**Chapter 17**

**Application to Benchmarking Clinicians: Switching Distributions**

# [H1] Learning Objectives

1. Benchmark clinicians against their peers

2. Simulate performance of clinicians on same set of patients by switching distributions

3. Estimate outcomes for synthetic cases and improve matching of all clinicians’ cases

# [H1] Key Concepts

* Overlap among clinicians and peers
* Distribution switch
* Event trees
* Expected outcomes
* Synthetic case

# [H1] Chapter at a Glance

In this chapter, we show how data balancing, in general, and stratified covariate balancing, in particular, can be used to benchmark clinicians. Data balancing enables the analyst to compare the performance of clinicians to their peer groups on the same set of patients. The comparison is done in four steps. First, the patients are described in terms of their features/comorbidities. Each combination of comorbidities is treated as a separate type of patients. Second, these patient types are used to calculate distributions of clinician’s and peer group’s patients. Naturally, these two groups will differ in who is taking care of sicker types of patients. Third, the expected outcomes are calculated for both the clinician and the peer group. An expected outcome is always the sum of product of probability of caring for a particular type of patient times the average outcome for that type. In calculating the expected outcome for the peer group a change is made. The distribution of peer group’s patients is switched with the distribution of the clinician’s patients. In this fashion, the analyst simulates the performance of peer group on the same clinician’s patients. The switch accomplishes the same goal as using propensity, or other types of weights, to balance the two distributions. Finally fourth, in reporting the outcomes of peer group on clinician’s patients, a problem arises when the peer group does not see patients seen by the clinician. To compensate, the peer group’s outcomes for these situations are constructed using synthetic cases. These synthetic cases replace missing cases and allow all clinician’s patients to have at least one match among peer group’s patients. The procedures described here can be applied easily to data in electronic health records and this chapter presents Standard Query Language for doing so.

# [H1] Introduction

Clinicians often complain that benchmarking procedures are unfair. They point out that they are taking care of sicker patients than their colleagues. Since sicker patients would naturally have worse outcomes, benchmarking blames them for patients’ conditions as opposed to the quality of their care. Since no two clinicians see the same frequency of sicker patients, clinicians have good reasons to be concerned. One way out is to randomly assign patients to clinicians. If patients were randomly assigned to their provider, then we can be reassured that the two groups are similar and differences in outcomes are not due to one group having sicker patients. Randomization is almost never done, although exceptions exist [Cebul 1991]. Patients choose their clinicians and resist giving up these choices. Another approach, one that we discuss in this tutorial, is to let the patients select their clinicians as they wish. Then data balancing procedure is used to simulate what would have happened if the peer group and the clinician had seen the same patients. Data balancing allows one to focus on quality of care independent of differences in patient conditions.

Data balancing refers to weighting the data so that sicker patients occur at the same rate among the patients seen by both the clinician and the peer group. Data balancing was first proposed in 1983 [Rosenbaum and Rubin 1983]. Since then, the approach has been repeatedly improved [Abadie and Imbens 2011; Abadie and Imbens 2006; Alemi, Elrafey, and Avramovic 2018; Hansen 2004; Ho et al. 2007; Imai and Ratkovic 2014; Hirano, Imbens, and Ridder 2003] and is in widespread use with several tutorials describing the nuances of propensity scoring, a type of data balancing [Austin 2013]. Most recently, “stratified covariate balancing” has been used as a method of balancing the data [Alemi et al 2013]. In this chapter, we use stratified covariate balancing to benchmark clinicians because it does not require statistical estimation of weights used in data balancing.

**[H2]** **Change in Terminology**

Throughout this chapter, we have replaced the terminology we used in previous chapters with terms that make sense in context of benchmarking clinicians. Sample is replaced with patient groups. Covariates is replaced with patient comorbidities. A stratum, or a unique combination of covariates, is replaced with patient types. In context of benchmarking, the clinician is the treatment variable. Treatment effect is replaced with clinician’s quality of care.

# [H1] Switching Probabilities

In data balancing, the types of patients seen in the peer group are weighted so that they occur at the same rate as the types seen by the clinician. The weights are derived through statistical models, usually regression. Stratified covariate balancing simplifies the calculation of the weights by deriving them analytically, without statistical modeling. To further simplify the data balancing procedures, we bypass the estimation of weights entirely. Since weights are the means to creating equal distributions; one can discard with the complex task of estimating the weight and simply switch the distribution of the peer group’s patients with the distribution of the clinician’s patients (see the box for the mathematics of how to do so). This switch in distributions sidesteps the awkward and often confusing modeling necessary to estimate the weights. It also helps the interpretation of the findings as the switch has a physical interpretation: it simulates how the peer group would have performed if they had seen the clinician’s patients.

|  |
| --- |
| **The Math of Switching Distributions**  If is the probability of observing patients in the severity group "i" and is the average outcome for the clinician for severity group "i", then the expected outcomes for the clinician is calculated as:  In this formula, the summation is over the index value i, which indicates low, medium or high severity of illness groups. The same calculation can be done for the peer clinicians:  The observed peer outcome, shown as is not a reasonable benchmark. It is calculated on a different set of patients than the clinician’s patients. The simulated peer outcome, shown as, adjusts for the differences in the patient populations. It is calculated by switching the peer’s with the clinician’s probability of caring for sick patients. It is calculated as:  This simulation estimates what would have happened if the peer group would have cared for the clinician’s patients. See Appendix for how to do this within electronic health records. |

Of course, the simulated values are counterfactual calculations, in the sense that these statistics report outcomes if the peer group had seen the patients of the clinician. The peer group has not really done so. We need to simulate what the situation would have been, if they had done so. If we compare the clinician’s observed outcomes to the peer group’s simulated outcomes, then we are comparing the two groups on the same types of patients. These comparisons are apple-to-apple comparisons. The comparison does not blame either group for taking care of sicker patients.

An example can demonstrate. Consider that a clinician and his/her peers have had the outcomes displayed in Exhibit 17.1. Is this clinician better or worse than the peer providers? To answer this question, the analyst must compare the expected outcomes for the clinician to the expected outcomes for the peer providers simulated on the same patients as the clinician. The calculation of expected outcomes has two components: probability of observing different types of patients and outcomes within different types of patients. Expected outcome is the sum of the product of these two components.

**Exhibit 17.1:  Severity Adjusted Comparison of Performance of Several Clinicians**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Severity of patients | **Clinician** | | **Peer Group** | | |
| Number of patients | Average length of stay of patients | | Number of patients | Average length of stay of patients |
| Low | 20 | 3.1 | | 40 | 4.1 |
| Medium | 30 | 3.4 | | 40 | 3 |
| High | 70 | 5.2 | | 5 | 4.5 |

The first step is to calculate the probability of finding a patient in a different severity groups. This is done by dividing the number of patients in a severity group by the total number of patients seen by the clinician being evaluated. For the clinician, the probability of having a low severity patient is 20/120, a medium severity patient is 30/120, and a high severity patient is 70/120. This clinician mostly sees severely ill patients. Once the probabilities are calculated, the second step is to calculate the expected outcome, in this example expected length of stay (LOS) for the clinician:

To understand whether 4.4 days is too high or too low, the analyst needs to compare this clinician's performance to that of his/her peer providers. But the peer providers do not see as many severely ill patients as does the clinician; the clinician sees 70 patients in high severity group while the peer group sees only 5. To simulate the performance of the peer providers on the patients seen by the clinician, the analyst uses the frequency of severity among clinician's patients to weigh the outcomes of the peer providers:

In the above calculation, the probabilities are from the clinician’s experience with different types of patients and the outcomes are from the peer group’s experiences with same types of patients. The expected outcome is calculated to be 4 days for the peer group seeing clinician’s patients. The clinician seems to be less efficient than the average of his peer group, when both are compared on the same set of patients. Because we equalized the frequency of low-, medium-, and high-severity patients, the differences cannot be of the result of the patient severity. Of course, the analysis can be misleading, if the classification of patients into various severity groups is done incorrectly. But if the classification of patients into severity groups is correct, switching of probabilities is an easy way to simulate the performance of the clinician and the peer group on the same set of patients.

# [H1] Example with Multiple Comorbidities

In the previous section, we divided patients in broad categories of severity (Low, Medium and High) and compared care provided within each category. To calculate these benchmarks, one needs access to a reliable and valid measure of the severity of illness. Sometimes, such a measure is not available. When no severity index is available, an analyst must match patients of clinicians and his/her peer group by history of their conditions/comorbidities. A great deal has been written on matching and many methods exist to match patients [Rose and Laan 2009; Rosenbaum 2007].

An example can demonstrate the use of patient matching. Exhibit 17.2 shows 20 patients of one clinician and 24 patients of his peer providers. These patients were admitted to a hospital for myocardial infarction (MI). In each case, we have recorded two features: existence of a previous MI, and presence of congestive heart failure (CHF). Obviously, a patient with a previous MI and with CHF has a worse prognosis than a patient without these features. The analyst needs to separate outcomes for patients with and without these characteristics.

**Exhibit 17.2** Patients of Clinician and Peer Providers May Differ in Significant Ways

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dr. A’s Patients | | | | Peer Provider’s Patients | | | |
| Case | Previous MI | CHF | Length of stay | Case | Previous MI | CHF | Length of stay |
| 1 | Yes | Yes | 6 | 1 | MI | CHF | 5 |
| 2 | Yes | No | 5 | 2 | MI | CHF | 5 |
| 3 | Yes | Yes | 6 | 3 | No MI | CHF | 4 |
| 4 | Yes | Yes | 6 | 4 | No MI | No CHF | 3 |
| 5 | Yes | Yes | 6 | 5 | No MI | CHF | 4 |
| 6 | Yes | No | 5 | 6 | No MI | CHF | 4 |
| 7 | Yes | Yes | 6 | 7 | MI | CHF | 5 |
| 8 | Yes | No | 5 | 8 | MI | CHF | 5 |
| 9 | Yes | Yes | 6 | 9 | MI | CHF | 5 |
| 10 | Yes | No | 5 | 10 | MI | CHF | 5 |
| 11 | Yes | Yes | 6 | 11 | MI | CHF | 5 |
| 12 | No | Yes | 4 | 12 | No MI | No CHF | 3 |
| 13 | No | Yes | 4 | 13 | No MI | CHF | 4 |
| 14 | No | Yes | 4 | 14 | No MI | CHF | 4 |
| 15 | Yes | Yes | 6 | 15 | No MI | CHF | 4 |
| 16 | Yes | Yes | 6 | 16 | No MI | CHF | 4 |
| 17 | Yes | Yes | 6 | 17 | No MI | CHF | 4 |
| 18 | Yes | No | 5 | 18 | No MI | No CHF | 3 |
| 19 | Yes | No | 5 | 19 | MI | No CHF | 4 |
| 20 | Yes | Yes | 6 | 20 | MI | CHF | 5 |
|  |  |  |  | 21 | MI | CHF | 5 |
|  |  |  |  | 22 | MI | CHF | 5 |
|  |  |  |  | 23 | MI | No CHF | 4 |
|  |  |  |  | 24 | No MI | CHF | 3 |

An event tree can organize and summarize the data in exhibit 17.2. Each feature of the patient (e.g. previous MI or CHF) can be used to create a new branch in the event tree. The branches end with the outcomes, the length of stay, presented to the right of the tree. A branch on the tree shows a particular combination of patient features. When data are in table format the branches, the combinations of features of the patient, indicate different patient types. The idea is to make sure that both the clinician and the peer group are matched on the same branches or same patient types. For example, the event tree for the patients seen by the clinician and peer group are provided in Exhibit 17.1. In this tree, previous MI is the first event; CHF is the second event. The length of stay (LOS) is given to the right of the tree. The probability of previous MI and the conditional probability of CHF given a previous MI is given on the arcs within the tree. To make it easier to read, since probabilities around a node must add up to one, some probabilities that can be deduced from the context are not reported.

For example, the event tree for the patients seen by the clinician and peer group are provided in exhibit 17.3. In this tree, previous MI is the first event, and CHF is the second event. The LOS is given to the right of the tree. The top branch in exhibit 17.3 is the combination of previous MI and CHF. The probability of previous MI and the conditional probability of CHF given a previous MI are given on the arcs. To make it easier to read the tree, and because in each node the probabilities add up to 1, the probabilities of negative events, such as not having CHF, are not provided. They can be easily derived.

**Exhibit 17.3** Decision Trees for the Clinician’s Practice and Peer Group Practice

.85

.65

1.0

**Length of Stay**

**6**

**5**

**4**

**?**

.50

.83

.75

**Length of Stay**

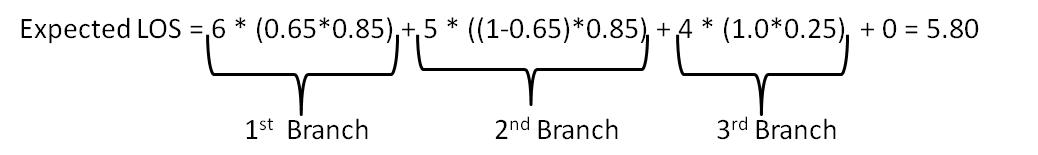
**5**

**4**

**4**

**3**

*Note:* Probabilities around each node add up to one. CHF is congestive heart failure, MI is myocardial infarction, and LOS is length of stay.  
The joint probability of events within each branch of the tree is calculated by multiplying the probabilities of each arc within the branch. So the joint probability of previous MI and CHF for the clinician is calculated as 0.85 times 0.65. The expected length of stay is calculated as the sum of the product of the probability of events on the branch times the outcome associated with the branch. The expected LOS for the clinician was 5.8 days:

****

To simulate how the same patients would have been cared for by peer clinicians, the clinician’s event tree is kept and the LOS of each patient grouping is replaced with the LOS of the patients in the peer group. This results in expected outcome for peer group to be calculated as 4.95 days:

Expected LOS = **5** \* (0.65\*0.85) + **4** \* ((1-0.65)\*0.85) + **4** \* (1.0\*0.25) + **3** \* (0) = 4.95

In above calculation, we are showing the days of hospital stay in bold and the probabilities within parenthesis. The probabilities come from the clinician’s tree, while the outcomes (i.e. the bold days of stay) come from the peer group’s tree. The clinician’s patients stay on average 5.80 - 4.95 = 0.85 days longer than if these same patients were treated by the peer group. Note that if we had not switched the probabilities, we could not have claimed that performance had been calculated based on same patients. Also note that even small difference in average performance adds up over large number of patients. In our example, in 100 patients, clinician’s patients stay 85 days longer in the hospital, which is not a minor cost.

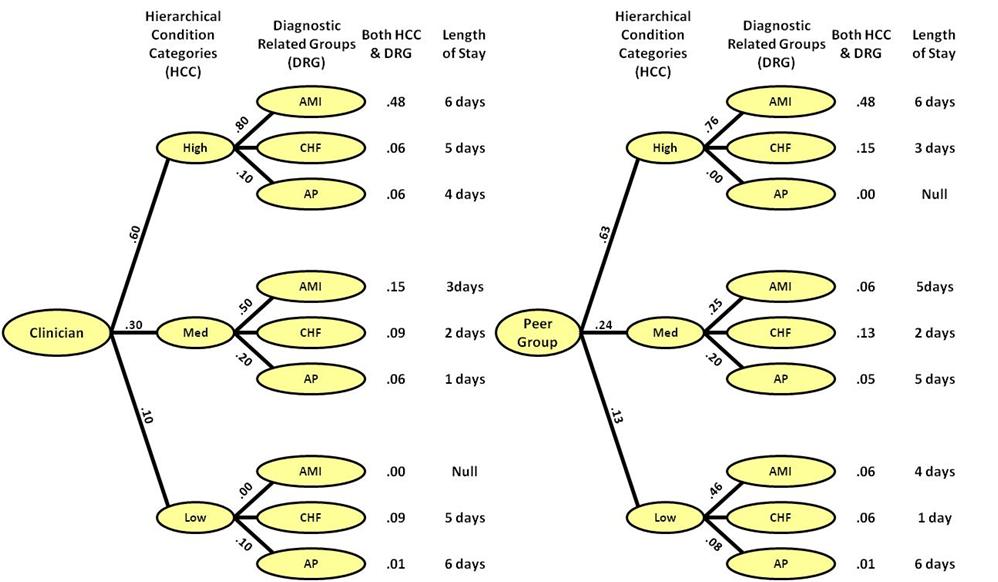
# [H1] Overlap of Clinicians and Peer Group’s Patients

So far, we have compared the clinician and the peer group by finding the same patients in the two groups. In essence, we have matched the patient types across the two groups and noted the differences in outcomes. Matching patients doesn’t always work. As the number of features increases, the number of data points that falls into each branch (each patient type) becomes smaller. Soon most branches will have no patients. Many peer providers' patients cannot be matched feature-by-feature, condition-by-condition, to the clinician's patients. A clinician may see patients never seen by his peer and vice versa. When the features available do not exactly match, the analyst can rely on partial matches. Obviously as fewer features are matched, the benchmarking conclusions become less defensible. One way out of this situation is to create synthetic patients whenever an exact match does not exist. In artificial intelligence literature, the procedure for constructing synthetic cases is known as Synthetic Minority Oversampling Technique (Blagus and Lusa 2013).

For example consider the data in Exhibit 17.4. For each patient managed by these providers we have information on the Centers for Medicare and Medicaid Service's (CMS) Hierarchical Condition Category (HCC), and the assigned Diagnostic Related Groups (DRGs). For simplicity, we have divided the HCC scores into three groups: low, medium and high. Also for simplicity, we have assumed there were only 3 DRGs shown as Acute Myocardial Infarction (AMI), Congestive Heart Failure (CHF), and Angina Pectoris (AP). Since HCC scores are always available for any patient (i.e. patients with no data are assumed to have HCC of zero), we will start the event tree using this variable. DRGs are not always occurring for all clinicians, so these values may be null. The tree shows the probability of various HCCs and DRGs. Note that the probabilities shown for the DRGs are conditional on HCC values. The probabilities at each node add to one, some values are not shown to keep the display simple. The product of probability of HCC and conditional probability of DRG shows the joint probability of the combination of HCC and DRG. These values are also shown in Exhibit 17.4 under the column named “Both HCC & DRG”. Also further suppose that the clinician and his peer have the length of Stay (LOS) indicated in Exhibit 17.4. A glimpse at the tree suggests that the clinician’s peer may be seeing patients with higher severity and thus longer stays. The tree for the clinician and the tree for his peer group are structurally different. There are some branches in clinician’s tree that are not in the peer group and vice versa. Given these differences, it is not possible to switch the probability events from one tree with another, without first making some adjustments.

**Exhibit 17.4: Probability and Outcomes for Dr. A and His Peer Group**

(Probabilities around each node add up to one. AMI=Acute Myocardial Infarction, CHF=Congestive Heart Failure, and AP= Angina Pectoris. Data are not real and for demonstration purposes.)



Notice how clinician does not see any patients who have acute myocardial infarction with low HCC scores. Likewise, the peer group does not see patients with angina pectoris with high HCC scores. In most benchmarking, some patient types are missing from one or the other group. In the literature the extent of the match in branches of the tree is referred to as “overlap” between the two trees. It is rare to have a perfect overlap. Procedures are needed on how to manage the data when the two trees have partial overlap. Of course, a portion of the data where the two trees do not overlap can be ignored; but doing so may throw away a lot of data including crucial data about clinicians who see rare but severely ill patients.

# [H1] Synthetic Cases

One solution is to construct synthetic cases to replace the missing patient outcomes. In synthetic cases, the missing outcome is constructed from the available information through assumption of independence among features of the patients. Not all missing outcomes need to be replaced with synthetic outcomes. When the clinician misses patients seen by the peer group, we can ignore it. The comparison of the clinician and the peer group is done on clinician’s patients. Therefore, patients not seen by the clinician do not affect the final conclusions. There is no need to replace missing outcomes for the clinician. De facto, these patients will have zero chance of occurrence and whatever the outcome it will make no difference. For example, in Exhibit 17.4 the clinician does not see patients with acute myocardial infarction with low HCC scores. Since we are examining the performance of both groups on clinician’s patients, we can safely ignore the cases missing from clinician’s practice. The missing information does not affect either the expected performance of the clinician or the simulated performance of the peer group on the clinician’s patients. Both groups are evaluated on cases seen by the clinician and missing cases from clinician’s practice do not matter.

The situation is different when the peer group does not see patients seen by the clinician. Then, we need to construct synthetic cases to estimate the missing outcome values. There are different ways to construct synthetic cases. One way is to use the independence among patient features. The missing case is broken into its components. The average for each component is calculated and referred to as the marginal average. Under assumption of independence, the product of averages for two complementary components of the missing case can be used to estimate the outcome. For example, Exhibit 17.5 shows the calculation for peer group in Exhibit 17.4. The peer group of the clinician is missing patients with high HCC scores and angina pectoris DRG. To estimate the length of stay for these types of patients, we divide the missing case into two components: (1) high HCC and (2) angina pectoris. For each component the marginal average is calculated. The missing outcome is estimated as the product of two marginal estimates divided by the average of all values for the peer group. See Exhibit 17.5 for calculation of marginal averages and synthetic case outcomes. The marginal average for angina pectoris is 5.5. The marginal average for high HCC scores is 4.5. The missing outcome for the combine angina pectoris and high HCC is calculated as 4.5 \* 5.5 / 4 to be 6.19 days.

**Exhibit 17.5: Estimating Missing Outcomes for Peer Group**

(Missing length of Stay is calculated as 4.5 \* 5.5 / 4 = 6.19)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DRG** | **HCC** | | | **Average** |
| **Low** | **Med** | **High** |
| **Angina Pectoris** | 6 | 5 | ? | 5.5 |
| **Congestive Heart Failure** | 1 | 2 | 3 |  |
| **Acute Myocardial Infarction** | 4 | 5 | 6 |  |
| **Average** |  |  | 4.5 | 4 |

Another way is to construct synthetic cases from a model of the data. First, for the peer group, a regression is done, estimating outcomes as a function of patient’s features. Suppose, the regression formula for predicting peer group’s outcomes is given as Then, the missing case with High HCC and AP is estimated to have an outcome of

With the addition of the outcome for the synthetic case, both the clinician’s and the peer group’s trees have the same structure. Since both trees have the same branches, we can now switch the probabilities and simulate the performance of peer group on clinician’s patients. The observed expected value for the clinician is calculated as, where probabilities are put in parenthesis and the outcomes are shown in bold:

Clinician’s Expected LOS = **6**\*(0.6\*0.8)+**5**\*(0.6\*0.1)+**4**\*(0.6\*0.1)+**3**\*(.3\*.5)+**2**\*(.3\*.3)+**1**\*(.3\*.2)+**5**\*(.1\*.9)+**6**\*(.1\*.1) = 4.62

The simulated expected value for the peer group seeing the clinician’s patients is calculated as:

Peer’s Expected LOS =

=**6**\*(0.6\*0.8)+**3**\*(0.6\*0.1)+**6.2**\*(0.6\*0.1)+**5**\*(0.3\*0.5)+**2**\*(0.3\*0.3)+**5**\*(0.3\*0.2)+**1**\*(0.1\*0.9)+**6**\*(0.1\*0.1) = 4.81

Based on these calculations, we conclude that on average, and on the same set of patients, the clinician is 4.81 – 4.62 = 0.19 days more efficient than the peer group.

# Limitations

Data balancing allows us to simulate the performance of clinicians and peer group on the same set of patients. Despite the availability of these new procedures, it is important to be cautious about benchmarking clinicians as it may lead to unintended consequences (Iezzoni 1997; Krumholz et al. 2002). Benchmarking may distort clinical goals; clinicians may improve their performance on one dimension but inadvertently deteriorate on another. Benchmarking may lead to defensive behavior. Clinicians may put their effort in defending their existing practices as opposed to improving them. An environment needs to be created where no one is blamed and all clinicians are encouraged to seek improvements as opposed to arguing about the results. Inadequate measure of severity may mislead the analysis. A poor severity index, one that is not predictive of the patients’ prognoses, might give a false impression of severity adjustment. Finally, too much measurement, may lead to too little improvement. Sometimes analysts who conduct benchmark studies take considerable time to collect information and analyze it. In these circumstances there may be too little time spent on discussing the results, selecting a new course of action, implementing the change, and following up to make sure that the change is an improvement. It is important to keep in mind that the goal of benchmarking is improvement. Conducting an accurate analysis is only helpful if it leads to improvement, otherwise it is a waste of time.

# [H1] Summary

This chapter described how clinicians and their peer group can be compared on the same set of comorbidities. When it comes to benchmarking performance, clinicians are concerned that they are being blamed for caring for sicker patients, who typically have worse outcomes. This tutorial describes how clinicians and their peer group can be compared on the same set of patients; thus removing concerns with differences in patient populations. First, the distribution of the clinician’s patient characteristics is measured. This distribution is the probability of observing various combinations of patient characteristics. Second the clinician’s and peer group are matched on same patient characteristics. When there is no overlap or match, a synthetic case is organized to make sure that all patients that are seen by the clinician have a comparable match in the peer group. Third, the probability distribution of patients of the peer group is switched with the distribution of the patients of the clinician. This switch allows the analyst to simulate what the outcome would have been if the peer group had seen the same patients as the clinician.

# [H1] Supplemental Resources

Problem set, solutions to problems, multimedia presentations, and other related material are on the course website. The following code shows how SQL can be used to create benchmarks:

/\*

This Code Simulates Peer Group’s Performance on Patients of the clinician

\*/

USE Benchmarking

/\*

The data used in this analysis consists of the following fields:

Each row corresponds to one patient’s outcome of care.

Comorbidities are in columns named DRG and HCC.

The DRG field contains many different values.

HCC field contains 3 different values for low, medium, and high severity.

The column Dr indicates patients was cared for by clinician or peer group.

Outcomes of care are in column LOS.

\*/

-- Calculate pattern of care for clinician

DECLARE @total as float

SET @total = (SELECT COUNT([ID])FROM [dbo].[clinician] WHERE Dr='Clinician')

SELECT [DRG] as DRGa

,[HCC] as HCCa

,Avg(CAST([LOS] as Float)) as LOSa

,COUNT([ID]) as NumA

,COUNT([ID])/@total as ProbA

INTO #Clinician

FROM dbo.Clinician

WHERE Dr='Clinician' -- Select the clinician

GROUP BY [DRG], [HCC]

-- Calculate pattern of care for peer group

DECLARE @totalb as float

SET @totalb = (SELECT COUNT([ID])FROM [dbo].[clinician] WHERE Dr='Peer')

SELECT [DRG] as DRGb

,[HCC] as HCCb

,Avg([LOS]) as LOSb

,COUNT([ID]) as Numb

,COUNT([ID])/@totalb as ProbB

INTO #Peer

FROM dbo.Clinician

WHERE Dr='Peer' -- Select peer group

GROUP BY [DRG], [HCC]

-- Match clinicians and peer group on common strata

SELECT CASE When HCCa IS null Then HCCb Else HCCa END as HCCa

, CASE When DRGa IS null Then DRGb Else DRGa END DRGa

-- Does not matter if outcomes for clinician is null

, CASE WHEN LOSa IS NULL Then -1 Else LOSa END AS LOSa

, CASE When NUMa IS null Then 0 Else NUMa END NUMa

, CASE When ProbA IS null Then 0 Else ProbA END AS ProbA

, CASE When HCCb IS null Then HCCa Else HCCb END as HCCb

, CASE When DRGb IS null Then DRGa Else DRGb END AS DRGb

, CASE When NUMb IS null Then 0 Else NUMb END NUMb

, CASE When ProbB IS null Then 0 Else ProbB END AS ProbB

, LOSb -- Null values require synthetic case calculations

INTO #Match

FROM #Clinician Full Join #Peer on DRGa=DRGb and HCCa = HCCb

-- Overlap between peer and clinician cases

SELECT Round(100.\*CAST (SUM(NUMa)

-SUM(CASE WHEN LOSb is null then NUMa else 0 end) AS float)/

CAST(SUM(NUMa) as Float),2) AS [Overlap without Synthetic Cases]

FROM #Match

-- Calculate peer group's performance, if it had clinician's patients

SELECT NumA

, HCCa AS HCC

, DRGa AS DRG

, ProbA

, LOSa

, ProbA AS ProbB -- Switch probabilities of peer group to clinician

-- For missing outcomes, calculate synthetic outcomes:

, CASE WHEN LOSb IS NULL THEN

(SELECT AVG(LOS) FROM dbo.clinician INNER JOIN #Match ON HCC=HCCb

WHERE Dr='Peer' and LOSb is null) \* --Average for a marginal

(SELECT AVG(LOS) FROM dbo.clinician INNER JOIN #Match ON DRG=DRGb

WHERE Dr='Peer' and LOSb is null) / --Average for complement marginal

(SELECT AVG(LOS) FROM dbo.clinician -- Average for entire set

WHERE Dr='Peer')

ELSE LOSb END AS LOSb

INTO #All

FROM #Match

-- Overlap between peer and clinician cases

SELECT Round(100.\*CAST (SUM(NUMa)

-SUM(CASE WHEN LOSb is null then NUMa else 0 end) AS float)/

CAST(SUM(NUMa) as Float),2) AS [Overlap with Synthetic Cases]

FROM #All

Select Round(SUM(ProbA

\*CASE WHEN LOSb is null then 0 else LOSa End),2) As [Clinician LOS]

, Round(SUM(ProbB\*LOSb),2) AS [Peer LOS]

, Round(((Cast(SUM(ProbB\*LOSb) as float)

-Cast(SUM(ProbA\*CASE WHEN LOSb is null then 0 else LOSa End) as float))\*100)

/Cast(SUM(ProbB\*LOSb) as float),2) AS [Percent More Efficient]

FROM #ALL

# [H1] References

# Abadie A., and G. W. Imbens. 2011. “Bias-Corrected Matching Estimators for Average Treatment Effects.” *Journal of Business and Economic Statistics* 29: 1–11.

# ———. 2006. “Large Sample Properties of Matching Estimators for Average Treatment Effects.” *Econometrica* 74 (1): 235–67.

# Alemi, F., A. H. Elrafey, and I. Avramovic. “Covariate Balancing Through Naturally Occurring Strata.” 2018 Feb; 3(1): 273-292.

# Austin, P. C. 2011. “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.” *Multivariate Behavior Research* 46 (3): 399–424.

# Blagus R, Lusa L. 2013. “SMOTE for high-dimensional class-imbalanced data.” BMC Bioinformatics. 2013 Mar 22; 14:106.

# Cebul, R. D. 1991. “Randomized, Controlled Trials Using the Metro Firm System.” *Medical Care* 29 (7): JS9–18.

# Hansen, B. B. 2004. “Full Matching in an Observational Study of Coaching for the SAT.” *Journal of the American Statistical Association* 99 (467): 609–18.

Heckman J. J., H. Ichimura, and P. Todd. 1998. “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme.” *Review of Economic Studies* 64 (4): 261–94.

Hirano, K., G. Imbens, and G. Ridder. 2003. “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score.” *Econometrica* 71 (4): 1307–338.

# Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2007. “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.” *Political Analysis* 15 (3): 199–236.

# Iezzoni, L. I. 1997. “The Risks of Risk Adjustment.” *Journal of the American Medical Association* 278 (19): 1600–1607.

# Imai, K., and M. Ratkovic. 2014. “Covariate Balancing Propensity Score.” *Journal of the Royal Statistical Society* 76 (1): 243–63.

# Krumholz, H. M., S. S. Rathore, J. Chen, Y. Wang, and M. J. Radford. 2002. “Evaluation of a Consumer-Oriented Internet Health Care Report Card: The Risk of Quality Ratings Based on Mortality Data. *Journal of the American Medical Association* 287 (10): 1277–287.

# Robins, J. M., M. A. Hernan, and B. Brumback. 2000. “Marginal Structural Models and Causal Inference in Epidemiology.” *Epidemiology* 11 (5): 550–60.

Robins, J. M., A. Rotnitzky, and L. P. Zhao. 1995. “Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data.” *Journal of the American Statistical Association* 90 (429): 106–21.

# Rose S, Laan MJ. Why match? Investigating matched case-control study designs with causal Rosenbaum PR. Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. Biometrics. 2007 Jun;63(2):456-64.

# Rosenbaum P.R. 2007 “Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies.” Biometrics Jun; 63(2):456-64

# Rosenbaum, P. R. 1991. “A Characterization of Optimal Designs for Observational Studies.” *Journal of the Royal Statistical Society* 53 (3): 597–610.

# ———. 1989. “Optimal Matching for Observational Studies.” *Journal of the American Statistical Association* 84 (408): 1024–32.

# ———. 1987. J*ournal of the American Statistical Association* 82 (398): 387–94.

# Rosenbaum, P. R., and D. B. Rubin. 1985. “Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score.” *American Statistician* 39 (1): 33–38.

# ———. 1984. “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score.” *Journal of the American Statistical Association* 79 (387): 516–24.

# ———. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70: 41–55.