**UNIT 8: Risk Adjustment**

Overview:

8A: The video reviews the logic of and major uses for risk adjustment (i.e., statistically accounting for patient characteristics that are independent of care delivery). I note the dangers of capitalizing on chance and of setting arbitrary cut-offs with cell-based methods.

8B: The video draws attention to problems that follow from risk adjustments based on information reported by providers. I also note difficulties with distinguishing up-coding from down-coding.

Learning Objectives:

1. Recognize the importance of risk adjustment for clinical research, for evaluation of providers, and for benchmarks for capitation payments.
2. Recognize threats to good risk adjustment from capitalizing on chance, from using cell-based methods that do not leverage all available (continuous variable) information, and from strategic self-reporting by providers (e.g., ‘upcoding’).

Role in Course: Risk adjustment is ubiquitous in care evaluation such that most of the concepts and interventions in this course assume competent, independent risk adjustment. Ryan’s Analysis of Healthcare Effectiveness and Outcomes course covers use of risk adjustment methods such as propensity scores, so Unit 8 of this courses focuses on issues with interpretation rather than methodology. (Still, you will see some overlap with Ryan’s content.)

**Risk Adjustment Logic**

We care about overall health outcomes: the overall benefits and costs of care. We also know that providers can only control the process quality of care (what they do). Ultimate benefit and cost outcomes are jointly determined by the process quality of care and by patient factors. So, we often can’t just look at outcomes such as mortality in order to understand the quality of a hospital, provider, or treatment. Instead, we must control for patient characteristics that are independent of current care. For instance, if older patients tend to have poorer outcomes in a clinical domain, then it would be unfair to compare a provider with a young patient base to one with an elderly patient base; we have to statistically control for patient age distributions when comparing provider quality. If age is all that matters, we can evaluate each provider on performance as compared to the age-related baseline outcome. Of course, we typically must control for much more than age.

More generally, there is a potential for selection error (or selection bias) whenever we compare groups that are not formed by random assignment. The vast majority of medical care and assessment is delivered outside of randomized, controlled trials, so subject to selection bias, as noted below:

* **Patient self-selection** behaviors can create selection bias. For instance, certain types of patients may prefer particular types of providers or care settings. For outcomes such as patient satisfaction or patient reported outcomes, we might similarly find that some types of patients are more likely to select to respond to questions, skewing patient-reported data towards the opinions associated with high-propensity responders.
* **Provider-generated selection** also exists. Providers may influence to distribution of patients they actually serve either directly (for instance through policies regarding admissions and qualifications for procedures) or indirectly (for instance through community outreach and advertising).

**Risk Adjustment Usage**

Value-based care requires risk adjustment for three main uses:

* Value-based care requires a solid evidence base evaluating treatment options. Risk adjustment is ubiquitous in the **Clinical Research** that provides this evidence base by estimating the benefits of various treatments (e.g., clinical care processes, medical procedures and other interventions, drugs, etc.). Essentially, if you are not running a randomized, controlled trial, you probably have to do some patient-level adjustment to understand the quality of various clinical options. In Ryan’s Outcomes course, you’ve discussed techniques for using risk adjustment to better understand the value of clinical interventions.
* We often also need risk adjustment for **Provider** **Quality Evaluation & Setting Quality Benchmarks**. We can’t really compare the costs and benefits of provider care unless we somehow adjust for patient factors. More generally, we cannot say much about what “high quality outcomes” look like for a provider unless we adjust for the characteristics of that provider’s patient base
* Finally, one common value-based payment initiative is to use **Capitation Payments**. Unless capitation payments are risk adjusted, they will create huge incentives for selection of healthier patients and avoidance of sicker. Capitation payment determination is often prospective, that is, we have to predict care needs in order to set fair payments. Hence, these models sometimes involve a bit of forecasting along with models of patient variability.

**Example Confounding Patient Factors**

Factors related to patient risk adjustment may exist in medical records or may be possible to measure with a bit of extra effort. Importantly, however, there are always some patient factors that are currently unobservable. These might be theoretically measurable factors that aren’t measured due to prohibitive cost or intrusiveness (such as known genetic factors or patient adherence to recommendations). This unobservable category might also include strictly unobservable variables such as currently un-discovered genetic factors that influence disease outcomes.

Below are some examples of the types of patient factors that are often used in risk adjustment.

* Basic Demographics, such as age, gender, etc. are commonly used in all sorts of observational work, including observations of medical outcomes.
* All or some of a patient’s Clinical Diagnoses are often relevant. For instance, Medicare’s system of hierarchical condition categories, or HCCs, attempts to categorize patients into similar diagnostic groups that will have similar needs for medical resources. Medicare then uses HCCs as a basis for its risk adjustment efforts.
* Patient Resources might also matter. These can include factors such as income or education and might extend to factors such as neighborhood or community.
* Health Behaviors, such as smoker status, might also matter.
* Insurance status is often also used in risk adjustment, as it might be directly relevant to care access and it might also be a proxy for other factors such as disability or resources.
* Particularly if we are measuring satisfaction with providers or any quality marker that is similar to satisfaction, then Patient Preferences or Expectations can be important.

The list above is not exhaustive. A particular challenge in risk adjustment is that we often have a seemingly unending number of possibilities for risk adjustment calculations, because we can use risk adjustment variables in any possible combination. This plethora of variables creates the potential problem of **capitalizing on chance**. Capitalizing on chance occurs if a researcher re-analyzes a dataset in many different ways until a desired result is found but that researcher doesn’t account for these multiple tests when conducting statistical inference. In this case, someone trying to support a particular outcome can simply try different sets of risk adjustment variables until that desired outcome is supported. Given any reasonable random variation, a researcher who is incorrectly failing to adjust for multiple tests will eventually be able to find a model that supports almost any result. Hence, we have to take care to monitor the methods and results of risk adjustment.

One way to deal with the problem of capitalizing on chance is to use theory in terms of what risk adjustment factors we consider. For instance, we should be more likely to use socioeconomic status for clinical situations where we know that patient level resources drive adherence and also that adherence drives outcomes. We don’t usually know enough to have a perfect theory-based determination, but as a first cut assessment it’s usually safe to conclude that risk adjustment models grounded in theoretical knowledge are less likely to have been the result of capitalizing on chance.

Another way to safeguard against these issues is to require consistency in risk adjustment. That is, if a provider argues in year 1 that capitation payments should be adjusted based on patient mental health status, that provider should be willing to stick with that source of adjustment over the longer term. Note that consistency doesn’t mean we can never learn more and adjust, it simply means that we stick with a methodology unless and until we have a statistically validated reason to change it.

A common way to ensure quality in risk adjustment is to have the adjustment done by third parties (e.g., neither payers nor providers creating a contract). These third parties would have motivations to adjust as accurately as possible rather than having motivations to make a particular treatment, provider, or patient population look good or bad. For instance, 3M has a clinical risk grouping software. This software is independently validated and also will be consistent across time and providers.

And, finally, it’s important to have appropriate confidence in risk adjustment. Competent risk adjustment is necessary and much more valuable than un-adjusted testing. However, “perfect risk adjustment” doesn’t typically exist.

**Cell Methods for Risk Adjustment**

Cell-based methods are a common approach to risk adjustment. These methods simply classify patients into mutually exclusive groups. For instance, providers might get an incremental capitation payment for each frail elderly patient they enroll. These groups can be stratified. For instance, we might stratify by age, creating groups of young adults, middle age, and elderly.

Cell based methods often seem transparent, easy to understand, and hence fair. However, as is true whenever we take continuous data and turn it into a cell-based classification scheme, we lose explanatory value when we combine patients into groups. The exact cutoffs we use for cell assignment can distort information; researchers sometimes refer to this issue as binning. That is, we wonder if the cutoffs for the cells (bins) are distorting our understanding. It’s more statistically powerful to use all of the relevant information by using model-based variables with continuous variable as input, rather than placing individuals into cells.

**Model-Based Methods for Risk Adjustment**

Model-based methods use statistical techniques to estimate the impact of patient factors on clinical outcomes. The complicating issue is that, as with any context where covariates are relevant to multiply-determined behavior, there are many specific, alternative models. You’ve learned about these models in Ryan’s class, so I offer only a brief review here.

For instance, consider the question of provider quality assessment. In this situation, one would observe some unadjusted, observed quality level for a provider. We might want to compare this provider’s quality to, say, the quality average of all providers in some country or region. We also have information on the average characteristics of the patients served by the provider being assessed as well as those served by all providers. What the risk adjustment model does, essentially, is adjust the observed provider quality up or down depending on the patients seen. If the provider saw higher risk patients that means the observed quality should be adjusted up because it’s more difficult to create high quality with riskier patients. If the provider saw lower risk patients on average, the observed quality should be adjusted down. There are many potential models in these cases, so it’s important to pay attention to model specification and model quality.

In this context, a very basic covariate approach is as follows.

* Step 1: We would first analyze the key dependent variable assessing quality, such as mortality, in a model where it is predicted by patient risk factors across all the data we have, specifically for all providers in whatever group is being assessed. So, for instance, we may model mortality as a function of patient age for an entire geographic region. We use this first model to derive weights. For instance, the model might tell us that for every additional decade in average patient age, we can expect 10% higher mortality.
* Step 2: Once we understand how patient risk factors predict outcomes overall, then we are ready to adjust for each provider. We do this simply by adjusting observed provider outcomes based on that provider’s measured patient characteristics and the weights from the overall model. For instance, using our simple example, if a provider is serving a population that is a decade younger than the average, we would adjust that provider’s observed mortality up by 10% to reflect the contribution of patient age (i.e., that younger patients have lower mortality). The adjustment would be based on the measured patient characteristics (here, age) multiplied by a weight derived from the initial, overall model run in step 1.

We usually have many relevant variables. This can sometimes cause issues due to collinearity among the patient variables and / or due to more parameters than a sample size can support. In this case, we can summarize the risk factors first, perhaps by running an initial factor analysis and then using summary factor scores.

We can leverage **propensity scores** for this problem as well. Recall from Ryan’s course that propensity scores are designed to summarize participant (patient) factors, so these scores are suited to risk adjustment applications. For instance, in assessing a provider, the propensity score would capture the probability that a particular type of patient would tend to be seen by the provider in question. In propensity model terms, the provider becomes the treatment. We can use inverse probability weighting where patients are weighted by the ratio of 1 divided by propensity score. This means that a patient who is very much like the provider’s other patients will get a lower weight because that patient has a high propensity score. So, a provider who serves particularly young patients would have lower weights for young patients and higher weights for older patients in the overall, adjusted assessment. Propensity scores can be part of a larger covariate-based model. Also, if we have sufficient data, we can match on propensity scores or stratify by propensity.

Overall, model-based methods are much more defensible from a statistical vantage point than cell based models are. On the other hand, providers don’t always love putting their own financial outcomes under the control of what can seem like a statistical black box. One final note is that these models will on average tend to perform worse for patterns of extreme patient factors. For example, if a provider serves a set of patients that have much higher average risk than the patients used to calibrate the original model used to derive weights for risk adjustment, we would have less confidence in the risk adjustment overall.

**Self-Reporting and Upcoding**

Under any incentive system, we have to worry about gaming, or agents (such as providers) using their superior knowledge and / or control of information flows to mis-represent their actions. One way to mis-represent actions in value-based care is to represent one’s patients as risker (e.g., sicker) than they actually are. We often talk about this issue as “upcoding” in diagnoses. Upcoding suggests knowingly and inaccurately exaggerating diagnoses which is, essentially, fraud. However, as we dig into what seems like evidence of upcoding, sometimes we find that the more aggressive diagnoses (or codes) actually seem more accurate. That is, it can be very difficult to distinguish upcoding from a reduction in downcoding; some providers might be better at accurately noting and recording all factors indicating patient disease. It is also sometimes difficult to distinguish variations in coding behavior from variations in patient populations. Any particular hospital might have sicker patients, and hence more generous or severe DRGs represented. This could occur because sicker patients choose the hospital and/or because that hospital encourages sicker populations to come to them for care.

Careful studies have illustrated evidence that seems consistent with changes in coding (independent of patient factors) across provider systems and geographies. Regulatory review can help address this issue. This is a major concern, as different diagnostic behavior can ultimately result in bias for value-based payment initiatives or other studies of comparative effectiveness of clinical treatment. When we conduct risk adjustment, we typically assume (at least implicitly) that patient risk factors are independent of provider coding behavior.

Of course, as you also learned in Ryan’s course, from a statistical perspective it is better to remove the need for risk adjustment by random assignment. When you can randomly assign, or even exploit random processes, you then have a much better window into causality.

Overall, it’s useful to avoid unverified self-report if possible. Either direct regulatory review, or the threat of review, or other third party methods or audit are ways to mitigate strategic behavior. Finally, it’s important to recognize the potential for bias through coding and adjust your level of confidence in results as needed.

**Summary**

In many aspects of value-based care, our conclusions are only as good as our risk adjustment. Not all risk adjustment methods are equivalent, so it is important to keep in mind:

* There are often many degrees of freedom, so research approaches must avoid capitalizing on chance. When possible, you should use theory to guide and monitor risk-adjustment variable construction. It is also often useful to be consistent over time.
* Cell-based approaches to risk adjustment are easy to understand but generally statistically inferior to model-based approaches.
* It’s also important to be cautious about self-report of risk-adjustment model inputs, given the various ways that strategic behavior such as upcoding could distort these inputs. Any unverified self-reporting creates the potential for bias through strategic behavior or gaming.