**Tutorial on Data Balancing: Application to Benchmarking Clinicians**

Sunday, November 04, 2018

Roshan Alemi, MA  
The Advisory Board

[roshan.alemi@gmail.com](mailto:roshan.alemi@gmail.com)

Amr Elrafey, MS

Health Informatics Program, George Mason University

[aelrafey@gmu.edu](mailto:aelrafey@gmu.edu)   
  
Duncan Neuhauser, PhD.  
Case Western Reserve University

[dvn@case.edu](mailto:dvn@case.edu)

Farrokh Alemi, PhD

Health Informatics Program, George Mason University

[falemi@gmu.edu](mailto:falemi@gmu.edu)

703 893 3799**Abstract**

In this tutorial, we show how stratified covariate balancing can be used to benchmark clinicians. Stratified covariate balancing is one of the growing methods for balancing data so unconfounded estimates of effectiveness of treatment can be estimated in observational data. The use of this method, and similar methods such as propensity scoring, has been growing steadily since 1983. This tutorial aims to explain the concepts behind these methods to readers who do not have a strong statistical background. The paper reports the use of covariate balancing in context of comparing the performance of clinicians to their peer groups on the same set of patients. To make understanding of these techniques easier, we take three steps. First, we use decision trees to demonstrate the concepts. This allows the reader to visualize the subset of patients that are being compared. Second, we replace statistical weighting, e.g. inverse propensity scoring or stratified covariate weights, with switching of probability distributions. The switch replaces the distribution of the patients managed by the peer group with the distribution of patients managed by the clinician. The switching simulates how the peer group would have performed if they were caring for clinician’s patients. Finally third, to improve the overlap between peer group and clinician’s patients, we introduce procedures for constructing synthetic cases. These synthetic cases replace missing cases and allow all clinician’s patients to have at least one match in patients cared for by the peer group. The procedures described here can be applied easily to data in electronic health records and we present Standard Query Language for doing so, enabling widespread use of data balancing in benchmarking clinicians.

**Introduction**

When clinicians are compared to each other, the common complaint is that the comparison is unfair. Clinicians are concerned that they see sicker patients than their colleagues. They are blamed them for having worse outcomes; when in reality these outcomes are expected in sicker patients. Since no two clinicians see 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 [[[1]](#endnote-1)]. Another approach, one that we discuss in this tutorial, is to let the patients select their clinicians as they wish. Then statistical procedures, in particular data balancing, are used to simulate the performance of peer group on patients of the clinicians.

Data balancing refers to weighting the data so that sicker patients occur at the same rate among clinician and peer group patients. Data balancing was first proposed in 1983 [[[2]](#endnote-2)]. Since then, the approach has been repeatedly improved [[[3]](#endnote-3)-[[4]](#endnote-4),[[5]](#endnote-5),[[6]](#endnote-6),[[7]](#endnote-7),[[8]](#endnote-8),[[9]](#endnote-9),[[10]](#endnote-10),[[11]](#endnote-11),[[12]](#endnote-12),[[13]](#endnote-13),[[14]](#endnote-14),[[15]](#endnote-15),[[16]](#endnote-16)] and is in widespread use with several tutorials describing the nuances of propensity scoring, a type of data balancing [[[17]](#endnote-17)]. Most recently, “stratified covariate balancing” has been used as a method of balancing the data [[[18]](#endnote-18)]. In this tutorial, we use stratified covariate balancing to benchmark clinicians.

|  |
| --- |
| **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. |

This tutorial is written for readers without a strong statistical background. We take several steps to explain complex statistical procedures in ways that clinicians and non-statisticians can build intuitive understanding of the issues involved. We separate the mathematical issues from the main ideas. We show short cuts that remove the need for estimating propensity weights. We also visualize the data using decision trees. These trees let the reader see the distribution of patients and thus be reassured that clinicians and their peer groups are compared on same patients. Finally, instead of using statistical terms such as sample, covariates, and treatment effect, we use terms that make more sense in our context. Sample is replaced with patient groups. Covariates are replaced with patient characteristics; and treatment effect is replaced with average outcome for clinician.

**Counterfactual Calculations & Switching of Probabilities**

In statistical balancing of observational data, weights are used to balance data so covariates occur at the same rate across the two groups. In our terminology, the patients of the peer group are weighted so that they have the same characteristics as the clinician’s patient. Often, the establishment of the weights is the first step in balancing the data. These weights are derived through statistical models, usually regression of treatment on covariates. In stratified covariate balancing, weights are derived analytically without statistical modeling. In this approach, all one needs to do is estimate the distribution of patient characteristics in the clinician and his peer group. Then, these data are used to assess weights that would make the peer distribution the same as the clinician’s distribution. In fact, since weights are simply the means to the end of making the two distributions the same; one can discard with the weighting procedure and simply switch the distribution of the peer group’s patient with the distribution of the clinician’s patients (see the box for the mathematics of how to do so). This switch in distributions by passes the awkward and often confusing steps necessary to estimate the weights. It also helps the interpretation of the findings: the switch simulates how the peer group would have performed if they had seen the clinician’s patients. One contribution of stratified covariate balancing to the data balancing field has been to show how combination of patient characteristics, called strata, can be used to calculate patient distributions and how with or without weighting these distributions can be switched so that peer group’s performance is simulated on clinician’s patients.

Readers not familiar with data balancing may be concerned that switching may distort the findings. These simulated values are counterfactual calculations, in the sense that these statistics report outcomes if the peer physicians 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. Since the observed clinician’s and the simulated peer’s outcomes are measured on the same set of patients, these are apple-to-apple comparisons. The difference between the two measures 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 Table 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.

**Table 1:  Severity Adjusted Comparison of Performance of Several Clinicians**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Severity of patients | **Clinician A** | | **Peer Group of A** | | |
| 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. 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 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 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 that clinician's patients to weigh the outcomes of the peer providers:

The clinician seems to be less efficient than the average of his peer group. 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.

**Comparing Clinicians on Matched Comorbidities**

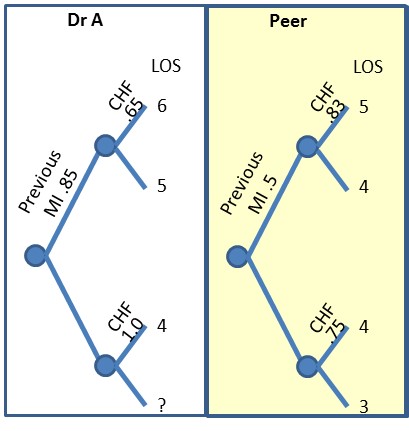
In the previous section, we divided patients in broad categories of severity (Low, Medium and High) and compared care provided within each category. To calculate these benchmarks, one needs access to a reliable and valid measure of the severity of illness. Sometimes, such a measure is not available or not trusted to adequately measure the full spectrum of severity of the patient's illness. When no severity index is available, an analyst must match patients of clinicians and his/her peer feature by feature and comorbidity by comorbidity. If we think of the clinician as cases and the peer group as controls, then this is similar to doing a matched case control study, where covariates in cases are matched to the same covariates in controls, and the average outcomes are examined. A great deal has been written on matching and many methods exist to match cases to controls [[[19]](#endnote-19),[[20]](#endnote-20),[[21]](#endnote-21),[[22]](#endnote-22),[[23]](#endnote-23),[[24]](#endnote-24)] or in our terminology patients of clinicians to patients of peer group.

An example can demonstrate the use of patient matching. Table 4 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 specific characteristics.

**Table 4: Patients of Clinician and Peer Providers May Differ in Significant Ways**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Clinician'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 | 4 |

An event tree can organize and summarize the data in the Table 4. Each feature of the patient (e.g. previous MI or CHF) can be used to create a new branch in the event tree. The tree then ends with the outcomes presented to the right of each branch. 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, are called strata. The idea is to make sure that both the clinician and the peer group are matched on the same branches or same strata. For example, the event tree for the patients seen by the clinician and peer group are provided in Figure 1. In this tree, previous MI is the first event, then 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. To make it easier to read the tree and because within each node the probabilities add up to 1, the probabilities of negative events, such as not having CHF, are not provided.

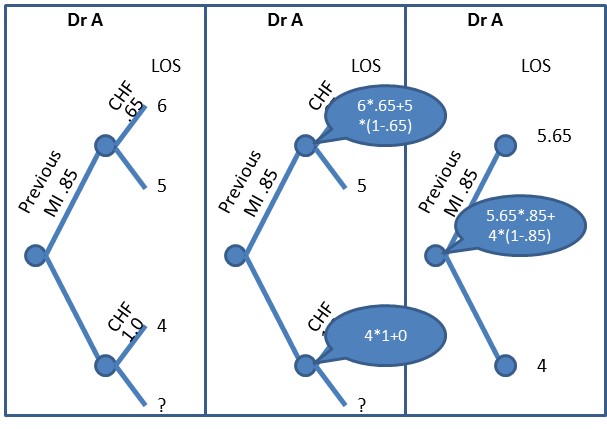
**Figure 1: Event Trees for the Clinician & His Peer Group**(CHF=Congestive Heart Failure, MI=Myocardial Infarction, LOS=Length of Stay)  


The expected LOS for the patients of the clinician being evaluated was: 5.4 days. This is obtained by folding back the tree to the root node as shown in Figure 2. Starting from the right upper corner of the tree, we first fold the event of occurrence of CHF. This is calculated as: 6 days \* 0.65 + 5 days \* (1 - 0.65). Next we do the same for CHF in the lower-right side: (4 days \* 1.0 + 0). The process of folding back continues by replacing the uncertain event with the calculated expected values. Each node is replaced with the expected LOS. This is done until we reach to the root of the tree. The expected value for the entire tree is calculated as:

Expected length of stay = (6 \* 0.65 + 5 \* (1 - 0.65)) \* (0.85) + (4 \* 1.0 + 0) \* (1-0.85) = 5.40

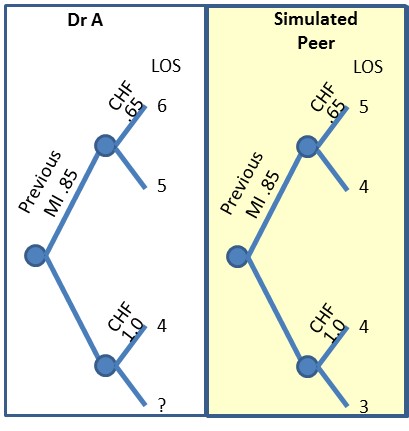
**Figure 2: Folding Back in 3 Steps**

(Probabilities around each node add up to one)

****

To simulate how the same patients would have been cared for by peer clinicians, the event tree is kept as before, but the average LOS of each patient grouping is replaced with the average LOS of the patients of peer clinicians. Figure 3 provides the resulting trees.

**Figure 3: Simulated Performance of Peer Group on Patients’ of the Clinician**(Probabilities around each node add up to one)



Note that the tree in Figure 3 is a hybrid. Part of it comes from the clinician’s frequency of seeing different types of patients. The other part, the outcomes, comes from the peer group. The expected Length of Stay (LOS) of the patients of Dr. A was calculated to be 5.40 days. If the peer group sees Dr. A’s patients, then the simulated expected LOS is 4.55:

Simulated LOS: (5\*.65+4\*(1-.65))\*.85 + (4\*1+3\*(1-1))\*(1-.85) = 4.55

Thus, Dr. A’s patients stay on average 0.85 days longer than his peer group treating the same group of patients. In 20 patients, Dr. A’s patients stay 17 days longer than if these patients had been treated by Dr. A’s peer group. Note that if we had not switched the probabilities, we could not claim that performance had been calculated based on same patients.

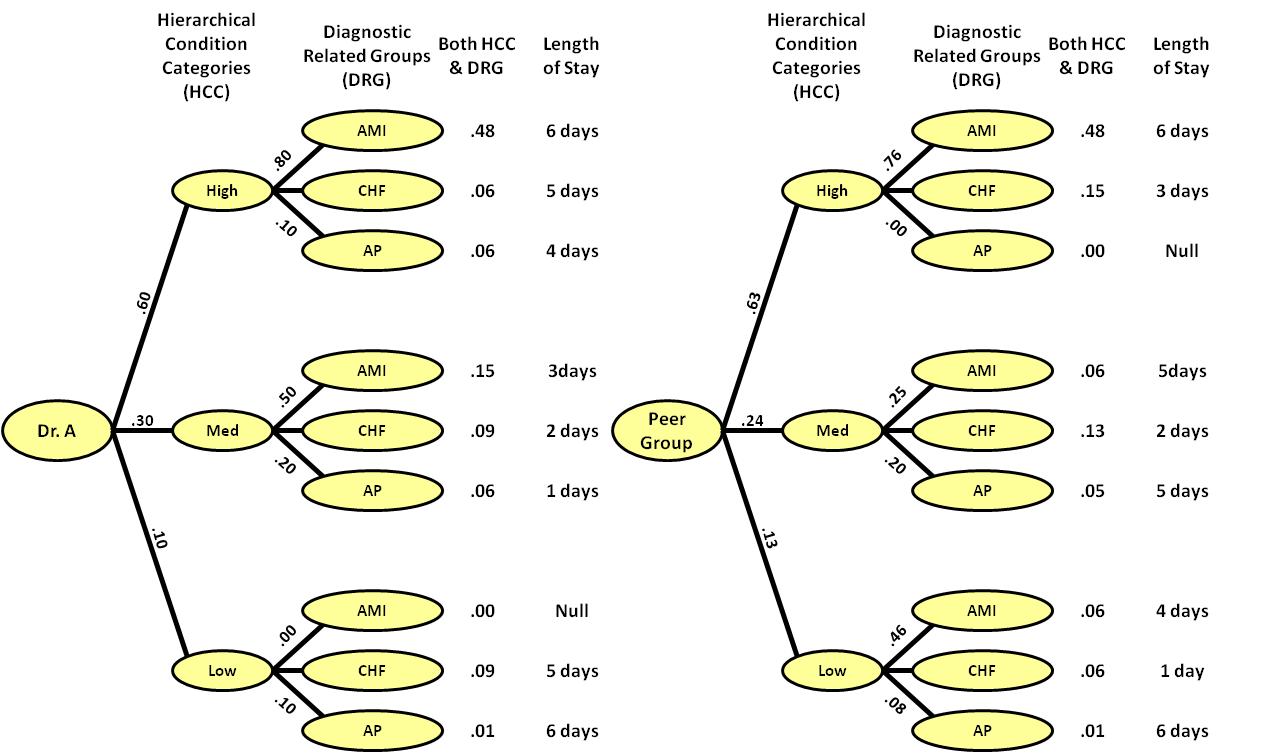
**Overlap of Cases among Clinicians and their Peers**

So far, we have compared the clinician and the peer group on same patients by finding the same patients in the two groups and noting how clinician and peer group differ. Matching patients doesn’t always work. As the number of features increases, the number of data points that falls into each branch on the decision tree becomes smaller. Soon most branches will have no patients. Many peer providers' cases cannot be matched feature by feature 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 study conclusions become less defensible. Stratified covariate balancing improves the match between clinicians and their peers by creating synthetic patients where actual patients are missing.

For example consider the following data on Dr. A and his peers. For each patient managed by these doctors 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 is always available, we will start the event tree using this variable. DRGs are not always occurring for all physicians 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 probability and conditional probability of DRG shows the joint probability of the combination of HCC and DRG. These values are also shown in Figure 4. Also further suppose that Dr. A and his peer have the length of Stay (LOS) indicated in Figure 4. A glimpse at the tree suggests that Dr. A’s peer may be seeing patients with higher severity and thus longer stays. The tree for Dr. A and the tree for his peer group are structurally different. There are some branches in Dr. A’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.

**Figure 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 Dr. A 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 general, in most benchmarking some strata are always 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.” 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.

One solution is to construct synthetic cases to replace the missing patients. Not all missing cases need to be replaced with synthetic cases. When the clinician does not see patients seen by the peer group, we can ignore the missing information. The comparison of the clinician and the peer group is done on clinician’s patients and therefore patients not seen by the clinician do not affect the final conclusions. De facto, these patients will have 0 probability of occurrence. For example, in Figure 4 the clinician does not see patients with acute myocardial infarction with low HCC scores. Since we are examining the performance of the peer group on clinician’s patients we can safely ignore these patients. It does not affect the performance of the clinician and neither the simulated performance of the peer group on the clinician patients.

The situation is different when clinician’s patients are not seen by the peer group. Then, we need to construct synthetic cases to estimate the missing outcome values. Marginal values are calculated by the average of outcomes that share the feature. For example, Table 5 shows the calculation for peer group in Figure 4. The peer group of Dr. A is missing patients with high HCC scores and angina pectoris DRG. To estimate the length of stay for these types of patients, we multiply the marginal estimate for angina pectoris DRG (i.e. the average of values for angina pectoris DRG) by the marginal value for high HCC patients (i.e. the average value for high HCC patients) and divide by the average of all values for the peer group. See Table 5 for calculation of marginal averages and synthetic case outcomes.

**Table 5: Estimating Missing Outcomes for Peer Group**

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

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

With the addition of the outcome for the synthetic case, both trees have the same structure but different distributions. Since both trees have the same branches, we can switch the probabilities and simulate the performance of peer group on patients of Dr A. The observed expected value for Dr. A is calculated as, where probabilities are put in bold:

**0.33**\*(**0.6**\*3 +**0.4**\*4+0) +**0.33**\*(**0.4**\*2 +**0.2**\*3 **+ 0.4**\*4) + **0.34**\*(**.8**\*1+**.1**\*1.5+**.1**\*2) = 2.503 days

The simulated expected value for the peer group seeing the patients of Dr. A is calculated as:

**0.33**\*(**0.6**\*6 +**0.4**\*5+0\*4) +**0.33**\*(**0.4**\*3 +**0.2**\*2 **+ 0.4**\*1) + **0.34**\*(**.8**\*6.19+**.1**\*1.5+**.1**\*6) = 4.552 days

Based on these calculations, we conclude that on average, and on the same set of patients, Dr. A is 4.552-2.503=2.036 days more efficient than his/her peers.

**Discussion**

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.

Despite the availability of these new procedures, it is important to be cautious about benchmarking clinicians as it may lead to unintended consequences [[[25]](#endnote-25),[[26]](#endnote-26)]: 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.

**Appendix**

/\*

This Appendix 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

**References**

1. Cebul RD. Randomized, controlled trials using the Metro Firm System. Med Care. 1991 Jul;29 (7 Suppl): JS9-18. [↑](#endnote-ref-1)
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983, 70, 41–55. [↑](#endnote-ref-2)
3. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am. Statistn 1985, 39, 33–38. [↑](#endnote-ref-3)
4. Rosenbaum PR. Optimal matching for observational studies. J. Am. Statist. Ass. 1989, 84, 1024–1032. [↑](#endnote-ref-4)
5. Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. Econometrica 2006, 74, 235–267. [↑](#endnote-ref-5)
6. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using sub-classification on the propensity score. J. Am. Statist. Ass. 1984, 79, 516–524. [↑](#endnote-ref-6)
7. Rosenbaum PR. A characterization of optimal designs for observational studies. J. R. Statist. Soc 1991. B, 53, 597–610. [↑](#endnote-ref-7)
8. Hansen BB. Full matching in an observational study of coaching for the SAT. J. Am. Statist. Ass. 2004, 99, 609–618. [↑](#endnote-ref-8)
9. Rosenbaum PR. Model-based direct adjustment. J. Am. Statist. Ass. 1987, 82, 387–394. [↑](#endnote-ref-9)
10. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000, 11, 550–560. [↑](#endnote-ref-10)
11. Hirano K, Imbens G, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 2003, 71, 1307–1338. [↑](#endnote-ref-11)
12. Heckman JJ, Ichimura H, Todd P. Matching as an econometric evaluation estimator. Rev. Econ. Stud. 1998, 65, 261–294. [↑](#endnote-ref-12)
13. Imai K, Ratkovic M. Covariate balancing propensity score J. R. Statist. Soc. B (2014), 76, Part 1, pp. 243–263. [↑](#endnote-ref-13)
14. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J. Am. Statist. Ass. 1995, 90, 106–121. [↑](#endnote-ref-14)
15. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit. Anal. 2007, 15, 199–236. [↑](#endnote-ref-15)
16. Abadie A, Imbens GW. Bias-corrected matching estimators for average treatment effects. J. Bus. Econ. Statist 2011, 29, 1–11. [↑](#endnote-ref-16)
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May;46(3):399-424. [↑](#endnote-ref-17)
18. Alemi F. Elrafey AH, Avramovic I. Covariate Balancing through Naturally Occurring Strata. In review and available through first author falemi@gmu.edu [↑](#endnote-ref-18)
19. Li H, Yuan Z, Su P, Wang T, Yu Y, Sun X, Xue F. A simulation study on matched case-control designs in the perspective of causal diagrams. BMC Med Res Methodol. 2016 Aug 20;16(1):102. [↑](#endnote-ref-19)
20. Rose S, Laan MJ. Why match? Investigating matched case-control study designs with causal effect estimation. Int J Biostat. 2009 Jan 6;5(1):Article 1. doi: 10.2202/1557-4679.1127. [↑](#endnote-ref-20)
21. Gyöngyösi M, Ploner M, Porenta G, Sperker W, Wexberg P, Strehblow C, Glogar D. Case-based distance measurements for the selection of controls in case-matched studies: application in coronary interventions. Artif Intell Med. 2002 Nov; 26(3):237-53. [↑](#endnote-ref-21)
22. Heckman JJ, Ichimura H, Todd P. Matching as an econometric evaluation estimator. Review of Economic Studies 1998, 65, 261-294. [↑](#endnote-ref-22)
23. Rosenbaum PR. Sensitivity analysis for matched case-control studies. Biometrics. 1991 Mar; 47(1):87-100. [↑](#endnote-ref-23)
24. Rosenbaum PR. Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. Biometrics. 2007 Jun;63(2):456-64. [↑](#endnote-ref-24)
25. Iezzoni LI. The risks of risk adjustment. JAMA. 1997 Nov 19;278(19):1600-7. [↑](#endnote-ref-25)
26. Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented internet health care report card: the risk of quality ratings based on mortality data. JAMA. 2002 Mar 13;287(10):1277-87. [↑](#endnote-ref-26)