Interaction of Risk Factors and the Incidence of Stroke

Exploratory Logistic Predictive Analytics

Results using a Simple 300k Random Sample

and 24 Risk Factors

Roger DeAngelis
7/6/2015

Preliminary

Abstract

A binary response of stroke was regressed on age,
atrial fibillation and hypertension accompanied by ischemic heart
disease, anemia and congestive heart failure.

A training dataset of 300k beneficiaries with a holdout of 100k
was used for development and verification respectively.

Due to the natural interpretaion some interactions
main effects were not included in the model

The logistic model provides a superior risk assessment
and can lead to better standards of care for potential
stroke victims.

Caveats

This exploratory analysis was done to demonstrate some algorithms and reports, useful when doing predictive analytics.

I am not a medical writer or skilled clinician so the results should be taken with skepticism.

It is possble to assess the interactions using a weighted logistc. The counts associated with each combination of the two dozen 'risks' would form the weights. Patients with no risk factors would be represented using 24 0s, 24 1s for patients with all risk factors. In this case a random sample would not be needed. The hold out sample could just be another year of medicare data. However adding other patient characteristics would be problematic

Dummification would be the next step, linearization was used here. Dummification generally improves the model.

I do plan on making all the SAS code available. The 'Gains' table code has already been posted

At this time of this presentation the model has not been verified using the holdout sample.

Agenda

- Overall Major Finding
- Major Non Specific Findings
- Major Findings Main Effects
- Major Findings Interactions
- Major Findings Zipcode Demographics
- Chi Square and Index Analysis of Character Variables
- Chi Square and Index Analysis of Numeric Variables
- Variables with highest Chi Square
- Levels with Greatest Indecies
- Gains or Lift table
- Model Betas
- Model Colinearity
- ROC
- Variable Contribution to highest probability
- Miscelaneous Diagnostics

Findings

Overall Major Finding

The logistic model has the potential for risk assessment and can lead to better standards of care for potential stroke victims.

Major Non Specific Findings

- Diabetes had the weakest association with stroke when compared to hypertension, IHD, COPD, CHF, Anemia, AMI and Age.
- Patients in long term care institutions had the highest incidence of stroke. This variable was not used in the model.
- Certain interactions of risk factors lead to an almost 30 percent probaility of a stroke in a given year.
- The model correctly placed almost 43 percent of the stroke victims in the top 10 percent of the gains table.
- The area under the receiving operating characterist curve was over 80 percent.
- The model is superior to the Medicare HCC score for stroke?

Major Findings Main Effects

- Patients in LTI have 5 times the rate of strokes(not used)
- Patients over 85 have twice the rate of strokes
- Patients over 85 have 3 times the rate of 70 year olds
- High HCC scores > 2 have 3 times the rate of strokes
- COPD or CKF have 3 times the rate of strokes
- Hypertension has twice the rate of strokes
- Anemia or IHD have 3 times the rate of strokes
- CHF have 4 times the rate of strokes
- Tobacco and Alcohol have twice the rate of strokes
- AFIB patients have 4 times the overall rate of Strokes
- IHD patients have 3 times the overall rate of Strokes

Major Findings Interactions

- **ANEMIA_AMI have a 6 times rate of strokes**
- **CHF_AMI have a 6 times rate of strokes**
- **ANEMIA_AFIB have a 6 times rate of strokes**
- **HYPERT_AMI have a 5 times rate of strokes**
- **CHF_AFIB have a 5 times rate of strokes**
- **IHD_AMI have a 4 times rate of strokes**
- **IHD_AFIB have 4 times rate of strokes**
- **HYPERT_AFIB have a 4 times rate of strokes**
- **CHF_ANEMIA have a 4 times rate of strokes**
- **IHD_ANEMIA have a 4 times rate of strokes**
- **HYPERT_CHF have a 4 times rate of strokes**
- **IHD_CHF have a 4 times rate of strokes**
- **HYPERT_ANEMIA have a 3 times rate of strokes**
- **HYPERT_IHD have a 3 times rate of strokes**

Major Findings Zipcode Demographics

- African American zipcodes have 20% increase strokes
- Large number of group homes have 10% increase in strokes
- Poor zipcodes have a 12% increase in strokes

Logistic Model Beta's positive signs are easy to explain effects

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
INTERCEPT	1	-5.4976	0.0309	31720	<.0001
_HYPERT_IHD	1	0.3847	0.0105	1353	<.0001
_AGE	1	0.4352	0.0208	436	<.0001
_AFIB	1	0.1987	0.00919	467	<.0001
_HYPERT_CHF	1	0.1243	0.00952	170	<.0001
_HYPERT_ANEMIA	1	0.2828	0.0105	729	<.0001

No Collinearity

Variable	DF	Pr > |t|	Variance Inflation
Intercept	1	<.0001	0
_HYPERT_IHD	1	<.0001	1.40852
_AGE	1	<.0001	1.06395
_AFIB	1	<.0001	1.15132
_HYPERT_CHF	1	<.0001	1.45781
_HYPERT_ANEMIA	1	<.0001	1.26584

Gains table or Lift Table

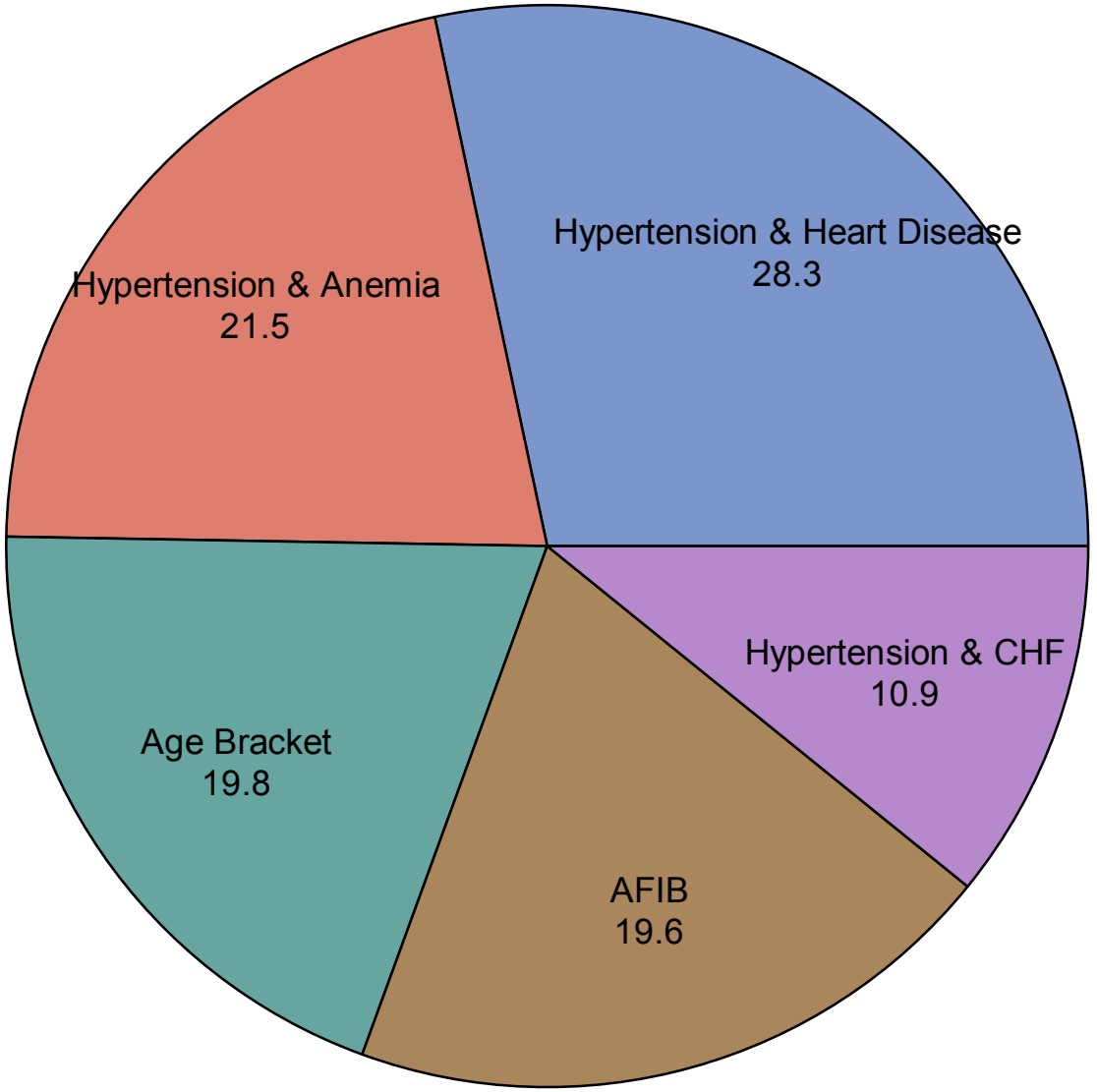
Gains or Lift Table

Index and Gains Table for Overall Stroke

The Overall Percentage of Patients with a Stroke is 2.5%
The index is the Percent Response in the Decile divided by the Overall Response.
For instance for the top 5% the Index=12.65/2.5 or 5.1
Patients in the top Decile are over 5 times as likely to have a Stroke
Average score times decile total is roughly equal to the Stroke

Probabilities			Decile Stats				Response and Non Response Counts				Percents				Index and Gain	
Quintile	Max Score	Mean Score	Min Score	Decile Count	Cum Decile Count	Cum Decile Percent	Healthy Patients	Cum Healthy Patients	Stroke	Cum Stroke	Stroke Percent of Total Stroke	Stroke Percent Total Response	Stroke Percent of Decile Total	Cum Stroke Percent of Decile Totals	Index	Cum Index
1	0.290	0.154	0.094	15141	15141	5	13226	13226	1915	1915	25.58	25.58	12.65	12.65	5.07	5.07
2	0.093	0.071	0.056	15356	30497	10	14023	27249	1333	3248	17.81	43.39	8.68	10.65	3.48	4.27
3	0.055	0.042	0.036	15121	45618	15	14147	41396	974	4222	13.01	56.40	6.44	9.26	2.58	3.71
4	0.035	0.030	0.026	14955	60573	20	14199	55595	756	4978	10.10	66.50	5.06	8.22	2.03	3.29
5	0.025	0.022	0.020	8953	69526	23	8603	64198	350	5328	4.68	71.17	3.91	7.66	1.57	3.07
6	0.020	0.020	0.020	17984	87510	29	17642	81840	342	5670	4.57	75.74	1.90	6.48	0.76	2.60
7	0.017	0.017	0.017	19265	106775	36	18958	100798	307	5977	4.10	79.84	1.59	5.60	0.64	2.24
8	0.016	0.016	0.015	1132	107907	36	1076	101874	56	6033	0.75	80.59	4.95	5.59	1.98	2.24
9	0.013	0.013	0.013	42823	150730	50	42365	144239	458	6491	6.12	86.71	1.07	4.31	0.43	1.73
10	0.011	0.011	0.011	50226	200956	67	49855	194094	371	6862	4.96	91.66	0.74	3.41	0.30	1.37
11	0.010	0.010	0.010	99044	300000	100	98420	292514	624	7486	8.34	100.00	0.63	2.50	0.25	1.00

Risk Factors Contribution to Stroke
Individuals Over 85 with thes conditions have over a 25% chance of a Stroke in any given year
Usually I don't use Interactions without main effects but these interraction have logical definitions
This was derived from real Mediacare data using a random 400k sample (300k Training and 100k holdout)
Over 30 thousand models where examined. Will release supporting analyses if authorized



Variables by Descending Chi-Square

Variables by Descending Chi-Square

Variable	Chi-Square	Prob > Chi_square
HYPERT	8235.1	0
HYPERT_IHD	8086.9	0
HYPERT_CHF	6819.5	0
IHD	6366.7	0
HYPERT_ANEMIA	6257.1	0
CHF	5876.7	0
IHD_ANEMIA	5802.0	0
IHD_CHF	5353.6	0
HYPERT_AFIB	5349.2	0
CHF_ANEMIA	5246.0	0
ANEMIA	5041.4	0
AFIB	4888.2	0
CKD	4705.8	0
IHD_AFIB	4098.0	0
ANEMIA_AFIB	3780.5	0
CHF_AFIB	3578.2	0
DIAB	2573.2	0
COPD	2388.2	0
AGE	2165.4	0
HYPERT_AMI	756.8	0
OBESITY	730.9	0
AMI	722.8	0
IHD_AMI	722.8	0
ANEMIA_AMI	700.6	0
TOBA	656.1	0
CHF_AMI	636.7	0
AMI_AFIB	332.9	0

Variables by Descending Chi-Square

Variable	Chi-Square	Prob > Chi_square
ALCO	205.7	0
DIV	167.4	0
RACE_CD	150.2	0
LNGC	131.3	0
CD113	74.0	0

Levels with the Highest Indecies

Levels with the Highest Indecies

Variable	Categories	Stroke Victims	Non Stroke	Total	Index	Percent Response
ANEMIA_AMI	1<=ANEMIA_AMI<=1	157	862	1019	6.175	15.41%
CHF_AMI	1<=CHF_AMI<=1	159	956	1115	5.715	14.26%
ANEMIA_AFIB	1<=ANEMIA_AFIB<=1	943	5752	6695	5.645	14.09%
HYPERT_AMI	1<=HYPERT_AMI<=1	215	1435	1650	5.222	13.03%
CHF_AFIB	1<=CHF_AFIB<=1	1028	7039	8067	5.107	12.74%
AMI	1<=AMI<=1	220	1549	1769	4.984	12.44%
IHD_AMI	1<=IHD_AMI<=1	220	1549	1769	4.984	12.44%
IHD_AFIB	1<=IHD_AFIB<=1	1222	8650	9872	4.961	12.38%
HYPERT_AFIB	1<=HYPERT_AFIB<=1	1626	11786	13412	4.859	12.12%
CHF_ANEMIA	1<=CHF_ANEMIA<=1	1693	12866	14559	4.660	11.63%
AFIB	1<=AFIB<=1	1734	14151	15885	4.375	10.92%
IHD_ANEMIA	1<=IHD_ANEMIA<=1	2195	18977	21172	4.155	10.37%
HYPERT_CHF	1<=HYPERT_CHF<=1	2550	22057	24607	4.153	10.36%
IHD_CHF	1<=IHD_CHF<=1	2095	18507	20602	4.075	10.17%
CHF	1<=CHF<=1	2698	27017	29715	3.639	9.08%
HYPERT_ANEMIA	1<=HYPERT_ANEMIA<=1	3099	32990	36089	3.441	8.59%
HYPERT_IHD	1<=HYPERT_IHD<=1	3887	41528	45415	3.430	8.56%
CKD	1<=CKD<=1	2672	30628	33300	3.216	8.02%
COPD	1<=COPD<=1	1690	21421	23111	2.931	7.31%
ANEMIA	1<=ANEMIA<=1	3329	42569	45898	2.907	7.25%
IHD	1<=IHD<=1	4149	54076	58225	2.856	7.13%
LNGC	1<=LNGC<=1	136	2009	2145	2.541	6.34%
ALCO	1<=ALCO<=1	242	3797	4039	2.401	5.99%
HYPERT	1<=HYPERT<=1	6567	106133	112700	2.335	5.83%
TOBA	1<=TOBA<=1	912	15623	16535	2.210	5.52%
DIAB	1<=DIAB<=1	3131	54102	57233	2.192	5.47%

Levels with the Highest Indecies

Variable	Categories	Stroke Victims	Non Stroke	Total	Index	Percent Response
OBESITY	1<=OBESITY<=1	1063	18619	19682	2.164	5.40%
AGE	86<=AGE<=115	1659	29778	31437	2.115	5.28%

Chi Square and Index Analysis of Character Variables

Response Decile from Low to High	Variable	Similar Stroke Levels	Responders Stroke Victims	Non-Responders All Others	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Strokes
Label= CD113 --> Chi-Square= 73.994493911 P-Value= 0.00000								
1	CD113	41 (5, 445, 1.1%) 42 (6, 489, 1.2%) 43 (7, 469, 1.5%) 44 (7, 445, 1.6%) 46 (5, 444, 1.1%) 48 (9, 650, 1.4%) 49 (5, 600, 0.8%) 51 (8, 518, 1.5%) 53 (8, 603, 1.3%)	60	4603	4663	1.55%	0.52	1.29%
2		20 (48, 2381, 2.0%) 24 (55, 2456, 2.2%) 25 (47, 2504, 1.9%) 26 (46, 2618, 1.8%) 32 (23, 1068, 2.2%) 34 (24, 1049, 2.3%) 35 (21, 1039, 2.0%) 36 (32, 1494, 2.1%) 37 (12, 534, 2.2%) 38 (13, 671, 1.9%) 39 (11, 589, 1.9%) 40 (7, 304, 2.3%) 45 (15, 654, 2.3%) 47 (12, 572, 2.1%) 50 (11, 612, 1.8%) 52 (10, 541, 1.8%)	387	18699	19086	6.36%	0.81	2.03%

Response Decile from Low to High	Variable	Similar Stroke Levels	Responders Stroke Victims	Non-Responders All Others	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Strokes
3		00 (124, 5175, 2.4%) 01 (763, 31163, 2.4%) 02 (768, 30705, 2.5%) 03 (658, 26542, 2.5%) 05 (503, 20667, 2.4%) 06 (440, 18269, 2.4%) 07 (389, 15876, 2.5%) 09 (284, 11739, 2.4%) 15 (103, 4337, 2.4%) 19 (72, 3019, 2.4%) 23 (60, 2539, 2.4%) 27 (70, 2905, 2.4%) 28 (27, 1157, 2.3%) 29 (21, 890, 2.4%) 31 (23, 979, 2.3%) 33 (31, 1243, 2.5%)	4336	172869	177205	59.07%	0.98	2.45%
4		04 (682, 25055, 2.7%) 08 (399, 14725, 2.7%) 11 (231, 8937, 2.6%) 12 (204, 7777, 2.6%) 14 (152, 5867, 2.6%) 16 (139, 5169, 2.7%) 18 (112, 4206, 2.7%) 21 (71, 2658, 2.7%) 22 (65, 2514, 2.6%) 30 (28, 1050, 2.7%)	2083	75875	77958	25.99%	1.07	2.67%
5		10 (264, 9041, 2.9%) 13 (203, 7152, 2.8%) 17 (132, 4369, 3.0%) 98 (21, 526, 4.0%)	620	20468	21088	7.03%	1.18	2.94%
			7486	*292514*	300000			
Label= Census Division --> Chi-Square= 167.38326324 P-Value= 0.00000								
1	DIV	8 (353, 19468, 1.8%) 9 (806, 42845, 1.9%)	1159	61154	62313	20.77%	0.75	1.86%
2		4 (443, 20673, 2.1%)	443	20230	20673	6.89%	0.86	2.14%

Response Decile from Low to High	Variable	Similar Stroke Levels	Responders Stroke Victims	Non-Responders All Others	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Strokes
3		1 (380, 15445, 2.5%) 2 (1061, 41762, 2.5%)	1441	55766	57207	19.07%	1.01	2.52%
4		3 (1292, 47036, 2.7%) 5 (1738, 61716, 2.8%) 6 (537, 19831, 2.7%) 7 (876, 31224, 2.8%)	4443	155364	159807	53.27%	1.11	2.78%
			7486	*292514*	300000			
Label= Beneficiary race code (modified using RTI algorithm) --> Chi-Square= 150.22367565 P-Value								
1	RACE_CD	0 (23, 2487, 0.9%)	23	2464	2487	0.83%	0.37	0.92%
2		3 (36, 2499, 1.4%) 4 (165, 8608, 1.9%) 5 (410, 22067, 1.9%) 6 (28, 1328, 2.1%)	639	33863	34502	11.50%	0.74	1.85%
3		1 (5831, 232144, 2.5%)	5831	226313	232144	77.38%	1.01	2.51%
4		2 (993, 30867, 3.2%)	993	29874	30867	10.29%	1.29	3.22%
			7486	*292514*	300000			

Chi Square and Index Analysis of Numeric Variables

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
Label= AFIB --> Chi-Square= 4888.2004556 P-Value= 0.00000									
1	AFIB	0<=AFIB<=0	0.00	5752	278363	284115	94.71%	0.81	2.02%
3		1<=AFIB<=1	1.00	1734	14151	15885	5.30%	4.37	10.92%
				7486	*292514*	300000			
Label= Age --> Chi-Square= 2165.3772771 P-Value= 0.00000									
1	AGE	1<=AGE<=68	60.10	1621	114267	115888	38.63%	0.56	1.40%
3		69<= AGE<=73	70.88	1150	61437	62587	20.86%	0.74	1.84%
4		74<= AGE<=80	76.75	1729	58073	59802	19.93%	1.16	2.89%
5		81<= AGE<=85	82.88	1327	28959	30286	10.10%	1.76	4.38%
6		86<= AGE<=115	90.32	1659	29778	31437	10.48%	2.11	5.28%
				7486	*292514*	300000			
Label= ALCO --> Chi-Square= 205.68910948 P-Value= 0.00000									
1	ALCO	0<=ALCO<=0	0.00	7244	288717	295961	98.65%	0.98	2.45%
3		1<=ALCO<=1	1.00	242	3797	4039	1.35%	2.40	5.99%
				7486	*292514*	300000			
Label= AMI --> Chi-Square= 722.7843164 P-Value= 0.00000									
1	AMI	0<=AMI<=0	0.00	7266	290965	298231	99.41%	0.98	2.44%
3		1<=AMI<=1	1.00	220	1549	1769	0.59%	4.98	12.44%
				7486	*292514*	300000			
Label= AMI_AFIB --> Chi-Square= 332.94802067 P-Value= 0.00000									

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
1	AMI_AFIB	0<=AMI_AFIB<=0	0.00	7420	292189	299609	99.87%	0.99	2.48%
3		1<=AMI_AFIB<=1	1.00	66	325	391	0.13%	6.76	16.88%
				7486	*292514*	300000			
Label= ANEMIA --> Chi-Square= 5041.3653112 P-Value= 0.00000									
1	ANEMIA	0<=ANEMIA<=0	0.00	4157	249945	254102	84.70%	0.66	1.64%
3		1<=ANEMIA<=1	1.00	3329	42569	45898	15.30%	2.91	7.25%
				7486	*292514*	300000			
Label= ANEMIA_AFIB --> Chi-Square= 3780.512207 P-Value= 0.00000									
1	ANEMIA_AFIB	0<=ANEMIA_AFIB<=0	0.00	6543	286762	293305	97.77%	0.89	2.23%
3		1<=ANEMIA_AFIB<=1	1.00	943	5752	6695	2.23%	5.64	14.09%
				7486	*292514*	300000			
Label= ANEMIA_AMI --> Chi-Square= 700.61610501 P-Value= 0.00000									
1	ANEMIA_AMI	0<=ANEMIA_AMI<=0	0.00	7329	291652	298981	99.66%	0.98	2.45%
3		1<=ANEMIA_AMI<=1	1.00	157	862	1019	0.34%	6.17	15.41%
				7486	*292514*	300000			
Label= CHF --> Chi-Square= 5876.7133084 P-Value= 0.00000									
1	CHF	0<=CHF<=0	0.00	4788	265497	270285	90.10%	0.71	1.77%
3		1<=CHF<=1	1.00	2698	27017	29715	9.91%	3.64	9.08%

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
				7486	*292514*	300000			
Label= CHF_AFIB --> Chi-Square= 3578.2398394 P-Value= 0.00000									
1	CHF_AFIB	0<=CHF_AFIB<=0	0.00	6458	285475	291933	97.31%	0.89	2.21%
3		1<=CHF_AFIB<=1	1.00	1028	7039	8067	2.69%	5.11	12.74%
				7486	*292514*	300000			
Label= CHF_AMI --> Chi-Square= 636.65463771 P-Value= 0.00000									
1	CHF_AMI	0<=CHF_AMI<=0	0.00	7327	291558	298885	99.63%	0.98	2.45%
3		1<=CHF_AMI<=1	1.00	159	956	1115	0.37%	5.71	14.26%
				7486	*292514*	300000			
Label= CHF_ANEMIA --> Chi-Square= 5246.0159809 P-Value= 0.00000									
1	CHF_ANEMIA	0<=CHF_ANEMIA<=0	0.00	5793	279648	285441	95.15%	0.81	2.03%
3		1<=CHF_ANEMIA<=1	1.00	1693	12866	14559	4.85%	4.66	11.63%
				7486	*292514*	300000			
Label= CKD --> Chi-Square= 4705.7959271 P-Value= 0.00000									
1	CKD	0<=CKD<=0	0.00	4814	261886	266700	88.90%	0.72	1.81%
3		1<=CKD<=1	1.00	2672	30628	33300	11.10%	3.22	8.02%
				7486	*292514*	300000			
Label= COPD --> Chi-Square= 2388.1962031 P-Value= 0.00000									

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
1	COPD	0<=COPD<=0	0.00	5796	271093	276889	92.30%	0.84	2.09%
3		1<=COPD<=1	1.00	1690	21421	23111	7.70%	2.93	7.31%
				7486	*292514*	300000			
Label= DIAB --> Chi-Square= 2573.2489518 P-Value= 0.00000									
1	DIAB	0<=DIAB<=0	0.00	4355	238412	242767	80.92%	0.72	1.79%
3		1<=DIAB<=1	1.00	3131	54102	57233	19.08%	2.19	5.47%
				7486	*292514*	300000			
Label= HYPERT --> Chi-Square= 8235.1175412 P-Value= 0.00000									
1	HYPERT	0<= HYPERT<=0	0.00	919	186381	187300	62.43%	0.20	0.49%
3		1<= HYPERT<=1	1.00	6567	106133	112700	37.57%	2.34	5.83%
				7486	*292514*	300000			
Label= HYPERT_AFIB --> Chi-Square= 5349.1830185 P-Value= 0.00000									
1	HYPERT_ AFIB	0<=HYPERT_ AFIB<=0	0.00	5860	280728	286588	95.53%	0.82	2.04%
3		1<=HYPERT_ AFIB<=1	1.00	1626	11786	13412	4.47%	4.86	12.12%
				7486	*292514*	300000			
Label= HYPERT_AMI --> Chi-Square= 756.81854985 P-Value= 0.00000									
1	HYPERT_AMI	0<=HYPERT_ AMI<=0	0.00	7271	291079	298350	99.45%	0.98	2.44%
3		1<=HYPERT_ AMI<=1	1.00	215	1435	1650	0.55%	5.22	13.03%
				7486	*292514*	300000			

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
Label= HYPERT_ANEMIA --> Chi-Square= 6257.0805538 P-Value= 0.00000									
1	HYPERT_ANEMIA	0<=HYPERT_ANEMIA<=0	0.00	4387	259524	263911	87.97%	0.67	1.66%
3		1<=HYPERT_ANEMIA<=1	1.00	3099	32990	36089	12.03%	3.44	8.59%
				7486	*292514*	300000			
Label= HYPERT_CHF --> Chi-Square= 6819.5302039 P-Value= 0.00000									
1	HYPERT_CHF	0<=HYPERT_CHF<=0	0.00	4936	270457	275393	91.80%	0.72	1.79%
3		1<=HYPERT_CHF<=1	1.00	2550	22057	24607	8.20%	4.15	10.36%
				7486	*292514*	300000			
Label= HYPERT_IHD --> Chi-Square= 8086.9060545 P-Value= 0.00000									
1	HYPERT_IHD	0<=HYPERT_IHD<=0	0.00	3599	250986	254585	84.86%	0.57	1.41%
3		1<=HYPERT_IHD<=1	1.00	3887	41528	45415	15.14%	3.43	8.56%
				7486	*292514*	300000			
Label= IHD --> Chi-Square= 6366.7216735 P-Value= 0.00000									
1	IHD	0<=IHD<=0	0.00	3337	238438	241775	80.59%	0.55	1.38%
3		1<=IHD<=1	1.00	4149	54076	58225	19.41%	2.86	7.13%
				7486	*292514*	300000			
Label= IHD_AFIB --> Chi-Square= 4097.9826132 P-Value= 0.00000									
1	IHD_AFIB	0<=IHD_AFIB<=0	0.00	6264	283864	290128	96.71%	0.87	2.16%

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
3		1<=IHD_AFIB<=1	1.00	1222	8650	9872	3.29%	4.96	12.38%
				7486	*292514*	300000			
Label= IHD_AMI --> Chi-Square= 722.7843164 P-Value= 0.00000									
1	IHD_AMI	0<=IHD_AMI<=0	0.00	7266	290965	298231	99.41%	0.98	2.44%
3		1<=IHD_AMI<=1	1.00	220	1549	1769	0.59%	4.98	12.44%
				7486	*292514*	300000			
Label= IHD_ANEMIA --> Chi-Square= 5801.9989622 P-Value= 0.00000									
1	IHD_ANEMIA	0<=IHD_ANEMIA<=0	0.00	5291	273537	278828	92.94%	0.76	1.90%
3		1<=IHD_ANEMIA<=1	1.00	2195	18977	21172	7.06%	4.15	10.37%
				7486	*292514*	300000			
Label= IHD_CHF --> Chi-Square= 5353.6464316 P-Value= 0.00000									
1	IHD_CHF	0<=IHD_CHF<=0	0.00	5391	274007	279398	93.13%	0.77	1.93%
3		1<=IHD_CHF<=1	1.00	2095	18507	20602	6.87%	4.08	10.17%
				7486	*292514*	300000			
Label= LNGC --> Chi-Square= 131.27462759 P-Value= 0.00000									
1	LNGC	0<=LNGC<=0	0.00	7350	290505	297855	99.29%	0.99	2.47%
3		1<=LNGC<=1	1.00	136	2009	2145	0.72%	2.54	6.34%
				7486	*292514*	300000			

Stroke Decile from Low to High	Variable	Similar Stroke Levels	Mean Value in Level	Responders Stroke	Non-Responders Stroke	Total	Percent Coverage	Index Response Rate divided by Overall 2.4953%	Percent with Stroke
Label= OBESITY --> Chi-Square= 730.86878109 P-Value= 0.00000									
1	OBESITY	0<= OBESITY<=0	0.00	6423	273895	280318	93.44%	0.92	2.29%
3		1<= OBESITY<=1	1.00	1063	18619	19682	6.56%	2.16	5.40%
				7486	*292514*	300000			
Label= TOBA --> Chi-Square= 656.07707495 P-Value= 0.00000									
1	TOBA	0<=TOBA<=0	0.00	6574	276891	283465	94.49%	0.93	2.32%
3		1<=TOBA<=1	1.00	912	15623	16535	5.51%	2.21	5.52%
				7486	*292514*	300000			

Diagnostics on 1.5% Sample - PDF Size issue

Full diagnostics on 1.5% sample of 300k

Model Information	
Data Set	WORK.REG_BUILDSAVALL
Response Variable	STROKE
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	4435
Number of Observations Used	4435

Response Profile

Ordered Value	STROKE	Total Frequency
1	1	106
2	0	4329

Probability modeled is STROKE=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1003.021	890.446
SC	1009.418	928.829
-2 Log L	1001.021	878.446

Full diagnostics on 1.5% sample of 300k

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	122.5752	5	<.0001
Score	197.8820	5	<.0001
Wald	138.0231	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
INTERCEPT	1	-5.4215	0.2581	441.1385	<.0001
_HYPERT_IHD	1	0.5126	0.0869	34.7672	<.0001
_AGE	1	0.3764	0.1799	4.3800	0.0364
_AFIB	1	0.1250	0.0793	2.4860	0.1149
_HYPERT_CHF	1	0.2351	0.0800	8.6415	0.0033
_HYPERT_ANEMIA	1	0.0482	0.0900	0.2871	0.5921

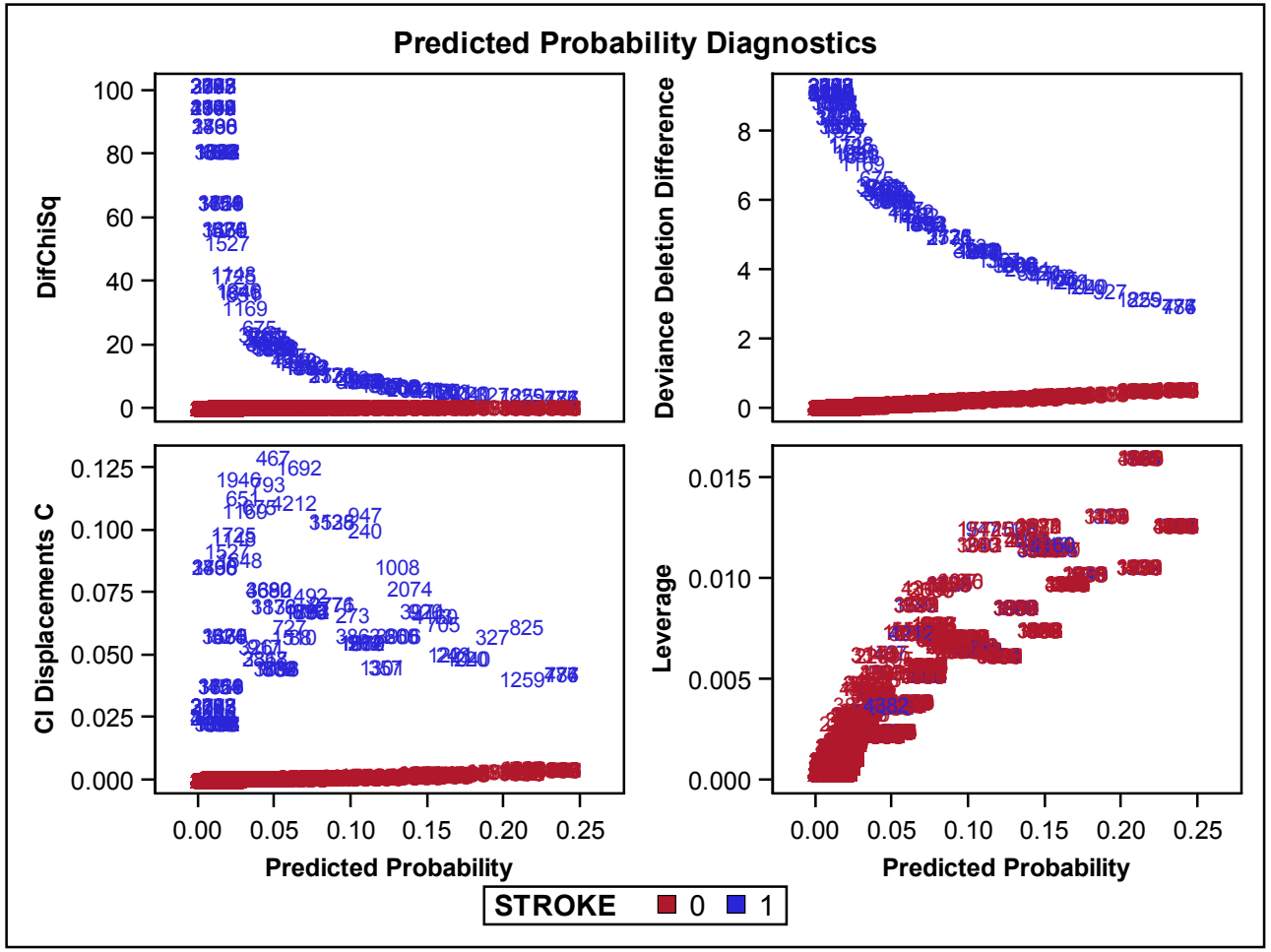
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
_HYPERT_IHD	1.670	1.408	1.980
_AGE	1.457	1.024	2.073
_AFIB	1.133	0.970	1.324
_HYPERT_CHF	1.265	1.081	1.480
_HYPERT_ANEMIA	1.049	0.880	1.252

Full diagnostics on 1.5% sample of 300k

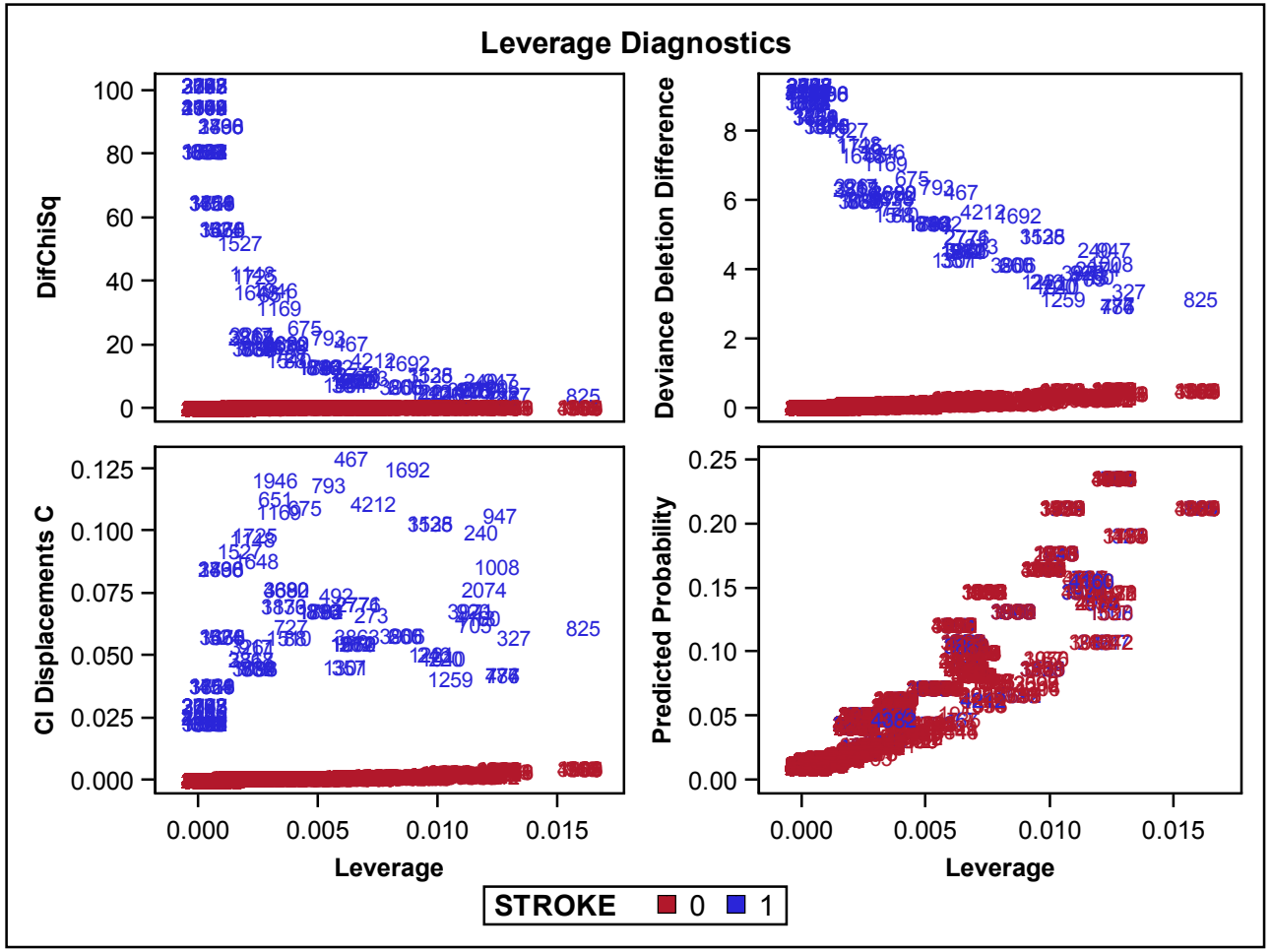
Partition for the Hosmer and Lemeshow Test					
		STROKE = 1		STROKE = 0	
Group	Total	Observed	Expected	Observed	Expected
1	1427	8	13.98	1419	1413.02
2	751	7	7.87	744	743.13
3	748	12	9.08	736	738.92
4	381	5	5.82	376	375.18
5	437	12	9.02	425	427.98
6	435	29	23.62	406	411.38
7	256	33	36.61	223	219.39

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.4720	5	0.2630

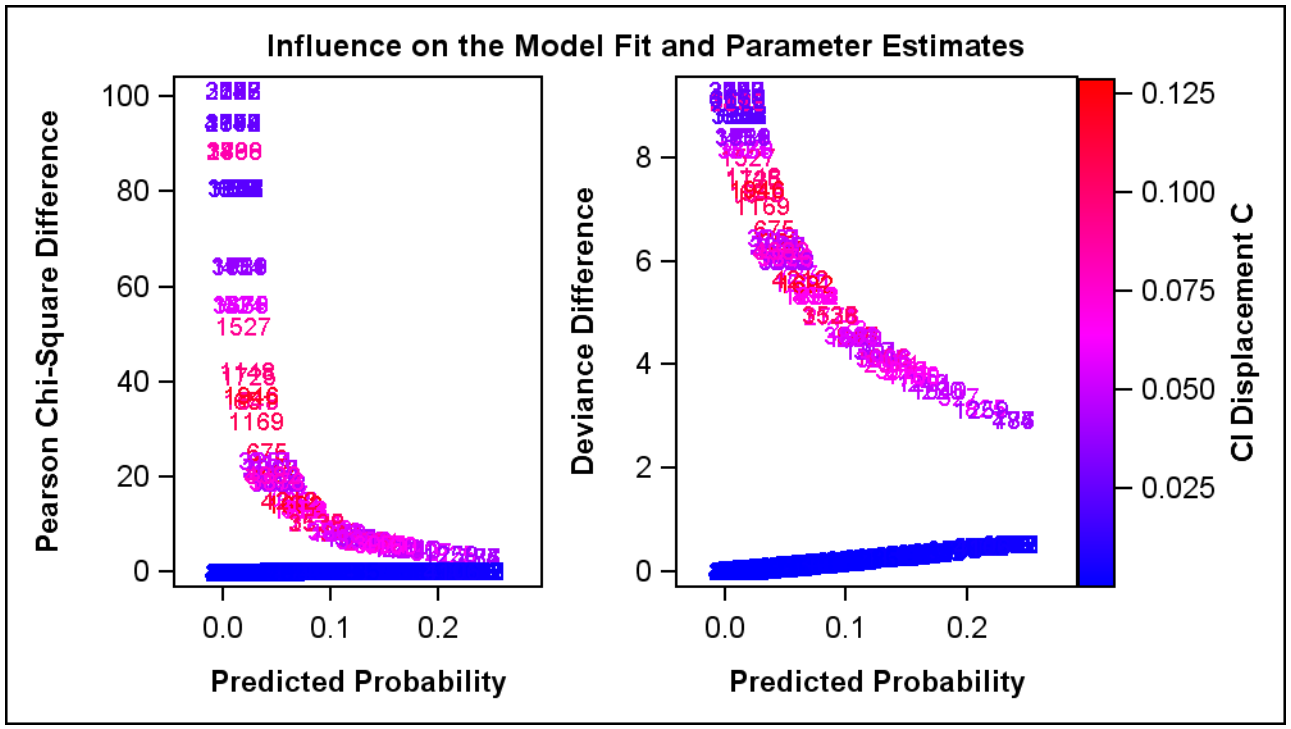
Full diagnostics on 1.5% sample of 300k



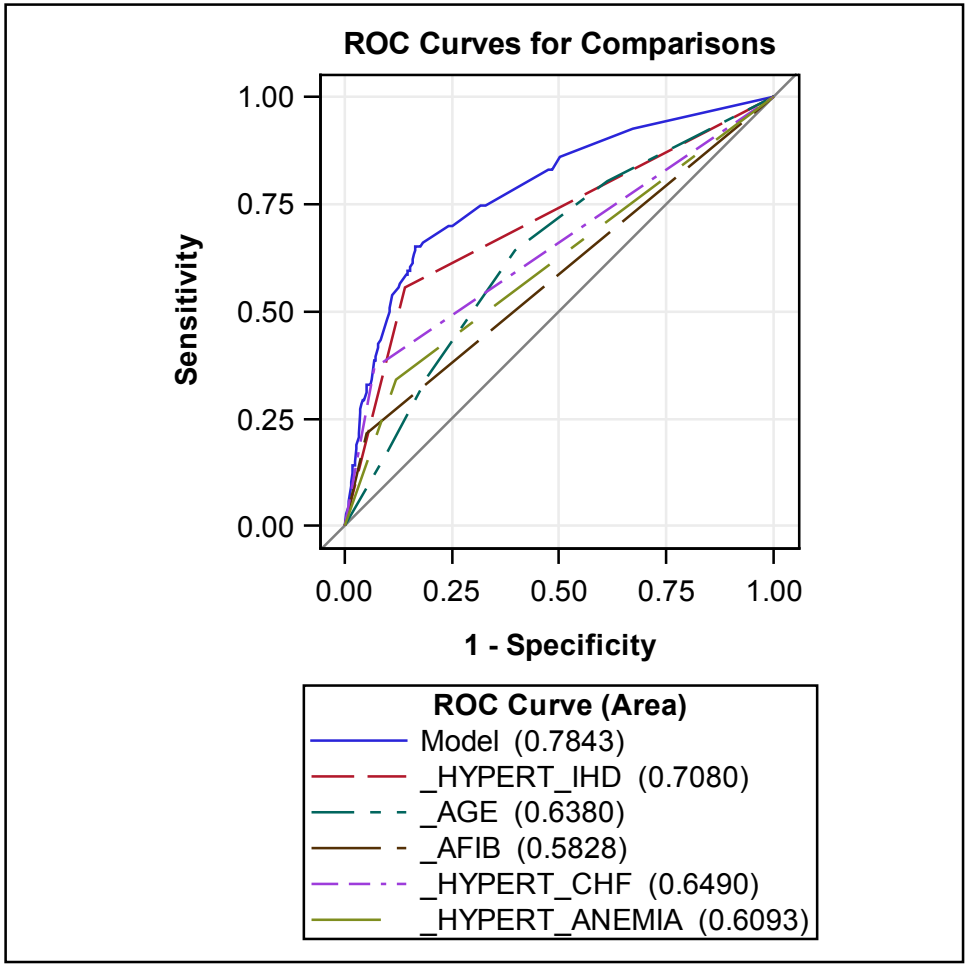
Full diagnostics on 1.5% sample of 300k



Full diagnostics on 1.5% sample of 300k



Full diagnostics on 1.5% sample of 300k



ROC Association Statistics

ROC Model	Area	Standard Error	Mann-Whitney 95% Wald Confidence Limits		Somers' D (Gini)	Gamma	Tau-a
Model	0.7843	0.0241	0.7372	0.8315	0.5686	0.6031	0.0265
_HYPERT_IHD	0.7080	0.0244	0.6602	0.7558	0.4159	0.7693	0.0194
_AGE	0.6380	0.0259	0.5872	0.6887	0.2759	0.3469	0.0129
_AFIB	0.5828	0.0202	0.5433	0.6224	0.1657	0.6735	0.00773

Full diagnostics on 1.5% sample of 300k

ROC Association Statistics							
		Mann-Whitney					
ROC Model	Area	Standard Error	95% Wald Confidence Limits		Somers' D (Gini)	Gamma	Tau-a
_HYPERT_CHF	0.6490	0.0236	0.6027	0.6952	0.2979	0.7710	0.0139
_HYPERT_ANEMIA	0.6093	0.0232	0.5637	0.6548	0.2186	0.5776	0.0102