Paper 168-2010

Medical Guidelines: Consensus Panel Versus Healthcare Analytics

Patricia B. Cerrito, University of Louisville, Louisville, KY

Abstract

Adherence to medical guidelines will become much more important given the recently enacted healthcare bill. However, there is often little evidence from studies using real data to define just what the guidelines should be. As a result, physician consensus panels are often used to define guidelines with the physicians relying primarily upon personal, anecdotal experience or upon general summary statistics. When this occurs, the average population value is used as an upper or lower limit, thereby defining half of the population in need of treatment, which then changes the average so that the guidelines are lowered or raised to reflect this new average. We can use available treatment records to investigate the relationship between the consensus parameters and other diseases using healthcare analytics. In this study, we examine hypertension, cholesterol, and BMI (body mass index). We use techniques in SAS/Stat, SAS Enterprise Miner, and SAS Text Miner. Results show that the relationship between the consensus of "high" levels versus co-morbidities is at best tenuous and at worse detrimental to overall health.

Introduction

Clinical trials are very costly. For this reason, they usually recruit high risk patients. If the trial indicates benefit for treatment, then the treatment is extrapolated beyond the high risk group to the general population. For example, studies of cholesterol medication can use a surrogate outcome to show that cholesterol levels drop. This is considerably different from looking at the relationship between cholesterol level and heart disease, or cholesterol level and survival. The clinical trial cannot give the optimal cholesterol level. The American Heart Association gives the following guidelines:

> Total cholesterol < 200 mg/dL is desirable
> Total cholesterol >200 but <240 mg/dL is borderline high risk
> Total cholesterol > 240 mg/dL is high risk

However, just how were these numbers arrived at? Are they based upon actual studies that show that those with cholesterol above 240 have more problems with heart disease, or have a reduced longevity? Are the numbers based upon averages and extrapolation down from patients with cholesterol levels that are even higher?

There are similar types of extrapolation. The definition of diabetes, and pre-diabetes has been at lower and lower values of fasting glucose levels. The definition of prolonged labor during childbirth has been decreasing in terms of time as well. Moreover, a relationship between salt intake in hypertensive patients has been translated to causation when statistical analysis cannot show causation but only correlation. (Sarafidis and Bakris 2008)

Because of this correlation, a similar attempt to define guidelines has occurred with respect to salt. The forced reduction in salt was proposed in a 2006 paper and implementation has started. (O'Shaughnessy 2006)The Food and Drug Administration is ready to regulate salt intake in foods, and to limit the amount of salt in foods served in restaurants and available in the grocery store. (Layton 2010) Some companies are already reducing salt rather than to give customers a choice. (Anonymous-CBS 2010) We need to examine the actual evidence concerning health and salt. In particular, we need to determine if the guidelines on salt were based on data from unhealthy people and extrapolated to healthy people.

As early as 1989, guidelines suggested reducing salt intake for patients with hypertension. (Hargreaves, Morgan et al. 1989) However, there were disagreements concerning the intake limits as there was little evidence and data to suggest that limiting salt was beneficial. (Hegsted 1991) Established guidelines excluded those without hypertension. (Fodor, Whitmore et al. 1999) More recent guidelines change the recommendations to a much stricter reduction in salt for everyone. (Touyz, Campbell et al. 2004) Long term results, however, do not show a real benefit to a reduction in salt intake for those without heart disease and hypertension. (Miura and Nakagawa 2005) Moreover, these guidelines are based upon reviews of studies rather than data analysis. (Dickinson, Mason et al. 2006) In spite of the lack of causation evidence, the amount of salt continues to be reduced as new guidelines are introduced. (Mohan and Campbell 2009; Titze and Ritz 2009)

In fact, salt is essential after exercise to restore electrolyte balance. (Maughan and Shirreffs 1997) Moreover, it was suggested that patients with co-morbid conditions should have blood pressure values that are lower than those generally recommended, along with even more reductions in salt. (Fodor, Whitmore et al. 1999) However, this, too, is an extrapolation to a subgroup of the population. This can mean an end to both bacon and country ham because of

the salt needed to cure both. A reduction in salt can also have the unintended consequence of increasing infection because the salt makes the meat resistant to food-borne illnesses. (Taubes 2006)

Extrapolation to Define Disease

We look at how the cholesterol levels were set in guidelines in order to define disease. Usually, a normal distribution is assumed and then we take a cut point that is the mean or one-standard deviation above the population mean. (Tonkin 2007) The use of the mean creates a treatable disease in roughly half the population (assuming that the assumption of normality is valid); use of the one-standard deviation will create a treatable disease in about 25% of the population. Similar use of these statistical values are used to define a lifetime risk, sometimes based upon one measurement. (Pencina, d'Agostino et al. 2007)

Other suggestions to define disease use low high-density lipoprotein levels rather than total cholesterol levels. (Rosenson 2005) A panel recommendation states that

> "Serum HDL-C levels of >40 mg/dL must be a therapeutic target in primary and secondary prevention. This goal is particularly important in patients with low serum HDL-C levels and ischemic heart disease (IHD) or its equivalents, even if the therapeutic target for serum LDL-C levels (<100 mg/dL) has been achieved."

These recommendations are based on a reduction in defined risk and not upon a difference in survival between those who achieve such levels and those who do not. (Ascaso, Fernandez-Cruz et al. 2004) This second definition and recommendation adds more individuals into the "needs treatment" category by looking at those with a total cholesterol of less than 200 but who have HDL levels that are not within recommended values. These guidelines are evaluated here largely because there is little evidence that reductions in cholesterol benefit the general population. (Moyer 2010) The results that show a differential survival are based largely upon a study of very unhealthy subjects and not the general population. (Ridker, Danielson et al. 2008) From these studies, the results were generalized to individuals with more moderate cholesterol levels with no evidence that such a reduction is beneficial.

We have a similar means of definition for overweight and obesity such that the acceptable range is for the BMI is 20-25. Obesity is taken to start at a BMI of 30 and gross obesity at 40. A BMI of 18-20 is defined as mild starvation and severe starvation begins when BMI falls below 16. The BMI is defined as an individual's weight in kilograms divided by the square of the individual's height squared where height is identified in meters. The problem with using the BMI is that it does not identify the actual body fat in the individual. Since muscle is heavier than fat tissue, someone with low body fat can have a high BMI and can be identified as overweight. The author of the book, The Obesity Myth, states: (Campos 2004; McArdle 2009)

> "Obesity is defined completely arbitrarily as a body mass index of 30 or higher (175 pounds for an average height woman). Now body mass follows more or less a normal distribution, which means if the the mean body weight is in the mid to high 20s, which it has been for many decades now, then tens of millions of people will have BMIs just below and just above the magic 30 line. So if the average weight of the population goes up by ten pounds, tens of millions of people who were just under the line will now be just over it."

However, older individuals with a BMI of 25-30 survive longer compared to individuals in the normal 20-25 category. (Zamboni, Mazzali et al. 2005; Boyles 2010) It is clear that the definition of overweight and obesity for the elderly are not related to a higher risk of mortality or morbidity. (Oreopoulos, Kalantar-Zadeh et al. 2009) Still other studies have used different methods to define obesity to show that the different definitions have different outcomes. (Garcia-Marcos, Valverde-Molina et al. 2008) Some studies are performed with the aim of showing that obesity is detrimental to health outcomes, but actually find a contrary outcome. (Lea, Crenshaw et al. 2009; Deglise, Bouchardy et al. 2010) Still other studies are inconclusive. (Davies, Smaldone et al. 2009; Siegel, Ulrich et al. 2010) While not yet a consensus, there is some suggestion that guidelines should be race-adjusted. (Anonymous-e! 2009) Yet, when the evidence shows that overweight is more beneficial compared to normal weight, there are physicians who are calling for the definition of overweight to be lowered to 22 BMI rather than 25 BMI.

Therefore, it is again a question of how these categories were defined and based upon what evidence. The line between normal and overweight, and between overweight and obesity is not that obvious. Therefore, the definition has relied upon consensus panels. (Crepaldi, Belfiore et al. 1991) Although there may be observational studies to define obesity, they often rely upon large samples and rare occurrences without modifying the logistic regression model to take both into account. (Komiya, Masubuchi et al. 2008) These problems with logistic regression are defined in detail in Cerrito. (Cerrito and Cerrito 2010) It is also a case of the outcome under study when attempting to determine the relationship between obesity and outcome. One such study examined the relationship between obesity and sick leave. However, it is not clear if the sick days were validated. (Neovius, Johansson et al. 2009) Moreover, the use of BMI does not examine the relationship of fitness to overall health and obesity. Those who are physically fit

do not necessarily have the same problems as those who are not physically fit given the same weight and BMI levels. (Weaver, Hayes et al. 2008)

In fact, though, rather than to rethink the concept that obesity is harmful and leads to worse outcomes compared to those of normal weight, the often encountered fact that obesity can lead to better outcomes is considered to be a "paradox". (Hastie, Padmanabhan et al. 2010) The extrapolation here is based upon the cutpoints used to define overweight and obese individuals. The definitions were created by consensus. While it seems reasonable that there should be a relationship between obesity and adverse outcomes, this does not necessarily occur. There have been calls to redefine the concept of overweight and obesity as information continues to be collected that those who are overweight tend to be healthy. (Rosman 2010) Moreover, evidence continues to accumulate that dietary recommendations are incorrect in that carbohydrates are more damaging to the heart than saturated fat. (Wenner 2010)

Data Investigation

Although claims data do not generally have laboratory results such as lipid values or glucose values, they often contain enough information to determine a patient's body mass index. We want to examine the relationship of body mass index in relationship to costs and conditions. We examine data from the Medical Expenditure Panel Survey (MEPS), which defines patient co-morbidities as well as the primary treatment condition. For more information concerning the MEPS, please refer to (http://www.meps.ahrq.gov/mepsweb/). We need to preprocess the data to isolate patients with specific conditions as there are multiple records per patient, each record with a unique patient condition. We do this using SAS Enterprise Guide 4.2. First, we use Proc Transpose to put all conditions related to one patient into one observation:

```
proc Transpose data=meps.h104
               out=work.tran (drop=_name_ _label_)
                   prefix=med_;
     var icd9codx ;
     by dupersid;
run;
data work.concat( keep= dupersid icd9codx ) ;
       length icd9codx $32767 ;
       set work.tran ;
       array chconcat We will med_: ;
       icd9codx = left( trim( med_1 )) ;
       do i = 2 to dim( chconcat ) ;
              icd9codx = left(trim(icd9codx)) || ' ' || left(trim( chconcat[i]
)) ;
       end ;
run ;
proc sql ;
       select max( length( icd9codx )) into :icd9codx_LEN from work.concat ;
quit ;
%put icd9codx_LEN=&icd9codx_LEN ;
data meps.icd9codes ;
       length icd9codx $ &icd9codx_LEN ;
       set work.concat ;
run ;
```

Once these have been transposed, we use Enterprise Guide to merge this datafile with a second datafile that includes patient characteristics such as age and weight. The next step is to isolate several conditions (high cholesterol, diabetes, and hypertension) and create 0-1 variables to indicate the presence or absence of the condition. We use the query builder in EG to recode the variable created in the code above, icd9codx.

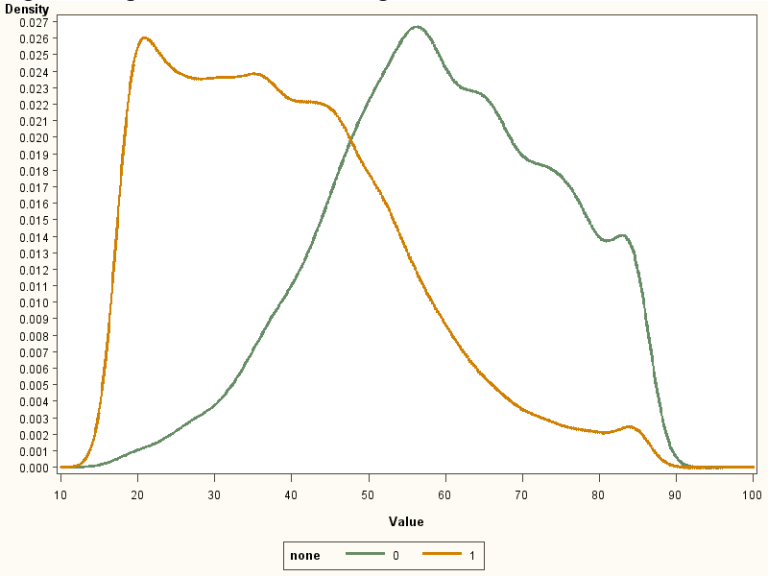
```
PROC SQL;
   CREATE TABLE SASUSER.filteredtable AS
   SELECT t1.variable1,
          t1.variable2,
      FROM sasuser.olddataset AS t1
      WHERE t1.ICD9CODX CONTAINS '401';
QUIT;
```

We first look at the condition of age. We use data visualization with PROC KDE in SAS/Stat so that we can examine the probability density of the different subgroups, and to make comparisons between subgroups. The basic code is

```
proc kde data=work.sort5;
univar age/gridl=10 gridu=800 out=hyper.kdeagenone method=srot bwm=1;
by none;
```

Figure 1 shows the age for those with no condition (hypertension, high lipids, diabetes) compared to those who have at least one condition out of the three.

Figure 1. Age Distribution for Diagnosed Conditions



Code 0 represents individuals with at least one specified condition; code 1 represents those with none of the diagnosed conditions. Note that the likelihood of having no diagnosed condition declines rapidly after the age of 45; in contrast, those with at least one diagnosed condition peaks at the age of 55. We also explore the nature of the different conditions. Figure 2 shows the distributions for those with just one specified condition compared to no condition.

Figure 2. Age Distribution for Just One Condition

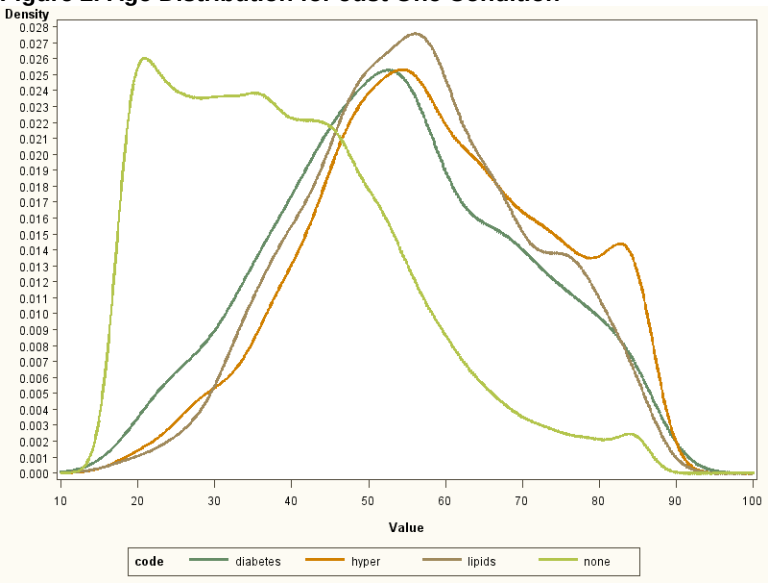


Figure 2 demonstrates that all three conditions have a peak age of 55 years. It demonstrates that the guidelines used to make these definitions tend to identify all those beyond a certain age. Figure 3 gives the distribution for those with two or more conditions.

Figure 3. Age Distribution for Two or More Conditions

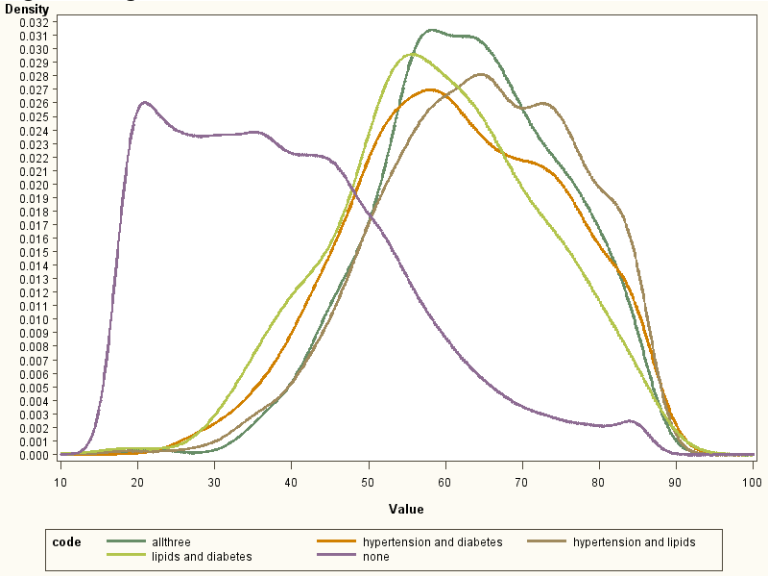


Figure 3 shows that a diagnosis of diabetes with either lipids or hypertension is more likely to occur at age 50 while a diagnosis of hypertension and lipids is more likely at age 60. It is also clear that the current guidelines almost guarantee a diagnosis of at least one of these conditions by age 55, meaning that the guidelines now define at least one treatable illness in the elderly. Similarly, obesity using the current definition is considered a risk factor for all three of these conditions; therefore, we consider the relationship between the body mass index and a diagnosis with at least one of these conditions. Figure 4 shows the outcome of body mass index compared to patients with at least one condition.

Figure 4. BMI by the Diagnosis of a Patient Condition

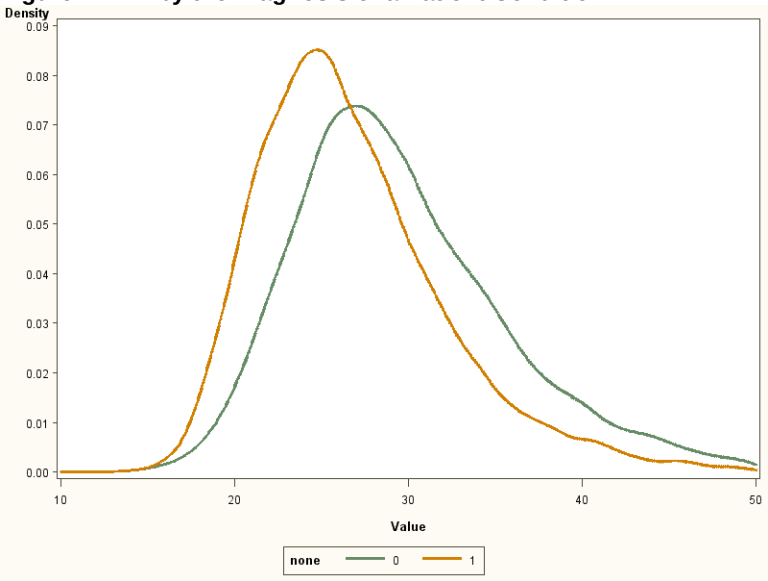


Figure 4 shows that individual patients in the overweight category (25-29) are marginally more likely to have a diagnosed condition (code 0) compared to those in the normal category where code 1 represents those with none of the diagnosed conditions. Figure 5 gives the relationship between BMI and individuals with just one of the specified conditions.

Figure 5. BMI and One Diagnosed Condition

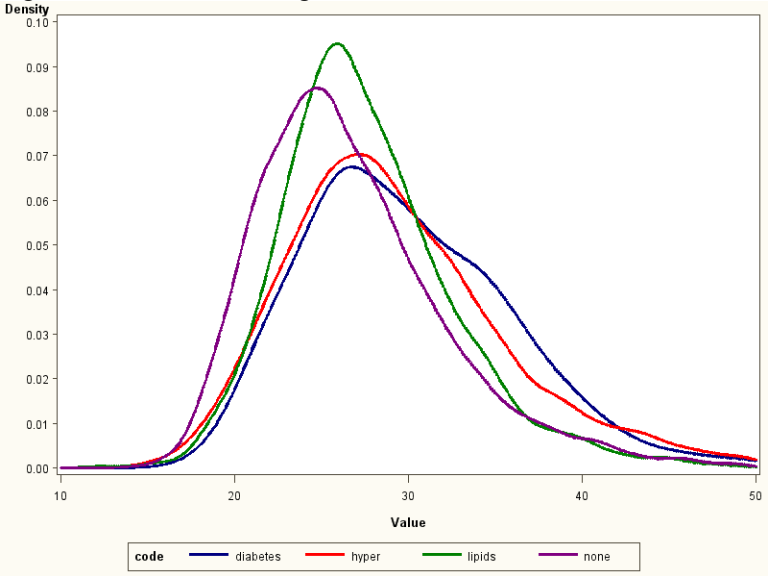
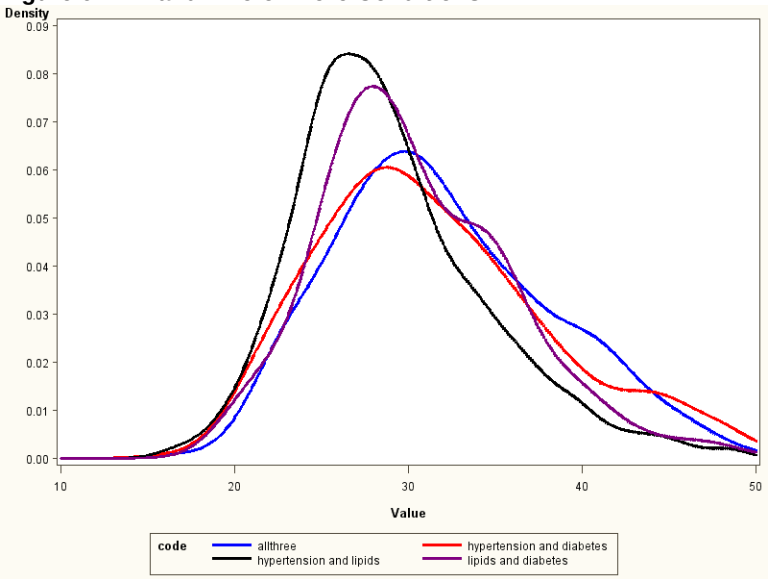


Figure 5 shows that those with no condition and those with high lipids only have a relatively similar distribution in relationship to BMI. Similarly, those with hypertension only and those with diabetes have similar BMI values, which peak in the overweight category. Lipids, on the other hand, peak in the normal weight category. Figure 6 shows the distributions for two or more conditions.

Figure 6. BMI and Two or More Conditions



Relationship of BMI, Hypertension, Lipids, and Diabetes to Co-Morbidities

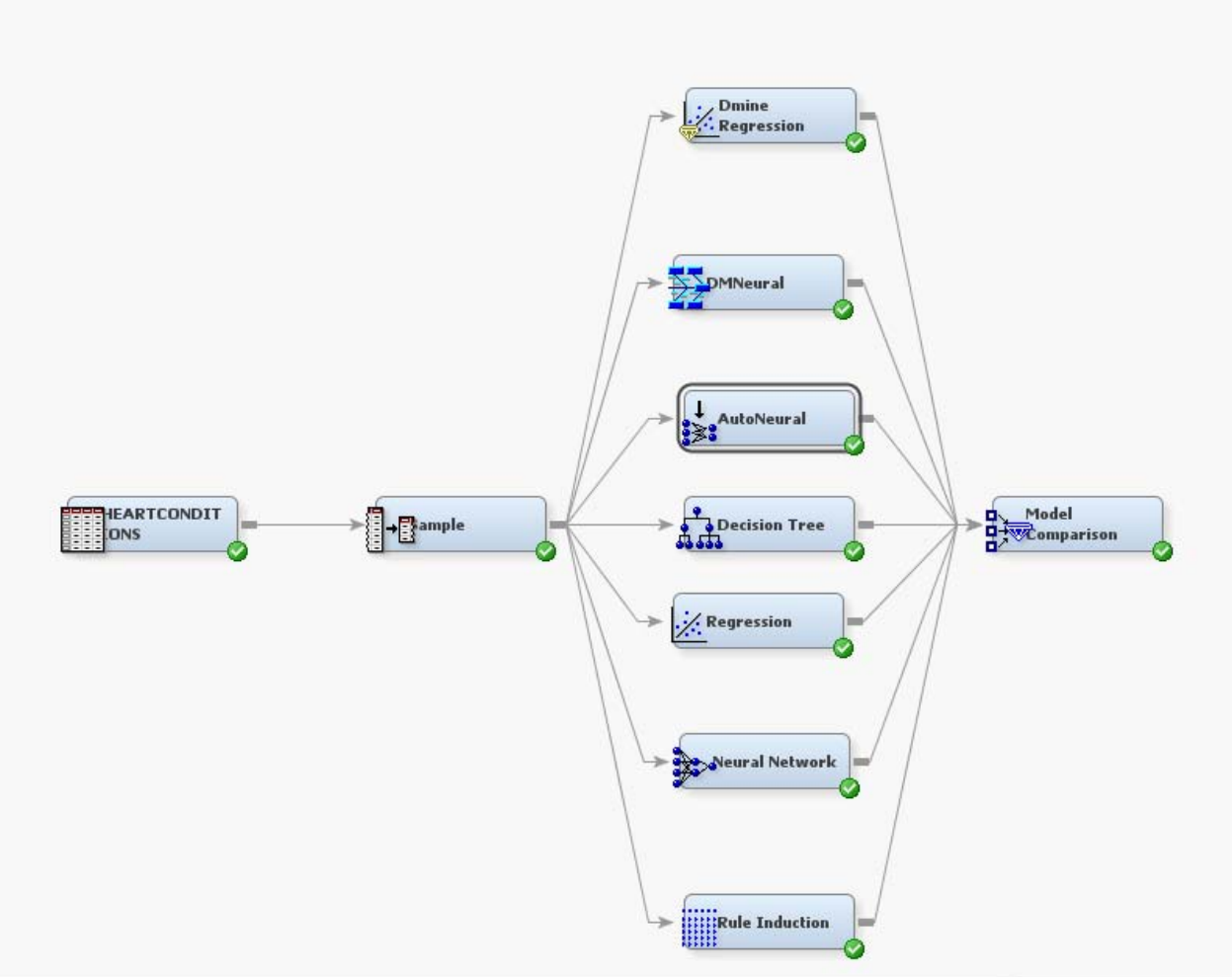
We next examine a diagnosis of heart disease to the specified conditions. We examine the condition of atherosclerosis, myocardial infarction, and peripheral vascular disease. Patients in the MEPS had a diagnosis of each condition. Out of 34,000+ individuals, there are 20 diagnosed with atherosclerosis, 322 diagnosed with a previous heart attack, and 46 diagnosed with peripheral vascular disease. These are all rare occurrences. We will first examine the outcome of atherosclerosis using logistic regression in SAS/Stat. Because of the disparity in the group sizes, the result is highly statistically significant. However, if we look at the classification table for predicting atherosclerosis, all of the observations are classified as non-occurrence (Table 1) Without looking at the false positives and false negatives, it can be overlooked that the model has no predictive ability.

Table 1. Classification Table for Logistic Regression

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.980	34116	0	29	0	99.9	100.0	0.0	0.1	.
1.000	0	29	0	34116	0.1	0.0	100.0	.	99.9

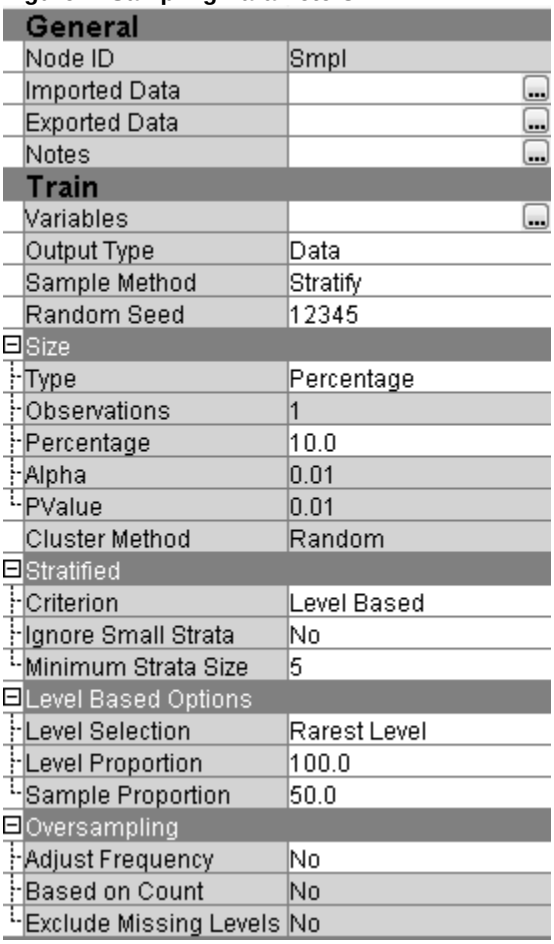
Because they are rare occurrences, logistic regression cannot be used without sub-sampling. Otherwise, the results will be statistically significant and accurate by predicting all observations as non-occurrences. Such a model will have no value in determining the relationship between co-morbidities. Therefore, we use predictive modeling in SAS Enterprise Miner. Figure 7 shows the predictive model.

Figure 7. Predictive Model for Sub-Sampling



We can use multiple models to make a comparison of the best fit. In addition, we use all of the rare occurrences and sample an equal number of non-occurrences. The inputs include BMI and whether exercise was recommended by the physician along with the presence of hypertension, diabetes, and high cholesterol. Figure 8 shows how to set the parameters for the sampling node in Enterprise Miner.

Figure 2. Sampling Parameters



Note that the sampling method is set to stratify with a level based criterion so that the rarest level is kept intact and the larger level is sampled to yield a 50/50 split.

Table 2 shows the results with atheroscleosis as the outcome.

Table 2. Logistic Regression Results for Atherosclerosis

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	P>ChiSq
Intercept	1	4.6560	182.8	0.00	0.9797
BMI	1	-0.0150	0.1079	0.02	0.8891
Exercise?	1	0.2359	1.0457	0.05	0.8215
Diabetes only	1	-6.3465	.	.	.
Hypertension and diabetes	1	7.7125	.	.	.
Hypertension and High Lipids	1	-6.3968	182.8	0	0.9721
Hypertension Only	1	0.2885	0.8361	0.12	0.7301
High lipids only	1	0.1628	0.8136	0.04	0.8414
None	1	1.6018	0.8119	3.89	0.0485

Note that BMI is not significant to the model, indicating that those with higher BMI are not necessarily more likely to have atheroscleosis. Only if the individual has all three conditions of diabetes, high cholesterol, and hypertension is there an increased likelihood of atheroscleosis. The analysis is based upon a reduced sample of size 58 with 4 false negatives, 25 true negatives, 4 false positives and 25 true positives. It would be preferable to reduce the number of false negatives at the risk of increasing the number of false positives. We can do this by assigning weights to each, with a higher cost for the false negative. Adding weights decreases the number of false negatives by one.

Tables 3 and 4 examine the occurrence of myocardial infarction and peripheral vascular disease respectively.

Table 3. Logistic Regression for Myocardial Infarction

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	P>ChiSq
Intercept	1	-3.1097	14.1449	0.05	0.8260
BMI	1	-0.00489	0.0135	0.13	0.7178
Exercise?	1	2.9494	14.1115	0.04	0.8344
Diabetes only	1	0.8637	0.3179	7.38	0.0066
Hypertension and diabetes	1	0.6303	0.2904	4.71	0.0299
Hypertension and High Lipids	1	-0.0743	0.2799	0.07	0.7908
Hypertension Only	1	0.3467	0.2659	1.70	0.1923
High lipids only	1	0.3231	0.2946	1.20	0.2727
High Lipids and diabetes	1	-0.0347	0.4337	0.01	0.9362
None	1	1.7251	0.2370	52.98	<0.0001

Only diabetes is statistically related to myocardial infarction. In particular, neither BMI nor the suggestion of exercise are related.

Table 4. Logistic Regression for Peripheral Vascular disease

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	P>ChiSq
Intercept	1	20.0010	340.4	0.00	0.9531
BMI	1	-0.0716	0.0394	3.30	0.0693
Exercise?	1	10.4700	329.5	0.00	0.9747
Diabetes only	1	-5.4600	.	.	.
Hypertension and diabetes	1	-0.0808	0.8343	0.01	0.9228
Hypertension and High Lipids	1	-5.1250	190.5	0.00	0.9785
Hypertension Only	1	0.5767	0.6766	0.73	0.3940
High lipids only	1	0.7488	0.7255	1.07	0.3020
High Lipids and diabetes	1	-5.6657	259.8	0.00	0.9826
None	1	1.6986	0.6402	7.04	0.0080

Peripheral vascular disease is related to having all three conditions of high cholesterol, hypertension, and diabetes. It is not correlated to any of the conditions individually.

We next use a logistic regression with BMI as the outcome variable, defined in terms of normal weight, overweight and obesity. Table 5 gives the number of individuals in each category.

Table 5. Level of Obesity

BMI	Frequency	Percent
Normal or Below Weight	8212	35.63
Overweight	8074	35.04
Obese	6759	29.33

As the categories have relatively similar proportions, no sub-sampling is necessary. Table 6 shows the results.

Table 6. Logistic Regression for BMI

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	**1: 0**	1	1.2179	0.3168	14.7803	0.0001
Intercept	**2: 1**	1	2.7567	0.3172	75.5452	<.0001
ather	**0**	1	-0.1818	0.1759	1.0680	0.3014
heart_attack	**0**	1	-0.0273	0.0560	0.2380	0.6257
pvd	**0**	1	-0.1127	0.1411	0.6377	0.4245
none	**0**	1	-0.8012	0.0389	424.4772	<.0001
hyperonly	**0**	1	-0.4149	0.0421	96.9551	<.0001
lipidonly	**0**	1	-0.6025	0.0479	158.4720	<.0001
diabetesonly	**0**	1	-0.2933	0.0550	28.4665	<.0001
hyperandlipid	**0**	1	-0.4241	0.0459	85.3621	<.0001
hyperanddiabetes	**0**	1	-0.1264	0.0537	5.5375	0.0186
lipidanddiabetes	**0**	1	-0.1943	0.0729	7.0971	0.0077

BMI is again not related to atherosclerosis, myocardial infarction, or peripheral vascular disease. The diagnoses of hypertension, high cholesterol, and diabetes are related to BMI, indicating that those who are overweight are more likely to be so diagnosed.

Side Effects of Statins

Statins are used to lower cholesterol levels. However, there are known side effects that are detrimental to health. We will examine complaints in the AERS database (located at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm) to determine some information concerning adverse events that could impact decisions on the use of statins. In particular, we want to look at the problem of memory loss and/or amnesia that are related to the use of statins. There are some observational studies that indicate the existence of such a relationship. (Golomb 2005; Cohen and Graveline 2008; Golomb and Evants 2008; Cohen 2009; Evans and Golomb 2009)

The AERS database has 24,718 voluntary reports of adverse events involving statins. There are 273 reports of memory loss, representing 1.1% of the total reports. More of the complaints indicate amnesia rather than memory. We look at the concept links related to the memory loss reports (Figure 1). There are 668 complaints, or 7.8% of the total complaints, that Lipitor caused some type of amnesia or memory loss and 190 complaints related to Zocor with just one complaint for generic statins. Statins are now the number one prescribed medication in the world. There is even discussion about prescribing it to all adults over a specified age. (Spatz, Canavan et al. 2009) Because of that, it is important to investigate the potential problems with the drugs so that patients can examine the risks versus the benefits. We extract the reports related to the statin medications. Some of the reported problems include myalgia, nausea, liver dysfunction, abdominal pain, and rectal hemorrhage.

We want to focus on severe problems. In the Medical Expenditure Panel Survey (MEPS), we have a cohort of patients taking many different statin medications. (Anonymous-MEPS 2007) In addition, all medical conditions that were diagnosed in these patients are also included in the database. We can investigate the relationship, for example, between statin drugs and the diagnosis of liver problems. We can filter down to the patients taking statins, and then look at their diagnosed conditions to see if any of them received a diagnosis of liver problems. For 2006, there were 11,500+ prescriptions identified for the statin medications in the MEPS medications dataset. These prescriptions were for a total of 1974 patients. We can then look to their conditions. We want to see if any of the patients have one of the following ICD9 codes:

573 Other disorders of liver
751.69 Other anomalies of gallbladder, bile ducts, and liver
571.8 Other chronic nonalcoholic liver disease
572 Liver abscess and sequelae of chronic liver disease
571.5 Cirrhosis of liver without mention of alcohol
751.62 Congenital cystic disease of liver
570 Acute and subacute necrosis of liver
794.8 Liver

These patients have a total of 24,718 conditions in the database. Restricted to the list above, there are just 9 occurrences of these conditions for patients taking statins. That can translate into a rate of 4 per 1000 or 400 per 100,000. There are 17,301 complaints involving Lipitor, 7333 complaints involving Zocor, and 84 for generic statins. The complaints for Lipitor account for 70% of the total complaints, indicating that Lipitor is more often prescribed compared to Zocor and other statins. This is an over-estimate of the adverse events because it is not known if these patients had liver disease prior to the start of statins, or developed the condition after statin therapy. Since there are so few occurrences, their records can be examined in detail to determine the amount of time each patient has been prescribed statins. The occurrence of liver problems from statins is documented in the medical literature. (Brown 2008) These were discovered through meta-analyses and review of clinical trials. (Law and Rudnicka 2006) To examine these adverse events, we use SAS Text Miner.

Figure 9. Introduction to Text Miner

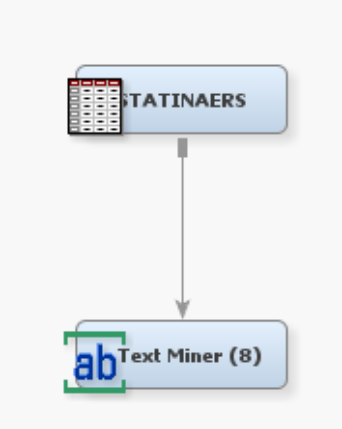


Figure 9 shows the complaints related to memory loss, one of the rare occurrences associated with statins. The settings for Text Miner are shown in Figure 10.

Figure 10. Settings for Text Miner

General	
Node ID	TEXT4
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Interactive	
Force Run	No
Parse	
Parse Variable	pt
Language	ENGLISH
Stop List	
Start List	STATIN.STARTLIST
Stem Terms	Yes
Terms in Single Docum	No
Punctuation	No
Numbers	No
Different Parts of Speec	Yes
Ignore Parts of Speech	
Noun Groups	Yes
Synonyms	Sashelp.engsynms
Find Entities	No
Types of Entities	

Transform	
Compute SVD	No
SVD Resolution	Low
Max SVD Dimensions	100
Scale SVD Dimensions	No
Frequency weighting	Log
Term Weight	Entropy
Roll up Terms	No
No. of Rolled-up Terms	100
Drop Other Terms	No
Cluster	
Automatically Cluster	No
Exact or Maximum Numl	Exact
Number of Clusters	3
Cluster Algorithm	EXPECTATION-MAXIMIZ
Ignore Outliers	No
Hierarchy Levels	.
Descriptive Terms	5
What to Cluster	Roll Up Terms

Text Miner can distinguish parts of speech, or it can be used to analyze text strings of nouns, which is the case for a list of adverse events. Text Miner can cluster the text strings into clusters, which can be used to identify the different types of adverse events and the links between adverse events. Figure 11 shows how we can find concept links, which demonstrate how different terms are linked in the same text string.

Figure 11. Finding Concept Links

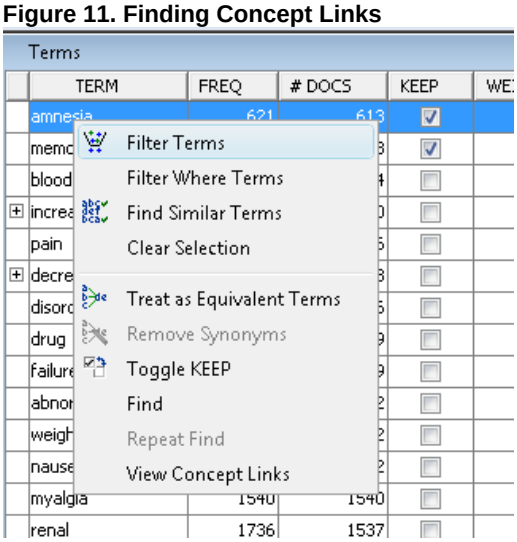
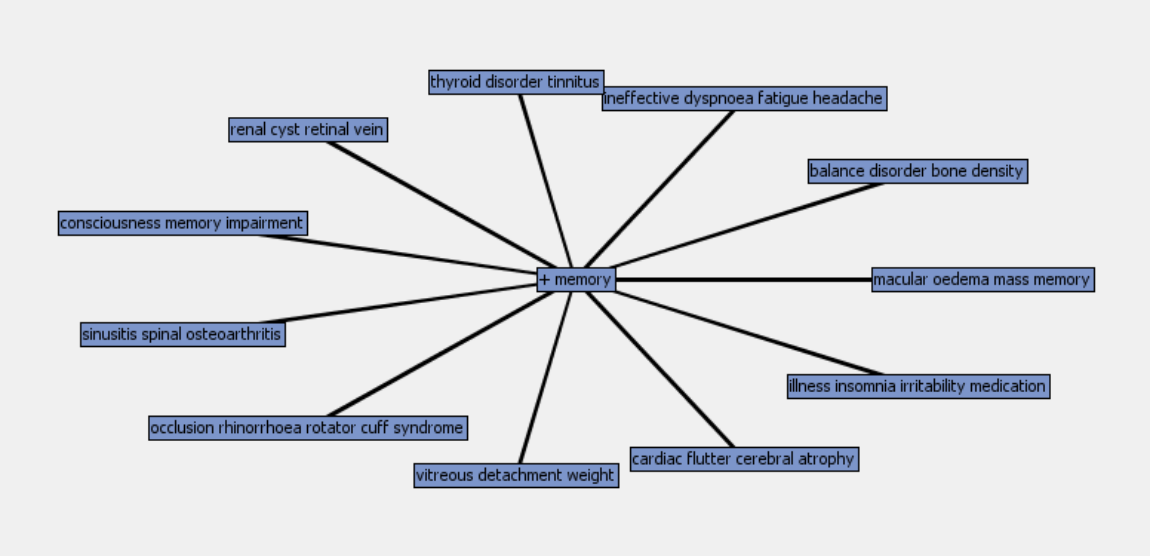


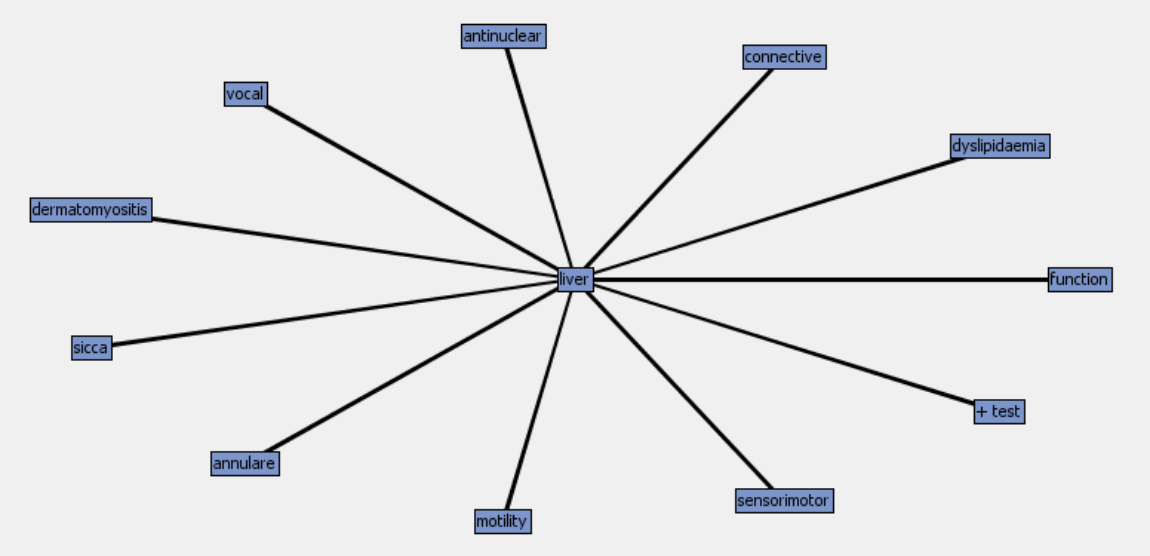
Figure 12 gives the nodes in Enterprise Miner for text analysis.

Figure 12. Adverse Events Related to memory Loss



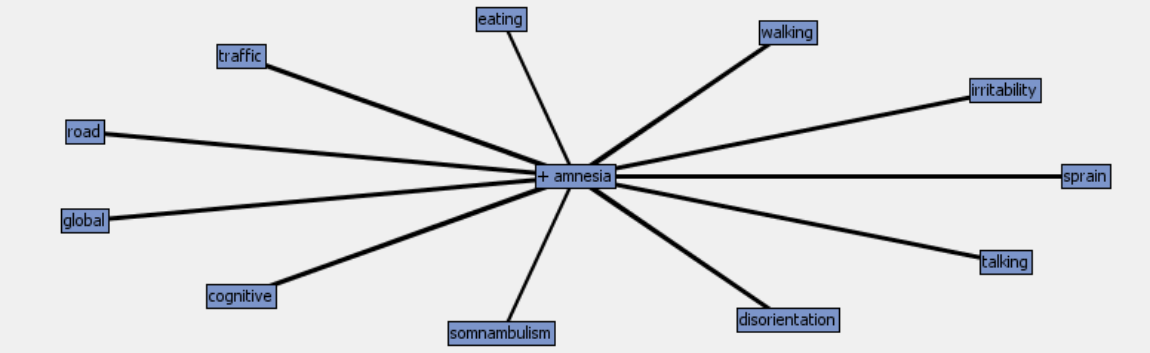
There are some specifics that identify memory loss. Tinnitus, headache, and insomnia can be related to head problems generally. Statins are also known to cause some liver damage in some people. We want to see how many are in the AERS database. There are a total of 589 records of adverse events concerning the liver, representing 2.4% of the reports in the AERS database. Figure 13 shows the concept links for the liver.

Figure 13. Concept Links for the Liver



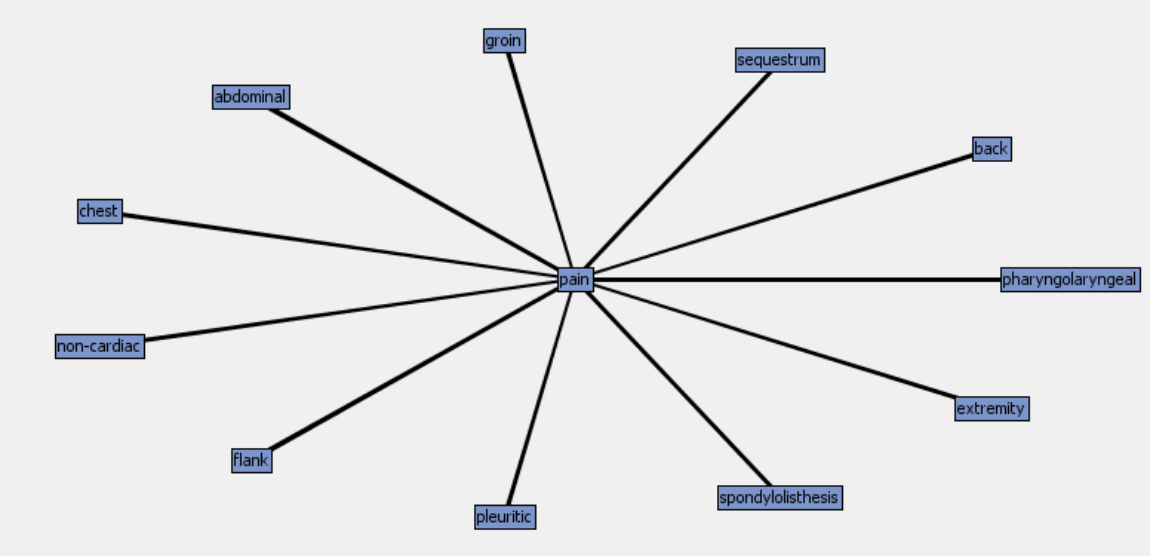
Some of the reports give the specific type of liver damage. One of the more prominent conditions is that of pain with 4099 complaints registered (16% of the database). Figure 14 shows the concepts related to pain.

Figure 14. Concept Links for Amnesia



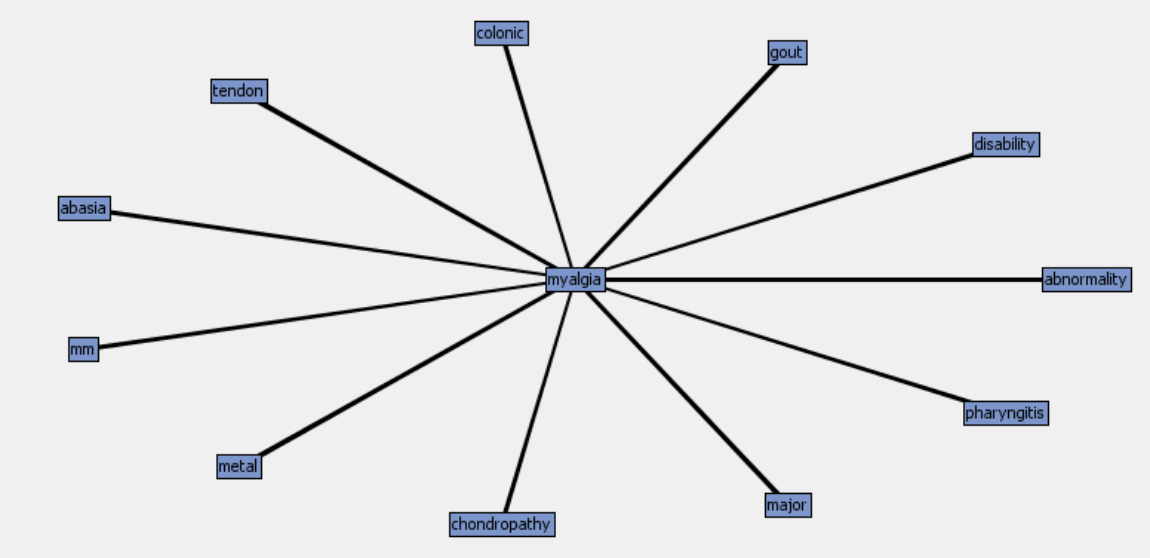
There are similar links for the term, amnesia, compared to those for memory loss. Amnesia is linked to disorientation and somnambulism as well as to the term, cognitive. The combined terms strongly indicate that a number of individuals suffer from memory problems after using statin drugs.

Figure 15. Concept Links Related to Pain



The pain appears to be in scattered locations, including the extremities, the back, the groin, and the abdomen. In many cases, the statin will be replaced by another statin. Related to pain is the adverse effect of myalgia (Figure 16) with 1540 reports, or 6.2% of the total. Myalgia can be extremely difficult to deal with as an adverse event.

Figure 16. Concept Links for Myalgia



These concept links show that there are some concerns with the use of statins, and the benefits should be considered along with the risks of the medications, especially for those who are at best marginally high in terms of cholesterol levels, especially for those with cholesterol values under 200.

Table 7. Text Clusters of Reported Adverse Events Related to Statins

Cluster #	Descriptive Terms	Frequency	Percentage
1	Dispense, ineffective, administer, error, overdose	779	0.0315
2	Blood, cholesterol, creatinine, glucose, weight		
3	Inadequate, mellitus, hypertension, diabetes, type	487	0.0663
4	Amnesia, asthenia, fatigue, depression, peripheral	1576	0.0628
5	Syndrome, dizziness, haemorrhage, fall, fracture	4525	0.1831
6	Chest, muscle, disturbance, myalgia, extremity	3054	0.1236
7	Erythema, injection, haemorrhage, bruise, site	314	0.0128

Cluster #	Descriptive Terms	Frequency	Percentage
8	Complaint, pharmaceutical, drug, product, ineffective	325	0.0131
9	Creatine, myalgia, spasm, muscle, muscular	1526	0.017
10	Weight, decrease, infection, count, blood	3361	0.1360
11	Breast, female, enlargement, mass, cancer	308	0.0125
12	Death, myocardial, coronary, renal, failure	1949	0.0788
13	Dry, skin, oedema, rash, eye	1993	0..806
14	Dizziness, disease, diarrhoea, vomit, upper	2032	0.0822
15	Hepatic, enzyme, test, liver, hepatitis	849	0.0343

Note that amnesia has its own category, and it is related to depression and fatigue. Liver problems also form a group in the list of adverse events in the AERS database. Hemorrhage is prominent in two groups, 5 and 7. In particular, hemorrhage is part of the largest cluster, indicating that the complaint is common.

Discussion

There is a heavy reliance upon the use of consensus panels to define cutpoints and disease. While these panels rely upon observational studies and clinical studies, it is rarely very clear cut just how these cutpoints can or should be defined. In some cases, the outcomes can be inversely related to these cutpoints, as it is in the case of overweight elderly patients. Such studies are generally not conducive to randomized, clinical trials, considered to be the "gold standard" in research medicine. The use of health outcomes data allows us to study the relationship between patient parameters when clinical trials cannot be performed.

References

Anonymous-CBS (2010) Heinz Ketchup's New Recipe has Less Salt. CBS News **May 14, 2010**,

Anonymous-e! (2009) Widely used body fat measurements overestimate fatness in blacks. e! Science News **June 11, 2009**,

Anonymous-MEPS (2007) Medical Expenditure Panel Survey.

Ascaso, J. F., A. Fernandez-Cruz, et al. (2004). "Significance of high density lipoprotein-cholesterol in cardiovascular risk prevention: recommendations of the HDL forum." American Journal of Cardiovascular Drugs **45**(5): 299-314.

Boyles, S. (2010) A few extra pounds may be a plus in old age, researchers say. WebMD, Healthy Aging Health Center

Brown, W. V. (2008). "Safety of statins." Current Opinion in Lipidology **19**(6): 558-562.

Campos, P. (2004). The Obesity Myth: Why America's obsession with Weight is Hazardous to Your Health. New York, Gotham.

Cerrito, P. B. and J. C. Cerrito (2010). Clinical Data Mining for Physician Decision Making and Investigating Health Outcomes: Methods for Prediction and Analysis. Hershey, PA, IGI Publishing.

Cohen, J. S. (2009) Interview: Duane Graveline. MedicationSense.com

Cohen, J. S. and D. Graveline (2008). Atorvastatin-associated memory loss:abaktsus if 662 cases of cognitive damage reported to Medwatch.

Crepaldi, G., F. Belfiore, et al. (1991). "Italian consensus conference-overweight, obesity and health." International Journal of Obesity **15**(11): 781-790.

Davies, B. J., M. C. Smaldone, et al. (2009). "The impact of obesity on overall and cancer specific survival in men with prostate cancer." The Journal of Urology **182**: 112-117.

Deglise, C., C. Bouchardy, et al. (2010). "Impact of obesity on diagnosis and treatment of breast cancer." Breast Cancer Research & Treatment **120**(1): 185-193.

Dickinson, H. O., J. M. Mason, et al. (2006). "Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials." Journal of Hypertension **24**(2): 215-233.

Evans, M. A. and B. A. Golomb (2009). "Statin-associated adverse cognitive effects: survey results from 171 patients." Pharmacotherapy **29**(7): 800-811.

Fodor, J., B. Whitmore, et al. (1999). "lifestyle modifications to prevent and control hypertension. 5. Recommendations on dietary salt. Canadian hypertension society, Canadian coalition for high blood pressure prevention and control, laboratory centre for disease control at health Canada, heart and stroke foundation of Canada." CMAJ **160**(9 Suppl): S29-34.

Garcia-Marcos, L., J. Valverde-Molina, et al. (2008). "Percent body fat, skinfold thickness or body mass index for defining obesity or overweight, as a risk factor for asthma in schoolchildren: which one to use in epidemiological studies?" Maternal & Child Nutrition **4**(4): 304-310.

Golomb, B. A. (2005). "Implications of statin adverse effects in the elderly." Expert Opinion: Drug Safety **4**(3): 389-397.

Golomb, B. A. and M. A. Evants (2008). "Statin adverse effects: A review of the literature and evidence for a mitochondrial mechanism." American Journal of Cardiovascular Drugs **8**(6): 373-418.
Hargreaves, M., T. Morgan, et al. (1989). "Exercise tolerance in the heat on low and normal salt intakes." clinical Science **76**(5): 553-557.
Hastie, C. E., S. Padmanabhan, et al. (2010). "Obesity paradox in a cohort of 4880 consecutive patients undergoing percutaneous coronary intervention." European Heart Journal **31**: 222-226.
Hegsted, D. (1991). "A perspective on reducing salt intake." Hypertension **17**(1 Suppl): 1201-1204.
Komiya, H., Y. Masubuchi, et al. (2008). "The validity of body mass index criteria in obese school-aged children." Tohoku Journal of Experimental Medicine **214**(1): 27-37.
Law, M. and A. Rudnicka (2006). "Statin safety: a systematic review." American Journal of Cardiology **97**(Suppl 8A): 52C-60C.
Layton, L. (2010). FDA plans to limit amount of salt allowed in processed foods for health reasons. Washington Post. Washington, DC, Washington Post. **April 20, 2010**.
Lea, J. P., D. O. Crenshaw, et al. (2009). "Obesity, end-stage renal disease, and survival in an elderly cohort with cardiovascular disease." Obesity **17**(12): 2216-2222.
Maughan, R. and S. Shirreffs (1997). "Recovery from prolonged exercise: restoration of water and electroylyte balance." Journal of Sports Sciences **15**(3): 297-303.
McArdle, M. (2009) America's moral Panic Over Obesity. **July 29, 2009**,
Miura, K. and H. Nakagawa (2005). "Can dietary changes reduce blood pressure in the long term?" Current Opinion in Nephrology & Hyperten14sion **14**(3): 253-257.
Mohan, S. and N. R. Campbell (2009). "Salt and high blood pressure." Clinical Science **117**(1): 1-11.
Moyer, M. W. (2010). "Static over status." Scientific American **302**(4): 26,28.
Neovius, K., K. Johansson, et al. (2009). "Obesity status and sick leave: a systematic review." Obesity Reviews **10**(1): 17-27.
O'Shaughnessy, K. M. (2006). "role of diet in hypertension management." Current Hypertension Reports **8**(4): 292-297.
Oreopoulos, A., K. Kalantar-Zadeh, et al. (2009). "The obesity paradox in the elderly: potential mechanisms and clinical implications." Clinics in Geriatric Medicine **25**(4): 643-659.
Pencina, M. J., R. B. d'Agostino, et al. (2007). "Estimating lifetime risk of developing high serum total cholesterol: adjustment for baseline prevalence and single-occasion measurements." American Journal of Epidemiology **165**(4): 464-472.
Ridker, P. M., E. M. Danielson, et al. (2008). "Rosuvastatin to prevent vascular events in men and women with elevated c-reactive protein." New England Journal of Medicine **359**(21): 2195-2207.
Rosenson, R. S. (2005). "Low high-density lipoprotein cholesterol disorders and cardiovascular risk: cnotribution of associated low-density lipoprotein subclass abnormalities." Current Opinion in cardiology **20**(4): 313-317.
Rosman, K. (2010). A case for those extra 10 pounds. Wall Street Journal. New York, Dow Jones, Inc.
Sarafidis, P. A. and G. L. Bakris (2008). "State of hypertension management in the United States: confluence of risk factors and the prevalence of resistant hypertension." Journal of Clinical Hypertension **10**(2): 130-139.
Siegel, E. M., C. M. Ulrich, et al. (2010). "The effects of obesity and obesity-related conditions on colorectal cancer prognosis." Cancer Control **17**(1): 52-57.
Spatz, E. S., M. E. Canavan, et al. (2009). "From here to Jupiter: identifying new patients for statin therapy using data from the 1999-2004 National Health and Nutrition Examination Survey." Circulation: Cardiovascular Quality and Outcomes **2**: 41-48.
Taubes, G. (2006) The (Political) Science of Salt. National Association of Science Writers
Titze, J. and E. Ritz (2009). "Salt and its effect on blood pressure and target organ damage: new pieces in an old puzzle." Journal of Nephrology **22**(2): 177-189.
Tonkin, A. M. (2007). "How should we define hypercholesterolemia." Current Atherosclerosis Reports **2**(4): 273-274.
Touyz, R., N. Campbell, et al. (2004). "The 2004 Canadian recommendations for the management of hypertension: Part III-Lifestyle modifications to prevent and control hypertension." Canadian Journal of Cardiology **20**(1): 55-59.
Weaver, N. F., L. Hayes, et al. (2008). ""Obesity" and "Clinical Obesity" Men's understandings of obesity and its relation to the risk of diabetes: a qualitative study." BMC Public Health **8**: 311.
Wenner, M. (2010). "Carbs against Cardio: More Evidence that Refined Carbohydrates, not Fats, Threaten the Heart." Scientific American **302 5**(19-21).
Zamboni, M., G. Mazzali, et al. (2005). "Health consequences of obesity in the elderly: a review of four unresolved questions." International Journal of Obesity **29**(9): 1011-1029.

Contact Information

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

Name Patricia Cerrito
Enterprise University of Louisville

Address Department of Mathematics
City, State ZIP 40292
Phone: 502-852=6826
Fax: 502-852-7132
E-mail: pcerrito@gmail.com