**UNIT 5: Unintended Consequences**

Overview:

Unit 5 is the third and final unit in course Section 2 (Incentives). This unit draws attention to two distinct issues related to implementing value-based incentives.

5A: Highlights three mechanisms whereby value-based care initiatives risk increasing health disparities: Treating to the test may particularly harm disadvantaged patients. Lower assessed quality and lower payment levels might form a vicious cycle. Selection bias might harm disadvantaged patients.

5B: Highlights assessment of screening tests, as these often represent a special case in value-based care initiatives due to the (often) long time-lag for impact on benefits. The video reviews sensitivity and specificity calculations assessing test quality, points to the role of prevalence and hence the usefulness of targeting tests towards higher-prevalence populations, and cautions against lead time biases that might cause an overestimation of the benefits of true positive results.

Learning Objectives:

1. In this unit, the most important objective is one regarding general perspective or approach. Specifically: *always look for unintended consequences when changing incentives*.
2. Understand that it is possible for value-based incentives to increase health disparities.
3. Understand how to evaluate screening tests (in terms of assessing measures of test accuracy and avoiding lead time bias in interpretation).

Role in course: Disparities and screening tests represent important examples to help illustrate the vast potential for unintended consequences. Unlike most course content, these specific examples might be more or less relevant to you depending on the populations you serve (for disparities) and the clinical areas you address (for screening). Regardless, I believe it is useful to think about heterogeneity in clinical populations and treatment and how this heterogeneity might mean that value-based care could possibly advantage some populations and interventions over other populations and interventions.

**Unintended Consequences: Overview of Unit 5**

This unit addresses distinct areas where value-based care incentives could create unintended consequences. One way to think about the content in unit 5 is that I am pointing towards undesirable ways to lower current costs of care, so that you will be thoughtful about avoiding them.

One undesirable way to make care seem higher value is for providers to leave out or avoid patients that will tend to either get less benefit from care, showing worse outcomes, or be costlier to serve. This can increase health disparities.

Another undesirable way to make care seem higher value is to leave out particular services that add cost, and that are beneficial in reality, but for which this positive impact doesn’t show up for a long time. Screening tests can fall into this category. If the time frame for showing overall benefits on health outcomes for screening is not the same as the time frame for provider incentives, screening might be neglected. That is, even if providers are focused on overall, high-value care for their covered populations, it’s often the case that patients and/or providers move on before the bulk of screening benefits would be realized. (So, a provider organization investing in screening this year may be creating benefits for different providers who will be serving these screened patients decades from now.)

Overall, then, unit 5 looks at how general incentives for value could cause certain patients (i.e., those of lesser advantage) or certain procedures (i.e., broad screening for maladies with a long lead time to realize benefits) to be neglected.

**Health Disparities**

A possible result of value-based care initiatives is the possibility of worsening health disparities. This is a challenging area to study, with few definitive findings.

Health disparities are preventable decrements in health services and/or outcomes that are experienced by socially disadvantaged populations (as compared to more socially advantaged populations). Race, ethnicity, education, socioeconomic status, sexual orientation, gender identity, and area of residence are some of the factors commonly used to define disparities. Note that these different sources of disparities might have differing specific effects, e.g., thorough lack of convenient access to care, through lower post-care adherence, through clinical guidelines being optimized for non-disadvantaged groups, etc.

Most generally, it’s important to realize that we can’t necessary assume that higher average healthcare quality means that there is higher quality for all groups of patients. An important goal to consider for incentive systems is therefore often to minimize unintended consequences in terms of increasing health disparities. Next, I will discuss three specific mechanisms whereby value-based incentives might worsen disparities, and present some partial solutions for each potential issue.

Note, also, that there is generally an underlying tradeoff between simplicity in incentive systems versus results for all patients. Extra efforts to address disparities may add complexity, so it is useful to keep the costs of complexity in mind and try for balance between goals to address disparities and goals to have simple and understandable systems.

**Increasing Disparities: Treating to the test**

*Issue*. As discussed in Unit 4 (with respect to process measures), directly incentivizing a subset of all clinical actions might cause **treating to the test** where incentivized actions are emphasized at the expense of other actions. This general phenomenon could worsen disparities in cases where the relevant measures are optimized for more advantaged groups, neglecting the unique needs of less advantaged groups. Further, disadvantaged patients may also be relatively less assertive in extracting care that isn’t incented (e.g., through asking questions, directly requesting referrals, etc.).

*Solutions*. One way to address this issue is to **minimize treating to the test overall** (see Unit 4). This could involve assessing overall health outcomes such as mortality (as these are determined by all care processes), randomly shifting the basis of process measures (so that providers cannot confidently ignore currently un-compensated processes), or tracking uncompensated measures for assessment (so that later adjustments can be made given negative spillovers on un-compensated processes).

A second way to address this issue is to **directly incentivize physicians to address disparity**. Disparities are difficult to fully understand, much less remove, so we don’t have definitive processes or measures to ensure this happens perfectly. However, even relatively soft requirements, such as asking providers to generate their own assessments of and remedies for disparity, can generate provider attention.

**Increasing Disparities: Vicious Cycle**

*Issue*. Financial rewards to quality might set up a **vicious cycle**. An initial resource base might tend to be lower for providers serving disadvantaged patients. If those lower initial resources make it harder to achieve quality benchmarks, then value-based incentive systems could lead to lower initial payments for providers serving less advantaged populations. These lower payments could then exacerbate the initial resource disparity, setting in motion a vicious cycle.

*Solutions.* The most common solution to this issue is to **compensate (or reward) improvements**; concerns about disparities are often a key motivator of incentives tied to improvements rather than or in addition to absolute measured quality levels. Two factors make it difficult to implement improvement-based compensation. First, there is a measurement challenge to ensure that baseline measures (from which improvement is calculated) are accurate and also to ensure that key metrics are measured with enough specificity to observe real improvement rather than random noise. Second, there can be a political challenge in that organizations that start off with high quality (and hence do not qualify for improvement incentives) might be resentful of or resist this system. For this reason, incentive systems will often mix some absolute and some improvement measures (either implementing both or allowing for choice[[1]](#footnote-1)).

**Increasing Disparities: Sorting / Selection**

*Issue*. The third and final mechanism whereby value-based incentives could exacerbate disparity is central to many course topics. Specifically, we might expect **sorting (or selection)** that harms disadvantaged patients. Physicians sometimes perceive disadvantaged patients as less compliant, for instance, and as generally more likely to have poor outcomes. This perception often seems to reflect reality, and in some cases the perception may be even stronger than reality. If providers expect disadvantaged patients to have poorer outcomes, and providers also expect to be compensated based on outcomes, then providers will have an economic incentive to avoid disadvantaged patients. This avoidance can be subtle, for instance, it could involve less outreach to these populations, a greater tendency to refer them elsewhere, etc.

*Solutions.* The key solution to this potential problem is **risk adjusting** incentives. This is our topic for Unit 8, so we’ll come back to it. (The U.K.’s NHS has a simple, related solution in that they allow providers to designate some patients as exceptions to outcome-based measures.). Another important solution is to directly **incentivize process measures,** particularly process measures that are perceived as under provider control. Provider-controlled processes (e.g., whether a test is recommended by a physician, whether a physician provides access to educational information) are often perceived as less influenced by patient disparities than are more outcome-based measures (e.g., whether the patient actually received the recommended test, whether patient shifts behavior based on education).

**Screening Tests: Overview**

Screening is often a special case in implementing value-based incentive plans. They are designed to detect asymptomatic, early stage disease. They add value if treatment is more beneficial and / or cheaper early on in the disease process.

With screening tests, the probability of health benefit for any one patient is low because most people don’t have whatever the screening is designed to find. Even for patients who benefit from screening (because something is found earlier), the lag time between expense outlay (for the screening test) and realized health benefit (avoiding more invasive treatment that would be needed later when more advanced disease is discovered) may be long. Thus, screening often costs more than not-screening in the shorter or even more moderate term. This is why we may worry about screening being de-emphasized under, for instance, capitation incentive systems.

**Evaluating Screening Tests**

Each binary screening test has 4 outcomes: true positive (the test accurately indicates disease), false positive (the test inaccurately indicates disease), true negative (the test accurately indicates health), and false negative (the test inaccurately indicates health). To evaluate a test, we have to think about the probability of each outcome, as well as the benefits and costs of each outcome.

One way to evaluate screening tests is the **accuracy for each type of person**, healthy or not.

First, **specificity** is the ability to accurately detect health, or equivalently correctness for a person without disease. It can be calculated from test validation data as the number of true negatives divided by the sum of true negatives and false positives. Second, **sensitivity** is the ability to accurately detect disease, or correctness for a person with disease. It can be calculated as the number of true positives divided by the sum of true positives and false negatives.

In setting test guidelines, we can often trade off sensitivity and specificity. For instance, if we know that a certain blood marker becomes more prevalent with a disease, we can set a higher or lower cutoff point to designate a particular level of marker as a positive (“disease present”) result. A higher (more stringent) cutoff means the test becomes less likely to be positive, increasing specificity and reducing sensitivity. Clinicians often set cutoffs by considering the costs of errors. False positive errors can cause harm from invasive, stressful, expensive follow up (e.g., surgical biopsies following imagery). False negative errors cause harm from missed disease (e.g., potentially delaying treatment or even causing others to be exposed to contagious disease). As false positives are relatively more expensive we want more specificity, but as false negatives are relatively more expensive, we want more sensitivity. If there are multiple, distinct testing options, a common overall strategy is to use a sensitive and cheap test but then follow up on all positives with a more specific and expensive test to identify false positives.

A second way to evaluate tests is to look at the **probability a particular result is accurate,** or the predictive value of a negative or positive test result. The **positive predictive value** is the probability that a person with a positive result actually has the disease and it is calculated as the number of true positives divided by the sum of true positives and false positives. The **negative predictive value** is the probability that a person with a negative result is actually healthy and it is calculated as the number of true negatives divided by the sum of true negatives and false negatives.

These predictive values depend on disease **prevalence** or the proportion of people in the screening population who actually have the disease (i.e., the sum of true positives and false negatives divided by the total population size). Positive predictive value increases with prevalence (equivalently the risk of false positives decreases with prevalence). Conversely, negative predictive value decreases with prevalence (equivalently, the risk of false negatives increases with prevalence).

Once can sometimes leverage prevalence to improve positive predictive value. Specifically, if it is possible to identify higher-prevalence sub-populations (e.g., based on age or other factors), then we can improve positive predictive value by focusing on these sub-populations.

**Finding Disease with Screening and Lead Time Bias**

While it is important to understand the costs and benefits of all test results (and to try to manage these), the benefit from screening is tied to one particular result: true positives. True positives allow for earlier disease treatment, potentially leading to better and/or cheaper care. For screening to have a real benefit, intervention must be more effective earlier. At the extreme, if there is no health benefit from treatment (or other intervention), then the medical benefits of screening just aren’t there.

The medical or clinical benefits of screening tests can often look larger than they really are, due to **lead time bias**. In the extreme, lead time bias occurs when the impact from screening is merely that the disease is identified earlier, but ultimate mortality is not changed. If the treatment and disease course is not carefully assessed, it can look as if screening lengthens survival when all it is really lengthening is the amount of time a person knows about the disease. Identifying a disease earlier with screening can easily make it seem as if early intervention works, because the patients for whom the disease is identified with screening live longer after diagnosis than the patients for whom the disease is identified through later, other means (such as symptoms). This can be an illusion in situations where screening simply moves up time of diagnosis but does not change mortality. Often there is both a lead time effect and a positive effect of early treatment on mortality, so these must be carefully parsed out.

**Summary**

Whenever there is a change to a system as vast and important as provider incentives, we should be concerned about unintended negative consequences. Two potential unintended negative consequences to value-based incentive systems are worsening health disparities and neglecting valuable screening.

* Disparities could increase through one of the following mechanisms:
  + Compensated outcomes or measures may be calibrated to advantaged populations such that “treating to the test” (i.e., emphasizing compensated processes or outcomes) increases disparities. Payers can try to minimize treating to the test overall (by expanding or shifting “test” measures or tracking outcomes) or they can direct physicians to try to assess disparities (thereby at least raising awareness of the issue).
  + Paying for higher measured quality can set up a vicious cycle if providers for the disadvantaged start with a lower resource base, therefore do relatively poorly on quality measures, and as a result receive lower payments exacerbating the initial resource-based disadvantage. Payers can address this issue by rewarding improvements either instead of or in addition to absolute measures.
  + Compensating positive health outcomes might encourage providers to prefer to serve more advantaged populations who are perceived as more likely to show positive outcomes (due to compliance, general access, etc.). Risk adjustment (see unit 8) is central for addressing this issue. Incentivizing process measures (that are more directly under provider control) can also help.
  + Note that efforts to address these or other issues regarding health disparities can potentially add a good deal of complexity to value-based incentive systems.
* Value-based care initiatives should also ensure sufficient incentives for valuable screening tests.
  + More accurate tests are more beneficial so we want tests with high sensitivity (accuracy given health) and specificity (accuracy given disease). In some cases, we can set screening cutoffs by balancing the costs of false positives versus false negatives; if false positives are particularly expensive, we may set high cutoffs for calling a particular result “positive” thereby increasing specificity at the cost of sensitivity.
  + We often target screening towards higher disease prevalence populations, as this will increase positive predictive value (i.e., the probability that a positive test result is a true positive).
  + We want to be sure we’re incentivizing screening for diseases where earlier identification results in clearly better outcomes. To do so, we have to avoid falling prey to lead time bias (i.e., we want to avoid mis-interpreting living longer with a diagnosis as surviving longer).

1. See Unit 4 regarding issues with choice in incentive systems. [↑](#footnote-ref-1)