Lanqin Zhao

Brandon Wong

Yun Kun Zhang

An Investigation into The Relationship Between Heart Disease and Hearing Impairment, NHANES 2003-2004

Introduction

According to the American Academy of Audiology, hearing loss is the third most common health problem in the United States, after heart disease and arthritis.[1] As of 2013, roughly 1.1 billion people are afflicted with hearing loss worldwide.[2] Hearing loss can be attributed to genetics, aging, medical conditions, and a variety of environmental factors, such as exposure to loud noises. Hearing loss can be qualified by varying levels of severity, culminating in deafness, which is defined as the degree of hearing loss such that a person is unable to comprehend speech even in the presence of amplification.

Many cases are progressive and irreversible (mostly due to aging or noise exposure), but it is estimated that roughly half of all cases of hearing loss are preventable. Patients who suffer from hearing loss but not deafness can sometimes be treated surgically or through the use of assistive devices, such as hearing aids. Of the three most common health problems in the United States, research shows that there may be a relationship between two of them - heart disease and hearing loss.

Hearing loss can be divided into multiple categories based on the location of the source of impairment within the ear. Hearing loss caused by a problem in the external or middle ear is called conductive hearing loss. The impairment is caused by the conduction of sound to the cochlea, the spiral cavity of the inner ear. Conductive hearing loss is typically medically correctable. Hearing loss that is caused by a problem in the inner ear or along the auditory

---

[1] http://www.audiology.org/sites/default/files/AAM%20Poster%20(40x60).pdf
[2] https://www.ncbi.nlm.nih.gov/pubmed/26063472

nerve is called sensorineural hearing loss. The impairment is caused by the inability of the cochlea to sense the sound or the inability of the auditory nerve to relay the message to the brain. Sensorineural hearing loss is typically not medically correctable. The effects of sensorineural hearing loss can be somewhat mitigated by the use of assistive devices, but the damage cannot be fully reversed.[3] The American Heart Association has published research linking dilated cardiomyopathy (DCM) and sensorineural hearing loss. Multiple loci have been located that, when mutated, cause dilated cardiomyopathy and sensorineural hearing loss.[4] This discovery sparked a deeper interest in the relationship between two seemingly unrelated conditions.

National Health and Nutrition Examination Survey (NHANES) 2003-2004

Our analysis uses data from the results of the 2003-2004 National Health and Nutrition Examination Survey (NHANES) administered by the Center for Disease Control and Prevention (CDC). NHANES is a population-based survey used to assess the general health and nutritional status of adults and children in America. The survey is unique in that it combines interviews and physical examinations, but for the purposes of this analysis, the data is only concerned with the results of the surveys. It was born out of the National Health Survey Act of 1956 and has surveyed over 190,000 people since 1960. Participants for NHANES were selected by multistage stratified sampling. All counties in the U.S. were divided into 15 groups based on their characteristics. One county was selected from each group and within each county 20-24 smaller strata were formed. About 30 households were chosen from these strata for NHANES interviewers to visit, and subsequently a computer algorithm randomly selected all, some, or none of the interviewees to be incorporated into the final data set.

This analysis required a merger of responses from three questionnaires issued during the 2003-2004 period - “Demographic Variables & Sample Weights”, “Audiometry”, and “Medical

[3] https://www.cdc.gov/nchs/data/nhanes/nhanes_09_10/audiometry_09.pdf
[4] http://circ.ahajournals.org/content/101/15/1812.abstract

Conditions” - all of which are formatted in SAS. Each data set was individually sorted by respondent sequence number before being merged to ensure that the same participants were present in all three questionnaires. This resulted in a sample size of 5040. Any instances in which a participant failed to respond or responded with “don’t know” were accordingly removed.

NHANES 2003-2004 Questionnaires: Response Variable and Explanatory Variables

The “Demographic Variables & Sample Weights” questionnaire includes questions outlining the basic demographic information of the participants, such as age, level of education, marital status, military veteran status, which can be used as explanatory variables. The age of each participant was listed in months and recoded to the more familiar format of years. A participant’s status as a military veteran was listed as a binary variable - either the participant served in the American armed forces or he/she did not. The demographics survey had 21 possible options to describe a participant’s level of attained education; this variable was recoded to have three outcomes - (1) - less than high school, (2) - high school diploma (including GED), (3) - more than high school. The survey had 6 possible options to describe a participant’s marital status - (1) - married, (2) - widowed, (3) - divorced, (4) - separated, (5) - never married, (6) - living with partner. The participants were given various ranges to describe their annual family income.

The “Medical Conditions” questionnaire asks a series of questions to create a comprehensive picture of a participant’s medical history, especially regarding chronic conditions. In particular, this analysis is concerned with chronic heart disease, specifically congestive heart failure, coronary heart disease, angina, and myocardial infarctions (also known as heart attacks). The participants were asked “Has a doctor ever told you that you had congestive heart failure?”, for example, and this format was repeated for each condition. These are binary categorical variables, also called dichotomous variables; there are only two possible responses - (1) - yes or (2) - no. The chronic conditions considered for this analysis are congestive heart failure, coronary heart disease, angina (also called angina pectoris), and emphysema. Congestive heart failure occurs when a person's heart does not pump blood as efficiently as it should. Coronary heart disease is the most common form of heart disease and can occur when a person's arteries are damaged or blocked, leading to poor blood flow to the heart. Angina is characterized by a feeling of pain or discomfort in the chest and can be a symptom of an underlying heart problem. Emphysema, also called chronic obstructive pulmonary disease (COPD), is a group of lung diseases that are characterized by long-term poor airflow, generally caused by deterioration of the lungs. Smoking is the leading cause of emphysema. This condition is unlike the others in that it is not directly an affliction of the heart, but it would be interesting to see its significance during the model building and model selection.

In contrast to chronic conditions, each participant was also asked whether he/she had ever had a heart attack before. This would, of course, be a binary variable. This question was repeated to discern whether each participant had ever had a stroke in the past. For the purposes of this analysis, due to lack of information provided in the NHANES questionnaire results, a distinction was not made as to how many heart attacks or strokes a participant had suffered. Whether a participant has had one attack or multiple, this analysis is only concerned as to whether a participant's medical history contains any single instance of a heart attack or stroke.

The response variable, general hearing condition, is taken from the "Audiometry" questionnaire. The participant was asked to describe the quality of his/her hearing and was asked the following questions - "Which statement best describes your hearing without a hearing aid? Would you say your hearing is good, that you have a little trouble, a lot of trouble or are deaf?" This response variable is an ordinal categorical variable, with its responses coded as (1) - good, (2) - little trouble, (3) lot of trouble, (4) - deaf.

In addition, two explanatory variables were taken from the "Audiometry" questionnaire. Each participant was asked, "Outside of work have you ever been exposed to firearms noise for an average of at least once a month for a year?" This variable had a binary response. Similarly, each participant was asked, “Outside of work, have you ever been exposed to other types of loud noise such as noise from power tools or loud music, for an average of at least once a month for a year? By loud noise, I mean noise so loud that you had to speak in a raised voice to be heard?” This variable also had a binary response.

In particular, these two variables highlight the somewhat unreliable nature of survey questionnaires. Surveys are often used in what are called retrospective studies, which are a form of observational study that investigate the effects of events in the past. In contrast with a prospective study, statisticians conducting a retrospective study do not follow up with participants after the initial data collection to analyze any changes that have transpired in the intervening time period. In the case of these two variables from the “Audiometry” questionnaire, it is distinctly possible that there may be a level of bias because the questions ask the participant to estimate exposure to loud noise monthly over the course of a year. Psychologically, suffering a heart attack or contracting a serious heart condition are major life events and would presumably not be difficult to recall, whereas exposure to loud noise is relatively banal and possibly more difficult to recall. Therefore, it is possible that these variables may not give a completely accurate picture of each participant’s exposure to loud noise. However, due to the constraints of the “Audiometry” questionnaire and the nature of retrospective studies, they will suffice for this analysis. Also, contrary to popular belief, a single instance of excessive noise can lead to permanent hearing damage.[5] So, like heart attacks, it is possible that repeated cases may not be pertinent for the purposes of this analysis.

The Cumulative Logit Model: Background

An ordinal response variable with more than two responses makes a cumulative logit model the ideal choice for this analysis. Unlike linear regression, which predicts a continuous outcome, logistic regression specifically predicts a response variable that has a finite number of values. Binary logistic regression predicts a response variable that only has two possible values

[5] http://www.audiology.org/sites/default/files/AAM%20Poster%20(40x60).pdf

- usually affirmative or negative (i.e. yes or no) - whereas ordinal logistic regression predicts a response variable with more than two possible outcomes that are ordered. Ordinal categorical variables differ from nominal categorical variables in that the latter do not have any natural order. In contrast with ordinary linear regression, which has its response written as a single variable (such as Y), the response for logistic regression is the log odds function, *logit p/(1-p)*, where the function's parameter represents a probability *p*.

A cumulative probability for a response Y is the probability that Y falls at or below a particular value. The cumulative probabilities reflect the ordering of the variable, with $P(Y \leq 1) \leq P(Y \leq 2) \leq \ldots \leq P(Y \leq J) = 1$. Cumulative probability models do not use the final $P(Y \leq J)$ because it tautologically equals to 1, so J-1 is used. Therefore, the model for a cumulative logistic regression for an explanatory variable x is:

$$\text{logit}\,[P(Y \leq j)] = \alpha_j + \beta_j x, \qquad j=1,\ldots,J-1$$

where α_j is the intercept and β_j is the effect of x on the log odds of the response in category j or below.

The data was fitted to a cumulative logit model in SAS. To recap, there were 12 explanatory variables: age, military veteran status, marital status, education, annual family income, PIR ratio, congestive heart failure, coronary heart disease, angina, heart attack, firearm noise exposure (outside of work), and loud noise exposure (outside of work). After merging the three datasets for the appropriate NHANES questionnaires ("Demographics & Sample Weights", "Medical Conditions", "Audiometry"), these variables were fitted to a cumulative logit model with general hearing condition from the "Audiology" questionnaire as the response variable. In SAS, the proc logistic function was used with a “clogit” (cumulative logit) link. The reference level for the response variable, general hearing condition, was (1) - “good” hearing.[6]

When constructing a cumulative logit model, it is important to note whether it is also a proportional odds model. A proportional odds model will have the following form:

$$\text{logit}\,[P(Y \leq j)] = \alpha_j + \beta_1 X_1 + \cdots + \beta_p X_p, \qquad j=1,\ldots,J-1$$

The ? term does not have the *j* subscript because the proportional odds model assumes that the effect of x is identical for all J-1 cumulative logits. If this is not the case, different sets of coefficients would be required to describe the relationship between each pair of outcome groups. This is not ideal. If a model fails the proportional odds assumption, the alternative models are:

1) Baseline category logit
2) Logistic regression

Logistic regression is the least ideal and should only be used if other options have been exhausted. To convert a model from a cumulative logit model to a logistic regression model, the response variable would have to be restructured from an ordinal categorical variable to a binary variable. This is not ideal because the ordered nature of the variable allows for a more comprehensive illustration of the relationship among the variables than could be achieved with just two responses. (In the case of this analysis, the response could be broadly divided into “mostly good” hearing and “poor” hearing.) A baseline category logit model takes the form:

$$\log(\pi_j/\pi_J) = \alpha_j + ?_j x \qquad j=1,\ldots,J-1$$

Baseline category logit models are typically used for models with nominal response variables. One of the nominal responses is chosen as a baseline category and each equation would describe the log-odds relationship between another of the nominal responses and the baseline

---

[6] For the reference level of each individual explanatory variable, please refer to the SAS code in the appendix.

response. Again, this is also unideal because it ignores the ordered nature of the original variable and treats each ordered response as nominal.

Whether a model passes the proportional odds assumption is determined by the p-value generated in the SAS output. In statistical hypothesis testing, if the p-value is less than the designated significance level, alpha, then the null hypothesis is rejected and the alternative hypothesis (also called the research hypothesis) is accepted. The significance level alpha is the probability of rejecting the null hypothesis given that it is true. In the case of this analysis, the significance level is $\alpha=0.05$. If the p-value for the proportional odds test is less than 0.05, the null hypothesis is rejected and the alternative hypothesis is accepted. If the p-value is greater than 0.05, the null hypothesis fails to be rejected and is therefore accepted. For the proportional odds assumption, the following set of hypotheses is used:

H_0: The model passes the proportional odds assumption.

H_a:The model does not pass the proportional odds assumption.

The p-value for the first model, using stepwise selection[7] in SAS to choose its variables, had a p-value of 0.0028 and a chi-square statistic of 28.4179 with 11 degrees of freedom. This p-value is less than 0.05. Therefore the null hypothesis, H_0, is rejected and the model fails the proportional odds assumption.

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
28.4179	11	0.0028

[7] Stepwise selection is a semi-automated process of model building that successively adds or removes variables based solely on the t-statistics of their estimated coefficients.

Subsequently, because the model failed the proportional odds assumption, the next alternative, a baseline category logit model was applied to the data. A goodness-of-fit statistic tests the following hypothesis:

*H*0: the model fits;

HA: the model does not fit.

For baseline category logit model, both Deviance and Pearson Statistics were examined. P-value under Deviance test was 0.9969 while P-value under Pearson test was less than 0.05. Under Pearson test and significance level of α=0.05, rejected *H*0 and accepted the alternative that the baseline category logit model does not fit.

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2711.2938	2916	0.9298	0.9969
Pearson	3279.2838	2916	1.1246	<.0001

Given that baseline category logit model did not fit, a simpler model, logistic regression, was attempted to fit the data. The response variable, General condition of hearing, was reclassified into Good and Trouble. However, again the model didn't pass the Goodness-of-Fit Test.

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	1870.7270	1460	1.2813	<.0001
Pearson	1526.1513	1460	1.0453	0.1115

Final Model:

In the very first cumulative logit model under the Stepwise selection method, military veteran status, firearms noise, congestive heart failure and coronary heart disease were identified to be most significant.

Summary of Stepwise Selection								
	Effect							
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	Veteran_Military		1	1	133.3825		<.0001	Veteran/Military Status
2	Congestive		1	2	73.6261		<.0001	Ever told had congestive heart failure
3	Firearm_Noise_Exp		1	3	32.6361		<.0001	Firearm noise exposure outside work
4	Coronary		1	4	27.7195		<.0001	Ever told you had coronary heart disease
5	Heart_Attack		1	5	11.0954		0.0009	Ever told you had heart attack
6	Loud_Noise_Exp		1	6	8.4642		0.0036	Loud noise exposure outside work
7	Angina		1	7	5.9046		0.0151	Ever told you had angina/angina pectoris

A new cumulative logit model with these four variables was built. Both proportional odds assumption and Goodness-of-Fit test passed. It should be noted the SAS reversed order of the values in the response variable, hearing condition with 1 being lot of trouble, 2 being little trouble, and 3 being good.

Response Profile		
Ordered Value	Hearing_Condition	Total Frequency
1	3	287
2	2	1110
3	1	3643

The model is written as

logit [P(Y ≤ 2)] = -1.2203 + 0.6245 (Firearm_Noise_exposure) + 0.8838 (Congestive) + 0.6970 (Coronary) + 0.71718 (Veteran Military Status)

logit [P(Y ≤ 3)] = -3.1452 + 0.6245 (Firearm_Noise_exposure) + 0.8838 (Congestive) + 0.6970 (Coronary) + 0.71718 (Veteran Military Status)

Both Proportional Odds Assumption and Goodness of Fit Test passed.

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
1.7752	4	0.7770

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	30.4855	26	1.1725	0.2480
Pearson	30.0404	26	1.1554	0.2659

Interpretation of the model:

1) the odds of hearing condition is in the worse direction rather than better direction for with firearm noise exposure is about 1.8647 times that without firearm noise exposure, given other variables fixed. 2) the odds of hearing condition is in the worse direction rather than better direction for with congestive heart failure is about 2.42 times that without congestive heart failure, given other variables fixed. 3) the odds of hearing condition is in the worse direction rather than better direction for with coronary heart disease is about 2.008 times that without coronary heart disease, given other variables fixed. 4) the odds of hearing condition is in the worse direction rather than better direction for with veteran military status is about 2.050 times that without being a veteran, given other variables fixed.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Firearm_Noise_Exp 1 vs 2	1.867	1.508	2.313
Congestive 1 vs 2	2.420	1.805	3.245
Coronary 1 vs 2	2.008	1.548	2.605
Veteran_Military 1 vs 2	2.050	1.745	2.408

According to CDC, Veterans were 30% more likely to have severely hearing damage than nonveterans after adjusting for age and current occupation, and veterans who served in the United States or overseas during September 2001--March 2010, the era of overseas contingency operations (including Operations Enduring Freedom and Iraqi Freedom), were four times more likely than nonveterans to have severely hearing damage[8].

Heart and Hearing connection can be explained at least two ways: 1) many studies have been show that multiple genetic loci, when mutated, cause dilated cardiomyopathy and sensorineural hearing loss[9]. 2) An unhealthy cardiovascular system results in an inadequate blood flow and trauma to the blood vessels of the inner ear and thus contributes to hearing loss[10].

If there is enough time, more variables can be explored and included in the model, such as different diseases, occupation, hearing treatment and etc. The interactions between different disease variables can be considered as well.

8: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6028a4.htm
9: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6028a4.htm
10: http://circ.ahajournals.org/content/101/15/1812.abstract

Appendix (SAS code):

```
/*input data*/
libname NH "\\dtchyb-caod008\C_MTGCRT_Users\wzhang3\My
Documents\Wilson\DATA";
libname XP xport "C:\NHANES\TEMP\AUQ_C.XPT";
proc copy in=XP out=NH;
run;

libname XP xport "C:\NHANES\TEMP\DEMO_C.XPT";
proc copy in=XP out=NH;
run;

libname XP xport "C:\NHANES\TEMP\MCQ_C.XPT";
proc copy in=XP out=NH;
run;

/*select variables of interest*/

proc contents data=NH.AUQ_C varnum;
run;

data nh.auq;
set NH.AUQ_C;
keep SEQN AUQ130 AUQ210 AUQ230;
run;

proc sort data=NH.AUQ;
by  SEQN;
run;

data nh.MCQ;
set NH.MCQ_C;
keep SEQN MCQ160B MCQ160C MCQ160D MCQ160E   ;
if MCQ160B;
run;

proc sort data=NH.mcq;
by  SEQN;
run;

data nh.DEMO;
set NH.DEMO_C;
keep SEQN RIDAGEMN DMQMILIT DMDEDUC DMDMARTL INDFMINC INDFMPIR ;
run;

proc sort data=NH.demo;
by  SEQN;
run;

data nh.hearing0 ;
merge  nh.Auq(in =a )  nh.Demo(in =b )  nh.Mcq(in =c);
by seqn;
if a and b and c;
run;
```

```
proc means data=nh.hearing0 N nmissing min max  ;
run;

proc print data=nh.hearing1 (obs=100);
run;
proc freq data=nh.hearing0;
tables AUQ130;
run;

proc sgplot data=nh.hearing1;
scatter y= auq130 x=Age;
run;
/*clean data*/
data nh.hearing1;
set nh.hearing0;
if AUQ130=4       then AUQ130=3; /* deaf cases are too few to use*/
if AUQ130=9       then AUQ130=1;
if AUQ210 ne 1            then AUQ210=2;
if AUQ230 ne 1            then AUQ230=2;
if MCQ160B ne 1           then MCQ160B=2;
if MCQ160C  ne 1           then MCQ160C=2;
if MCQ160D  ne 1           then MCQ160D=2;
if MCQ160E  ne 1           then MCQ160E=2;
if MCQ160F  ne 1           then MCQ160F=2;
if MCQ160G   ne 1           then MCQ160G=2;
if MCQ160K   ne 1           then MCQ160K=2;

/*recode age*/
if RIDAGEMN=.                then RIDAGEMN=500;
Age= RIDAGEMN/12;
if Age<=30       then Age1=0;
else if Age<=40       then Age1=1;
else if Age<=50       then Age1=2;
else if Age<=60       then Age1=3;
else if Age<=65       then Age1=4;
else if Age<=70       then Age1=5;
else if Age<=75      then Age1=6;
else Age1=7;
/*based on effect of age1, recode age again by combine levels that have
similar effect*/
if Age<=40      then Age2=0;
else if Age<=65  then Age2=1;
else if Age<=75  then Age2=2;
else Age2=3;
/*recode marital*/
/*based on effect of Marital, recode again by combine levels that have
similar effect*/
if DMDMARTL=1 or DMDMARTL=2      then Marital1=DMDMARTL;
else if DMDMARTL=5               then Marital1=3;
else  Marital1=4;
/*recode Famlily_income */
if INDFMINC<6                then Family_Income=1;/*lower than 25k*/
else if INDFMINC<9           then Family_Income=2;/*lower than 55k*/
else if INDFMINC<12          then Family_Income=3;
else Family_Income=1;

if DMQMILIT ne 1           then DMQMILIT=2;
```

```
if DMDEDUC not in (1,2,3)  then DMDEDUC=1;
if INDFMPIR=.              then INDFMPIR=2.5;
if INDFMPIR<=1             then FamilyIncome_PIR=1;
else if INDFMPIR<=2             then FamilyIncome_PIR=2;
else if INDFMPIR<=3            then FamilyIncome_PIR=3;
else if INDFMPIR<=4             then FamilyIncome_PIR=4;
else FamilyIncome_PIR=5;
run;

proc print data=nh.hearing1 (obs=100);
run;
proc means data=nh.hearing1 N Nmissing min max  ;
run;

/*Rename*/
data nh.hearing2;
set nh.hearing1;
RENAME
       AUQ130 = Hearing_Condition
         AUQ210= Firearm_Noise_Exp
         AUQ230 = Loud_Noise_Exp
         MCQ160B= Congestive
       MCQ160C= Coronary
       MCQ160D= Angina
       MCQ160E= Heart_Attack
       DMQMILIT= Veteran_Military
       DMDEDUC=Edu
       DMDMARTL=Marital
       INDFMPIR= Family_PIR
        ;
run;
proc print data=nh.hearing2 (obs=100);
run;


/*build models*/
/*since we have a few variables, use stepwise  select variables*/


PROC LOGISTIC DATA = nh.hearing2 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="0")
            Marital1(ref="1")
            Family_Income(ref="1")
         FamilyIncome_PIR(ref="1")
        /PARAM = REF;
MODEL  Hearing_Condition  = Firearm_Noise_Exp  Loud_Noise_Exp  Congestive
Coronary  Angina Heart_Attack
                            Veteran_Military Age2  Marital1 Edu Family_Income
FamilyIncome_PIR
```

```
                            /LINK= CLOGIT AGGREGATE SCALE = NONE
selection=stepwise ;
OUTPUT OUT = PROB PREDPROBS=I;
RUN; quit;
proc contents data=prob;
run;
proc freq data=prob ;
tables _FROM_*_INTO_ / nocol nopercent;
run;

PROC LOGISTIC DATA = nh.hearing2 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="0")
            Marital1(ref="1")
            Family_Income(ref="1")
         FamilyIncome_PIR(ref="1")
        /PARAM = REF;
MODEL  Hearing_Condition  = Firearm_Noise_Exp  Loud_Noise_Exp  Congestive
Coronary  Angina Heart_Attack
                            Veteran_Military
                            /LINK= CLOGIT AGGREGATE SCALE = NONE
selection=stepwise ;
OUTPUT OUT = PROB1 PREDPROBS=I;
RUN; quit;

PROC LOGISTIC DATA = nh.hearing2 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="0")
       Marital1(ref="1")
         Family_Income(ref="1")
        /PARAM = REF;
MODEL  Hearing_Condition  = Firearm_Noise_Exp  Congestive Coronary
                            Veteran_Military Angina Coronary*Angina
                            /LINK= CLOGIT AGGREGATE SCALE = NONE
selection=stepwise ;
OUTPUT OUT = PROB PREDPROBS=I;
RUN; quit;


PROC LOGISTIC DATA = nh.hearing2 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
```

```
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
/*       Stroke(ref="2")*/
/*       Emphysema(ref="2")*/
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="0")
       Marital1(ref="1")
         Family_Income(ref="1")
        /PARAM = REF;
MODEL  Hearing_Condition(ref="3")  = Firearm_Noise_Exp   Congestive Coronary
                             Veteran_Military
                             /LINK= CLOGIT AGGREGATE SCALE = NONE  ;
OUTPUT OUT = PROB PREDPROBS=I;
RUN; quit;


/* Proportional odds assumption fails.
baseline-category logits*/
PROC LOGISTIC DATA = nh.hearing2 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="0")
       Marital1(ref="1")
         Family_Income(ref="1")
         FamilyIncome_PIR(ref="1")
        /PARAM = REF;
MODEL  Hearing_Condition(ref="1")  = Firearm_Noise_Exp  Loud_Noise_Exp
Congestive Coronary  Angina Heart_Attack
                             Veteran_Military  Age2 Edu  Marital1
Family_Income FamilyIncome_PIR
                             /LINK= gLOGIT AGGREGATE SCALE = NONE
selection=stepwise ;
OUTPUT OUT = PROB PREDPROBS=I;
RUN; quit;


PROC LOGISTIC DATA = nh.hearing2 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="1")
       Marital1(ref="1")
         Family_Income(ref="1")
         Family_PIR(ref="1")
```

```
        /PARAM = REF;
MODEL  Hearing_Condition(ref="1")  = Firearm_Noise_Exp  Loud_Noise_Exp
Congestive Coronary  Angina Heart_Attack
                            Veteran_Military  Age2 Edu
                            /LINK= gLOGIT AGGREGATE SCALE = NONE
selection=stepwise;
OUTPUT OUT = PROB PREDPROBS=I;
RUN; quit;

data nh.hearing3;
set nh.hearing2;
if Hearing_Condition=1 then Hearing_Condition1=0;
else Hearing_Condition1=1;

PROC LOGISTIC DATA = nh.hearing3 order=data;
CLASS  Firearm_Noise_Exp(ref="2")
         Loud_Noise_Exp(ref="2")
         Congestive(ref="2")
       Coronary(ref="2")
       Angina(ref="2")
       Heart_Attack(ref="2")
       Veteran_Military(ref="2")
       Edu(ref="1")
        Age2(ref="0")
       Marital1(ref="1")
         Family_Income(ref="1")
         FamilyIncome_PIR(ref="1")
        /PARAM = REF;
MODEL  Hearing_Condition1(ref="0")  = Firearm_Noise_Exp  Loud_Noise_Exp
Congestive Coronary  Angina Heart_Attack
                            Veteran_Military  Age2 Edu  Marital1
Family_Income FamilyIncome_PIR
                            / AGGREGATE SCALE = NONE  selection=stepwise ;
OUTPUT OUT = PROB1 PREDPROBS=I;
RUN; quit;

proc freq data=prob1 ;
tables _FROM_*_INTO_ / nocol nopercent;
run;
```