The Effect of Missing Data in the Analysis of a Bariatric Surgery Program

Katharine Fischer Berry

A thesis presented to the faculty of Mount Holyoke College

in partial fulfillment of the requirements for the degree of

Bachelor of Arts with Honors.

Department of Mathematics and Statistics

South Hadley, Massachusetts

May 1, 2007

Acknowledgements

First and foremost, I owe a world of thanks to Professor George Cobb, not only for being a patient and dedicated thesis advisor, but for his enthusiasm and encouragement from the very beginning. George allowed me to work with him on a discrimination trial which brought to life what I'd been learning in the classroom and solidified my passion for the field of statistics. I have the utmost appreciation for his wonderful sense of humor and calm disposition, which allowed me to find excitement and joy in writing this thesis.

I would also like to thank Professor Janice Gifford, whose title as my academic adviser is an understatement; life adviser might be more fitting. I thank her for her guidance and support throughout my four years here (and in the years to come!).

Thank you to Sandy Skinner who provided access to such fascinating data, without which my thesis would likely have been dry and dull. I thank her, and Jo-Anne Gibely, for their readiness to answer the many questions I had about the data.

Lastly, I dedicate this thesis to my parents, who are not only my biggest supporters in the classroom, but my biggest fans on the field and my best friends.

Abstract

As obesity rates in the United States continue to rise, numerous programs across the nation have been created in order to educate the population about obesity and cardiovascular disease, and to reduce their prevalence. The purpose of the first part of my research is to investigate the effects of a particular bariatric program at North Shore Medical Center (Salem, MA).

Although 397 patients underwent the procedure before September of 2004, only 191 patients have returned for their two year follow-up. Are patients who returned systematically different in some way than patients who did not return? Conclusions based on available-case analysis are only valid if the missing data are missing at random—that is, observed values are a random subsample of the complete dataset (Rubin and Little 1987). When data are missing at random, the observed distributions match the underlying distributions, and the missing data mechanism is *ignorable* (Rubin and Little 1987). When the data are not missing at random, available-case analysis underestimates and/or overestimates effects, and thus the missing data mechanism must be taken into account in order to reach valid conclusions.

Unfortunately, we cannot know the missing data mechanism for the bariatric surgery dataset and so we cannot assume that the data are missing at random. The second part of my investigation is therefore dedicated to

exploring the missing data mechanism, imputing values based on hierarchical loglinear models, and generating new, improved estimates.

Table of Contents

Table of Figures

Chapter V.

Table of Tables

Appendix A

I. Introduction

1.1 Cardiovascular Disease and Bariatric Surgery

Cardiovascular disease is the number one killer in the United States today, accounting for over 910,000 deaths each year and afflicting another 70 million Americans who live with it.[1] While obesity is a recognized risk factor of cardiovascular disease, obesity rates in the United States continue to rise—around 30% of adults 20 years and older are obese (where obesity is defined at having a Body Mass Index of 30 kg/m^2 or more).[2] Consequently, numerous programs across the nation have been created in order to educate the population about, and reduce the prevalence of, obesity and cardiovascular disease.

The purpose of the first part of this thesis is to investigate the effects of a particular bariatric program at North Shore Medical Center (Salem, MA). The bariatric program involves bariatric surgery coupled with a 13-week long program on cardiovascular risk reduction. The North Shore Medical Center (NSMC) offers two types of bariatric surgery: the Roux-en-Y gastric bypass procedure (both open and laproscopic) and vertical banded gastroplasty (laproscopic only).

---

[1] Gerberding, Julie Louise. "Chronic Disease Prevention." *Center for Disease Control and Prevention*. May 2006. Department of Health and Human Services. 5 Dec. 2006 <http://www.cdc.gov/nccdphp/publications/aag/cvh.htm>.

[2] Center for Disease Control and Prevention. "Overweight and Obesity." Center for Disease Control and Prevention. November 2006. Department of Health and Human Services. 5 Dec. 2006 <http://www.cdc.gov/nccdphp/dnpa/obesity/>.

The Roux-en-Y gastric bypass surgery involves creating a small new pouch in the upper portion of your stomach. The pouch can hold only about one cup of food. The pouch is connected directly to the jejunum (the middle section of the small intestine), bypassing the duodenum (the upper section of the small intestine where most calories are absorbed). [3] The Roux-en-Y gastric bypass procedure thus reduces caloric intake as well as caloric absorption.

The vertical banded gastroplasty (also known as gastric banding) serves to reduce caloric intake, but has no effect on caloric absorption. A gastric band is placed around the top of your stomach, inhibiting the tolerated amount of food intake.[4] Patients with gastric banding feel full quickly, and vomiting can occur if they eat too much or too fast.

After surgery, the Cardiovascular Risk Reduction Program serves to educate patients about nutrition, exercise, and stress management. Patients are expected to maintain healthy eating habits and exercise routines for the rest of their lives. During the program, patients receive personalized assessments of their eating habits and their cardiovascular disease risk factors (e.g. high blood pressure). Exercise is monitored by evaluating blood pressure, heart rate and response to exercise.

---

[3] Kassel, Karen. "Roux-en-Y Gastric Bypass." *Health Library*. November 2006. North Shore Medical Center. 18 Dec. 2006 <http://healthlibrary.epnet.com/GetContent.aspx?token=c969dc7d-0aa7-43de-ba12-bfeedab0944f&chunkiid=96212>.

[4]Kassel, Karen. "Vertical Banded Gastroplasty." *Health Library*. November 2006. North Shore Medical Center, 18 Dec. 2006 <http://healthlibrary.epnet.com/GetContent.aspx?token=c969dc7d-0aa7-43de-ba12-bfeedab0944f&chunkiid=96213>.

The Executive Director at North Shore Cardiovascular Associates, Sandra Skinner, is particularly interested in:

(a) Studying the overall benefits of the bariatric program, including its effect on weight, blood pressure, cholesterol, and quality of life.

(b) Comparing results for females and males. Do results differ between genders?

(c) Studying the effect of compliance. Is compliance with exercise and nutritional recommendations associated with more positive results?

(d) Studying preliminary results from five year follow-up.

1.2 Missing Data

Although 525 patients have undergone the procedure since 1999, many have been lost to follow-up:

Follow-Up	N
1 month	323
4 months	285
8 months	211
12 months	177
24 months	191
36 months	40
5 years	27

These dwindling numbers are not only due to patients' failure to return for a follow-up, but also a result of patients who recently had surgery. For instance, for patients who underwent surgery in January of 2006, we could only have up to 8-month follow-up data since it has not even been one year

since their surgery. Figure 1a shows the percentage of patients who returned at each follow-up, out of the patients who could have returned.

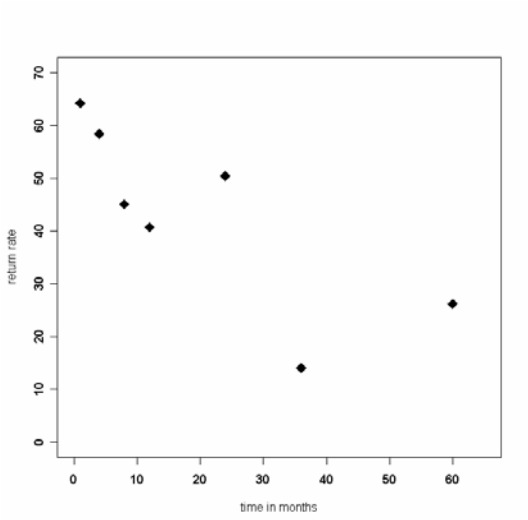


Figure 1a. Follow-up exam versus return rate. The general trend is that the percentage of patients who return decreases as time increases.

In the face of such extensive missing data, it is difficult to assess whether conclusions made are valid. After all, ordinary estimates and inferences take into account only the data which is available. It is plausible that patients who returned for follow-up differ considerably in various aspects from patients who did not return. For example, it could be that patients who did not lose much weight were less likely to return for follow-up; consequently, estimates on weight loss (and probably other risk factors) would be overestimates. Alternatively, it could be that patients who lost the most weight were less likely to return (maybe they were satisfied with their success and did not feel they needed a check-up); if such were the case, available data would lead to underestimates. A third (and most unlikely) situation is that the

missing data are missing at random (MAR), meaning the observed values are a random subsample of the complete data set (Rubin and Little, 1987). In this last situation, the observed distributions match the underlying distributions, and, as Rubin and Little note, the missing data mechanism is *ignorable* (1987). In other words, conclusions made from available data are valid.

Unfortunately, we cannot know the missing data mechanism for the gastric bypass surgery (GBP) dataset, and so we cannot assume that the data are MAR. Therefore, the second part of this paper is dedicated to exploring the various statistical methods that try to account for missing data. The methods studied include hierarchical loglinear models, imputation and weighted averages.

II. Introduction to the Data

2.1 Variables

The GBP dataset includes more than fifty variables, each followed over time. Albumin, B_{12}, calcium, FBS, iron, glucose, cholesterol, HDL (high-density lipoprotein), LDL (low-density lipoprotein), triglyceride, HgbA1C, potassium, total protein, and magnesium levels were measured pre-surgery, as well as 1, 4, 8, 12, 24, 36, and 60 months after surgery. Weight, diastolic blood pressure, systolic blood pressure, and heart rate were also recorded at each follow-up.

Several risk factors for cardiovascular disease were looked at pre-operation, two years post-operation (24 month follow-up), and five years post-operation: asthma, binge eating, coronary artery disease, cellulitis, depression, hyperlipidemia, hypertension, insulin-dependent diabetes mellitus (IDDM), non-insulin-dependent diabetes mellitus (NIDDM), incontinence, joint pain, morbid obesity, stress/anxiety, sedentary lifestyle, sleep apnea, and smoking. These risk factors are self-reported.

Any medications patients were on—including antidepressant, antihypertensive, insulin and lipid lowering drugs—were recorded pre-surgery, two years post-surgery, and five years post-surgery.

The cardiovascular risk reduction program monitored compliance with exercise, compliance with vitamins, whether patients continued food records, and whether they continued to practice meditation/relaxation techniques

(which they were taught during the cardiovascular risk reduction program following surgery).

Finally, subjective measures to gage success were also considered: quality of life pre-surgery and two years post-surgery, emotional improvement, and physical improvement.

See Table 2a for a summation of variables and when they are measured.

Table 2a. Summary of what variables are measured at each follow-up exam.

Variable	Pre-Operative	Post-Operative						
		1 mo	2 mo	8 mo	12 mo	24 mo	36 mo	5 year
Primary Response	√	√	√	√	√	√	√	√
Weight								
Diastolic BP								
Systolic BP								
Heart rate								
Blood Chemistry	√	√	√	√	√	√	√	√
Albumin								
B_{12}								
Calcium								
FBS								
Iron								
Glucose								
Cholesterol								
HDL								
LDL								
Triglyceride								
HgbA1C								
Potassium								
Total Protein								
Magnesium								
Risk Factors	√					√		√
Asthma								
Binge eating								
CAD								
Cellulitis								
Depression								
Hyperlipidemia								
Hypertension								
IDDM								
Incontinence								
Joint Pain								
Morbid Obesity								
NIDDM								
Stress/Anxiety								
Sedentary Lifestyle								
Sleep Apnea								
Smoking								
Compliance	√					√		√
Exercise								
Vitamins								
Food records								
Relaxation techniques								
Medications	√					√		√
Antidepressant								
Antihypertensive								
Insulin								
Lipid Lowering								
Other	√					√		√
Quality of life								
Emotional improvement								
Physical improvement								

2.2 Baseline Characteristics

Patients were between 20 and 65 years old, and weighed between 180 and 560 pounds at the time of surgery. Neither age nor weight follow a normal distribution, although neither has a very skewed distribution either.

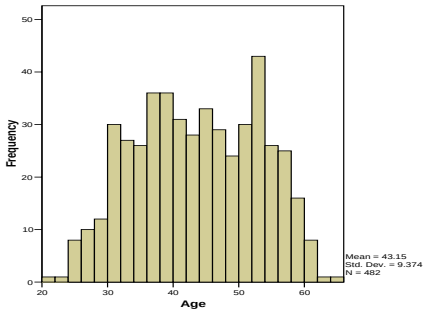


Figure 2a. Baseline age

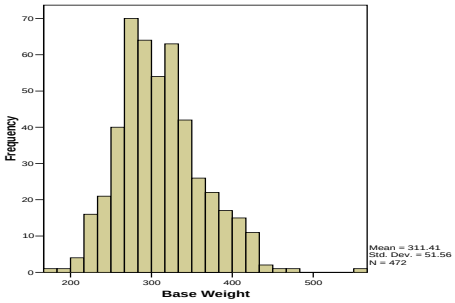


Figure 2b. Baseline weight

The majority of patients were female (85.3%). Although national rates of obesity are currently higher for women (34%) than for men (30%), the percentages in this study are not reflective of those in the larger population.[5] Skinner notes that men, in addition to feeling less societal pressure than women to lose weight, are also more successful in losing weight non-surgically.

Many patients exhibited risk factors for cardiovascular disease above and beyond obesity. For example, 61.3% suffered from hypertension while 49% had been diagnosed with hyperlipidemia. Patients also exhibited high rates of depression (61.7%), stress/anxiety (66.8%), and joint pain (83.6%).

---

[5] Center for Disease Control and Prevention. "Health, United States, 2006." Department of Health and Human Services. 2006. National Center for Health Statistics. 5 Dec. 2006. <http://www.cdc.gov/nchs/data/hus/hus06.pdf>.

Patients rated their quality of life pre-surgery from 0 (being the lowest) to 100 (being the highest). Of the patients with ratings, the mean quality of life rating was 35.6 (see Figure 2c). Almost a quarter of the patients rated the quality of their life below 10.

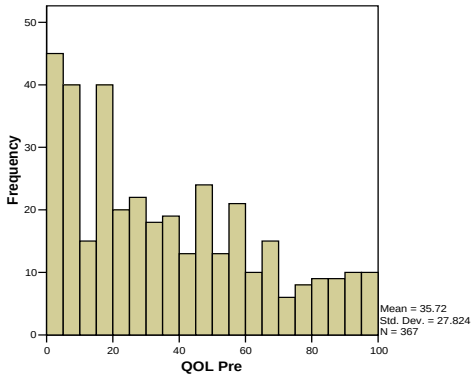


Figure 2c. Quality of life rating pre-surgery

For more baseline characteristics, see Table 2b on the next page.

Table 2b. Baseline Characteristics

Characteristic		Mean	SD	Range	n
Age		43.2	9.4	20-65	482
Weight		311.4	51.6	180-560	472
BMI		51.4	7.2	39-80	416
Total Cholesterol		201.3	39.0	101-417	431
	LDL	124.2	34.4	37-269	372
	HDL	44.1	11.6	24-98	398
SBP		132.3	15.9	92-180	485
DBP		82.3	10.1	50-118	485
Heart Rate		81.3	9.4	54-120	481
		N	**%**		
Diabetes					
	NIDDM	85	16.6		512
	IDDM	28	5.5		512
Hypertension		314	61.3		512
Hyperlipidemia		251	49.0		512
Cellulitis		15	2.9		512
Asthma		135	26.4		512
Joint pain		428	83.6		512
Sedentary lifestyle		457	89.3		512
Sleep apnea		163	31.8		512
Stress/anxiety		342	66.8		512
Depression		316	61.7		512
Smoking		68	13.3		512

III. Analyses of the Data

3.1 Overall Results

For the 191 patients who returned for a follow-up 24 months post-surgery, results are positive. On average, patients lost 35.7% of their original weight ($p < .001$); the mean weight loss is113.6 pounds. Mean blood pressure, heart rate and cholesterol also decreased significantly (see Table 3a).

Table 3a. Comparison of means pre-surgery and 24 months post-surgery

	n	Mean(SD) Pre	Mean(SD) Post	Difference in Means	p-value[a]
Weight	189	316.5(52.0)	202.9(44.5)	113.6	<.001
SBP	181	133.9(15.5)	124.8(17.9)	9.1	<.001
DBP	181	82.2(10.1)	76.5(11.8)	5.7	<.001
Heart Rate	177	81.7(9.3)	69.4(10.1)	12.3	<.001
Cholesterol	113	207.8(39.0)	178.9(35.3)	28.9	<.001
QOL	169	34.7(27.3)	58.9(30.4)	-24.2	<.001

a.Paired-samples t test used to calculate p-values. Distributions of the difference in means were roughly normal (none were considerably skewed), and pre and post SDs were similar. Therefore, a t-test is appropriate. Each difference is highly significant.

The average rating for quality of life rose 24.2 points. Prior to surgery, half of the patients (out of the subgroup that returned two years post-surgery) rated the quality of their life below 30. Two years post surgery, the median quality of life rating doubled: half the patients rated the quality of their life higher than 61. As shown in Figure 3a, the distribution for quality of life pre-surgery is skewed to the right, indicating more patients gave their life low ratings. Conversely, the distribution for quality of life two years post-surgery (Figure 3b), is skewed to the left.

Recovery rates for most cardiovascular risk factors were found to be statistically significant (see Table 3b). In particular, 55 out of the 63 people

who had sleep apnea prior to surgery reported no sleep apnea two years post-surgery. Other factors that saw the most success are sedentary lifestyle, morbid obesity, and joint pain.

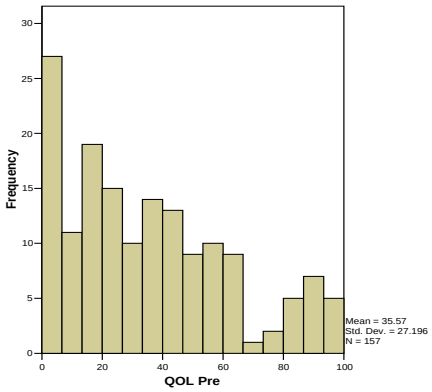


Figure 3a. Quality of life ratings pre-operation. Only a quarter of the respondents rated the quality of their life over 53.

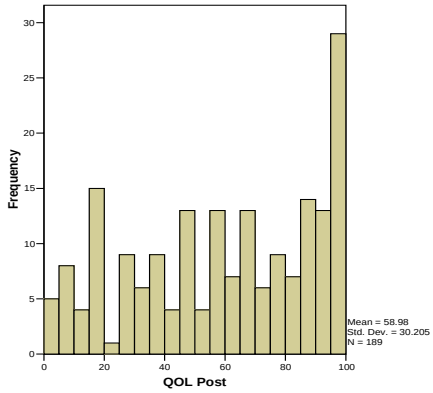


Figure 3b. Quality of life ratings post-operation. A quarter of the respondents rated the quality of their life over 89.

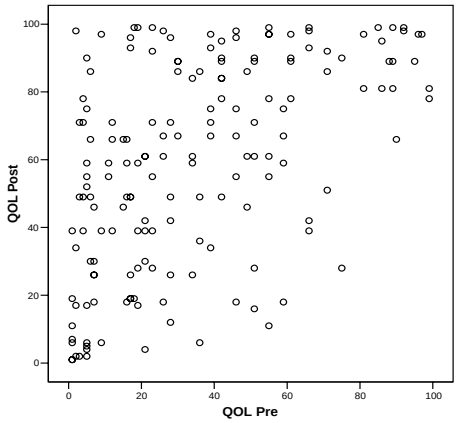


Figure 3c. Scatterplot of quality of life pre-surgery versus quality of life-post surgery. The identity line corresponds to no change in quality of life rating. Most points are above the line, indicating that the quality of life improved after surgery for the majority of the patients.

Table 3b. Recovery rates for several risk factors associated with obesity (n=191). Recovery rate is the percentage of patients who do not report a risk factor two years post-operation, out of the patients who did report the risk factor pre-operation.

Risk Factor	% Yes Pre	Recovery Rate	p-value[a]
NIDDM*	16.2	80.6	.001
IDDM*	9.9	63.2	<.001
Hypertension*	68.1	75.4	<.001
Hyperlipidemia*	56.0	74.8	<.001
Cellulitis	4.2	75.0	.031
Asthma*	29.3	71.4	<.001
Joint pain*	92.7	70.1	<.001
Sedentary Lifestyle*	95.8	85.2	<.001
Morbid Obesity	100.0	82.7	
Sleep Apnea*	33.0	87.3	<.001
Stress/Anxiety*	80.1	24.2	<.001
Depression*	78.0	33.6	<.001
Incontinence*	16.2	67.7	<.001
Smoking	11.0	42.9	.021

a. McNemar's test (binomial distribution) used to calculate p-values.
* Significant at the .01 level

The recovery rates for stress/anxiety and for depression were not as high; around a quarter of the patients "recovered" from stress/anxiety and a third "recovered" from depression. ("Recovered" meaning that patients who reported these risk factors prior to surgery did not report them two years post-surgery.) While the recovery rates for smoking and cellulitis are encouraging (43% and 75%, respectively), they are not significant at the .01 level. Only a small number of patients reported these risk factors at baseline, and thus results are less conclusive than for risk factors that were initially more frequent.

Many patients in the bariatric program were able to reduce or get off their medication by two years after surgery: 76.4% of patients on insulin had either reduced or were no longer taking insulin; 62.5% of patients on lipid lowering drugs were off the drugs completely; 73.5% of patients on

hypoglycemic medication were off the medication completely; 79.4% of patients on hypertensive medication had either reduced or were no longer taking hypertensive medication. Of the 191 patients who returned, 94 patients were on antidepressant medication at intake. Two years post-surgery, the majority (almost 80%) of those patients were still on antidepressants. It appears that the bariatric program is able to help patients reduce or get off medications related to metabolism, but not medication related to mood.

Consistent with the positive findings described above, 96.3% of patients recorded that they saw physical improvement two years post-surgery. The majority also experienced emotional improvement (88.5%).

3.2 Gender Comparisons

Females and males showed similar baseline values for many of the risk factors related to obesity. However, on average, males weighed significantly more than females. Males also saw considerably higher rates of cellulitis and sleep apnea. On the other hand, females had a significantly higher mean HDL level than males and were also more likely to report depression.

Out of the 191 patients who returned for a follow-up two years post-surgery, 159 (83.2%) were female and 32 (16.8%) were male. These percentages are consistent with the percentages of females and males in the beginning of the study. Compliance rates are similar across the two genders—about half of the females and half of the males complied with exercise recommendations, while a little over 80% of each gender complied with taking

their vitamins and Tums.[6] Rates of physical improvement and emotional improvement are also about equal for females and males, around 95% and 89%, respectively.

Males lost more weight on average than females, although the difference was *not* statistically significant (see Table 3d). (Remember that males also weighed more at baseline, which means they had more weight to lose. In fact, if we look at the percentage of weight lost, the relationship is reversed: the mean weight loss percentage for females is 36.3 whereas that for males is 32.9; this difference is also *not* statistically significant). Females did see a significantly greater reduction in heart rate. Reductions in blood pressure and cholesterol were similar for males and females, as were rates of reduction in the number of patients who had hypertension, hyperlipidemia, sleep apnea, joint pain, etc. (refer to Tables 3a and 3b for overall rates).

The mean quality of life rating increased for both genders, although it increased about 10 points higher for females than for males (from 35 to 50.2 and 60.8, respectively). See Figure 3d.

---

[6] Tums is taken for calcium.

Table 3c. Baseline characteristics, by gender. For the number of patients measured for each variable, see Table 3.i in Appendix A.

Characteristic		Female		Male		p-value
		n=434 (85%)		n=76 (15%)		
		Mean(SD)		Mean(SD)		
Age		42.8(9.3)		44.9(9.4)		.086[a]
Weight*		302(46)		359(54)		<.001[a]
BMI		51.1(7.0)		53.1(8.3)		.037[a]
Total Cholesterol		201.8(38.9)		198.9(39.9)		.721[a]
	LDL	124.8(34.3)		120.9(35.0)		.473[a]
	HDL*	45.1(11.7)		39.0(10.0)		<.001[b]
SBP		131.8(15.9)		135.3(15.7)		.079[a]
DBP		82.3(10.1)		81.8(10.3)		.647[a]
Heart Rate		81.6(9.1)		79.6(11.1)		.132[b]
		N	%	N	%	
Diabetes						
	NIDDM	67	15.4	18	23.7	.094[c]
	IDDM	21	4.8	7	9.2	.165[c]
Hypertension		260	59.9	54	71.1	.074[c]
Hyperlipidemia		205	47.2	45	59.2	.062[c]
Cellulitis*		8	1.8	7	9.2	.003[c]
Asthma		122	28.1	13	17.1	.049[c]
Joint pain		366	84.3	61	80.3	.400[c]
Sleep apnea*		108	24.9	54	71.1	<.001[c]
Stress/anxiety		295	68.0	46	60.5	.234[c]
Depression*		279	64.3	36	47.4	.007[c]
Smoking		58	13.4	10	13.2	1.00[c]

a. Independent-samples t test used to calculate p-values. Equal variances assumed.
b. Independent-samples t test used to calculate p-values. Equal variances not assumed.
c. Fisher's Exact Test used to calculate p-values.
* Statistically significant at the .01 level

Table 3d. Comparing mean weight loss percentage and reduction in heart rate and cholesterol for females and males.

	Female	Male	p-value[a]
Weight Loss Percentage	36.3	32.9	.08
HR Reduction(SD)*	13.6(10.6)	6.2(15.7)	.002
Cholesterol Reduction(SD)	28.9(34.6)	25.8(41.9)	.715

a. Independent-samples t test used to calculate p-values.
* Significant at the .01 level

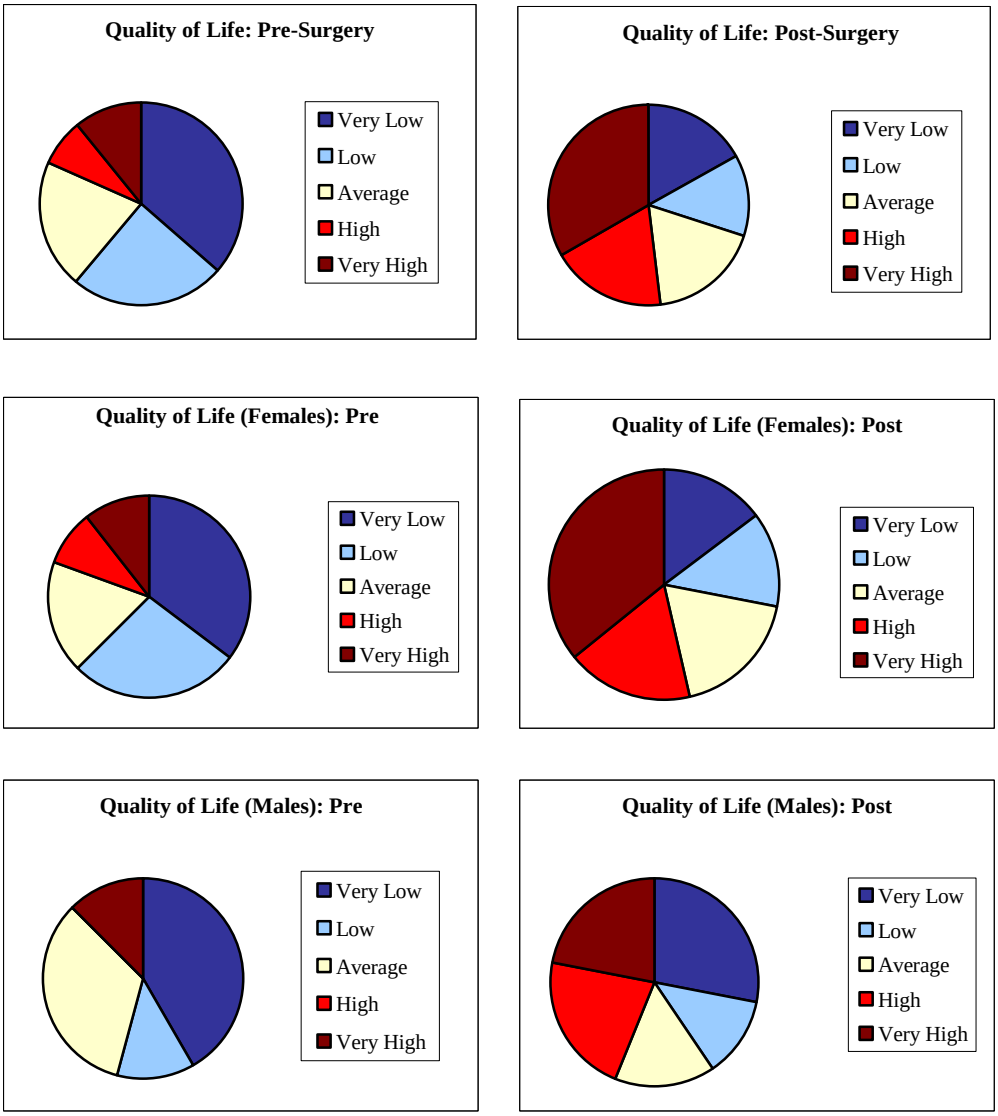


Figure 3d. The top two pie charts represent all patients who returned for follow-up two years post-surgery. The majority of these patients rated the quality of their life below average (either low or very low) before surgery; two years after surgery, the majority rated the quality of their life above average (either high or very high), with the largest percent of patients rating the quality of their life as very high. The distribution for females (the middle two pie charts) follows this same pattern. The distribution for males (the bottom two pie charts) is a little different: prior to surgery, the majority of males rated the quality of their life below average. After surgery, less than half of the males rated the quality of their life as "high" or "very high"; still, the percentage that gave above average ratings increased considerably. [Very Low=0-20, Low =21-40, Average=41-60, High=61-80, Very High=81-100]

3.3 Compliance with Exercise

Out of the 191 patients who returned for a follow-up 24 months post-operation, 50.3% reported compliance with exercise guidelines.

Compliance with exercise is associated with greater weight loss. Two-years post-surgery, those who had complied with exercise recommendations had lost 37.3% of their baseline weight, on average, while those who had not complied with exercise lost 34.1% of their baseline weight on average. While this 3% difference is statistically significant at the .05 level ($p=.028$), it is not as large of an effect as might be expected.

Compliance with exercise did not seem to have an impact on systolic or diastolic blood pressure, heart rate, or cholesterol.

Exercise does not appear to have as much an effect as expected on weight loss and other risk factors relating to obesity. However, it should be noted that responses are subject to not only non-response bias, but also to response bias. Whether or not one complied with exercise was reported by the patient, and not monitored closely by doctors once the 13-week Cardiovascular Risk Reduction Program ended. Patients who did not exercise regularly might have indicated that they had complied with exercise, because they know that is what their doctors expected of them. If such were the case, the average weight loss for those who exercised regularly might be underestimated.

3.4 Compliance with Vitamins

Compliance with vitamins is very important for bariatric surgery patients. Because patients will be drastically reducing their amount of food intake, malnutrition is a serious risk. This is especially true for patients who received the gastric bypass operation; bypassing the duodenum not only reduces caloric absorption but also reduces nutrient absorption. Patients must carefully monitor their nutrient levels and take the required supplemental vitamins as advised by their doctor.

Of the patients returning two years post-operation, 82.2% were still taking the recommended vitamins and Tums. Compliance with taking vitamins and Tums was not associated with greater weight loss; both groups lost about 112 pounds, on average. It was also not found to be associated with greater reduction in diastolic blood pressure, cholesterol or heart rate. However, those who took vitamins and Tums regularly did see a greater decrease in their systolic blood pressure, with an average reduction of 10.8 mmHg compared to a reduction of only 1.3 mmHg for those who did not take vitamins and Tums (95% confidence interval: -17.2 to -1.9).

Since compliance with vitamins should closely relate to levels of B_{12}, potassium, magnesium, etc., I decided to look at these variables as well. Those who continued to take vitamins regularly had higher levels of glucose (93.3 versus 81.8), magnesium (4.8 versus 2.2), iron (86.7 versus 82.0), and B_{12} (352.9 versus 260.8) (although none of these differences are statistically

significant). Average potassium, protein, and calcium levels were similar for the two groups.

Again, findings may be biased due to self-reporting and non-response.

3.5 Other Measures of Compliance

Because the number of patients who reported using meditation and relaxation techniques was small (n=26, 13.6%), as was the number who reported having continued their food records (n=13, 6.8%), it is difficult to assess the effects of either.

3.6 Five Year Preliminary Results

The five year follow-ups are currently being conducted. To date we have 27 patients who have reported for their five year follow-up. Preliminary results on these patients are not as positive as two year results, but they are more encouraging than the expected relapse rate. Skinner says that many bariatric programs find that their patients end up maintaining only 50% of their initial weight loss.

Therefore, it is not surprising that, for the 27 patients that returned, their mean weight, systolic blood pressure, diastolic blood pressure, and heart rate have all increased since their two year follow-up. However, none of the means have reached their baseline values. For example, the mean weight five years post-surgery is 214.4 pounds. This is 26 pounds heavier than the mean

weight for those patients at two years post-surgery (188.4 pounds). It is still not nearly as heavy, though, as their mean baseline weight of 305.3 pounds. In other words, patients were between 31 and 160 pounds lighter than their baseline weight, but were between 6 pounds lighter (one person) to 53 pounds heavier than their two year post-surgery weight.

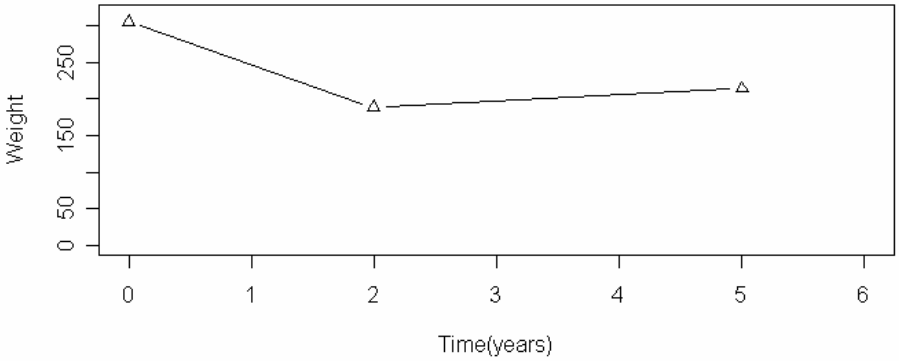


Figure 3e. Patients for whom we have five year post-op results have seen a slight increase in their weight since the two-year exam.

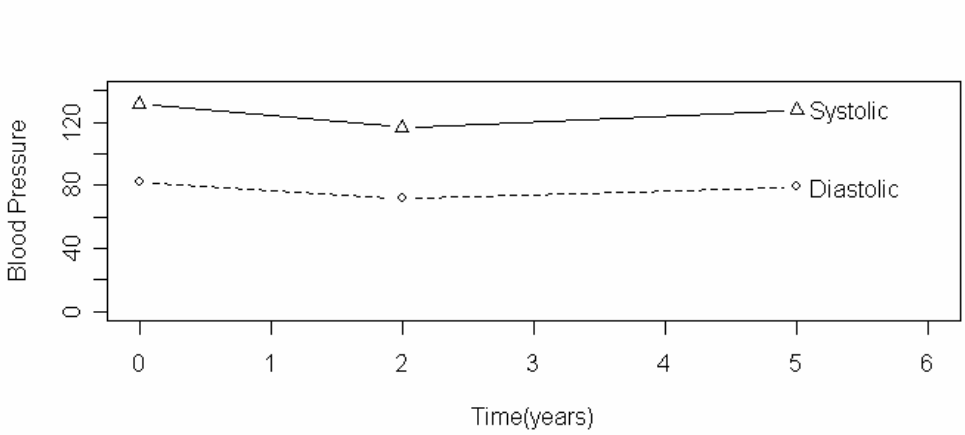


Figure 3f. Patients for whom we have five year post-op results have seen an increase in their blood pressure since their two-year exam, although means have not surpassed baseline values.

It is important to note that the 27 patients who have returned for their five year post-surgery exams may not be a good subsample of all patients. Their mean weights (both baseline and two years post-op) are noticably lower than those for all patients who returned for a two year post-op exam. They also vary somewhat in systolic blood pressure, diastolic blood pressure and cholesterol reductions. See Table 3e.

Table 3e. Patients who returned for a five year follow-up exhibited weight gain, as well as increases in blood pressure and heart rate, since their two year follow-up exam. These patients, however, differed in these variables from the overall patients who returned for a two-year follow-up.

	5 Yr Follow-Up Patients			**2 Yr Follow-Up Patients**		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
Base Weight	27	305.3	42.6	189	316.5	52.0
Weight 24 mos	17	188.4	39.6	189	202.9	44.5
Weight 5 yr	27	214.4	43.2			
Systolic BP	27	131.9	14.2	181	133.9	15.5
Systolic BP 24 mos	17	116.6	16.8	181	124.8	17.9
Systolic BP 5 yr	26	127.5	14.4			
Diastolic BP	27	82.2	11.8	181	82.2	10.1
Diastolic BP 24 mos	17	71.5	7.6	181	76.5	11.8
Diastolic BP 5 yr	26	79.0	9.9			
HR	26	84.5	6.8	177	81.7	9.3
HR 24 mos	17	68.9	7.3	177	69.4	10.1
HR 5 yr	25	71.8	7.5			
Cholesterol	13	216.5	37.3	175	206.8	40.2
Cholesterol 24 mos	6	163.3	26.8	128	177.9	34.0
Cholesterol 5 yr	14	190.9	46.5			

Five years post-surgery compliance rates were lower (as could be expected). Only four of the 27 people (about 15%) were still complying with exercise recommendations, down from around 50% compliance for the two-

year follow-up. The majority was still taking their vitamins and Tums, about 70% (19 of the 27) compared to 80% three years ago. Four patients were continuing to use meditation and relaxation techniques. None of the patients had continued their food records five years post-surgery.

Rates of depression and stress/anxiety still remain high. About half the patients indicated that their lifestyles were sedentary, and over half the patients are experiencing joint pain. Rates of hyperlipidemia and sleep apnea remain low.

Patients were slightly less happy with their physical and emotional improvement than they were three years ago, with 81.5% (vs. 96.3%) recording physical improvement and 70.4% (vs. 88.5%) indicating emotional improvement. To see the specific changes, see Tables 3.ii and 3.iii in Appendix A.

Interestingly enough, the five year follow-up patients do not differ significantly in their pre-surgery and two year post-surgery quality of life ratings from the entire sample. Their mean quality of life rating pre-surgery is 35.7 (which is also the baseline mean). Their mean quality of life rating two years post-surgery is 61.0 (which is close to the 58.8 mean rating that 189 patients gave). Three years later, their mean quality of life rating is down ten points to 51.1 (SD=31.4). See Figure 3g for the distribution.

Rather than continue to lose weight and improve on other cardiovascular risk indicators, many of the patients who have returned for a

five year follow-up have more or less regressed (as can be expected). Even so, none are worse off (on the risk factors studied here) than they were before the surgery. Nevertheless, it is important to remember, as I said before, that, in addition to being a small sample, these 27 patients are unlikely to represent a random sample of all patients, and so preliminary results could be biased. It will be interesting to see how findings change as more patients return for their five year follow-up exam.

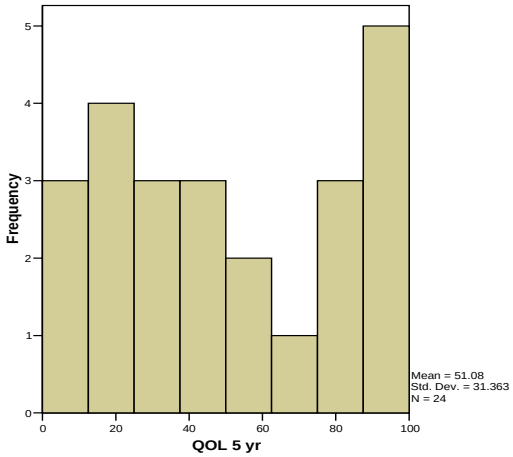


Figure 3g. Histogram of the ratings for quality of life, five years post-operation. Half of the patients rated the quality of their life below 45.5. A quarter of the patients rated the quality of their life above 78.

IV. Missing Data

4.1 Introduction to Missing Categorical Data

While we have information for 512 patients at baseline, we have two year follow-up data on only 191 patients, suggesting a very low return rate. However, only 397 patients had surgery more than two years ago.[7] After accounting for patients who received surgery less than two years ago (and therefore whose 24 month follow-up would not have occurred yet), there are only 188 patients who failed to return for a follow-up, or about a 50% return rate.

Unfortunately, this is still a substantial amount of missing data. Furthermore, as stated in the introduction, we do not know the missing data mechanism. For example, we do not know the way in which patients who returned differ from those who did not return (e.g. did they lose less weight? More weight? About the same?). Although the analyses done in Part I use only the available data, available-case analysis can lead to serious biases if the data are not missing at random (MAR) (Little and Rubin 1987). If, for instance, patients who lost less weight were less likely to return for a follow-up, then an available-case analysis would lead to overestimating two-year weight loss.

---

[7] There are 397 patients who received surgery before September 1, 2004. I chose this date as the cut-off date because I obtained the data in October of 2006. Assuming it takes time for data to be entered into the system, I estimated that the data I received was updated as of September 1, 2006.

In addition, it is unlikely that the missing data are MAR; rather, the missing data mechanism of the GBP dataset is probably nonignorable. In order to understand the true underlying distribution of any specific variable, it is necessary to try to impute the missing values. In this paper, I will be using model-based procedures to estimate missing categorical data. Before I begin estimating missing values, however, it is important to first understand contingency tables and hierarchical loglinear models.

4.2 Contingency Tables

A contingency table is used to summarize the relationship between two or more categorical variables. Since I will be working with a three-way table to model return status, I will start with a concrete, three-dimensional example using pre-operation variables (so that there is no missing data).[8] Let Y_1 = Depression, Y_2 = NIDDM, and Y_3 = Stress:

Depression	**NIDDM**	
	0	**1**
0	66	20
1	15	1

Stress = 0

Depression	**NIDDM**	
	0	**1**
0	20	4
1	212	36

Stress = 1

If $Y_1 = 0$, then the patient does not report depression; $Y_1 = 1$ indicates that the patient does report depression. If $Y_2 = 0$, then the patient does not have

[8] I use the term "return status" in place of Rubin and Little's term "missingness" (1987).

NIDDM; $Y_2 = 1$ indicates that the patient has NIDDM. If $Y_3 = 0$, then the patient does not report stress; $Y_3 = 1$ indicates that the patient does report stress. For example, there are 66 patients who fall into cell (0, 0, 0), meaning there are 66 patients who are not depressed, do not have NIDDM and are not stressed. Likewise, there are 15 patients who fall into cell (1, 0, 0), indicating that there are 15 patients who are depressed but do not have NIDDM and are not stressed. To save space, I will use Y_{ijk} to refer to the cell frequency of row i, column j, table k, where i, j, k = 0, 1 (e.g. Y_{000}=66, Y_{100}=15). The total number of patients is Y = 374. [9]

The *marginal distributions* refer to the total number of patients in Y_{i++}, Y_{+j+}, and Y_{++k}, where the plus sign indicates summing over that index. Thus, Y_{0++} is the total number of patients who are not depressed (86 + 24 = 110), Y_{1++} is the total number of patients who are depressed (16 + 248 = 264), Y_{+0+} is the total number of patients who do not have NIDDM (81 + 232= 313), etc.

Depression	**NIDDM**		
		0	**1**
	0	66	20	*86*
	1	15	1	*16*
		81	*21*	*102*

Stress = 0

Depression	**NIDDM**		
		0	**1**
	0	20	4	*24*
	1	212	36	*248*
		232	*40*	*272*

Stress = 1

[9] I have used data on the 397 patients who had surgery before 9/1/04. We have complete information regarding depression, NIDDM and stress status for 374 of these patients; thus, the total of the contingency table is 374.

4.3 Hierarchical Loglinear Models

To determine the relationship among the three variables—are they mutually independent? Does knowing one variable help predict another's response? etc.—we need to test how well various models fit the data. For clarity, I will divide the process into three steps. First, assume a model. Second, calculate the expected cell frequencies based on the assumed model. Third, run a chi-square goodness-of-fit test to see how well the model fits the data.

4.3.1 The Model

A patient selected at random from the population will fall into cell (i,j,k) with probability π_{ijk}, where $\sum_i \sum_j \sum_k \pi_{ijk} = 1$. Hierarchical loglinear models express the logarithm of cell probabilities ($\ln \pi_{ijk}$) as a sum of effects. The fullest three-dimensional model includes a constant, the main effects of each variable, and all two- and three-way interactions:

$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1 Y_2} + \lambda_{ik}^{Y_1 Y_3} + \lambda_{jk}^{Y_2 Y_3} + \lambda_{ijk}^{Y_1 Y_2 Y_3}$.[10] This model is known as the *saturated* model because it has as many parameters as there are cells in the table, and thus fits the data perfectly. However, we do not want to include all eight terms if a smaller model also fits the data. We want the model that

---

[10]Agresti (1990) defines the λ -terms differently than Bishop, Feinberg and Holland (1975). For a detailed mathematical interpretation of model parameters, see Appendix B.

has the fewest terms and still fits the data. There are nineteen different models that can be derived from setting certain parameters in the saturated model to zero. Bishop, Fienberg and Holland explain the hierarchical property of these loglinear models: "The family of hierarchical models is defined as the family such that if any [λ-term] is set equal to zero, all its higher-order relatives must also be set equal to zero. Conversely, if any [λ-term] is not zero, its lower-order relatives must be present in the loglinear model" (Bishop, Fienberg and Holland 1975). See Figure 4a for a partial ordering of model parameters.

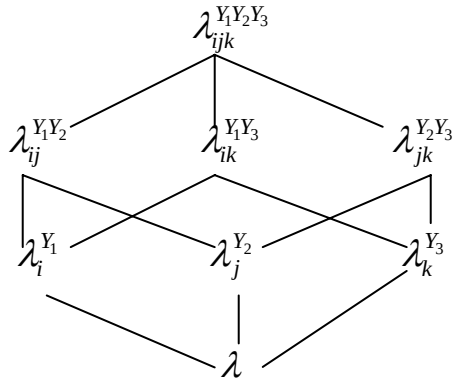


Figure 4a. A partial ordering of model parameters. In a hierarchical model, if a λ-term is in the model, then each λ-term below it in the partial ordering must also be in the model. Furthermore, if a model with certain λ-terms fits the data, then models involving λ-terms above them in the partial ordering also fit the data.

Table 4a lists each comprehensive model for three-dimensional tables, along with a brief explanation of the model.[11] I am going to concentrate on three specific models—{Y_1, Y_2, Y_3}, {Y_3, Y_1Y_2}, and {Y_1Y_2, Y_1Y_3}—for a more detailed understanding. I am not concerned so much with whether these

[11] Comprehensive models are models which include at least the main effects (Bishop, Fienberg, and Holland 1975).

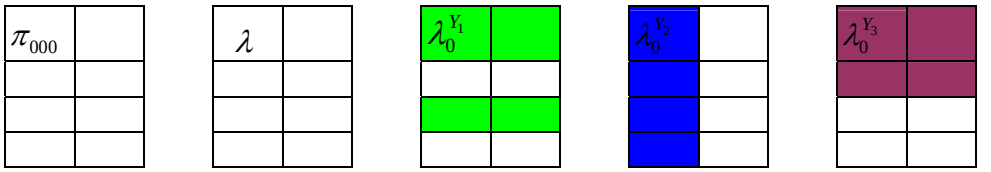
particular models make sense to fit the depression/NIDDM/stress relationship; rather, they are models which will be useful later in my analysis involving missing data.

Table 4a. Hierarchical loglinear models for three-way contingency tables.

Label	Description	Model
$\{Y_1Y_2Y_3\}$	The *saturated* model has as many parameters as there are cells in the table	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} + \lambda_{ijk}^{Y_1Y_2Y_3}$
$\{Y_1Y_2, Y_2Y_3, Y_1Y_3\}$	There is no three-factor interaction; the two-factor interactions are the same at each level	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3}$
$\{Y_1Y_2, Y_1Y_3\}$	Y_2 and Y_3 are "conditionally independent, given Y_1"; NIDDM is independent of stress, given depression	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3}$
$\{Y_1Y_2, Y_2Y_3\}$	Depression is independent of stress, given NIDDM	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jk}^{Y_2Y_3}$
$\{Y_1Y_3, Y_2Y_3\}$	Depression is independent of NIDDM, given stress	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3}$
$\{Y_1, Y_2Y_3\}$	Y_1 is "jointly independent" of Y_2 and Y_3; Depression is independent of NIDDM and stress	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{jk}^{Y_2Y_3}$
$\{Y_2, Y_1Y_3\}$	NIDDM is independent of depression and stress	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ik}^{Y_1Y_3}$
$\{Y_3, Y_1Y_2\}$	Stress is independent of depression and NIDDM	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2}$
$\{Y_1, Y_2, Y_3\}$	The three variables are "mutually independent"	$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3}$

a. $\{Y_1, Y_2, Y_3\}$ denotes the independence model:

$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3}$. The independence model assumes mutual independence among the three variables and, accordingly, does not contain any higher-order interactions. In the GBP example, this model implies that depression, NIDDM and stress are not associated with each other; knowing that a patient is depressed will not help predict her NIDDM status or her stress status, and vice versa. The natural logarithm of the probability of each cell can be decomposed into a constant plus the sum of main effects, depression status, NIDDM status and stress status. Visually, this can look like the decompositions used in analysis of variance (ANOVA):[12]



b. $\{Y_3, Y_1Y_2\}$ indicates the model where Y_3 is jointly independent of Y_1 and Y_2:

$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2}$. In terms of the GBP dataset, this

---

[12] I arranged the cells in a 2x2x2 table for this factor diagram, thinking it would be less confusing than if the data were arranged in two tables of 2x2, split by the stress variable, as I have done earlier in the paper. The array here is:

		NIDDM	
		No	Yes
No Stress	No Depression		
	Depression		
Stress	No Depression		
	Depression		

model assumes that depression and NIDDM are associated but that stress is independent of these two variables. That is, there is interaction between depression and NIDDM, but that interaction remains constant across all levels of k (i.e. stress or no stress). To help clarify the meaning of this model, I will draw upon methods of ANOVA again and construct analogous (fictional) interaction graphs. In the interaction graph on the next page on the left you can see that depression and NIDDM are related; if a patient is depressed, they are more likely to have NIDDM. This relationship is true at all levels of k (independent of their stress status) which is made visible by the single line indicating that "stress" and "no stress" are the same line. Thus, even if we know that a patient is depressed and has NIDDM, this does not help us determine her stress status. In contrast, the interaction graph on the right exemplifies what it looks like when the relationship between depression and NIDDM is not held constant for all levels of k, and thus the lines for "stress" and "no stress" are separate, non-parallel lines. In this case, knowing a patient's depression status and NIDDM status can help predict her stress status. For instance, if we know that a patient is depressed and has NIDDM, then she is also more likely to be stressed.

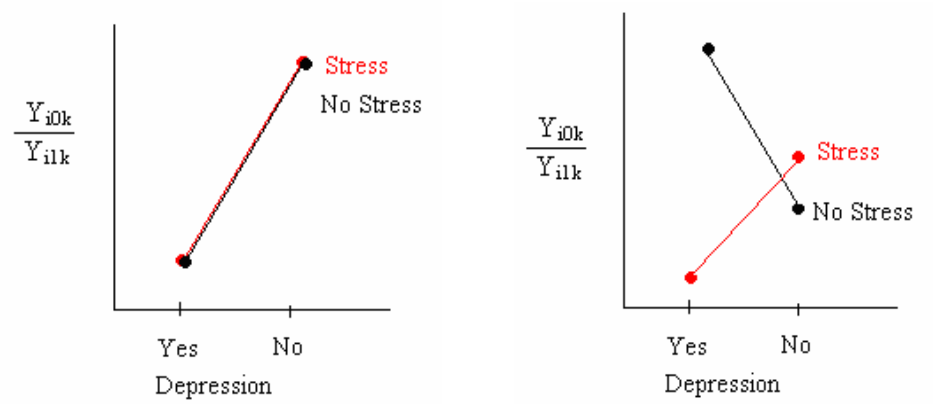


Figure 4b. No interaction with stress present (left) versus interaction with stress present (right). The y-axis is the odds of *not* having NIDDM. Lower values of (Y_{i0k}/Y_{i1k}) indicate a higher prevalence rate. For instance, (Y_{i0k}/Y_{i1k}) = 2 indicates that for every person with NIDDM, there are two people without it. Whereas, if (Y_{i0k}/Y_{i1k}) = 10, then there is one person with NIDDM for every ten people without NIDDM.

The interaction graph for the actual data can be seen in Figure 4c. This interaction graph would seem to suggest that depression and NIDDM are associated with stress. However, I will continue to use the $\{Y_3, Y_1Y_2\}$ model as an example because it will be necessary to understand this model when I begin to work with contingency tables involving missing data.

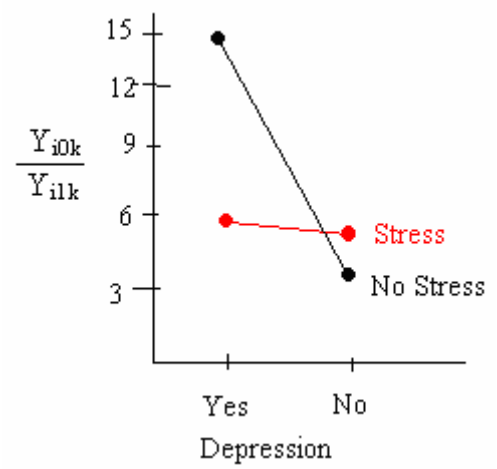


Figure 4c. The interaction graph for the GBP data suggests that depression and NIDDM are associated with stress.

c. $\{Y_1Y_2, Y_1Y_3\}$ denotes a model of conditional independence, where Y_2 and Y_3 are conditionally independent, given Y_1:

$\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jk}^{Y_2Y_3}$. In other words, NIDDM is independent of stress, given depression. Out of the three models, this model appears to be the best fit to the GBP data. The interaction graph of the real data (Figure 4c) shows that there is an association between depression and NIDDM as well as an association between depression and stress.

4.3.2 The Expected Values

The expected cell frequencies are calculated according to which model is assumed. Let $\hat{e}_{ijk}$ equal the expected cell frequency, and $\hat{p}_{ijk}$ the expected cell probability, of row i, column j, table k, where i, j, k = 0, 1.

a. I will start with the independence model. Mathematically, if P(ABC) = P(A)P(B)P(C) is true, then A, B and C are mutually independent. Therefore, if the independence model $\{Y_1,Y_2,Y_3\}$ is assumed, then the expected cell probabilities would have to satisfy the equation: $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$.[13] Since $\hat{e}_{ijk} = Y\hat{p}_{ijk}$:

---

[13] For the relationship between the equation $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$ and the equation of the hierarchical loglinear model of independence ($\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3}$), see Appendix B.

$$\hat{e}_{ijk} = Y(\hat{p}_{i++})(\hat{p}_{+j+})(\hat{p}_{++k})$$
$$= Y(\frac{Y_{i++}}{Y})(\frac{Y_{+j+}}{Y})(\frac{Y_{++k}}{Y})$$
$$= \frac{Y_{i++}Y_{+j+}Y_{++k}}{Y^2}$$

Thus, the maximum likelihood (ML) estimate for the expected cell frequencies is $\hat{e}_{ijk} = (Y_{i++}Y_{+j+}Y_{++k})/Y^2$. See table 4b for the expected values for the GBP data.

b. For stress to be jointly independent of depression and NIDDM (model $\{Y_3, Y_1Y_2\}$), the cell probabilities must satisfy the equation: $\pi_{ijk} = \pi_{ij+}\pi_{++k}$. For instance, the probability of a patient falling into cell (0,1,1)—i.e. not being depressed, but having NIDDM and stress—is equal to the probability of a patient not being depressed and having NIDDM multiplied by the probability of a patient having stress:

$$P(Depression = 0, NIDDM = 1, Stress = 1) = P(Depression = 0, NIDDM = 1)P(Stress = 1)$$

The ML estimate for the expected cell frequencies, found using the same method as above, is: $\hat{e}_{ijk} = (Y_{ij+}Y_{++k})/Y$.

c. Lastly, the third model of conditional independence $\{Y_1Y_2, Y_1Y_3\}$ takes on the probabilistic form of: $\pi_{ijk} = \frac{\pi_{ij+}\pi_{i+k}}{\pi_{i++}}$. This translates directly to the ML estimate for the expected cell frequencies: $\hat{e}_{ijk} = \frac{Y_{ij+}Y_{i+k}}{Y_{i++}}$.

Table 4b. Comparing observed values with the expected values calculated under the assumed models $\{Y_1, Y_2, Y_3\}$, $\{Y_3, Y_1Y_2\}$, and $\{Y_1Y_2, Y_1Y_3\}$. Neither the independence model nor the joint probability model are a good fit for the data; the conditional probability model fits quite well, implying that, given depression, NIDDM and stress are independent.

				Model		
Depression	**NIDDM**	**Stress**	**Observed**	**$\{Y_1, Y_2, Y_3\}$**	**$\{Y_3, Y_1Y_2\}$**	**$\{Y_1Y_2, Y_1Y_3\}$**
Yes	**Yes**	**Yes**	36	31.32	26.91	34.76
		No	1	11.74	10.09	2.24
	No	**Yes**	212	160.68	165.09	213.24
		No	15	60.95	61.91	13.76
No	**Yes**	**Yes**	4	13.05	17.45	5.24
		No	20	4.89	6.55	18.76
	No	**Yes**	20	66.95	62.55	18.76
		No	66	25.11	23.45	67.24
	χ^2			213.39	204.25	1.33
	d.f.			4	3	2
	p-value			<.0001	<.0001	.5143

4.3.2 The Goodness-of-Fit Test

Now that we have observed and expected values, we can run a χ^2 goodness-of-fit test to see whether a particular model is a good fit for the data. The null hypothesis (H_0) states that the particular model you are testing is a valid model for your data, and so any observed departures from the expected values are due to chance variation. The alternative hypothesis is that the particular model you are testing is not a valid model for your data—observed departures from expected values are *not* due to chance error but to real differences. Therefore, although statisticians usually want to reject H_0, this is not necessarily the case here. Rejecting H_0 means you have not found a good model for your data, and you are left with the saturated model.

I turn to Agresti to explain degrees of freedom: "The degrees of freedom (df) for goodness-of-fit tests equal the difference in dimension between the alternative and null hypotheses. This equals the difference between the number of parameters in the general case and when the model holds" (Agresti, 175). For a 2 x 2 x 2 contingency table, the number of parameters in the general case (the saturated model) is eight.[14] Subtracting the number of parameters in the model being tested from eight yields the appropriate degrees of freedom.

a. The independence model $\{Y_1, Y_2, Y_3\}$

H_0: $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$

H_1: $\pi_{ijk} \neq \pi_{i++}\pi_{+j+}\pi_{++k}$

df: (# of parameters in saturated model) – (# of parameters in independence model)

= 8 – 4 = 4

Even before calculating the χ^2 statistic, you can see that this model is not a good fit for the data because the expected values are so far off from the observed values. Unsurprisingly, the χ^2 statistic is large and the p-value tells us that the chance of observing our data, given that depression, NIDDM and

[14] Note that π_{1++} and π_{0++} count as one parameter (π_{i++}) because of the constraint that $\pi_{1++} = 1 - \pi_{0++}$.

stress are mutually independent, is less than one in one million (the actual p-value is in the 10^{-45} range!). Therefore, we reject the model and conclude that depression, NIDDM and stress are not mutually independent.

b. The joint probability model $\{Y_3, Y_1Y_2\}$

H_0: $\pi_{ijk} = \pi_{ij+}\pi_{++k}$

H_1: $\pi_{ijk} \neq \pi_{ij+}\pi_{++k}$

df: (# of parameters in saturated model) – (# of parameters in H_0 model)

= 8 – 5 = 3

The expected values for this model are also far off from the observed values, and a goodness-of-fit test tells us to reject the null hypothesis—stress is not independent of depression and NIDDM.

c. The conditional probability model $\{Y_1Y_2, Y_1Y_3\}$

H_0: $\pi_{ijk} = (\pi_{ij+}\pi_{i+k})/(\pi_{i++})$

H_1: $\pi_{ijk} \neq (\pi_{ij+}\pi_{i+k})/(\pi_{i++})$

df: (# of parameters in saturated model) – (# of parameters in H_0 model)

= 8 – 6 = 2

The expected values for the conditional probability model fall close to the observed values, and so we can anticipate that this will be a good fit to the data, at least certainly better than the other two models. It is! The p-value is not

even close to being significant (.5143), and so we tentatively accept the model associated with the null hypothesis. Given depression, NIDDM and stress are independent. This result is consistent with our earlier observation of the interaction graph.

Model $\{Y_1Y_3, Y_2Y_3\}$ was also found to be a good model for the GBP data (p=.29). This model indicates that NIDDM is independent of depression, given stress. Practically speaking, this model seems to make more sense than the model $\{Y_1Y_2, Y_1Y_3\}$ for this particular data. Studies have shown that high levels of stress are associated with type 2 diabetes (NIDDM).[15] Thus there is reason to believe that stress should not be independent of NIDDM as the first model indicates. Obviously, stress and depression are also related (and, indeed, model $\{Y_1Y_2, Y_2Y_3\}$ proved to be a horrible fit to the data). The association found between depression and NIDDM may be attributed to a confounding factor, namely stress. Controlling for stress, the two variables are shown to be independent of each other.

4.4 Iterative Proportional Fitting

The models I have worked with so far have all had closed form solutions, i.e. a direct way to find the ML estimates $\hat{e}_{ijk}$. However, the expected values for some three-dimensional loglinear models cannot be

[15]American Diabetes Association. "Stress." *American Diabetes Association*. 11 Feb. 2007 <http://www.diabetes.org/type-2-diabetes.jsp>.

calculated so easily. When a closed form solution does not exist, we must turn to iterative methods. The *iterative proportional fitting* (IPF) algorithm is one of the more straightforward iterative procedures used to compute expected values. I will use the model of no three-factor interaction $\{Y_1Y_{2,}\ Y_2Y_{3,}Y_1Y_3\}$, which does not have direct estimates, to demonstrate how to apply the IPF algorithm.

The IPF algorithm sequentially adjusts cell estimates so that fitted marginal distributions approach the sufficient statistics of the model being tested.[16] In the case of model $\{Y_1Y_{2,}\ Y_2Y_{3,}Y_1Y_3\}$, this means that estimates are adjusted so that fitted marginal distributions Y_{ij+}, Y_{i+k}, and Y_{+jk}, are sufficiently close to their respective observed distributions. (What exactly "sufficiently close" implies is discussed at the end of this section.)

The only restriction on preliminary estimates for cell values is that their structure cannot have more interaction terms than that of the model being fitted. That is, if we are fitting the model of no three-factor interaction, we cannot begin with estimates that have three-way interaction. Perhaps the easiest and most common way to begin is with preliminary estimates $\hat{e}_{ijk}^{(0)} = 1$.

Next, we want to consecutively adjust the estimates to fit the marginal distributions. For $\{Y_1Y_{2,}\ Y_2Y_{3,}Y_1Y_3\}$, this involves repeating the same three steps over and over again, namely:

---

[16] Sufficient statistics contain all the information necessary for estimating a parameter. Here, the marginal distributions are sufficient, and knowing the full data will not contribute any more to fitting the model $\{Y_1Y_{2,}\ Y_2Y_{3,}Y_1Y_3\}$.

(1) The first step is to match the fitted and observed marginal distributions Y_{ij+}: $\hat{e}_{ijk}^{(1)} = \hat{e}_{ijk}^{(0)} \frac{Y_{ij+}}{\hat{e}_{ij+}^{(0)}}$.

(2) The second step is to match the fitted and observed marginal distributions Y_{+jk}: $\hat{e}_{ijk}^{(2)} = \hat{e}_{ijk}^{(1)} \frac{Y_{+jk}}{\hat{e}_{+jk}^{(1)}}$.

(3) The third step is to match the fitted and observed marginal distributions Y_{i+k}: $\hat{e}_{ijk}^{(3)} = \hat{e}_{ijk}^{(2)} \frac{Y_{i+k}}{\hat{e}_{i+k}^{(2)}}$.[17]

We repeat this cycle using $\hat{e}_{ijk}^{(3)}$ to estimate $\hat{e}_{ijk}^{(4)}$, etc. until each fitted marginal distribution is sufficiently close to the observed distributions. "Sufficiently close" is a rather ambiguous term, but there are various stopping rules we could use. Bishop, Fienberg and Holland (1975) suggest stopping when, after a complete cycle, no cell values change more than some predetermined quantity (e.g. $\varpi = .05$). In other words, they stop once the statement $\left|\hat{e}_{ijk}^{(3r)} - \hat{e}_{ijk}^{(3r-3)}\right| < \varpi$ is true for all i,j,k (Bishop, Fienberg and Holland, 85). Alternatively, Agresti stops once all fitted marginal frequencies are within .02 of their respective observed frequencies (Agresti, 186).

Interestingly enough, I did not have to use a stopping rule when fitting the model of no three-way interaction to the GBP data. After the first two steps—not even completing a whole cycle—the fitted marginal distributions

---

[17] See Bishop, Fienberg and Holland (1975), p.85.

were exactly equal to the observed marginal distributions (see Table 4c). A χ^2 goodness-of-fit test for this model proved not significant (p=.25), meaning the model of no-three-way interaction is a good fit for the data. However, since the models $\{Y_1Y_3, Y_2Y_3\}$ and $\{Y_1Y_2, Y_1Y_3\}$ are also good fits and have fewer parameters, we should accept one of these models as the best fit for the GBP data. As reasoned at the end of Section 4.3, model $\{Y_1Y_3, Y_2Y_3\}$ is a workable model—both statistically and practically speaking; given stress, depression is independent of NIDDM.

Table 4c. Iterative Proportional Fitting of model $\{Y_1Y_2, Y_2Y_3, Y_1Y_3\}$ to GBP data

			Fitted Values		
Depression	NIDDM	Stress	$\hat{e}_{ijk}^{(0)}$	$\hat{e}_{ijk}^{(1)}$	$\hat{e}_{ijk}^{(2)}$
Yes	Yes	Yes	1.0	18.5	34.76
		No	1.0	18.5	2.24
	No	Yes	1.0	113.5	213.24
		No	1.0	113.5	13.76
No	Yes	Yes	1.0	12	5.24
		No	1.0	12	18.76
	No	Yes	1.0	43	18.76
		No	1.0	43	67.24

4.5 Investigating the Relationship between Binary Variables and Return Status

Return status (i.e. the status of a patient's two-year follow up exam: returned or did not return) may be dependent upon any single variable or a combination of variables, both pre- and post-operative. For some variables, namely the pre-operative variables, we are able to determine whether or not return status depends on that variable. On the other hand, the relationship

between post-operative variables and return status cannot be known; we do not have information about the health status of patients who did not return for their two year follow-up. In this paper I will be concentrating on the relationship between the categorical, binary variables (both pre- and post-operative) and return status.

I started my investigation by creating 2 x 2 contingency tables, where Y_1 denotes a binary variable (e.g. asthma) and Y_2 corresponds to whether or not a patient returned for his two year follow-up. For example:

		Two Year Follow-up?			
		0	**1**	**Total**	**% Return**
Asthma?	**0**	135	135	270	50%
	1	48	56	104	54%

Judging by this contingency table, it appears that asthma and return status are unrelated: 50% of the patients without asthma pre-surgery returned for their two year follow-up exam and about 53.8% of patients with asthma returned. A patient's pre-operative asthma status does not affect her probability of returning for the two year follow-up.

In fact, the only binary variables that were found to be associated with a patient's chance of returning are stress and depression (p=.001 for both).[18] See Figure 4d. While 56.3% of patients with pre-operative stress returned two

---

[18] P-values are from χ^2 tests for independence. Hyperlipidemia and IDDM were also found to be significant at the .05 level (p=.03 and p=.02, respectively). However, due to time constraints, it is unlikely I will get to work with each variable. See Figure 4d for my plan of attack.

years later, only 37.3% of patients without stress returned. Likewise, 56.4% of patients with depression returned, but only 38.2% of those without depression returned. Thus, patients who reported either stress or depression (or both) before surgery were much more likely to return for their two year follow-up. This is an interesting finding, as I would have assumed that patients who reported stress are very busy and would have been *less* likely to make time for their two-year follow up exam. Retrospectively, perhaps patients who reported no stress are more laid back in general and were lackadaisical about returning for their follow-up exams, whereas patients with stress and/or depression were more inclined to look to people of authority for help and assurance.

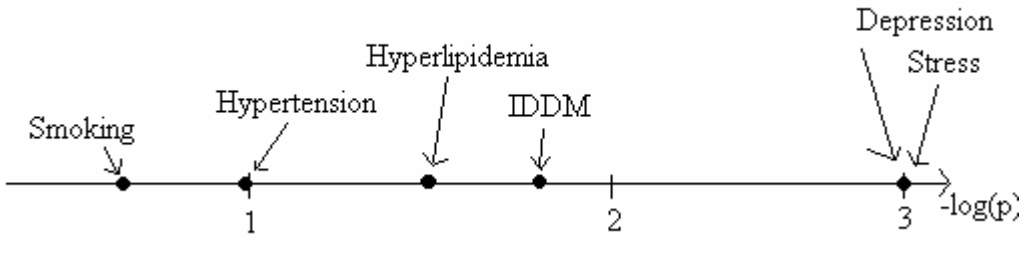


Figure 4d. The negative logarithm of the p-value for binary variables with –log(p) > .50. A lower p-value (and thus a higher –log(p)) indicates higher significance. The conservative approach in this situation would involve setting a higher level of significance (say, α =.10) since we would rather commit type one error (wrongly rejecting the null hypothesis) than type two error (wrongly failing to reject the null hypothesis). In other words, it is better to investigate further the relationship between return status and a variable even if the two are actually independent of each other, then to overlook an important association between return status and a variable. Therefore, I will begin by exploring the two most significant variables, depression and stress. I will then continue down the line exploring the other variables.

A three-way contingency table relating depression status, stress status and return status can be seen in Figure 4e on the next page.

Table 4d. A three-way contingency table to illustrate the relationship between depression (D), stress (S) and return status (M).

Depression	Stress			
		0	**1**	
	0	32	10	*42*
	1	6	143	*149*
		38	*153*	*191*

Miss = 0

Depression	Stress			
		0	**1**	
	0	54	14	*68*
	1	10	105	*115*
		64	*119*	*183*

Miss = 1

After testing various hierarchical loglinear models, I found that both {DS,DM} and {DS,SM} were good fits to the data (p=.267 and p=.321, respectively). The first model implies that, given depression, return status is independent of stress. The second model implies the parallel alternative—given stress, return status is independent of depression. The model of three two-way associations {DS, DM, SM} was also a good fit to the observed data (p=.371). How can we determine which model is the best fit then?

Taking a look at the interaction graph can perhaps shed some more light on the relationship:

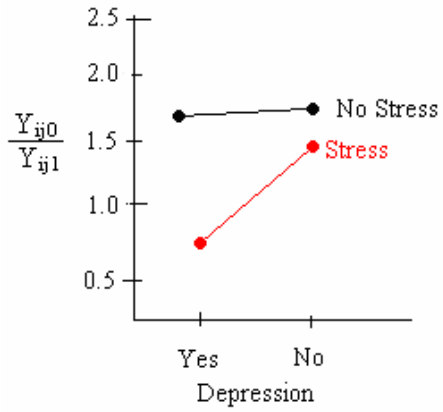


Figure 4e. An interaction graph where the response equals the odds of not returning.

Figure 4e shows that the relationship between depression and return status is not the same for each level of stress status. For patients who reported stress, the odds of returning depend upon their depression level. For patients who reported no stress, the odds of returning is about equal whether or not they reported depression. In other words, for people who reported stress, there is interaction between depression and return status, whereas, for people who reported no stress, there is no interaction between depression and return status. Let us take a closer look at these relationships, via partial and marginal odds ratios.

4.6 Odds Ratios

First I will explain odds and odds ratios for 2 x 2 contingency tables, and then I will relate them back to Figure 4f and the relationship between depression, stress and return status.

The odds for a 2 x 2 table is defined to be the probability of falling into cell (i,0) divided by the probability of falling into cell (i,1): $\frac{\pi_{i0}}{\pi_{i1}}$. We can estimate these probabilities using the observed cell frequencies ($\frac{Y_{i0}}{Y_{i1}}$). Take, for example, the 2 x 2 contingency table relating pre-operative asthma and two year follow-up status:

Asthma?	Two Year Follow-up?		
		0	**1**
	0	135	135
	1	48	56

For patients who reported no asthma, the estimated odds of not returning for their two-year follow up is Y_{00}/Y_{01} = 135/135 = 1. For patients who reported having asthma, the estimated odds of not returning is 48/56 = .857; for every one person that returned, .857 did not return. We could also look at the odds of returning—which is just the inverse of the odds of not returning—which tells us that, for every one person that did not return, 1.2 people did return (among patients with asthma).

The odds ratio compares the odds of returning for those with and without asthma. We estimate the odds ratio by calculating $(\frac{Y_{00}}{Y_{01}})/(\frac{Y_{10}}{Y_{11}}) = \frac{(Y_{00} * Y_{11})}{(Y_{10} * Y_{01})}$. Estimated odds ratios close to one indicate independence between the variables. For the asthma example, the odds ratio equals 1.16, meaning the odds of returning is 1.16 times as high for patients with asthma as for patients without asthma. The odds ratio is close to 1, suggesting independence between pre-operative asthma and return status. This evidence is consistent with the conclusion made in section 4.5 using a χ^2 test of independence.

For a 2 x 2 x 2 table, there are marginal odds ratios and partial odds ratios. Marginal odds ratios express the association between two variables,

ignoring the third variable. Marginal odds ratios are calculated, after summing observed frequencies over the third variable, in the same way odds ratios for 2 x 2 tables are found. Partial odds ratios express the association between the other two variables while controlling for the third variable. Thus, for each pair of variables, there are two partial odds ratios, one for when the third variable is zero and one for when the third variable is one. See Table 4e for marginal and partial odds ratios relating depression, stress and return status.

As shown in Table 4e, the odds ratios relating depression and stress are very high, confirming what we already know—depression and stress are very much related! Ignoring return status, the estimated odds of having stress is around 56 times higher for patients with depression than for patients without depression. Interestingly enough, the partial odds differed from each other: when a patient did not return, the estimated odds of stress were 40 times higher for those with depression than those without depression, whereas, if a patient did return, the estimated odds of stress were 76 times higher for those with depression than for those without depression.

Furthermore, the partial odds ratios between depression and return status correspond to the observations made regarding the interaction graph in Figure 4e. Earlier I noted that the interaction graph shows that, for patients who reported stress, the odds of returning depend upon their depression level; for patients who did not report stress, on the other hand, the odds of returning are about equal whether or not they reported depression. Indeed, if we

condition on stress, we find that the odds ratio is different for patients with and without stress. For patients who reported pre-operative stress, the odds of returning were almost twice as high for those with depression than for those without depression. For patients who did not report pre-operative stress, the odds ratio is one—that is, the odds of returning do not depend on depression.

Table 4e. Odds ratios for depression (D), stress (S), and return status (M).[19]

		Variables		
Association		**D-S**	**D-M**	**S-M**
Marginal		55.54	2.09	2.31
Partial	**Third Variable=0**	40.50	1.01	1.21
	Third Variable=1	76.27	1.91	2.27

4.7 Model Selection

By the end of section 4.5, the decision of which model best fits the data had come down to three different models: {DS,DM}, {DS,SM}, and {DS,DM,SM}. Taking into consideration the goodness-of-fit tests, the interaction graphs and the odds ratios, the model {DS,SM} emerges as the best model for relating depression, stress and return status.

In rationale for accepting the conditional independence model, (a) its goodness-of-fit test gives a higher p-value—i.e. observed values are closer to

[19] Table 4d is of the same form as that found in Agresti's *Categorical Data Analysis*, page 137.

expected values—than model {DS,DM}; (b) grounds to believe that the association found between depression and return status is due to the confounding variable stress, as laid out earlier; (c) the p-value for the model of no three-factor interaction automatically must be higher than that of the conditional independence model because it has more parameters. The p-value, however, is only .05 higher (.371 versus .321), suggesting marginal improvement in fit—an improvement so minor that it is probably just due to the fact that the no three-factor interaction model has more parameters; (d) the interaction graph and odds ratio reveal dependence between depression and stress as well as dependence between stress and missingness.

Therefore, I conclude that, given pre-operative stress status, pre-operative depression status and return status are independent. The findings of this investigation will help us as we move on to the heart of the matter: imputing missing values.

V. Imputing Missing Values

5.1 Overview

There are three kinds of variables in my analysis: return status, variables used to predict return status, and variables used to assess the effect of surgery. There are several variables I have and/or will use to predict return status, including stress, depression, IDDM, hyperlipidemia, and hypertension (all post-operative). The variables used to assess the effect of surgery—the response variables—will include two-year weight loss, systolic blood pressure, diastolic blood pressure, and cholesterol reduction. I hope to adjust these estimates by taking into account the missing data.

Before delving into the technical aspects of imputing missing values, I would first like to illustrate the importance of missing data via a hypothetical example. Let us imagine that we have a study of 200 patients but post-operative information for only 100 of these patients (i.e. 100 patients did not return). Furthermore, let's say we are looking at return status, asthma as the predictor of return status, and weight loss as the response variable. First, we find an association between pre-operative asthma and return status: whereas 55% of patients with asthma returned for their follow-up, only 40% of patients without asthma returned. The available-case analysis yields an unadjusted weight loss estimate of 161 pounds. We then notice, however, that, not only is return status associated with asthma, but patients without post-operative asthma lost much less weight than patients with post-operative asthma (100

pounds versus 200 pounds). Since patients without asthma not only returned at a lower rate, but also lost a lot less weight, we should be concerned that our unadjusted estimate for weight loss is biased. Adjusting for the missing data (and assuming model {PrePost,PostMiss}), we get a mean weight loss of 134 pounds. The adjusted estimate is considerably lower than the unadjusted estimate!

Surely we could also imagine the reverse situation, where available-case analysis leads to underestimating weight loss. [In fact, if we switch the mean weight loss for the two groups—so patients without asthma lose 200 pounds on average and patients with asthma lose 100 pounds on average—the adjusted estimate (166 pounds) is considerably higher than the unadjusted estimate (139 pounds).][20]

5.2 Contingency Tables with Partially Classified Margins

In the example relating depression, stress and return status, we were able to create a fully classified 2 x 2 x 2 table because we were working with pre-operative variables and we know each patient's pre-operative status. However, in working with post-operative variables, we do not know every patient's status and thus are unable to complete each cell in a 2 x 2 x 2 contingency table. We can, however, create a contingency table with a supplemental margin. Since return status was found to be dependent upon pre-

[20] See Appendix B for more details.

operative stress status in Chapter 4, I will work with the variables pre-operative stress status (Pre), post-operative stress status (Post), and return status (Miss).

Table 5a. A 2 x 2 contingency table with a supplemental margin.

Pre	Post		
		0	**1**
	0	34	4
	1	37	116

f = 191

Miss = 0

Pre	Post		
		0	**1**
	0	?	?
	1	?	?

m_0=64

m_1=119

m = 183

Miss =1

Since future calculations may get confusing using Y_{ijk}s, let f_{ij} denote the fully classified frequencies (Miss=0 cells) and m_i refer to the partially classified margins (for Miss=1, where we do not know the values for Post); $\hat{m}_{ij}$ denotes imputed (estimated) values.

Given our investigation in Chapter 4, we can rule out certain hierarchical loglinear models that we know would not be reasonable models for the missing data. For instance, the model of mutual independence can be ruled out since we've already established that return status is dependent upon pre-operative stress (and it is also sensible to presume that pre- and post-operative stress are related). For that same reason, we can also assume that the model of joint independence—that pre- and post- operative stress are jointly independent of return status, {PrePost,Miss}—is also not a reasonable model. See table 5b for a list of reasonable models that could be used to estimate missing values.

Table 5b. Reasonable models for relating pre- and post-operative stress and return status. The "Estimable" column indicates whether the model has parameters that can be estimated without additional information.

Model	**Interpretation**	**Estimable?**
{PrePost,PreMiss}	Return status does not depend upon post-operative stress status, only upon pre-operative stress status. Data are MAR; missing data is ignorable.	Yes
{PrePost,PostMiss}	Return status depends on post-operative stress.	Yes
{PrePost,PreMiss,PostMiss}	There is no three-factor interaction; two-factor interactions are the same at each level.	No
{PrePostMiss}	There is three-factor interaction.	No

The first model in Table 5b assumes that return status is dependent only upon pre-operative stress status, but not dependent on post-operative stress status. Another way to interpret the model is to say that, given pre-operative stress status, knowing a patient's post-operative stress status will not help any more in predicting whether or not the patient will return for her two year follow-up. If this model is correct, then the data are MAR and the missing data mechanism is ignorable. In other words, estimates of recovery rates, weight loss, etc. using imputed values would be equal to the estimates found using available-case analysis. In actuality, it is improbable that return status is not related to post-operative stress status, since pre- and post-operative stress statuses are highly related.

However, the second model in Table 5b—which states that, given post-operative stress, return status is independent of pre-operative stress—is a more plausible model. Having concluded earlier that return status and pre-operative

stress are related, how can this be a reasonable model? A parallel alternative to the first model, this model implies that if we know a patient's *post*-operative stress status, then knowing her pre-operative status is of no help in predicting whether or not she returns for her two-year follow-up exam; that is, the patient's pre-operative status does not give us any more information above and beyond that which we get from her post-operative status.

Although we are unable to investigate associations between post-operative variables and return status (as explained previously), it is logical to assume that if a patient's pre-operative status is related to her return status, then her post-operative status will also be related to her return status. Moreover, since post-operative status is recorded at the same time as return status, it is likely that post-operative status would be an even better predictor of return status. To help recognize why this is true, take a more extreme example: if you are trying to guess Jane's weight today, would you rather know what her weight was a week ago or what her weight was at birth? Similarly, would you rather know what her weight was one year ago or two years ago? In both situations, you would want to choose the first option—knowing the most up-to-date status of Jane's weight would be the most helpful in predicting her weight today.

Likewise, knowing a patient's most up-to-date stress status would be most helpful in anticipating her return status. Furthermore, just as knowing the aforesaid Jane's birth weight in addition to knowing her weight yesterday

would not offer any more helpful information, knowing a patient's pre-operative stress status would also probably offer no more helpful information. Therefore, the second model with two two-way associations is a plausible assumption: pre- and post-operative stress are associated and post-operative stress and return status are associated. (See Figure 5a for a diagram of the relationship of the three variables.)

The last two models of Table 5b involve more interactions between variables. However, because they are inestimable (that is, model parameters cannot be estimated without additional information), I will not be working with them in this paper.[21]

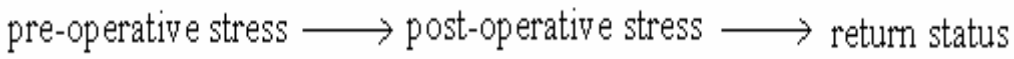


Figure 5a. The relationship between pre-operative stress, post-operative stress, and return status. Pre-operative stress is predictive of post-operative stress which, in turn, is predictive of return status. Thus, we are able to find an association between pre-operative stress and return status because that is the relationship we are able to study. However, had we been able to look at post-operative stress status and return status, it is likely we would have not only found an association between the two variables, but also found that, given post-operative stress, pre-operative stress is no longer helpful in predicting return status.

5.3 Imputing Values

Since model {PrePost,PostMiss} implies that return status is dependent upon post-operative stress, we want the column distributions of the fully

---

[21] The model of no three-factor interaction and the model of three-factor interaction are inestimable because there are more parameters than there are degrees of freedom in these two models. Parameters are unknowns and degrees of freedom are a set of linear equations; having more unknowns than equations makes the problem unsolvable.

observed values and the imputed values to match. That is, we want $\frac{f_{ij}}{f_{+j}} = \frac{\hat{m}_{ij}}{\hat{m}_{+j}}$, or in terms of the GBP dataset, $\frac{37}{34} = \frac{\hat{m}_{10}}{\hat{m}_{00}}$ and $\frac{116}{4} = \frac{\hat{m}_{11}}{\hat{m}_{01}}$ (see table 5ai below).[22]

Table 5ai. A 2 x 2 contingency table with a supplemental margin. (Copy of Table 5a)

Pre	Post		
		0	**1**
	0	34	4
	1	37	116

f = 191

Miss = 0

Pre	Post		
		0	**1**
	0	$\hat{m}_{00}$	$\hat{m}_{01}$
	1	$\hat{m}_{10}$	$\hat{m}_{11}$

m_0=64

m_1=119

m = 183

Miss =1

We also know what the row totals of the missing data must be: $\hat{m}_{00} + \hat{m}_{01} = m_0$ ($\hat{m}_{00} + \hat{m}_{01} = 64$) and $\hat{m}_{10} + \hat{m}_{11} = m_1$ ($\hat{m}_{10} + \hat{m}_{11} = 119$). Solving so that an estimated $\hat{m}_{ij}$ term involves only known (observed) values yields

$$\hat{m}_{00} = \frac{m_1 - (m_0)(\frac{f_{11}}{f_{01}})}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})} = \frac{119 - (64)(\frac{116}{4})}{(\frac{37}{34}) - (\frac{116}{4})} = 62.23.$$[23] The imputed values for the GBP data can be seen in Table 5c, and the column distributions do indeed match: $\frac{37}{34} = \frac{67.72}{62.23} \approx 1.09$ and $\frac{116}{4} = \frac{51.28}{1.77} \approx 29$.

The imputed values of Table 5c estimate that, out of the 183 patients who did not return for their two year follow-up, 129.95 (or 71%) of them did

[22] Recall that the new notation is meant to reduce confusion: the f_{ij} terms stand for the fully-

not have post-operative stress. We would expect the majority of patients who did not return to not have post-operative stress (as is the case here), since the assumption is that patients are less likely to return if they do not have stress.

Table 5c. 2 x 2 x 2 contingency table filled in based on model {PrePost,PostMiss}. Imputed values are in grey.

Pre	Post		
		0	**1**
	0	34	4
	1	37	116
		71	*120*

38			
153			
191			

Miss = 0

Pre	Post		
		0	**1**
	0	**62.23**	**1.77**
	1	**67.72**	**51.28**
		129.95	***53.05***

64
119
183

Miss = 1

The available-case analysis found that 37% of patients are stress-free two years after the operation (a 24.2% recovery rate). In contrast, using the imputed values finds that about 53.7% of patients have no stress two years post-op (a 38.5% recovery rate).

Using the imputed values we can also find adjusted estimates for two-year weight loss as well as systolic blood pressure, diastolic blood pressure, and cholesterol reductions.

5.4 Adjusted Estimates

---

classified data and the m_{ij} terms stand for the missing data.

[23] See Appendix B for the proof.

In using the imputed values to adjust estimates for weight loss, there are two main assumptions being made. First, weight loss depends on stress. Second, given stress, weight loss is independent of return status.

To check the first assumption, we can compare the two-year weight loss of patients with post-operative stress to that of patients without stress (out of those patients who returned). In fact, the values do differ, although not considerably: patients with post-operative stress lost, on average, 110 pounds while their stress-free counterparts lost around 120 pounds.

The adjusted weight loss estimate is found by taking into account the proportion of patients with and without post-operative stress and their respective mean weight loss. Table 5d relates return status and post-operative stress. The "No" column values are observed frequencies, while values in the "Yes" column are estimated from the three-way analysis of Section 5.2. Also, whereas the column totals (Y_{+j}) are known, the row totals (Y_{i+}) are estimated.

Table 5d. A 2 x 2 contingency table relating return status and post-operative stress. Observed values are in black; imputed values are in grey.

		Missing?		
		No (Obs.)	**Yes (Est.)**	**Total**
Post-Operative Stress?	**No**	71	129.95	200.95
	Yes	120	53.05	173.05
	Total	191	183	374

To find a weighted averaged we will use the estimated row totals Y_{i+}:

[(Number of patients with post-operative stress)*(The observed mean weight for patients who returned with post-operative stress)] + [(Number of patients without post-operative stress)*(The observed mean weight for patients who returned without post-operative stress)] all divided by the total number of patients. This comes out to $\frac{(200.95)(119.6)+(173.05)(110)}{374}=115.2$ pounds.

The two-year weight loss estimate based on a weighted average is only the slightest bit higher than that found in Part I using the available-case analysis (113.6 pounds).

This conclusion is not surprising. Patients with and without post-operative stress varied only slightly in their mean weight loss, and the weighted average has to be between their two means. That is, the weighted average was going to be somewhere between 110 and 120 pounds, falling closer to one or the other depending on whether there were more or less patients with post-operative stress.

Patients with and without post-operative stress also did not vary much in their blood pressure and cholesterol reductions. Thus, estimates found using weighted averages are again quite close to those found using available-case analysis. See Table 5e below for a comparison of averages.

Table 5e. Comparison of estimates found using available-case analysis versus weighted averages (based on stress model).

Estimate	***Available-Case Average***	***Weighted Average***
Weight Loss	113.6	115.2
SBP Reduction	9.1	8.8
DBP Reduction	5.7	5.9
Cholesterol Reduction	28.9	28.2

5.5 Adjusted Estimates Based on Other Models

Figure 4d from Section 4.5 ordered the strength of the relationship between return status and binary variables (based on p-values from χ^2 tests of independence), with stress and depression having the strongest relationship with return status. It is safe to assume for the other variables (IDDM, hyperlipidemia, etc.) that which was assumed for the stress model: given post-operative status, pre-operative status is no longer any help in determining return status. Again, it is likely that pre-operative status is predictive of post-operative status which, in turn, is predictive of return status.

Interestingly enough, whereas we did not find a large difference between the mean weight loss of patients with and without post-operative stress, there is a large difference in the mean weight loss for patients with and without post-operative hyperlipidemia as well as those with and without IDDM. Patients without hyperlipidemia lost, on average, 117.4 pounds; patients with hyperlipidemia only lost 87 pounds on average (n=161 and n=27, respectively). Likewise, patients without IDDM lost 114.1 pounds on average while patients with IDDM lost an average of 83.4 pounds (n=181 and n=7, respectively).[24]

I imputed values based on models of the form {PrePost,PostMiss}—where Pre is pre-operative status, Post is post-operative status and Miss is return status—for variables depression, IDDM, hypertension and

[24] I also look at the effect of the combination of IDDM and hyperlipidemia in a 4 x 4 analysis; see section 5.6 for more details.

hyperlipidemia. See Table 5f for a comparison of adjusted estimates based on these different models.

All of the adjusted estimates are quite close to the unadjusted estimates. This finding is neither necessarily good nor necessarily bad; rather, this investigation shows that we can be confident in the available-case analysis. For this particular study, the unadjusted estimates are sound estimates.

Table 5f. Comparison of available-case averages and weighted averages, based on various models.

Estimate	***Available-Case Average***	***Weighted Average***			
		Depression Model	**IDDM Model**	**Hyperlipidemia Model**	**Hypertension Model**
Weight Loss	113.6	115.6	113.5	115.2	114.2
SBP	9.1	9.2	8.7	9.0	10.2
DBP	5.7	6.1	5.5	6.1	6.1
Cholesterol	28.9	29.7	29.3	31.5	30.0

5.6 Models Involving Two Predictor Variables

I have found that the adjusted estimates based on models involving a single predictive variable are quite similar to the unadjusted estimates. What about adjusted estimates based on models involving more than one predictive variable? We can, for example, look at the relationship between return status and the combination of depression and IDDM. To summarize this relationship, we can make a 4 x 2 contingency table (see Table 5g).

While 36.3% of the patients without either depression or IDDM returned, 77.8% of the patients with depression and IDDM returned. This is, in fact, the pattern that each of the ten models with two predictive variables

follows: the return rate is lowest for patients with neither risk factor and highest for patients with both risk factors.

Table 5g. A 4 x 2 contingency table relating depression, IDDM and return status. (0,0) indicates the patient had neither depression nor IDDM; (1,0) indicates the patient had depression, but not IDDM, etc.

		Missing?		
(Depression, IDDM) Pre		**No (%)**	**Yes**	
	(0,0)	37 (36.3)	65	*102*
	(1,0)	135 (54.9)	111	*246*
	(0,1)	5 (62.5)	3	*8*
	(1,1)	14 (77.8)	4	*18*
		191	*183*	*374*

With two predictive variables, we now have a 4 x 4 contingency table with a partially classified margin (See Table 5h). Furthermore, there are six zeros in the completely classified table of Miss=0. We cannot, therefore, use the same equations to impute missing values as we did for the 2 x 2 contingency tables. Instead, we need to use the EM algorithm, which "always converges to a solution (when IPF is used for the M step), even if the solution lies of the boundary of a parameter space," i.e. an imputed value is equal to zero (Baker and Laird, 63).

The maximization step, or M-step, involves fitting the nonresponse model to the data (Baker and Laird 1988). Since we are continuing to assume that, given post-operative status, pre-operative status and return status are independent (model {PrePost,PostMiss}), the maximum likelihood estimates

are obtained by making $Y_{ijk} = \frac{(Y_{ij+})(Y_{+jk})}{Y_{+j+}}$. Let M_{ijk} equal the maximum likelihood estimate at the present iteration for cell i, j, k, where i, j, k = 1, 0.

Table 5h. A 4 x 4 contingency table with a partially classified margin.

		(Depression, IDDM) Post				
(Depression, IDDM) Pre		**(0,0)**	**(1,0)**	**(0,1)**	**(1,1)**	
	(0,0)	37	0	0	0	*37*
	(1,0)	46	89	0	0	*135*
	(0,1)	2	1	2	0	*5*
	(1,1)	3	6	1	4	*14*
		88	*96*	*3*	*4*	*191*

Miss = 0

		(Depression, IDDM) Post				
(Depression, IDDM) Pre		**(0,0)**	**(1,0)**	**(0,1)**	**(1,1)**	
	(0,0)	?	?	?	?	*65*
	(1,0)	?	?	?	?	*111*
	(0,1)	?	?	?	?	*3*
	(1,1)	?	?	?	?	*4*
		?	*?*	*?*	*?*	*183*

Miss = 1

The expectation step, or E-step, revises the expected missing values by setting $E(Y_{ij0}) = Y_{ij0}$ (since values of Y_{ij0} are the known, observed values), and $E(Y_{ij1}) = \frac{(Y_{i+1})(M_{ij1})}{M_{i+1}}$ (where, again, Y_{i+1} are the observed counts of the incompletely classified table Miss=1, and M_{i+1} are the mles at the present iteration) (Baker and Laird 1988).

Using the stopping criterion as suggested by Baker and Laird—to stop once the consecutive change in the log-likelihood is less than .001—the EM

algorithm for the depression and IDDM model converges after 228 steps (1988). The resulting table with imputed values is shown in Table 5i.

Table 5i. A 4 x 4 x 2 contingency table. All values are imputed. If the model {PrePost,PostMiss} is indeed a good model, the maximum likelihood estimates will be the best estimates.

(Depression, IDDM) Pre		(Depression, IDDM) Post				
		(0,0)	(1,0)	(0,1)	(1,1)	
	(0,0)	36.82	0	0	0	*36.82*
	(1,0)	47.37	89.42	0	0	*136.79*
	(0,1)	1.71	0.97	2	0	*4.68*
	(1,1)	2.09	5.61	1	4	*12.7*
		87.99	*96*	*3*	*4*	*190.99*

Miss = 0

(Depression, IDDM) Pre		(Depression, IDDM) Post				
		(0,0)	(1,0)	(0,1)	(1,1)	
	(0,0)	65.18	0	0	0	*65.18*
	(1,0)	83.86	25.36	0	0	*109.22*
	(0,1)	3.04	0.28	0	0	*3.32*
	(1,1)	3.71	1.59	0	0	*5.3*
		155.79	*27.23*	*0*	*0*	*183.02*

Miss = 1

After imputing values, we follow the same process used before to find the new adjusted estimates using a weighted average. Once again, adjusted estimates based on this particular model (depression and IDDM) are very similar to the unadjusted estimates. In fact, the adjusted estimates resulting from each of the ten models involving two predictive variables are all close to the unadjusted estimates: the adjusted weight loss, systolic blood pressure, diastolic blood pressure and cholesterol estimates are within, respectively, five pounds, 1.6 mmHg, 3.5 mmHg, and 4 mg/dL of the unadjusted estimates.

These results are consistent with previous findings and support the conclusion that the available-case analysis is reliable.

VI. Conclusions

6.1 Further Research

As the number of people opting for bariatric surgery increases, it is important that prospective patients are fully informed about the benefits and risks involved before proceeding with the surgery. This thesis has concentrated on the benefits of a particular bariatric program at North Shore Medical Center. The two year results are quite positive and, although the preliminary five year results show a small relapse, they are still more encouraging than what might be expected. Five year weights are still much lower than baseline weights and, as I said before, these five-year figures are only preliminary results based on a small sample of 27 patients. I look forward to completing a more thorough analysis of the five year results as more patients return for their five year follow-up exam.

Furthermore, I would like to study the risks involved with the surgery. In light of the evidence suggesting gastric bypass patients have a reduced tolerance for alcohol post-surgery, it would be worthwhile recording whether patients noted a heightened sensitivity toward alcohol post-surgery (if consumed).[25] Also, what is the risk of death for this particular program? What is the risk of other complications associated with the surgery? Such information could help potential patients decide if the benefits of the surgery

[25] Klockhoff, H., Naslund, I., and Jones, A.W. (2002), "Faster Absorption of Ethanol and Higher Peak Concentration in Women after Gastric Bypass Surgery," *British Journal of Clinical Pharmacology*, 54, 587-591.

outweigh the risks. In addition, for North Shore Medical Center, a study involving risks could help identify areas of the program that deserve increased attention, care and/or improvement.

6.2 Conclusions Regarding Missing Data

For this particular study, the unadjusted (available-case) and adjusted analyses yielded similar estimates. Despite what one might expect, the fact that only 50% of the patients returned for their two year follow-up exam does not introduce much bias. Insofar as my investigation is concerned, the available-case analysis can be trusted.

Appendix A

Table 3.i Number of patients per sample. (From Table 3c on Page 17)

Characteristic		Female	Male
		N (%)	N (%)
Age		411 (85.3)	71 (14.7)
Weight*		400 (84.7)	72 (15.3)
BMI		350 (84.3)	65 (15.7)
Total Cholesterol		364 (84.5)	67 (15.5)
	LDL	315 (84.7)	57 (15.3)
	HDL*	336 (84.4)	62 (15.6)
SBP		410 (84.7)	74 (15.3)
DBP		410 (84.7)	74 (15.3)
Heart Rate		405 (84.4)	75 (15.6)
Diabetes		434 (85.1)	76 (14.9)
	NIDDM	434 (85.1)	76 (14.9)
	IDDM	434 (85.1)	76 (14.9)
Hypertension		434 (85.1)	76 (14.9)
Hyperlipidemia		434 (85.1)	76 (14.9)
Cellultis		434 (85.1)	76 (14.9)
Asthma		434 (85.1)	76 (14.9)
Joint pain		434 (85.1)	76 (14.9)
Sleep apnea		434 (85.1)	76 (14.9)
Stress/anxiety		434 (85.1)	76 (14.9)
Depression		434 (85.1)	76 (14.9)
Smoking		434 (85.1)	76 (14.9)
Addictive behavior		434 (85.1)	76 (14.9)

Table 3.ii Crosstabs with the number of patients indicating physical improvement two-year post-op across rows, and the number of patients indicating physical improvement five-years post-op across columns. Only two patients improved from two years to five years, while four patients regressed.

		Physical Improvement5yr		Total
		0	1	
Physical Improvement	0	1	2	*3*
	1	4	20	*24*
Total		*5*	*22*	*27*

Table 3.iii Crosstabs with the number of patients indicating emotional improvement two-year post-op across rows, and the number of patients indicating emotional improvement five-years post-op across columns. One patient improved while six patients regressed.

		Emotional Improvement 5 yr		
		0	1	Total
Emotional	0	2	1	*3*
Improvement	1	6	18	*24*
Total		*8*	*19*	*27*

Appendix B

4.3 Hierarchical Loglinear Models

(Page 29)

Interpretations of loglinear model parameters are taken from Agresti's *Categorical Data Analysis* (page 143). Agresti defines the saturated model slightly differently than Bishop, Fienberg and Holland (1975), setting the equation equal to the logarithm of the expected cell frequency rather than the logarithm of the cell probability. The two are closely related, and I will describe the relationship after defining model parameters for the Agresti model.

Let $l_{ijk} = \ln(e_{ijk})$, the natural logarithm of the expected cell frequency. As before, a plus sign in the subscript indicates summing over that index. For example, l_{00+} denotes the sum of l_{000} and l_{001}. Agresti defines the saturated model as follows: $l_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jk}^{Y_2Y_3} + \lambda_{ik}^{Y_1Y_3} + \lambda_{ijk}^{Y_1Y_2Y_3}$, with the following model parameters:

λ = grand average = the overall mean of the log of the expected frequencies

$$= \frac{l_{+++}}{8} = (\sum_i \sum_j \sum_k l_{ijk})/8$$

$\lambda_i^{Y_1}$ = the main effect of Y_1 = (average over Y_1 – grand average)

$$= \frac{l_{i++}}{4} - \lambda = \frac{l_{i++}}{4} - \frac{l_{+++}}{8}$$

The main effects of Y_2 ($\lambda_j^{Y_2}$) and Y_3 ($\lambda_k^{Y_3}$) have analogous equations. Since the main effects are deviations from the mean, $\sum_i \lambda_i^{Y_1} = \sum_j \lambda_j^{Y_2} = \sum_k \lambda_k^{Y_3} = 0$. In other words, $\lambda_0^{Y_1} + \lambda_1^{Y_1} = 0$ and the same is true for Y_2 and Y_3.

The two-factor interaction terms take on the form:

$$\lambda_{ij}^{Y_1Y_2} = \frac{l_{ij+}}{2} - \lambda_i^{Y_1} - \lambda_j^{Y_2} + \lambda = \frac{l_{ij+}}{2} - \frac{l_{i++}}{4} - \frac{l_{+j+}}{4} + \frac{l_{+++}}{8}$$

$\lambda_{ik}^{Y_1Y_3}$ and $\lambda_{jk}^{Y_2Y_3}$ have analogous equations.

Lastly, the three-factor interaction term:

$$\lambda_{ijk}^{Y_1Y_2Y_3} = l_{ijk} - \lambda_{ij}^{Y_1Y_2} - \lambda_{jk}^{Y_2Y_3} - \lambda_{ik}^{Y_1Y_3} + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} - \lambda$$

$$= l_{ijk} - \frac{l_{ij+}}{2} - \frac{l_{+jk}}{2} - \frac{l_{i+k}}{2} + \frac{3l_{i++}}{4} + \frac{3l_{+j+}}{4} + \frac{3l_{++k}}{4} - \frac{7l_{+++}}{8}.$$

Now we need to relate $\ln e_{ijk}$ (Agresti's model) to $\ln \pi_{ijk}$ (Bishop, Fienberg and Holland's model). We know that $\pi_{ijk} = \frac{e_{ijk}}{e_{+++}}$. Therefore, $\ln \pi_{ijk} = \ln(\frac{e_{ijk}}{e_{+++}}) = \ln(e_{ijk}) - \ln(e_{+++})$. The two models differ by the logarithm of the overall mean of the expected values.

(Page 36)

Let us take the variables depression, NIDDM and stress to illustrate the relationship between the equation $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$ and the equation of the hierarchical loglinear model of independence ($\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3}$). Fitting the independence model to the data yields the following expected values (taken from Table 4b):

Depression	NIDDM	
	0	**1**
0	25.11	4.89
1	60.95	11.74

Stress = 0

Depression	NIDDM	
	0	**1**
0	66.95	13.05
1	160.68	31.32

Stress = 1

Taking the natural logarithm of the expected values yields:

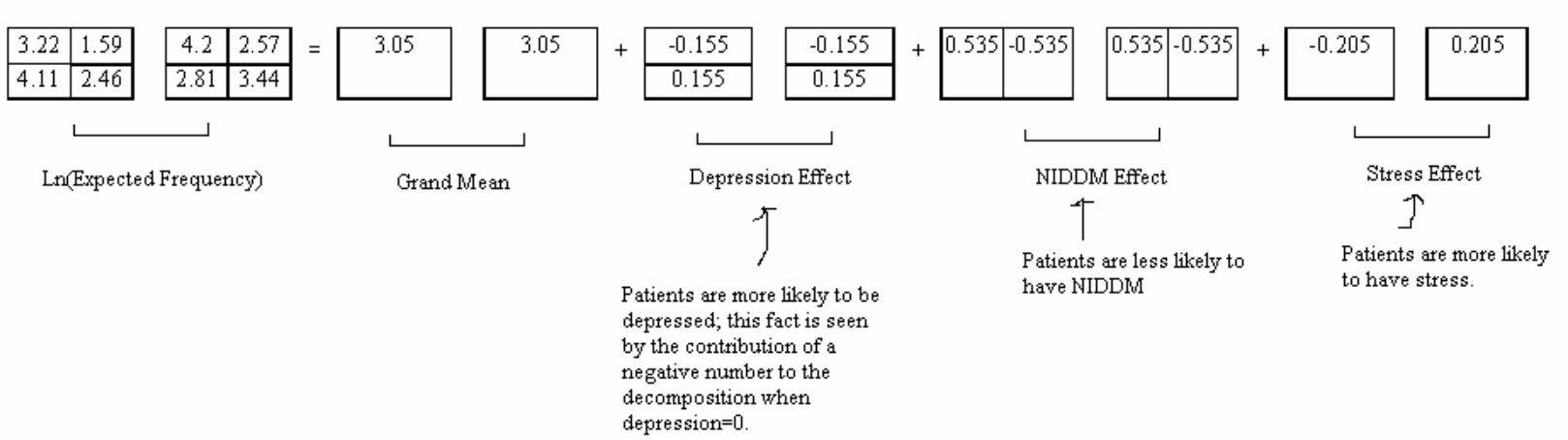
Depression	NIDDM	
	0	**1**
0	3.22	1.59
1	4.11	2.46

Stress = 0

Depression	NIDDM	
	0	**1**
0	4.2	2.57
1	2.81	3.44

Stress = 1

Next, we decompose the data according to Agresti's model.



Let us look at one specific cell to see how the equations $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$ and $\ln \pi_{ijk} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3}$ relate to each other.

We can see from the decomposition that $l_{000} = 3.05 + (-.155) + .535 + (-.205) = 3.22$. Furthermore, $l_{000} = \ln \pi_{000} + \ln(e_{+++})$. To check this statement, we first can find $\ln \pi_{000}$. According to the independence model, $\pi_{000} = \pi_{0++}\pi_{+0+}\pi_{+0+}$. Thus we have:

$$\ln \pi_{000} = \ln \pi_{0++} + \ln \pi_{+0+} + \ln \pi_{++0} = \ln(\frac{110}{374}) + \ln(\frac{314}{374}) + \ln(\frac{103}{374}) = (-1.22) + (-.176) + (-1.29) = -2.7$$

Now, $\ln \pi_{000} + \ln(e_{+++}) = (-2.7) + \ln(374) = (-2.7) + 5.9 = 3.22 = l_{000}$.

5.1 Overview

(Page 54)

The 2 x 2 contingency table with a partially classified margin for the hypothetical example can be seen below:

Pre \ Post	0	1	
0	17	11	*28*
1	22	50	*72*
	39	*61*	*100*

Miss = 0

Pre \ Post	0	1	
0	?	?	*42*
1	?	?	*58*
	?	*?*	*100*

Miss = 1

From the observed data, we have 39 patients who are asthma-free post-operatively and 61 patients who have asthma two years after the operation. Given that patients without asthma lost 100 pounds on average and that patients with asthma lost 200 pounds on average, we find that the unadjusted weight loss estimate is 161 pounds ([(39)(100)+(61)(200)]/100).

However, assuming the model {PrePost,PostMiss}, we impute missing values, resulting in the following table:

Pre \ Post	0	1	
0	17	11	*28*
1	22	50	*72*
	39	*61*	*100*

Miss = 0

Pre \ Post	0	1	
0	40.9	1.1	*42*
1	52.9	5.1	*58*
	93.8	*6.2*	*100*

Miss = 1

The adjusted weight loss estimate is now 134 pounds ([(39+93.8)(100)+(61+6.2)(200)]/200).

5.3 Imputing Values

(Page 59)

Proof:

$$\frac{f_{10}}{f_{00}} = \frac{\hat{m}_{10}}{\hat{m}_{00}} \text{ and } \frac{f_{11}}{f_{01}} = \frac{\hat{m}_{11}}{\hat{m}_{01}}$$

$$\Leftrightarrow (\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}}) = (\frac{\hat{m}_{10}}{\hat{m}_{00}}) - (\frac{\hat{m}_{11}}{\hat{m}_{01}})$$

$$\Leftrightarrow \hat{m}_{00}[(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})] = \hat{m}_{10} - \hat{m}_{00}(\frac{\hat{m}_{11}}{\hat{m}_{01}})$$

$$\Leftrightarrow \hat{m}_{00} = \frac{\hat{m}_{10} - \hat{m}_{00}(\frac{\hat{m}_{11}}{\hat{m}_{01}})}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})}$$

$$\Leftrightarrow \hat{m}_{00} = \frac{\hat{m}_{10} - \hat{m}_{00}(\frac{f_{11}}{f_{01}})}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})}$$

$$\Leftrightarrow \hat{m}_{00} = \frac{(m_1 - \hat{m}_{11}) - (m_0 - \hat{m}_{01})(\frac{f_{11}}{f_{01}})}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})}$$

$$\Leftrightarrow \hat{m}_{00} = \frac{(m_1 - \hat{m}_{11}) - (m_0)(\frac{f_{11}}{f_{01}}) + (\hat{m}_{01})(\frac{f_{11}}{f_{01}})}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})}$$

$$\Leftrightarrow \hat{m}_{00} = \frac{m_1 - \cancel{\hat{m}_{11}} - (m_0)(\frac{f_{11}}{f_{01}}) + \cancel{\hat{m}_{11}}}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})}$$

$$\Leftrightarrow \hat{m}_{00} = \frac{m_1 - (m_0)(\frac{f_{11}}{f_{01}})}{(\frac{f_{10}}{f_{00}}) - (\frac{f_{11}}{f_{01}})}$$

R Computer Code[26]

5.6 Models Involving Two Predictor Variables

The EM algorithm used to impute missing values for the 4 x 4 x 2 contingency table:

```
# Notation (follows Baker & Laird, 1988)
###################################
#        Z = cell counts and expected cell counts, given M
#        M = mles for cell counts, given the model and expected
cell counts Z
#        x = 1, 2, ..., I:  pre-op categories; I = number of
categories
#        y = 1, 2, ..., J:  post-op categories; J = number of
categories
#        r = 1, 2:                 Response:  1 = Yes, 2 = No
#     Z[x,y,1] are observed (pre/po, for those who respond)
#        Z[x,+,2] are observed (pre-op totals, for those who
don't return)
#
# Data
#####################################################################
#####
I <- 4
J <- 4
Z <- array(0,dim=c(I,J,2),dimnames=c("Pre","Post","Return"))
#
# Pre-op counts for those who return
#
Z[1,,1] <- c(37,0,0,0)
Z[2,,1] <- c(46,89,0,0)
Z[3,,1] <- c(2,1,2,0)
Z[4,,1] <- c(3,6,1,4)
#
# Pre-op totals for those who don't return
#
Z2 <- c(102,246,8,18)                # Pre-op totals for all
Zx_2 <- Z2 - rowSums(Z[,,1])        # Pre-op totals for those
who don't return
#
# M Step:  Function to update mles Mxyr for cell counts
###################
MStep <- function(Z){
   Zxy <- apply(Z,c(1,2),sum)
   Zyr <- apply(Z,c(2,3),sum)
   Zy <- apply(Z,2,sum)
   Zxy_ <- outer(Zxy,rep(1,2))
```

[26] Much thanks to George Cobb for writing this R code!

```
   Z_yr <- outer(rep(1,J),Zyr)
   Z_y_ <- outer(outer(rep(1,J),Zy),rep(1,2))
   Mxyr <- Zxy_*Z_yr/Z_y_
   return(Mxyr)
   }
#
# E Step:  Function to update expected missing values
######################
EStep <- function(M,Z){
   Zx2 <- apply(Z[,,2],1,sum)
   Zx_2 <- outer(Zx2,rep(1,J))
   Mx2 <- apply(M[,,2],1,sum)
   Mx_2 <- outer(Mx2,rep(1,J))
   Z[,,2] <- Zx_2*M[,,2]/Mx_2
   return(Z)
   }
#
# Starting values
############################################################
ZNew <- array(0,dim=c(I,J,2),dimnames=c("Pre","Post","Return"))
ZNew[,,1] <- Z[,,1]
ZNew[,,2] <- outer(Zx_2,rep(1,J))/J
diff <- 1                          # Starting value for change in
expected counts
Steps <- 0                          # Number of steps
eps <- .001                          # Convergence criterion
#
# Iterate M and E steps
#####################################################
while(diff > eps){
   Steps <- Steps + 1
   ZOld <- ZNew
   MNew <- MStep(ZOld)
   ZNew <- EStep(MNew,ZOld)
   diff <- sum(abs(ZOld-ZNew))
   }
# Summarize results
#########################################################
Steps
diff
round(MNew/sum(MNew),digits=2)
round(ZNew,digits=2)
round(MNew,digits=2)
```

References

Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley & Sons.

American Diabetes Association. "Stress." *American Diabetes Association.* 11 Feb. 2007 <http://www.diabetes.org/type-2-diabetes.jsp>.

Baker, S., and Laird, N. (1988), "Analysis for Categorical Variables with Nonresponse," *American Statistical Association,* 83, 62-69.

Bishop, Y., Fienberg, S., and Holland, P. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MA: The MIT Press.

Center for Disease Control and Prevention. "Overweight and Obesity." *Center for Disease Control and Prevention.* November 2006. Department of Health and Human Services. 5 Dec. 2006 <http://www.cdc.gov/nccdphp/dnpa/obesity/>.

Center for Disease Control and Prevention. "Health, United States, 2006." *Department of Health and Human Services*. 2006. National Center for Health Statistics. 5 Dec. 2006. <http://www.cdc.gov/nchs/data/hus/hus06.pdf>.

Cobb, G.W. (1998), *Introduction to Design and Analysis of Experiments*, New York: Springer-Verlag New York.

Fienberg, S.E. (2000), "Contingency Tables and Log-Linear Models: Basic Results and New Developments," *Journal of the American Statistical Association*, 95, 643-647.

Fingleton, B. (1984), *Models of Category Counts*, Cambridge, Great Britain: Cambridge University Press, 1984.

Gerberding, Julie Louise. "Chronic Disease Prevention." *Center for Disease Control and Prevention*. May 2006. Department of Health and Human Services. 5 Dec. 2006 <http://www.cdc.gov/nccdphp/publications/aag/cvh.htm>.

Goodman, L.A. (1964), "Simple Methods for Analyzing Three-Factor Interaction in Contingency Tables," *Journal of the American Statistical Association*, 59, 319-352.

Greenlees, J.S., Reece, W.S., and Zieschang, K.D. (1982), "Imputation of Missing Values when the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association*, 77, 251-261.

Horton, N.J., and Kleinman, K.P. (2007), "Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models," *The American Statistician*, 61, 79-90.

Horton, N.J., and Switzer, S.S. (2005), "Statistical Methods in the Journal," *New England Journal of Medicine*, 353, 1977-1979.

Kassel, Karen. "Roux-en-Y Gastric Bypass." *Healthy Living*. November 2006. North Shore Medical Center. 18 Dec. 2006 <http://healthlibrary.epnet.com/GetContent.aspx?token=c969dc7d-0aa7-43de-ba12-bfeedab0944f&chunkiid=96212>.

Kassel, Karen. "Vertical Banded Gastroplasty." *Health Library*. November 2006. North Shore Medical Center, 18 Dec. 2006 <http://healthlibrary.epnet.com/GetContent.aspx?token=c969dc7d-0aa7-43de-ba12-bfeedab0944f&chunkiid=96213>.

Klockhoff, H., Naslund, I., and Jones, A.W. (2002), "Faster Absorption of Ethanol and Higher Peak Concentration in Women after Gastric Bypass Surgery," *British Journal of Clinical Pharmacology*, 54, 587-591.

Little, R., and Rubin, D. (1987), *Statistical Analysis with Missing Data*, New York: John Wiley & Sons.

Little, R., and Rubin, D. (2002), *Statistical Analysis with Missing Data* (2nd ed.), New York: John Wiley & Sons.