

# Reporting Healthcare Data: Understanding Rates and Adjustments

Gregory S. Nelson

ThotWave Technologies, Chapel Hill, North Carolina

## Abstract

In healthcare, we often express our analytics results as being “adjusted”. For example, you may have read a study in which the authors reported the data as “age-adjusted” or “risk-adjusted.” The concept of adjustment is widely used in program evaluation, comparing quality indicators across providers and systems, forecasting incidence rates and in cost-effectiveness research. In order to make reasonable comparisons across time, place or population, need to account for small sample sizes and case-mix variation – in other words – we need to level the playing field and account for differences in health status and accounts for uniqueness in a given population.

If you are new to healthcare it may not be obvious what that really means to adjust the data in order to make comparisons. In this paper, we will explore the methods by which we control for potentially confounding variables in our data. We will do so through a series of examples from the healthcare literature in both primary care and health insurance. Included in this survey of methods, we will discuss the concepts of rates and how they can be adjusted for demographic strata (such as age, gender and race) as well as health risk factors such as case mix.

<b>ABSTRACT .....</b>	<b>1</b>
<b>INTRODUCTION .....</b>	<b>2</b>
UNDERSTANDING RATES.....	3
WHY SHOULD WE CARE ABOUT ADJUSTING DATA? .....	4
ADJUSTMENT TYPES .....	6
STATISTICAL ADJUSTMENTS .....	10
<b>REAL-WORLD EXAMPLES.....</b>	<b>11</b>
COMPARING QUALITY MEASURES (ASSOCIATIONS).....	11
AGE-ADJUSTMENTS (PROPORTIONS).....	13
RISK-ADJUSTMENTS (FOR PATIENT LEVEL ANALYSIS) .....	15
CASE-MIX ADJUSTMENTS (FOR ORGANIZATIONAL LEVEL ANALYSIS) .....	16
QUALITY-ADJUSTED LIFE YEARS (QALY).....	19
<b>SUMMARY.....</b>	<b>20</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>20</b>
<b>REFERENCES .....</b>	<b>20</b>
BIOGRAPHY .....	21
CONTACT INFORMATION: .....	22

# Introduction

Like many industries, healthcare is in a state of continuous improvement. We are looking for better ways of treating patients and achieving better outcomes. We see examples in drug discovery (comparative effectiveness of drugs/ therapies), healthcare providers (variations in practice), and equality among hospitals (quality, cost, outcomes) and in health economics (cost-effectiveness of various treatment options.) In each of these scenarios, we want to be sure that when we make these comparisons, the differences in the measure or rate is comparable across time, place or patient population and not due to some other factor (known or unknown.)

There are examples in both academic publications, as well as in the press, where these comparisons have stirred controversy. Both health payers and healthcare providers often argue that their patients are poorer, older/ younger and even sicker and any analysis that does not take this into account is open to scrutiny. One popular press article in fact reported data concerning the chance of a patient dying after bypass surgery. Here, the raw number of deaths was adjusted to take into account the fact that some hospitals and surgeons do take care of sicker patients. In this article, they show, for each hospital, the death rate and the 95% confidence intervals. From this you can see that there are two hospitals that are significantly below the statewide rate. This article was challenged not only by the hospitals themselves but also the industry group that represents surgeons.



## Heart Bypasses Are Safer, Study Shows

### But a Few Hospitals Fall Below Average

By ELISABETH ROSENTHAL

The safety of cardiac bypass surgery in New York State hospitals has continued to improve, although a few hospitals and surgeons had track records that were significantly worse than the average, state health officials reported yesterday.

In releasing its annual survey of the bypass operations, the State Health Department described the results as extremely good, since the rate of dying after bypass surgery in New York is lower than that reported from any other area of the state.

Patients in New York are lucky," Dr. Barbara DeBuono, the State Health Commissioner, noting that the chance of dying from the procedure had decreased almost 50 percent in the last five years. At the study, which each year audits 31 hospitals and dozens of cardiac surgeons performing bypass surgery in the state, also raised red flags regarding some hospitals and surgeons, pointing to areas that need improvement. Three update hospitals — Buffalo General, Strong Memorial and University Hospital — showed performances that were statistically worse than the statewide norm.

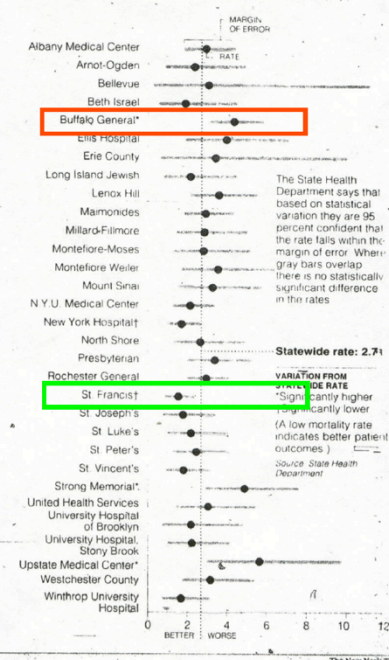
Based on 1993 data, the state found two hospitals, Saint Francis Hospital in Roslyn, L.I., and New York Hospital-Cornell Medical Center in Manhattan, had unusually low statistical ratings this year, according to the report. Although the state's officials said that the state's statistics were outdated, that the study did not adequately take into account how sick patients were before surgery and that the state's data of statistical analysis could directly identify a hospital as a performer by chance.

For example, Dr. Robert Panzer,

### RATES

#### How the State's Hospitals Fared

Risk-adjusted mortality rates of New York State hospitals performing coronary artery bypass surgery. This rate is adjusted for the fact that some hospitals have higher-risk patient populations than others.



The state health commissioner cautioned "that the latest statistics were not primarily intended to be used by patients in choosing a hospital or a surgeon but instead a tool to motivate or—embarrass—those with poor records to clean up their acts."

In other contexts, such as financing and in evaluating financial returns, we see the notion of “adjusting” for time or risk. It is pretty common that we adjust for inflation and discount based on the time-value of money or the riskiness of a portfolio. We learned that the present value of money is greater than the same amount of money in the distant future (time value of money.) To account for this, we use discounted cash flows when both costs and consequences accrue over time – such as a therapy that will extend over a long period of time. For example, let’s imagine that we have a healthcare intervention that will cost us \$50,000 and the money will be spent over twenty years. If we want to compare two or more interventions (say “medication alone” versus “surgery + medication”) and one takes 20 years and another takes only 10 years, we have to make sure that we adjust our valuation (cost-effectiveness) of both therapies using the “present value” of money so that the comparisons we make are on a level playing field. Similarly, we want to level the playing field as it relates to other factors beyond cost, such as the patient’s quality of life (QALY) or Disability-Adjusted Live Year (DALY) [see for example (WHO, 2014).]

If we compare quality outcomes between two hospitals, it would be prudent to normalize the data to account for the fact that the two hospitals may treat a larger number of patients or perhaps they really do treat sicker patients. For example, let’s say that we want to compare mortality rates for two hospitals. One hospital had 28 patients die and the other had only 9 patients die. On first blush, we may say that the first hospital had worse outcomes than the second. The most obvious question we should be asking – is how many total patients did they treat in the time period under review? If the hospital that had 9 patients die, but only treated 10 patients, that would be a problem! We will go through this example in detail later in this paper – but suffice it to say, this is a critical measure to ensure that we are comparing the two entities in a responsible manner.

In this paper, we will provide real-world examples and discuss techniques that can be used to adjust our “raw” data so that reasonable comparisons can be made. While this paper is not intended to provide a primer on statistical methods for dealing with confounding variables, rather we hope to provide a brief introduction to the concepts and survey the landscape of approaches.

### ***Understanding Rates***

In healthcare, and in particular epidemiology, we often express our data as a rate. Rate is usually expressed as the total number of events occurring for a given population over some period of time. For example, we may want to express the teen pregnancy rate as the number of pregnancies for women between 10 and 19 over some time period and geography (United States for 2013.) The numerator in this case is the number of women that become pregnant and the denominator would be the total number of women in the population. Because the rates can often be very small, we can express this as the number of cases per 100,000.

For the most part, there are three types of rates:

- **Crude Rate** – The rate usually expressed as the total number of events occurring for a given population over some period of time. For example, the annual death rate is simply the number of deaths that occurred in a given time period (say for the year 2013) divided by the total population under examination (the population of the United States) times a multiplier, usually 100,000 for this measure.

$$\frac{\text{Number of US deaths in 2013}}{\text{Estimated US population in June, 2013}} \bullet 100,000$$

- **Specific Rate** – Similar to crude rate except this is the rate for a given subgroup such as age, race or gender. For example, we can think of a specific mortality rate as the number of deaths in a specific sub-population such as the elderly or those with cardiovascular disease. To calculate the age-specific mortality rate, we would calculate the deaths in person of a certain age range divided by the midyear population in that same age range x 100,000.

$$\frac{\text{Number of deaths for specific age range}}{\text{Estimated population in that age group}} \bullet 100,000$$

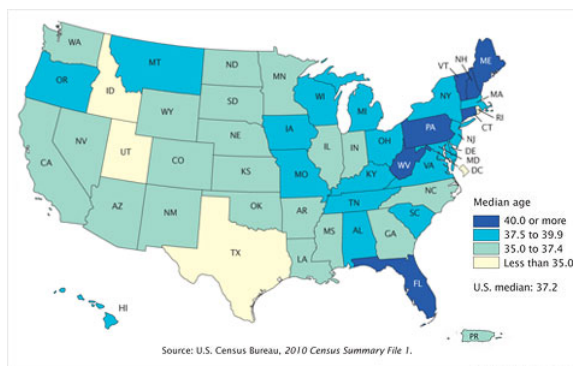
- **Adjusted Rate** – A summary rate that is statistically adjusted to remove the effect of a variable (such as age, race or gender), to allow for an unbiased comparison between groups that may differ with respect to these variables. For example, as we look at mortality rates, we know that the mortality rates for 110-year olds is perhaps a bit higher than say 38 year olds. To account for these differences we need to adjust our crude rate to account for the differences that age brings.

A simple example here illustrates the point. Let's say that we want to create a weighted average based on age-specific incidence and age distribution of population. To do this, we calculate the incidence for each age group and multiply it by the proportion of people in that age group

$$\begin{aligned} \text{Age-adjusted incidence} = & \\ & \text{Age-specific incidence 0-4 years} \times \text{Proportion 0-4 years} + \\ & \text{Age-specific incidence 5-9 years} \times \text{Proportion 5-9 years} + \dots \\ & \text{Age-specific incidence 95-99 years} \times \text{Proportion 95-99 years} \end{aligned}$$

### Why should we care about adjusting data?

Intuitively, we know that in order to understand what a measure means, we have to have some context. For example, if we wanted to compare the birth rates for two different states, we might want to first get a clear



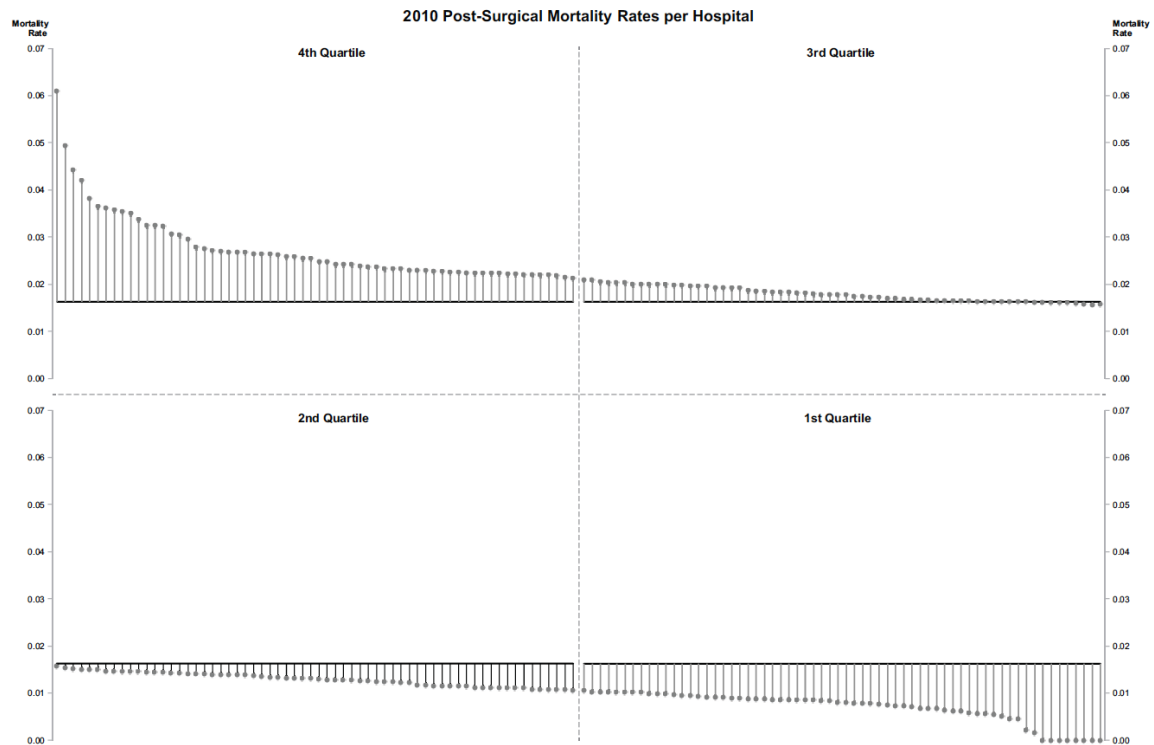
picture of the average age. Saying that Maine, for example, has a low birth rate (9.8 per 1,000) as compared with the national birth rate of 13 (Foundation, 2014), might be misleading unless we first understand the population of the state and the number of potential women capable of giving birth in the state. In the map shown here, we can see that Maine has one of the oldest populations which might help explain the difference (Christie, 2011).

Another example comes straight from the headlines like the ones shown here. Sadly, because most people interpret raw numbers (or crude rates) as evidence, the lose sight of the context where more accurate comparisons can be made.

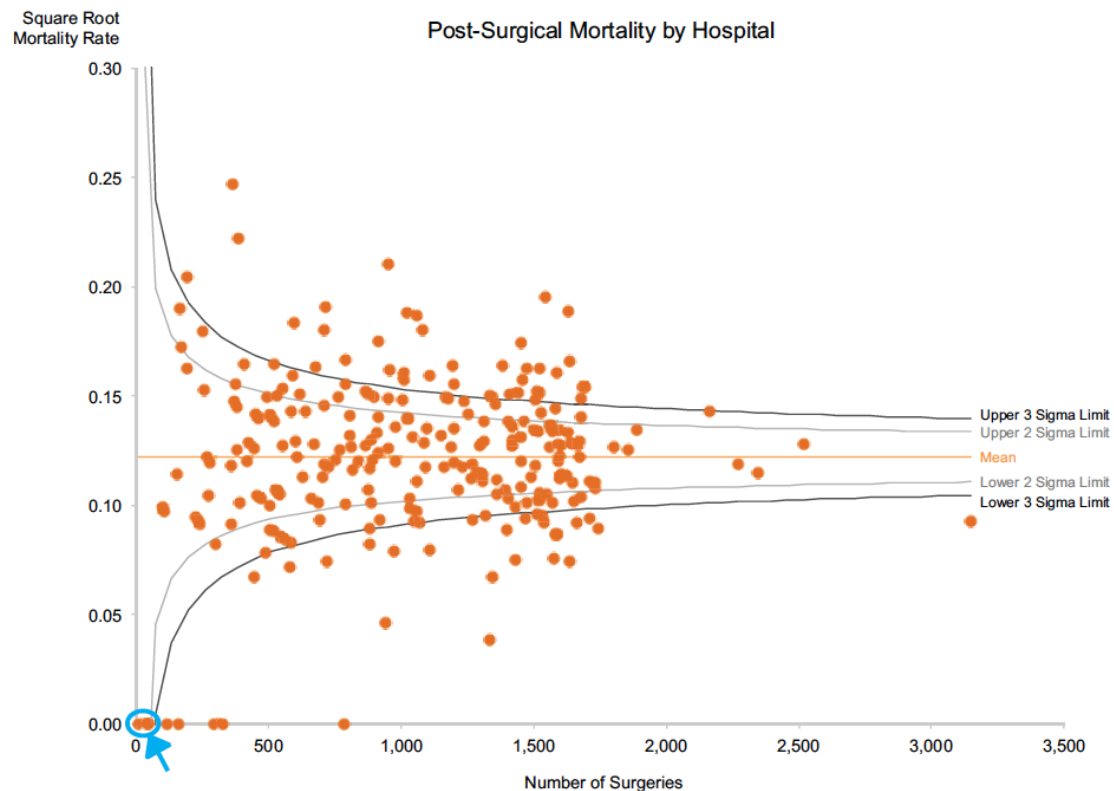
Recently, Stephen Few and Katherine Rowell demonstrated this by showing us surgical mortality rates for a number of hospitals (Few & Roswell, 2013). In the figure below, each data point represents a hospital's

mortality rate ranked order from the highest (4<sup>th</sup> quartile) to those hospitals with the lowest rate. The values are mortality rates following surgeries and the solid horizontal line represents the mean.

Note, there are a number of hospitals with mortality rates of over 3% (and even one over 6%!) while there are several hospitals that have a 0% mortality rate. The challenge, Few and Roswell argue, is that the sample size for the hospitals represented here range from 7 patients to over 3,000 and go on to say “For this reason it isn’t appropriate to rank the hospitals by mortality rate. The ranking suggests a relationship of relative performance that cannot be determined by the data.”



Arguably, the data should be adjusted to account for both the sample sizes as well as the varying levels of risk as some surgeries are more risky than others and some patients, due to various factors including age, comorbidities, disease severity, are more at risk than others. Few and Roswell offer the funnel chart as an alternative way to show the same data while controlling for sample size.



The authors highlight the fact that what was once seen as a very low mortality rate can now be viewed in the context of the sample size – that is, the number of surgeries conducted. While Few and Rowell admit that this still doesn't take into account (adjust for) risk or case-mix, it does demonstrate the idea that adjusting for context is critically important.

### ***Adjustment Types***

So by now, you understand that data need to be presented in context to tell the complete story. Considering factors such as health status, age, gender, income and other demographic variables help us adjust for the impact of these factors. Organizations such as health systems (hospitals, integrated delivery networks, Accountable Care Organizations (ACOs), Patient Center Medical Homes (PCMHs)) and health payers (self-insured organizations, government, health insurers/ plans), often take into account different types of adjustments including case-mix, risk adjustments and other segmentation or stratification methodologies to “level the playing field”.

Here are some “adjustments” plucked from the medical literature:

- “**Risk adjusted** 30-day mortality after a bypass surgery”
- “Standard error of the **age-adjusted** death rate”
- “**Risk adjusted** readmission rate after a hospitalization for acute decompensated heart failure”
- “Impact of 24 hour critical care physician staffing on **case-mix adjusted** mortality in paediatric intensive care”
- “The **risk-adjusted** rate of in-hospital hip fracture among acute care inpatients aged 65 years and over”



- “Willingness to pay for a **quality-adjusted life year**: implications for societal health care resource allocation”
- “Improved **comorbidity adjustment** for predicting mortality in Medicare populations”
- “The relative risks were similar for fatal and nonfatal disease and were unaltered after **adjustment for cigarette smoking, hypertension, diabetes, high cholesterol levels, a parental history of myocardial infarction, past use of oral contraceptives, and obesity**”
- “Crude and **Age-Adjusted** Death Rates: United States, 1960-2004”
- Impact of hip and vertebral fractures on **quality-adjusted life years**

The table below summarizes some of the most common adjustment types along with a working definition and how this might be used/ implemented in the real world.

Term	Description
Age-adjusted	<p><i>Definition:</i></p> <p>Also called age standardization, is a technique used to better allow populations to be compared when the age profiles of the populations are quite different. Standardizations can be used on other demographic variables such as race, geographic location or gender.</p> <p><i>Reason:</i></p> <p>Allows for fairer comparisons between groups with different age distributions. For example, a county having a higher percentage of elderly people may have a higher rate of death or hospitalization than a county with a younger population, merely because the elderly are more likely to die or be hospitalized. Age adjustment can make the different groups more comparable.</p> <p><i>Method:</i></p> <p>A "standard" population distribution is used to adjust death and hospitalization rates. The age-adjusted rates are rates that would have existed if the population under study had the same age distribution as the "standard" population. Therefore, they are summary measures adjusted for differences in age distributions. Adjustments can be made using direct or indirect standardization.</p> <p>Source: (Services, 2014)</p>
Risk Adjustment	<p><i>Definition:</i></p> <p>Risk adjustment refers to the modifications made to measurement to reflect the health status of patients. It is the “use of patient-level information to explain variation in health care spending, resource utilization and health outcomes over a fixed time period.” (Ellis, 2007) A risk adjustment model refers to an actuarial tool used to predict health care costs based on the relative risk of enrollees (say for example in a health plan.)</p> <p><i>Reason:</i></p> <p>Risk adjustment is a tool that we can use to aid in the comparability of populations, place</p>

Term	Description
	<p>and time. It creates a level playing field when comparing rates in that we can account for differences in health status and accounts for uniqueness in a given population. For example, Medicare uses risk-adjustment efforts to determine the reimbursement rates for Medicare Advantage patients and assessing penalties for readmission rates.</p> <p>Common examples include variables such as demographics variables (age, gender, family size, category of assistance, geography), prescription drugs/ medications, functional status, self-reported health status, prior utilization or expenditures, diagnosis code and other interventions. The goal is to create categorizations of patients that are similar in their health status and resource needs. Risk adjustment is commonly used in quality outcomes measurement, case and disease management, predictive modeling and historical comparative reporting. We also see this in risk-based payment approaches that are becoming more commonplace with ACOs and PCMH [see for example, (Siegel, 2012).]</p> <p><i>Method:</i></p> <p>Just like case-mix adjustment, we can use the techniques we saw above – our focus in just slightly different as we really do care about the individual level of analysis. This is often useful for care coordination or setting case-management priorities. Once such example is the <b>King’s Fund Combined Model</b> – used primarily in the National Health Service (NHS) in the United Kingdom as a methodology to generate risk scores for patients susceptible for readmission based on a variety of clinical, administrative and demographic data. For an article that highlights the practical application of this in reducing readmissions, see (Billings et al., 2012) and (Billings, Georghiou, Blunt, &amp; Bardsley, 2013). We at ThotWave implemented this model in support of the NHS to reduce chronic disease readmissions for one of the largest primary care trusts (PCTs) in London.</p>
Case-Mix Adjustment	<p><i>Definition:</i></p> <p>An adjustment technique that takes into account the relative numbers of various types of patients being treated as categorized by disease-related groups, severity of illness, rate of consumption of resources, and other indicators; used as a tool for managing and planning health care services.</p> <p><i>Reason:</i></p> <p>Case-mix adjustment is an attempt to control for a particular type of confounding, namely that some patients have a set of covariate risk factors that predispose them to the outcome. Case mix adjustment is used primarily to evaluate the relative performance of a “population” of patients. For example, when comparing the incidence of surgical errors at two hospitals, we might want to control for the fact that one is a major regional teaching hospital that gets very complex cases.</p> <p>Case-mix adjustment and risk-adjustment are often used interchangeably but we think a useful distinction can be made with case-mix being applied at the organizational level while risk-adjustment is often made at the patient level.</p>



Term	Description
Quality Adjusted Life Years (QALY)	<p><i>Method:</i></p> <p>There are a number of ways to adjust for case-mix. We can stratify with small numbers, match using something like a propensity score, include some measure of case-mix or risk (e.g., particular covariates or risk scores (APACHE II) or indexes (Charlson Comorbidity Index) as a covariate in a regression model, or ever more sophisticated techniques to try to arrive at an unbiased estimate.)</p> <p>There are a hundred of methods that are used to perform risk-adjustments – including models tailored to a unique set of patients. A few of the common risk-adjustment methodologies include:</p> <ul style="list-style-type: none"> <li>• <b>HCC (Hierarchical Condition Categories)</b> – which calculates patient sickness by determining how many diagnoses patients have received and the number of physician visits patients have made. See (Pope et al., 2011)</li> <li>• <b>CDPS (Chronic Illness and Disability Payment System)</b>– classify patients into conditions and severities and to attach risk scores [see (University of California, 2014)] One of the challenges with this model is that it tends to underperform when there is not a lot of patient claims history.</li> </ul>
	<p><i>Definition:</i></p> <p>Quality Adjusted Life Years involves placing a lower value on time spent with impaired physical and emotional function than time spent in full health. A similar concept is the disability-adjusted life year (DALY). This is a measure of overall disease burden, expressed as the number of years lost due to poor health, illness, disability or early death.</p> <p><i>Reason:</i></p> <p>These approaches are particularly useful when alternative treatments produce out comes of different types, or when increased survival is bought at the expense of reduced quality of life. In cost-effectiveness research, health improvement is generally measured in quality adjusted life years (QALYs) and the results are expressed as cost per QALY gained. If a patient has to endure a painful or time-consuming treatment and they only extend their life by a short time, this allows us to adjust for the quality of life and compare both treatments on a level playing field.</p> <p><i>Method:</i></p> <p>To calculate QALYs, the patient’s utility score in various health states and the time spent in each health state needs to be measured. Utility is the preference or worth assigned to a particular health state on a scale for which 0 represents death and 1 represents perfect health – the greater the impairment, the lower the value of a particular health state. The utility scores are then transferred into QALYs.</p>

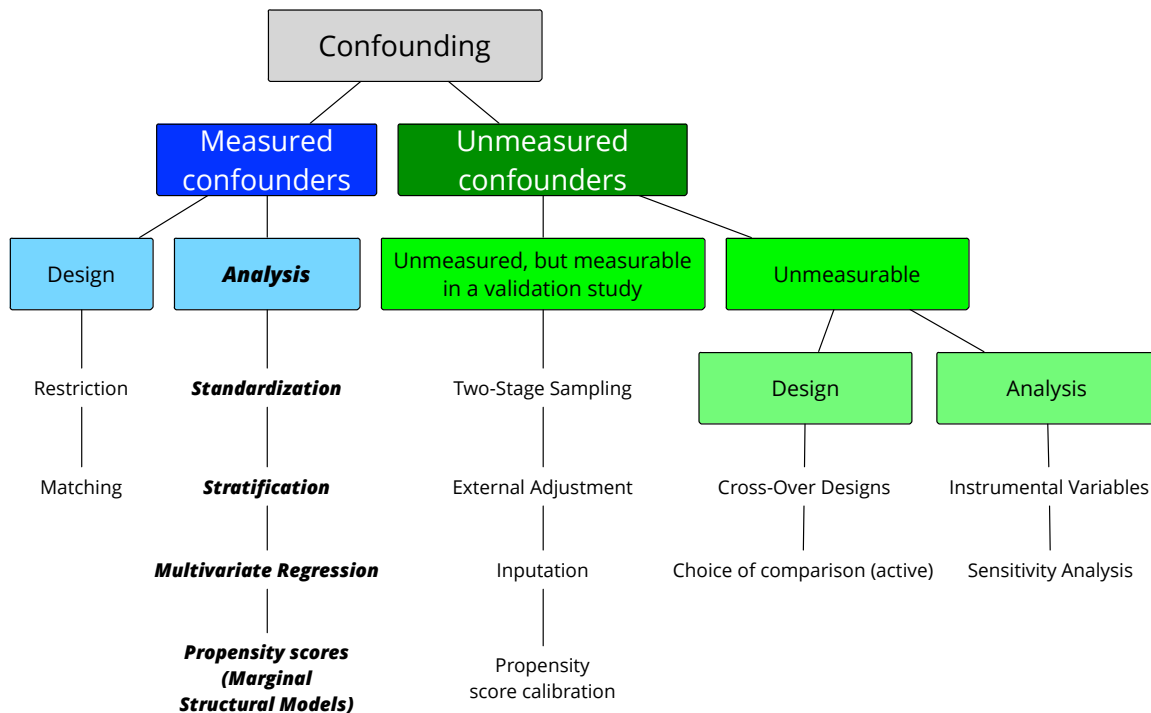
## Statistical Adjustments

As we have highlighted thus far, there are a number of techniques that are useful in controlling for confounding variables in our data and adjusting for them in our analysis. If we are being asked to rate two hospitals on their 30-day readmission rates, for example, we need to make sure that we are establishing the relationship between the care being provided (Hospital A versus Hospital B) and their measures while controlling for potentially confounding characteristics of the environment, the patients and the providers. To accomplish this, we can put together a list of factors that are independent predictors of treatment outcome and may be associated with the provider. A few examples of the types of data for each category includes:

Patient	Physician	Hospital
Severity of primary diagnosis	Level of training	Nurse/bed ratio
Age	Years in practice	Equipment
Comorbidities	Degrees/ certifications	HIMSS EHR Adoption
		Maturity Level

This list can then be used to guide our analysis and in statistical methods to control for those external factors or confounders.

Sebastian Schneeweiss (Schneeweiss, 2006) has provided a nice way to visualize the techniques that we can use in controlling for these confounding factors. Soko Setoguchi, MD, DrPH has adapted this model (personal communication) and we have included her adaptation here.



As we have highlighted above (***bold/ italic***), there are a variety of methods that we can use to control for confounding variables when we have actually measured them in our study/ observation. Throughout this paper, we assume that we have the detail in our data. If we want to age-adjust our mortality rates, we must, of course, have the breakdown of age distributions or use population benchmarks to estimate these. If we don't have them in our data, we can often validate our adjustment methodologies using a separate study (*unmeasured, but measurable in a validation study*).

Let's now turn our attention to practical examples.

## Real-World Examples

### ***Comparing Quality Measures (Associations)***

As discussed in the introduction, we often like to make comparisons between across time, place or population. In this example, we want to compare two hospitals on a quality measure – mortality rate for patients that were hospitalized for congestive heart failure (CHF.) Here have a total of 1035 patients that were hospitalized across the two hospitals and 37 of the patients died. We want to know whether the differences in the mortality rates between the first two hospitals are significant.

#### **CHF MORTALITY RATES FOR EACH HOSPITAL**

Since we are interested in whether the mortality rate between the two hospitals is different (i.e. is there an association between the hospital and the mortality rate), we can perform a frequency analysis using both hospital and a derived variable that indicates whether the patient died or not.

For our quality indicator, we are interested in the following for each hospital:

$$\frac{\# \text{ of deaths}}{18 + \text{ year old CHF Patients}}$$

For each of the hospitals we note the following mortality rates:

Hospital A	Hospital B
$\frac{28}{435} = 0.0644$	$\frac{9}{600} = 0.015$

At first blush, it would be easy to say that Hospital A has a significantly higher mortality rate as compared to Hospital B. However, bear with me and let's see the proper method to compare the two hospitals using a statistical measure of association where we calculate something called the Odds Ratio and the Risk Ratio. (For a great paper on how to do this in SAS®, please see (Waller & Johnson, 2013)). Here we look to see if the row percentages differ by the column category. We can, of course, do this by hand, or with SAS.

	Died	Did Not Die	Total
Hospital A	28	407	435

Hospital B	9	591	600
	37	998	1035

The general process for conducting the analysis is outlined here<sup>1</sup>:

- We are going to use the Chi-squared test to evaluate whether two categorical variables (Hospital and Mortality) are independent
- Our Null hypothesis,  $H_0$ : Mortality is independent of Hospital – that is, the expected and observed values are equal and p-value for a *chi square* = probability of observing test statistic at least as large given no association
- To calculate the chi-squared statistic, we build a contingency table that counts numbers of patients for each combination of levels of Mortality and hospital
- If Mortality is independent of hospital, then the probability distribution of Mortality is the same for each Hospital (and vice versa)
- Next, we calculate the expected number of subjects in each cell if Mortality is independent of Hospital: 
$$\text{Hospital: } \frac{\text{expected value} = (\text{row count} \times \text{column count})}{\text{total count}}$$
- The chi-square statistic is the sum of  $\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ , over all cells in the table
- The null distribution of the chi-square statistic is the chi-square distribution with  $(2 - 1) \times (2 - 1)$  df, (the number of rows and the number of columns are both 2 in our contingency table).
- If the observed chi-square statistic is more extreme than some chosen theoretical value of the null distribution, then reject the hypothesis of independence (e.g. for  $\alpha = 0.05$ ,  $\chi_1^2 = 3.96$ , if  $\chi_1^2 = 3.96$  or equivalently p-value  $< 0.05$  then reject  $H_0$ ).

We can calculate these by hand:

- Risk Ratio =  $\text{Risk}_{\text{exposed}} / \text{Risk}_{\text{unexposed}} = \frac{28}{435} / \frac{9}{600} = 0.0644 / .015 = 4.2912$
- Odds Ratio =  $\text{Odds}_{\text{exposed}} / \text{Odds}_{\text{unexposed}} = \frac{28 \times 591}{9 \times 407} = 16548 / 3663 = 4.5176$
- As we can see CHF Mortality is over 4 1/2 times as high in Hospital A compared to Hospital B.

### **CALCULATE RELATIVE RISK**

Another concept that we often see in the healthcare literature is the notion of relative risk. Above, we saw that CHF Mortality is over 4 1/2 times as high in Hospital B compared to Hospital A. We may also want to calculate the relative risk for mortality in hospitalized CHF patients. If we use the lower quality indicator

---

<sup>1</sup> Process adapted from <https://onlinecourses.science.psu.edu/stat504/node/88>

(QI) hospital as the reference and calculate 95% confidence interval for the relative risk comparing higher QI hospital to lower QI hospital. To calculate the relative risk, we can also use SAS to generate the estimate of relative risk.

### **INTERPRETATION**

One might look at CHF mortality rate as this indicates how well a hospital might do relative to treating those patients. Higher mortality rates might indicate that the hospital is not following standards of care or best practice. On the other hand, it could be argued that higher deaths from CHF are because the patients are sicker or there are exogenous variables that are not helping to explain the difference in the mortality rate.

We can perform basic inferential statistical tests such as chi-square to assess **unadjusted** associations. This helps guide us and provides directionality for additional analysis but does not determine causation. Note that chi-square is not a measure of the strength of the association, just the frequencies.

This helps provide evidence that, *caeteris paribus* (all things considered equal), one hospital has a higher mortality rate but not necessarily better quality of care. We have not performed case-mix adjustment or examined whether or not there are additional demographic variables that might help explain the difference.

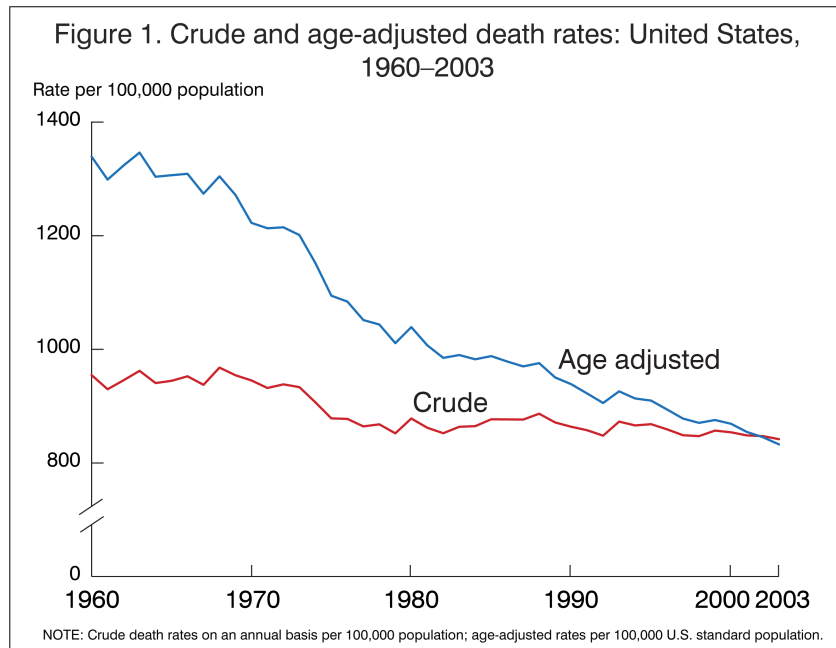
To account for this in our analysis we would have to control for the effect of other variables using a multivariate analysis such as Logistic regression where we add additional variables such as demographic and case-mix variables. This would help control for the effects of these other variables.

In the next example, we will look at how we can begin to model differences in rates by adjusting for subgroup proportions.

### ***Age-Adjustments (proportions)***

As we described earlier, age-adjustment is a statistical process applied to rates of death (mortality), birth, disease states and other health outcomes that allows for comparisons when the underlying population may be different among the comparison groups. We know, for example, that some diseases affect older people at higher rates. Standardizations can be done on other demographic variables beyond age such as gender or race but for our purposes, we will use age – generally, the same principles apply.

The graph from the CDC (Hoyert, Heron, Murphy, & Kung, 2003) below shows just how dramatic the crude rate can differ from an adjusted rate. Here, we are looking at the age-adjusted mortality rates (per 100,000 population) for the United States over a 43-year period. This demonstrates the importance of correcting for real differences in mortality rates. Notice that the crude rate shows a modest decline, whereas if we adjust for age, the decline in death rates has dropped significantly.



Let's explore this in more detail using childhood mortality data adapted from John Hopkins Fundamentals of Epidemiology online course (Lecture 7) for two fictitious populations (Kanchanaraksa & Diener-West, 2008). Here, they present a table that summarizes the mortality rates for two hospitals.

**Community Hospital A**

Age	Population Size	Number of Deaths	Death Rate per 100 <sup>2</sup>
0-4	100	5	5
5-14	90	10	11
15-19	110	15	14
<b>Total</b>	300	30	

**Teaching Hospital B**

Population Size	Number of Deaths	Death Rate per 100
165	10	6
75	10	13
60	10	17
300	30	

Note that the crude rates are identical for both hospitals (30 deaths/300 population = 0.10 = 10 deaths per 100 population), but if you look at the distribution of the age categories between the two hospitals you will notice that the teaching hospital has far more younger patients than the community hospital (165 0-4 year olds as compared with 100 for Hospital A.) Instead of simply looking at the crude rate, we need to look at the age-specific death rates and will notice that the risk of dying in each age group is very different (17 deaths per 100 for 15-19 year olds in the teaching hospital!) In order to say something about the difference between

<sup>2</sup> Death rate per 100 = number of deaths in the specific age group divided by the population for that specific age group.

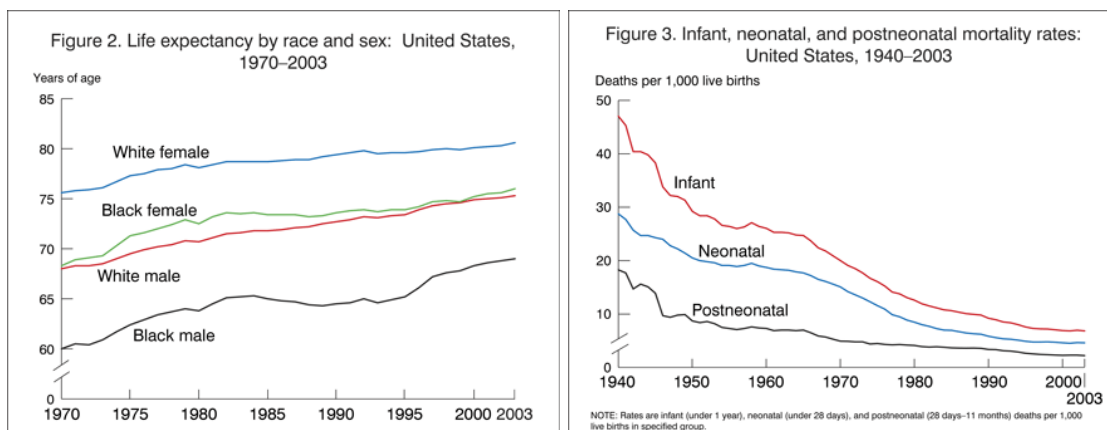


the two populations as a whole we need to adjust our data so that the comparisons are appropriate (level playing field!) That's where adjustments come into play.

Researchers use two methods of standardization:

- **Direct Standardization** – age-specific rates from the two groups (populations) are applied to a standard (e.g., the United States population). For example, we take our age-specific rates that we saw in the observed in our two hospitals and compare to our reference so that we can estimate the number of deaths expected in our reference population. From there, we calculate an adjusted rate based on expected number of deaths in the reference population.
- **Indirect Standardization** – age-specific rates from a standard population are applied to the specific age groupings of the two comparison groups. Apply stratum-specific reference rates to the populations of interest to obtain the number of expected deaths in each of those populations. Compare the observed number of deaths to the expected number of deaths for each population of interest.

While we have used age-adjustments, we can adjust for as many strata as relevant (for example, race/ethnicity or gender.) The graphics below show two variations on this (Hoyert et al., 2003).

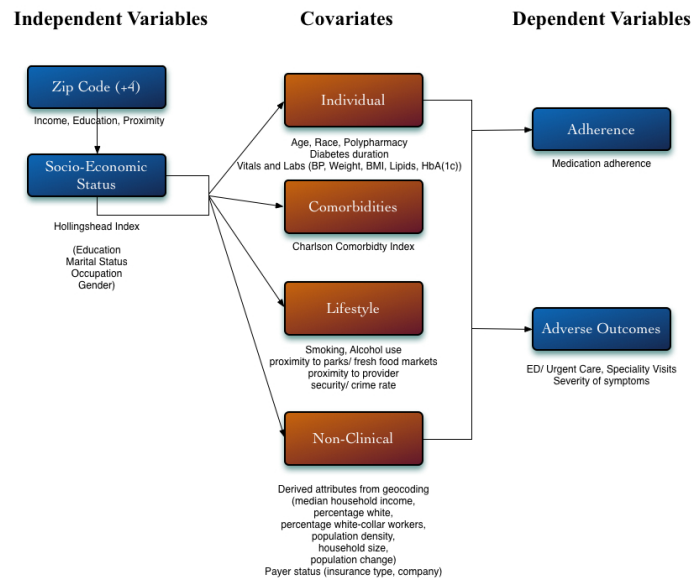


### ***Risk-Adjustments (for patient level analysis)***

Risk-adjustment is somewhat of an extension of age-adjustment techniques in that we are trying to “level the playing field” as we make comparisons. (Note the pattern here!) When we adjust for risk, we are saying that our outcome measure (e.g., mortality) is a function of age (as we saw above) and some other set of factors. In most modern risk adjustment models, we use data from lots of data sources and can include:

- **Demographic information** – age, sex, cultural, racial, ethnic and socioeconomic factors. Recently, we are seeing more interest in geographical associations such as proximity to public parks, transportation and access to health care (clinics, hospitals, etc.)
- **Self-reported health status** – patient reported outcomes, surveys (smoking, drinking status), health risk assessments
- **Administrative data** – health costs, diagnosis data, prescription data (most often from claims data)
- **Clinical Attributes** – current health state (i.e., primary diagnosis, comorbidities, and medication adherence)

The graphic below provides an example of some of the factors that might be used to predict medication adherence in diabetic patients.



As we outlined earlier, there are a myriad of approaches to generating risk adjustments including contingency tables, multivariate statistics, propensity scores and predictive modeling – to name a few.

While somewhat controversial, risk-adjustment has been used for decades in helping insurance companies adequately understand their “per member, per month” costs. As we discussed earlier when we defined the term risk-adjustment, risk-adjustment are often used interchangeably with the term case-mix adjustment. We make the distinction between risk-adjustment and case-mix adjustment, where the former is applied to an individual patient whereas case-mix is used when evaluating organizational performance such as variation in practice between two provider practices.

### ***Case-Mix Adjustments (for organizational level analysis)***

So if risk-adjustments are calculated on for an individual, how does case-mix adjustment work? Thanks to a colleague here at ThotWave, we’ve come up with an example that will hopefully demystify this a bit.

Case-mix adjustment allows comparison of efficiency of care across providers and health plans and is a corrective tool used to level the playing field with regard to reporting patient outcomes and costs. Comparing unadjusted rates for different providers would unfairly penalize those with higher risk patients (older, younger, sicker).

Case-mix adjustment methods take into account the patient population “mix” to allow for practice or plan comparisons. This allows for a comparable analysis where we can focus on the practice pattern variation instead of the differences in the patient population and their unique risk profile.

We often see these tools used rate setting, physician evaluation, physician incentive plans, capitation payment development, quality reporting, incidence rates analysis, patient stratification and cost-effectiveness research.

### **HOW CASE-MIX ADJUSTMENT WORKS**

Demographic and claims data is collected from all enrollees in a health plan or provider organization in a particular market. The data is converted to a score using a risk assessment tool or methodology based on

each individual's age, sex, and/or diagnoses. Individual risk scores are aggregated to an overall score for the practice or plan. This score is used as a multiplier on metrics to allow for cross practice or plan comparison.

#### RISK SCORE CREATION

Risk scores can be used to measure the relative expenditure at the patient level or aggregated to the provider or practice level.

Risk score = predicted cost/ average cost

$$\text{Risk Score} = \frac{\text{Predicted Cost}}{\text{Average Cost}}$$

Where we see the following:

Score > 1: Higher cost than the average population

Score < 1: Lower cost than the average population

We can use these risk scores to help us compare performance of physicians, practices, and health plans and can be applied to financial, utilization, and quality measures.

#### *Example #1 Using the age\sex risk adjustment to create an enrollee level relative risk score*

A simple method to create a risk factor for provider analysis is to use the relative cost of different age\sex categories of the overall population to create relative risk factors. This method requires age and sex data from the enrollment file and the corresponding medical expenses of the enrollee.

$$\text{Risk Score} = \frac{\text{Rate Cell PMPM}}{\text{Overall PMPM}}$$

PMPM Case-Mix Adjustment by Age/ Sex				
	Member Months	Total Claims	PMPM	Risk Score
Female < 18	5,000	\$625,000	\$125	0.42
Female 19 - 39	20,000	\$5,500,000	\$275	0.92
Female 40 - 65	15,000	\$5,625,000	\$375	1.25
Female 65+	5,000	\$2,750,000	\$550	1.83
Male < 18	4,500	\$675,000	\$150	0.50
Male 19 - 39	25,000	\$4,375,000	\$175	0.58

Male 40 - 65	20,000	\$7,000,000	\$350	1.17
Male 65+	6,500	\$3,737,500	\$575	1.92
Total	101,000	\$30,287,500	\$300	1.00

*Example #2 Using commercial risk adjustment tools to create enrollee level risk scores*

Popular commercial tools include Adjusted Clinical Groups (ACGs), Chronic Disease and Disability Payment System (CDPS), Diagnostic Cost Groups (DCGs), Clinical Risk Groups (CRGs), and Episode Risk Groups (ERGs). In addition to demographic data these tools include diagnosis codes when building a risk score. These tools build diagnosis clusters that are used to build a create a risk score. These scores are assigned at the enrollee level similar to the age\sex method.

If the average risk score for the overall population is defined as 1.0, a healthy 20-year-old male might receive a score of 0.5 based on the cluster groups he is assigned, while a 25 year old female woman with asthma might be scored at 1.5, and a 60 year old male with COPD might be scored at 3.0.

**USING RISK SCORES TO COMPARE PROVIDERS**

Once the individual risk scores have been calculated for all enrollees in the plan, these values can be averaged across the plan to calculate the average risk score at the provider and plan level. The average risk score, which is a weighted average of all enrollees' individual risk scores can be use to estimate the provider or plan's predicted expenses.

$$\text{Expected Cost} = \text{Actual Cost} \bullet \text{Risk Score}$$

$$\text{Adjusted Cost} = \frac{\text{Actual Cost}}{\text{Risk Score}}$$

	Actual	Average Risk	Risk	Expected	Adjusted
Level	Cost	Score	Status	Cost	Cost
Plan Total	\$200	1.0	N/A	\$200.00	\$200.00

Practice	Cost	Score	Status	Cost	Cost
Practice A	\$100	0.5	Younger\healthier	\$50.00	\$200.00
Practice B	\$200	2.0	Older\sicker	\$400.00	\$100.00
Practice C	\$150	1.5	Older\sicker	\$225.00	\$100.00

Practice A appears to be more cost efficient than Practice B and C without the use of case-mix adjustment. When taking into account patient risk factors and using the adjusted cost we see that Practice B and C are actually the more efficient providers.

With the use of risk adjustment, comparison among providers and health plans can evaluate practice patterns instead of health status of the underlying population.

### ***Quality-Adjusted Life Years (QALY)***

The final adjustment technique that we wanted to touch upon is quality-adjusted life years. There are a number of variations on this concept, but essentially this adjustment technique takes into account the impact of “quality of life” when making comparisons. Used extensively in health effectiveness, health outcomes and health economics as a way to ensure that treatment options are not only being compared on relative costs, but also the true impact to the patient.

When we compared the costs relative to the effectiveness of strategies, there are typically three different approaches commonly used. These include:

- **Cost-minimization** – the assumption is that when the outcomes are identical, we should pick the least expensive option. That is, we are minimizing our costs.
- **Cost-benefit** – We measure both the costs and outcomes in monetary units. While this is difficult to accomplish in practice, this allows us to make comparisons across any number of subjects – for example, we could compare health to education, roads, defense, etc.
- **Cost-effectiveness** – This is the most common evaluation method where our costs are in monetary units and our outcomes discrete health units such as quality-adjusted life expectancy. When the outcome is quality of life years (QALY) this is also referred to as cost-utility analysis

When making comparisons on cost-effectiveness, we often use the incremental cost-effectiveness ratio (ICER) as a way to compare the net benefit (cost or medical) of alternative strategies.

Here we show an example of an incremental cost-effectiveness ratio (ICER) scenario where we have identified three different policies, their costs and the QALY.

<b>Policy</b>	<b>Net Costs</b>	<b>Net Quality-adjusted Life Expectancy</b>
A	\$1000	65 years
B	\$2500	70 years
C	\$2750	70.5 years

Where:

- Comparing B to A:  $ICER = (\$2500 - \$1000) / (70 - 65) = \$300/\text{year of life}$
- Comparing C to B:  $ICER = (\$2750 - \$2500) / (70.5 - 70) = \$500/\text{year of life}$

Where the cost per life year is decision that you have to make regarding whether you would be willing to pay an additional \$300 per life year. Most of us that are patients or family members or patients don't like to talk

about the additional cost per life year, but as policy makers, this helps quantify the value of various approaches. The “optimal” strategy is that which yields the greatest life expectancy with ICER below some threshold of “cost-effectiveness” within available budget.

## Summary

Here we have covered the larger landscape and focused our examples on risk-adjustment, case-mix adjustment, quality of life adjustment and age-adjustment techniques. While we have provided a number of references throughout this text, we did not provide substantive commentary around a number of issues that you will no doubt encounter including the “small numbers problem” (see (Zaslavsky, 2001)) when comparing quality.

While we have just scratched the surface of adjustment and their use in healthcare, we hope that this has been a useful introduction. With tremendous national focus on healthcare, health insurance, comparative effectiveness for drug companies and increased interest in the application of analytics in primary healthcare, we wanted to provide this introductory material so when you present or evaluate results, you will understand the types of adjustments that may be required to accurately compare results.

## Acknowledgements

The author would like to thank Todd Burkard for his contribution of the examples use in explaining case-mix adjustments. In addition, we are very fortunate to be surrounded by an ecosystem of incredibly generous people who reviewed and provided commentary on this paper prior to publication. These contributors include: Carol Sanders, Jack Shoemaker and Todd Burkard.

## References

- Billings, J., Blunt, I., Steventon, A., Georghiou, T., Lewis, G., & Bardsley, M. (2012). Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open*, 2(4). doi: 10.1136/bmjopen-2012-001667
- Billings, J., Georghiou, T., Blunt, I., & Bardsley, M. (2013). Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. *BMJ Open*, 3(8), e003352. doi: 10.1136/bmjopen-2013-003352
- Christie, Les (2011). Where people are oldest - Maine tops Florida. *CNN Money*. Retrieved 2/15/2014, 2014, from [http://money.cnn.com/2011/05/26/real\\_estate/americas\\_oldest\\_states/](http://money.cnn.com/2011/05/26/real_estate/americas_oldest_states/)
- Ellis, Randall. (2007). Risk adjustment in health care markets: concepts and applications. In I. o. H. E. (IHE) (Ed.), *International Health Care Financing*. Boston University.
- Few, Stephen, & Roswell, Katherine. (2013). *Variation and Its Discontents: Funnel Plots for Fair Comparisons*. Perceptual Edge Newsletter. Newsletter. Perceptual Edge. Retrieved from [http://www.perceptualedge.com/articles/visual\\_business\\_intelligence/variation\\_and\\_its\\_discontents.pdf](http://www.perceptualedge.com/articles/visual_business_intelligence/variation_and_its_discontents.pdf)
- Foundation, The Kaiser Family. (2014). Number of Births per 1,000. Retrieved 2/15/2014, 2014, from <http://kff.org/other/state-indicator/birth-rate-per-1000/>
- Hoyert, Donna L., Ph.D., Heron, Melonie, Ph.D., Murphy, Sherry L., B.S., & Kung, Hsiang-Ching, Ph.D. (2003). Deaths: Final Data for 2003. from <http://www.cdc.gov/nchs/data/hestat/finaldeathso3/finaldeathso3.htm>
- Kanchanaraks, Sukon, & Diener-West, Marie. (2008). Fundamentals of Epidemiology I. from <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/FundEpi/coursePage/index/>
- Pope, Gregory C. MS, Kautter, John PhD, Ingber, Melvin J. PhD, Freeman, Sara MS, Sekar, Rishi BA, & Newhart, Cordon MA. (2011). Evaluation of the CMS-HCC Risk Adjustment Model. In C. f. M. M.



- Services, M. P. P. Group & D. o. R. A. a. P. Policy (Eds.). Research Triangle Park, NC: RTI International.
- Schneeweiss, S. (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*, 15(5), 291-303. doi: 10.1002/pds.1200
- Services, Missouri Department of Health and Senior. (2014). Age-Adjusted Rate. *Data, Surveillance Systems & Statistical Reports* Retrieved 2/15/2014, 2014, from [http://health.mo.gov/data/mica/CDP\\_MICA/AARate.html](http://health.mo.gov/data/mica/CDP_MICA/AARate.html)
- Siegel, M. (2012). Managing the risks of healthcare reform: Using risk adjustment and predictive methodologies to drive capitation arrangements, population health management and network performance initiatives. *Health Manag Technol*, 33(6), 14-15.
- University of California, San Diego. (2014). Chronic Illness and Disability Payment System (CDPS). Retrieved 02/15/2014, from <http://cdps.ucsd.edu/>
- Waller, Jennifer, & Johnson, Maribeth. (2013, April 28 - May 1, 2013). *Chi-Square and T-Tests Using SAS®: Performance and Interpretation*. Paper presented at the SAS Global Users Group, San Francisco.
- WHO. (2014). Disability weights, discounting and age weighting of DALYs. *Health statistics and health information systems*. Retrieved February 20, 2014, 2014, from [http://www.who.int/healthinfo/global\\_burden\\_disease/daly\\_disability\\_weight/en/](http://www.who.int/healthinfo/global_burden_disease/daly_disability_weight/en/)
- Zaslavsky, A. M. (2001). Statistical issues in reporting quality data: small samples and casemix variation. *Int J Qual Health Care*, 13(6), 481-488.

## Biography

### Greg Nelson, President and CEO

Greg is a global healthcare and business intelligence and analytics executive with over two decades of experience and leadership in the field. Greg is a prolific writer and speaker interested in healthcare analytics and the strategic use of information technology.

He received his BA in Psychology from the University of California at Santa Cruz and advanced his studies toward a PhD in Social Psychology and Quantitative Methods at the University of Georgia. Recently, Greg completed his masters degree from Duke University in Clinical Informatics from the Fuqua School of Business. His academic and professional interests include helping organizations mature their analytic capabilities. Founder, President, and CEO of ThotWave Technologies, a niche consultancy specializing in healthcare analytics, Greg is particularly interested in how lessons from across other industries can be applied to help solve the challenges in healthcare.

Before founding ThotWave, he worked for i3/Ingenix, the Gallup Organization and a boutique research firm in Palo Alto, California.

With certifications in Healthcare IT, Project Management, Six Sigma and Balanced Scorecard, Greg is also a prolific writer and has presented over 200 professional and academic papers in the United States and Europe. He won the Best Paper Award in 2013 at the Pharmaceutical SAS Users Group Conference and sits on the board of the SAS Global Users Group. In 2011, Greg was selected by SAS into their loyalty partner group. "This program acknowledges and supports individuals who are recognized experts in their fields and have a long-term relationship with SAS."

Married to wife Susan and living on a small “farmlet” in rural North Carolina, Greg is an avid woodworker, enjoys photography, rides a Harley-Davidson Motorcycle, and strives to be a lifelong learner.

## ***Contact information:***

Your comments and questions are valued and encouraged. Contact the authors at:

Greg Nelson      [greg@thotwave.com](mailto:greg@thotwave.com)

ThotWave Technologies, LLC

1289 Fordham Boulevard #241

Chapel Hill, NC 27514 (800) 584 2819

<http://www.thotwave.com>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

*thinking data*® is registered trademark of ThotWave Technologies, LLC.

Other brand and product names are trademarks of their respective companies.