However, with Drug B, nearly 9,000 fewer patients receive treatment that does not benefit them – assuming that both firms achieve $1B in revenues and receive an ICER based reimbursement.

In clinical practice, Firm A might find it difficult to market and compete versus Firm B due to the difference in expected patient overall survival from the use of the companion diagnostic. From a public health perspective, an ordered market in which more patients choose Drug B with the companion diagnostic could make the realized benefits from Drug A even lower. Since both work on the same biological target, Drug A may 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 may 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 may therefore show Drug A value substantially below that seen in its clinical trials while Drug B benefit would align with the clinical trial observation.

Firm C chooses to employ a companion diagnostic approach which 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 non-responders will receive a negative test score (CDx-). As shown by the far right vertical line in Figure B1A, the high cut-off also excludes some patients who would benefit from treatment. In this hypothetical case the corresponding sensitivity is 64% – 36% of patients who might benefit 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 longer than Drug A (10.3 months increased overall survival 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 likely will be 54% longer (3.6 months longer than 6.7 months) than that for Drug B.

The fraction of the disease patient population testing positive (the selected population), however, is much smaller than for Drug B, only 24,000 of the 100,000 patients. The smaller selected population results both from better excluding non-responders (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 possess 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 $138,582 ICER/QALY oncology rate and the 10.3 month overall survival benefit, Firm C might expect $119,000 Drug C reimbursement. At that price, to generate $1 billion in annual revenues, Drug C needs to achieve 35% market share of the selected population. As with the other two drugs, approximately 7,200 patients will respond, but due to the high PPV, only 8,400 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 $1 billion.

However, because of the high CDx cut-off, Drug C could never reach about 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 33,333 potential responders will receive a negative score and be denied Drug C resulting in about 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 about 1,500 patients and so 1,500 QALYs. Since Drug A does not use a CDx, it could reach ALL potential responders. Not knowing what alternative treatments might exist for the non-responders we do not estimate the off-setting QALY losses from non-responders 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 non-responder harm) for providing ubiquitous access to all 100,000 patients via a single drug formulary with no patient co-payment. Choosing Drug A would cost $4.6B per year and provide 33,333 QALYs per year. Drug B would cost $4.3B per year and provide ~32,000 QALYs per year. Drug C would cost $2.9B but only provide ~21,000 QALYs. Given the assumed ICER based pricing, the payer cost per QALY is identical at about $138,000/QALY.

A savvy payer or integrated provider might recognize that the drugs are 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/QALY to about $81,000 and total cost to $2.6B to achieve the nearly perfect health benefit of ~32,000 QALY/year and save the payer $1.7B, over 35% compared to 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, non-stratified value.

The companion diagnostic developer faces much lower revenue prospects. Even assuming a high reimbursement to the clinical laboratory of $400 of which the CDx developer receives 50% for the test kit, the entire CDx testing market is only $20 million (for a selection, but not monitoring CDx) compared to a market measured in billions of dollars for the therapeutics. If the CDx uses a standard technology such as immunoassay whose kit prices are often effectively limited to $25 or less per test the total market falls to merely $2.5 million. Due to 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 $2,000 per patient with 100% testing captured by the firm only produces $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 while 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 its stratified medicine. As part of the stratified medicine combination, companion diagnostics select patients more likely to benefit, encourage patient confidence to initiate and adhere to treatment, reduce patient non-response treatment opportunity costs, 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 to the associated surgeon fees, hospital bed day costs or branded specialty medicines. While companion diagnostics create the 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 Foundation Medicine) and academic medical centers mostly use non-proprietary 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 pre-competitive activities which they then perform through consortia and make broadly available. While pre-competitive 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* [*25*], and *Mayo Collaborative Services v. Prometheus Laboratories* [26] cases that some molecular diagnostics are ineligible for patent protection because, admittedly oversimplifying the rulings, the molecular phenomena underlying them are laws of nature not inventions.

Government regulators such as the Food and Drug Administration have employed their regulatory discretion to require first-to-market companion diagnostics to incur the costs of pre-market applications (PMAs) but to date have not removed competing non-approved diagnostics from the market.

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 for diagnostic tests and greatly reduce incentives for their development [27].

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[28,29,30]. Legislation was introduced in the U.S. House of Representatives in 2013 to address concerns about diagnostic innovation but was not enacted [31]. In January 2015, the Energy and Commerce Committee issued draft legislation for 21st Century Cures which, 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 (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 [32]. These admirable efforts usually focus on developing the science for biomarkers for new precision/stratified medicine therapeutics or to improve drug

development generally. They typically do not, however, address precision/stratified 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 due to increases in the number of high priced drugs as diseases fragment into orphan and ultra-orphan indication population sizes [33]. As the NNT analysis of cardiovascular disease demonstrated above, however, the cost of a single avoided event from a broadly empiric treatment, despite low per patient per year costs, may not be much less than the cost of new stratified medicines. The discussion in Box 1 demonstrates how improved selection of patients 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 approval [34]. 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. Incremental improved public health, unsurprisingly likely has incremental costs, even if a stratified medicine approach proves reasonably efficient at identifying those who will benefit.

Israel is an example nation 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 [35,36]. 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 which 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 cost for similar public health benefit. Drug manufacturers have little or no incentives to stratify generic or soon-to-be generic drugs. Diagnostic companies have neither resources nor incentives to invest in the required development of clinical evidence.

Payers, or the governments that often fund them, would seem to have the 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) [37,38,39]. Payers today generally do not invest in creating technologies to improve public health. But it is clearly in government payers' interest to do so. Private payers may fear free rider or scale issues. Free rider issues could likely be addressed either through pre-competitive consortia efforts or by using these technologies to help differentiate their insurance product offerings just as their data analytics and disease management capabilities already do today. Regarding scale, the number of lives covered by the largest U.S. insurers are comparable to the populations of European nations [40].

The contrast in government behavior between military (and space) technology development and health technology development is striking. For military technology, major governments coordinate an integrated requirements specification through development and deployment supply chain. 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 plays crucial, and often profitable roles throughout that chain.

In healthcare, government behavior is far more fragmented and uncoordinated even though governments fund basic research at the beginning, pay for the technology at the end and regulate its development throughout. Even in the more privatized healthcare of the United States, the federal government now provides healthcare directly for 34% of the population through Medicare, Medicaid, active military healthcare, and veterans' healthcare [41]. Not included in those 2013 Census figures are lives covered under federal employee health benefits (2.7 million employees plus their families) [42], 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 U.S. 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.

Conclusion

Stratified medicine tightens the links among science, the clinic and the marketplace. Setting the companion diagnostic cut-off value is the critical 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.

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, unfortunately, unlikely for stratified medicine in the near term. Proponents of changes already occurring such as alternative payment methods and accountable care organizations hope 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, although they do not address the issues underlying the economics of stratified medicines.

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 values and so unmet medical needs to therapeutic and drug developers or to the discovery scientists at the very beginning of the innovation value chain.

[1] US Food and Drug Administration. Remicade (infliximab) label, revised: January 2015. US Food and Drug Administration. Available at: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/ (2015).
[2] Helwick, C. Pathologic complete response: Understanding the subtleties. *The ASCO Post* **5**(16) October 15, 2014. accessed on 5/4/2015. Available at http://www.ascopost.com/issues/october-15,-2014/pathologic-complete-response-understanding-the-subtleties.aspx (2014).
[3] Newman, David. Statin Drugs given for 5 years for heart disease prevention (with known heart disease). *TheNNT.com*. Published November 2, 2013. Accessed April 17, 2015. http://www.thennt.com/nnt/statins-for-heart-disease-prevention-with-known-heart-disease/.
[4] Ridker, Paul M. The JUPITER Trial: Results, controversies and implications for prevention *Circulation: Cardiovascular Quality and Outcomes.* **2:** 279-285 doi: 10.1161/CIRCOUTCOMES.109.868299 2009. The Crestor 5 year NNT for the primary endpoint of all vascular events is 25 and 32 for the end point of myocardial infarction, stroke or death.
[5] Costco Pharmacy Drug Directory. Generic atorvastin, lovastatin and simuvastin prices range from $9.77 to $25.41 for 30 day supply. Accessed May 7, 2015. Available at http://www.costco.com/Pharmacy/drug-directory-main?storeId=10301&catalogId=10701&langId=-1 (2015).
[6] Consumer Reports. Choosing a statin-our Best Buy picks. Evaluating statin drugs to treat: High cholesterol and heart disease/Consumer Reports Best Buy Drugs 2014. Consumers Union (2014). Accessed April 21, 2015. Available at: http://www.consumerreports.org/health/resources/pdf/best-buy-drugs/StatinsUpdate-FINAL.pdf
[7] Schabert, V. F. et al. Annual costs of tumor necrosis factor inhibitors using real-world data in a commerecially insured population in the United States. *Journal of Medical Economics* **15**(2):264-275 (2012). Cost data combined with TNFα NNT estimate.
[8] Simon, R. Designs and adaptive analysis plans for pivotal clinical trials of therapeutics and companion diagnostics. *Expert Opin. Med. Diagn.* **2**(6):721-729 (2008).
[9] Taube, S.E., et al. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J. Natl. Cancer Inst.* **101**(21):1453-1463 (2009).
[10] Phillips, K. A. Diagnostics and biomarker development: priming the pipeline. *Nature Reviews Drug Discovery.* **5**:463:469 (June 2006).
[11] Trusheim, M. R., Berndt, E. R. Economic challenges and possible policy actions to advance stratified medicine. *Personalized Medicine.* **9**(4):413-427 (2012).
[12] Howard, D. H., Bach, P. B., Berndt, E. R., Conti, R. M. Pricing in the market for anticancer drugs. *Journal of Economic Perspectives.* **29**(1):139-162. (2015).
[13] Phillips K.A., et al. Clinical practice patterns and cost-effectiveness of HER2 testing strategies in breast cancer patients. *Cancer. 115*(22):5166-5174. (2009).
[14] Ferrusi, I.L. et al. Closing the personalized medicine information gap: HER2 Test documentation practice. *American Journal of Managed Care.* **19**(1):838-844 (2013).
[15] Trusheim et al. Quantifying factors for the success of stratified medicine. *Nature Reviews Drug Discovery.* **1**0:817-833 (November 2011).
[16] Trusheim et al. The Janus initiative: A multi-stakeholder process and tool set for facilitating and quantifying adaptive licensing discussions. *Health Policy and Technology* **3**:241-247 (2014).
[17] UK NICE. 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 Clinical Excellence (UK NICE).* Available at: http://www.nice.org.uk/guidance/dg10/documents/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-diagnostics-co2 (February 2013).
[18] Faulkner E. 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 in Health.* **15**:1162-1171 (2012).

[19] Annemans, L., Redekop, K., Payne, K. Current methodological issues in the economic assessment of personalized medicine. *Value in Health.* **16**:S20-S26 (2013).
[20]Jeter, E., Dickson D., Girish, P. MolDx clinical test evaluation process (CTE)P M00096 Version 2.0 *Palmetto GBA* September 2014.
[21] Deverka, P., Messner, D., Dutta, T. Evaluation of clinical validity and clinical utility of actionable molecular diagnostic tests in adult oncology. *Center for Medical Technology Policy*. Baltimore, MD. (May 2013).
[22] Bonner K et al. *Viral load monitoring as a tool to reinforce adherence: a systematic review.* J Acquir Immune Defic Syndr, online edition. DOI: 10.1097/QAI.0b013e31829f05ac, 2013.
[23] Bae Y.H. Mullins C. D. Do value thresholds for oncology drugs differ from nononcology drugs? *J. Manag Care Pharm,* **20(11):**1086-92. (Nov 2014).
[24] Westwood M, van Asselt T, Ramaekers B, et al. Health Technology Assessment, No. 18.62. Chapter 3 Table 8 Accuracy of KRAS mutation testing for the prediction of response to treatment with cetuximab in addition to chemotherapy. Southampton (UK): NIHR Journals Library; 2014 Oct
[25] Supreme Court of the United States. Ruling No. 12-398. Association for Molecular Pathology et al., *v.* Myriad Genetics Inc. Decided June 13, 2013.
[26] Supreme Court of the United States. Ruling No. 10-1150. Mayo Medical Laboratories, et al., *v.* Prometheus Laboratories, Inc. Decided March 20, 2012.
[27] Trusheim, M.R. Finley, M.J., Rausch, C. Berndt, E.R. Uncertain prognosis for high-quality diagnostics: clinical challenges, economic barriers and needed reforms. *Pharmacogenomics* **14**(3):325-334 doi:10.2217/pgs.13.3 (2013).
[28] Aspinall, M. A. Hamermesh, R. G. Realizing the promise of personalized medicine. *Harvard Business Review* 85:108-117 Reprint R0710F (2007).
[29] Cohen J. P., Felix, A. E. Personalized Medicines's bottleneck: Diagnostic test evidence and reimbursement. *Journal of Personalized Medicine.* 4:163-175 doi:10.3390/jpm4020163 (2014)
[30] Personalized Medicine Coalition. The future of coverage and payment for personalized medicine diagnostics. Washington, DC available at http://www.personalizedmedicinecoalition.org/Resources/The_Future_of_Coverage_and_Payment_for_Personalized_Medicine_Diagnostics .
[31] Roskam, P. J. Diagnostic innovation testing and knowledge advancement act of 2013. H.R. 2085. Accessed on April 21, 2015. Available at https://www.congress.gov/bill/113th-congress/house-bill/2085.
[32] The Biomarker Consortium. What we do. The Biomarker Consortium. Accessed 21 April 2015. Available at http://www.biomarkersconsortium.org/whatwedo.php .
[33] Aronson, N. Making personalized medicine more affordable. *Annals of New York Academy of Sciences* published on line February 27, 2015. DOI: 10.1111/nyas.12614. (2015).
[34] Hu S. X., Aitken, M. L., Epstein, A. M., Trusheim M. R., Berndt, E. R. Defining and quantifying the use of personalized medicines. *Nature Reviews Drug Discovery* **12**:896-897 (December 2013).
[35] Sax, P. The shaping of pharmaceutical governance: the Israeli case. *Isr J Health Policy Res,* **3**(16): published online May 27, 2014. Doi: 10.1186/2045-4015-3-16.
[36] Golan, O. Hansen, P. Which health technologies should be funded? A prioritization framework based explicitilly on value for money. . *Isr J Health Policy Res,* **1**(44) Published online Nov 26, 2012. doi: 10.1186/2045-4015-1-44.
[37] Randal, J. After 40 years, mammography remains as much emotion as science. *J. Natl. Cancer Inst.* **92**(20):1630-1632. doi: 10.1093/jnci/92.20.1630 (2000).
[38] Gold, R.H., Bassett, L.W., Widoff B. E. Radiologic History Exhibit, Highlights from the history of mammography. *RadioGraphics* **10**(6):1111-1131 (1990).
[39] Shapiro, S. Periodic screening for breast cancer: the HIP randomized controlled trial. *J.Natl Cancer Inst. Monogr.* **(22):**27-30 (1997).
[40] UnitedHealth Group serves 85 million lives (http://www.unitedhealthgroup.com/About/Default.aspx) about the same as the population of Germany. Anthem serves 37 million (Anthem Fact Sheet 2014 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 serves 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) slightly more than the populations of Switzerland or Austria

[41] U. S. Census Bureau. Table 1 Health Insurance coverage rates by type of health insurance: 2013. United States Census Bureau. Accessed 21 April, 2015 http://www.census.gov/hhes/www/hlthins/data/incpovhlth/2013/tables.html

[42] Historical Federal Workforce Table. US Office of Personnel Management. Accessed April 21, 2015. https://www.opm.gov/policy-data-oversight/data-analysis-documentation/federal-employment-reports/historical-tables/total-government-employment-since-1962/