

Evaluating Factors Associated with Unknown Stage

DURC Data Assessment Workgroup

Qingzhao Yu, PhD
Mei-Chin Hsieh, MSPH
Xiao-Cheng Wu, MD, MPH
Ying Fan, MS
Patricia Andrews, MPH
Brad Wohler, MS
Bin Huang, DrPH
Baozhen Qiao, PhD
Ahmedin Jemal, PhD
Umed Ajani, MD, MPH

NAACCR Meeting, Louisville KY, June 23, 2011

Background and Objective

- Valid stage is critical for cancer treatment and research
- The proportion of unknown stage cases reflects not only quality of abstraction but also quality and availability of source data
- The purpose of this study is to identify non-abstractor controlled factors contributing to the variations in the proportion of unknown stage by registry

Methods

- Data Source:

- NAACCR's CINA analytic file 1995-2007

- Eligibility Criteria:

- U.S. Cancer registries that met NAACCR's criteria of high quality data (45 cancer registries)
- Invasive colorectal, lung, female breast, cervix, and prostate cancer
- Diagnosis year: 2004 to 2007
- Not DCO and autopsy

Methods cont'd

- **Dependent Variable:**
 - Proportion of unknown derived stage at registry level
- **Independent Variables:**
 - Race: white, black, and others
 - Gender: male, female
 - Age group: Cancer screening age or distribution of unknown stage
 - Diagnostic year: 2004, 2005, 2006, and 2007
 - Diagnostic confirmation: microscopic, non- microscopic, and unknown
 - Type of reporting source: Hospital and non-hospital
 - Metropolitan: metro (population \geq 200,000), non-metro (population in 0-199,999)

Methods cont'd

■ **Statistic Analysis:**

- Multiple linear regression was used to evaluate the relationship between % unknown cancer stage and predictors
- Standardized residuals were used to evaluate data quality
- Relative importance measures the variable contribution to the variance explained by the model
- Statistical significant level: $p\text{-value} = 0.10$

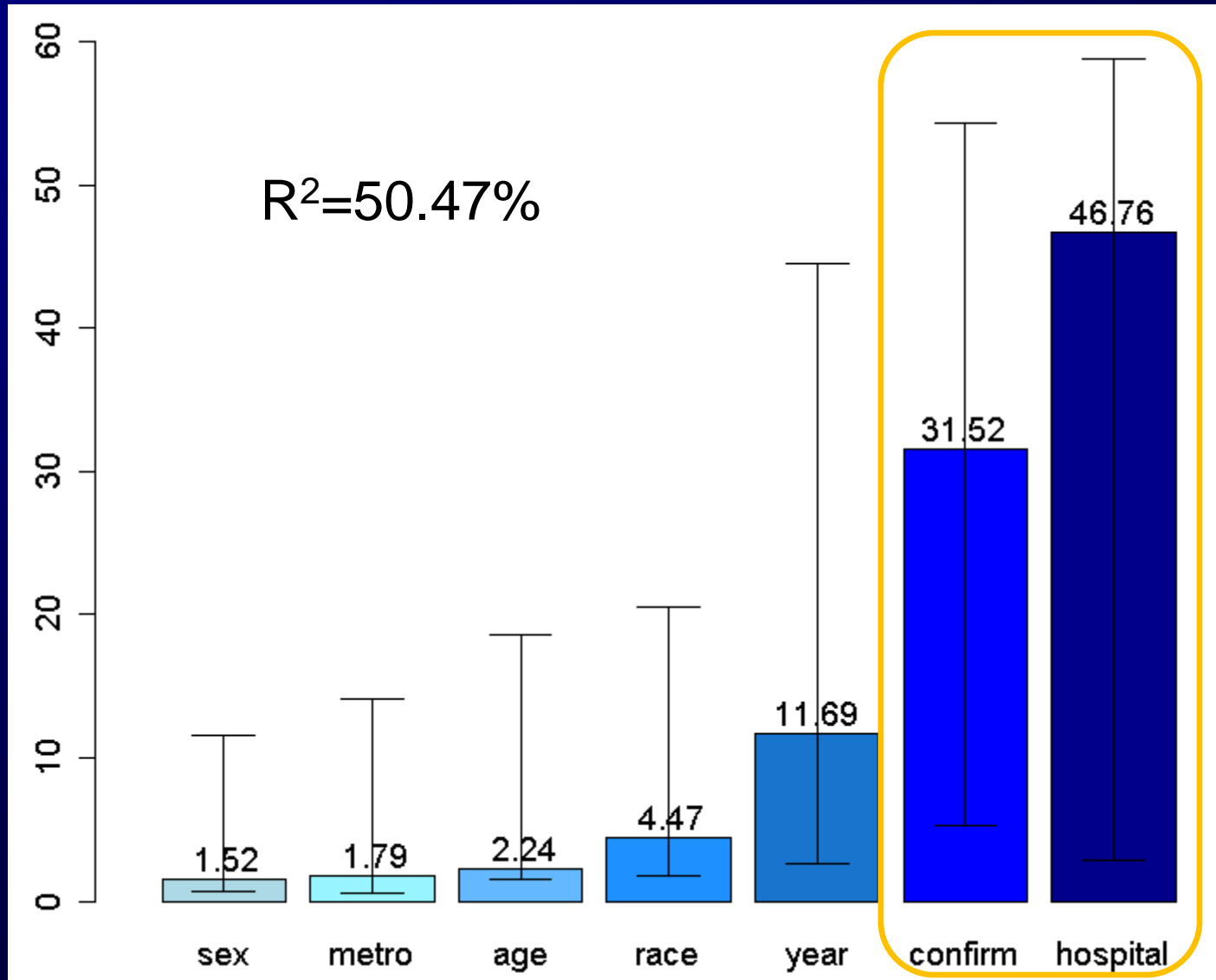
Methods cont'd

- **Relative Importance of Variables :**
 - Independent variables were sequentially added in the model based on their R^2 contribution
 - Permuted the order of variables to be added in regression and calculated the variable contribution at each permutation
 - The final variable relative importance was the averaged variable contribution from the permutation, explained as the proportion of R^2 that is attributable to the variable

Methods cont'd

- **Reduced model included:**
 - Variables that were significantly associated with % unknown stage in the full main effects model
 - Variables have higher percentage of explaining variance from the variable relative importance test

Colorectal Cancer: Variable Relative Importance (95 % CI) in Full Main Effects Model

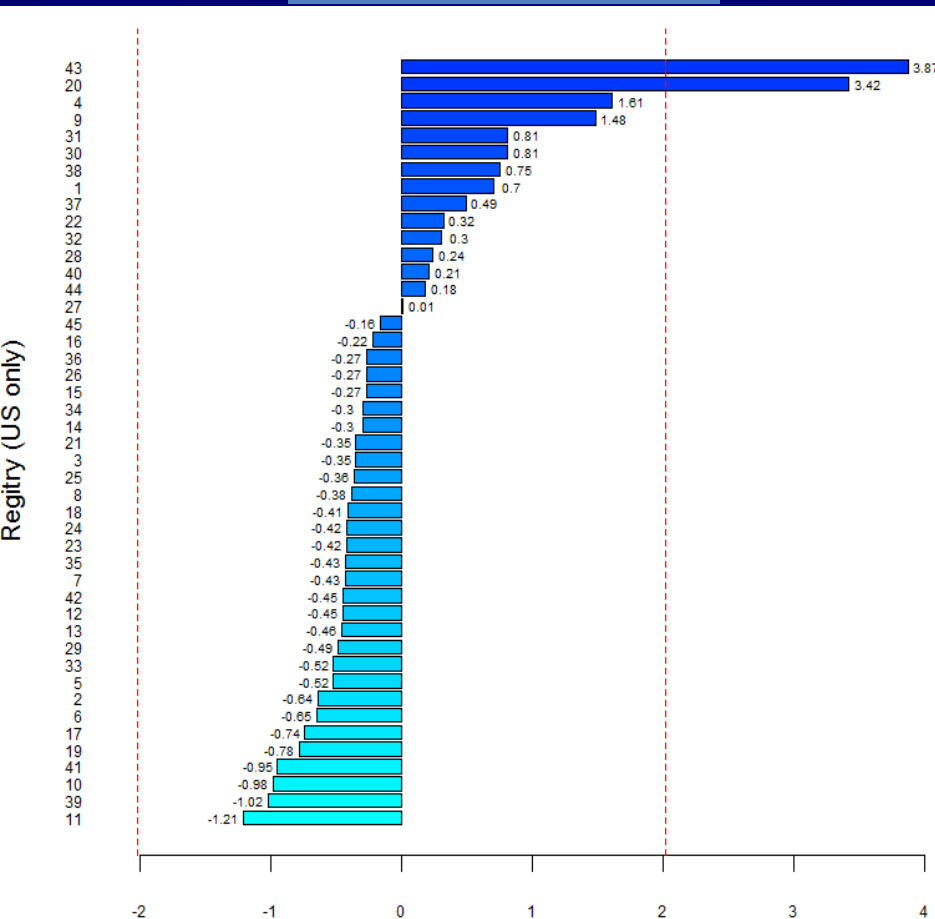


Colorectal Cancer: Model selection

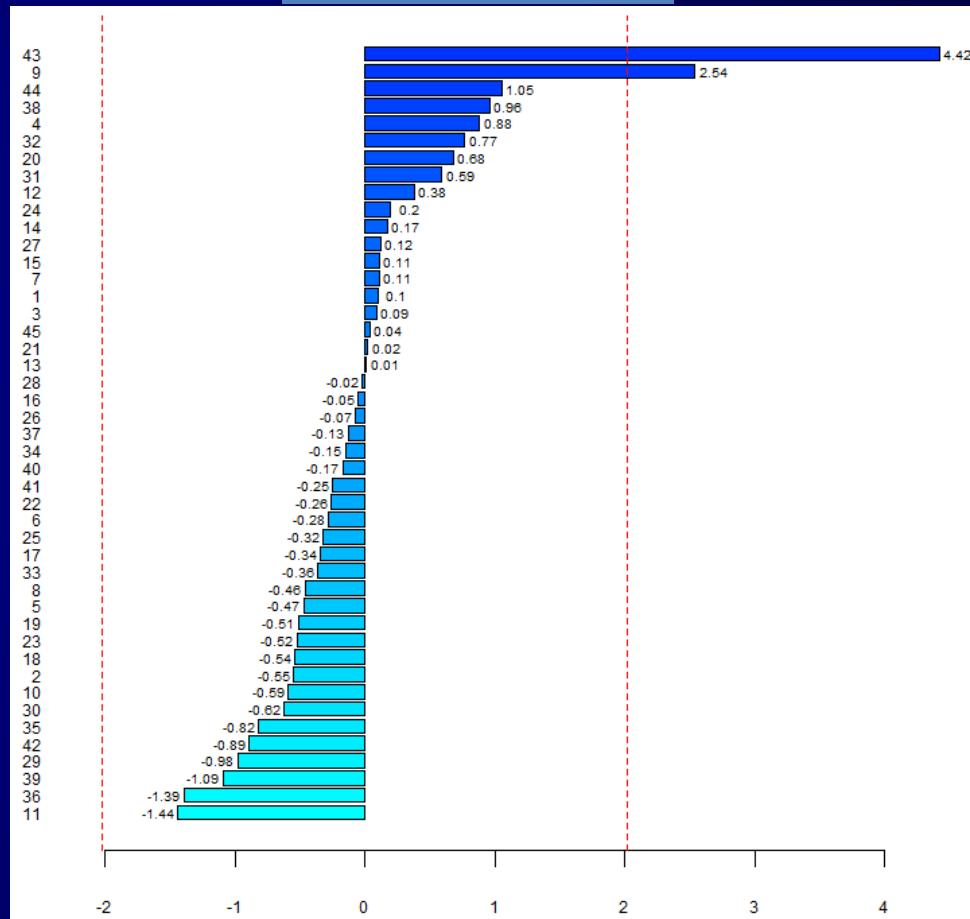
	Full Model		Reduced Model	
Predictor	Estimate(SE)	P value	Estimate(SE)	P value
Intercept	0.253 (0.40)	0.53	0.042 (0.01)	0.0025
Male	0.052 (0.35)	0.88		
Y2005	-0.299 (0.43)	0.49		
Y2006	0.234 (0.64)	0.72		
Y2007	-0.763 (0.63)	0.23		
Age ≤49	-0.254 (0.55)	0.64		
Age 80+	-0.0097 (0.23)	0.97		
Nonmetro	-0.0049 (0.02)	0.82		
Nonmicr_confirm	-1.085 (1.02)	0.30	-0.818 (0.81)	0.32
Unknown_confirm	2.009 (0.93)	0.04	2.277 (0.85)	0.01
Black	-0.036 (0.08)	0.66		
Other_race	-0.053 (0.04)	0.23		
nonhospital	0.340 (0.10)	0.001	0.304 (0.09)	0.001

Colorectal Cancer: Plots of Standardized Residuals of % Unknown Stage

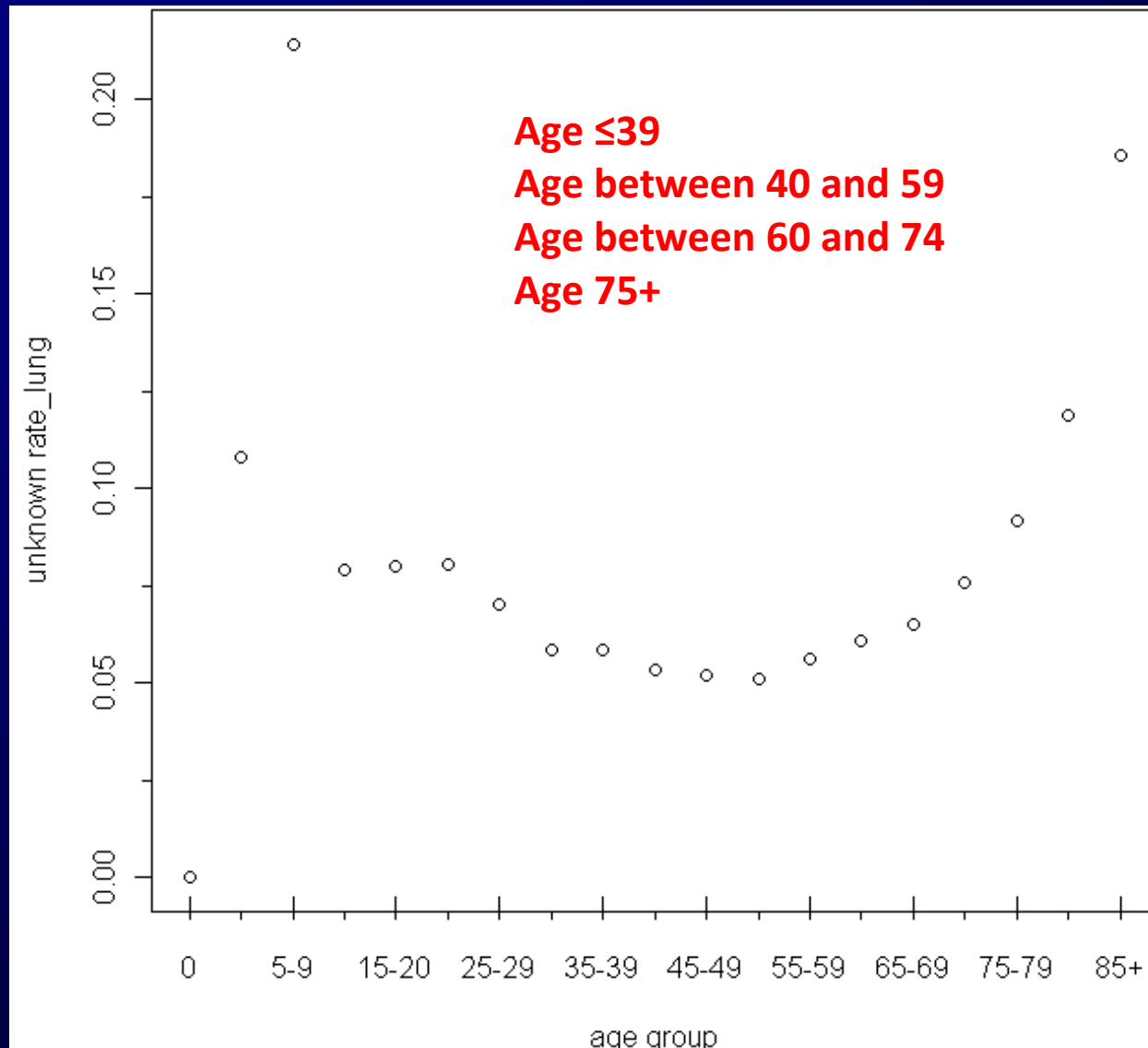
Before adjustment



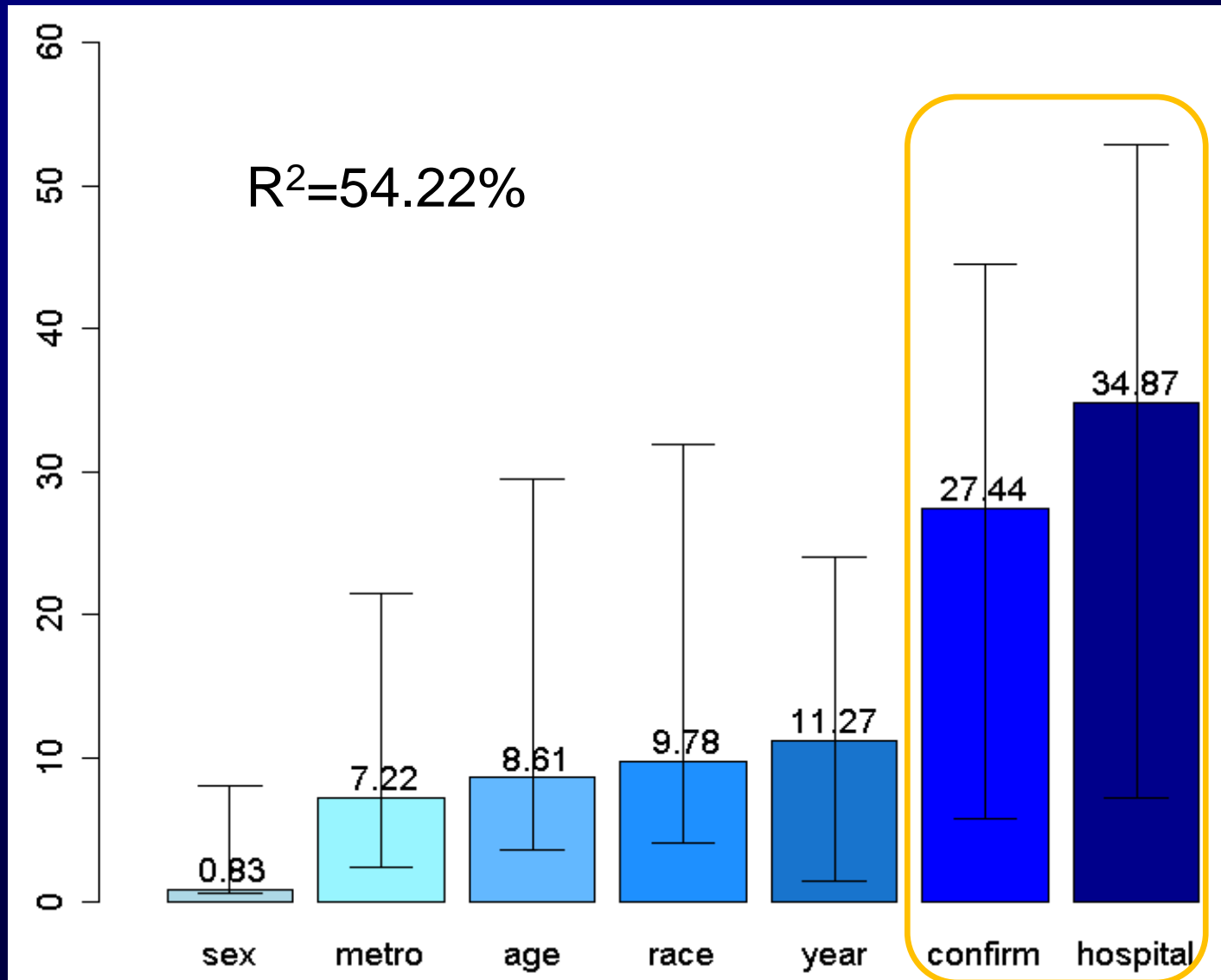
After adjustment



Lung Cancer: Age Grouping



Lung Cancer: Variable Relative Importance (95 % CI) in Full Main Effects Model

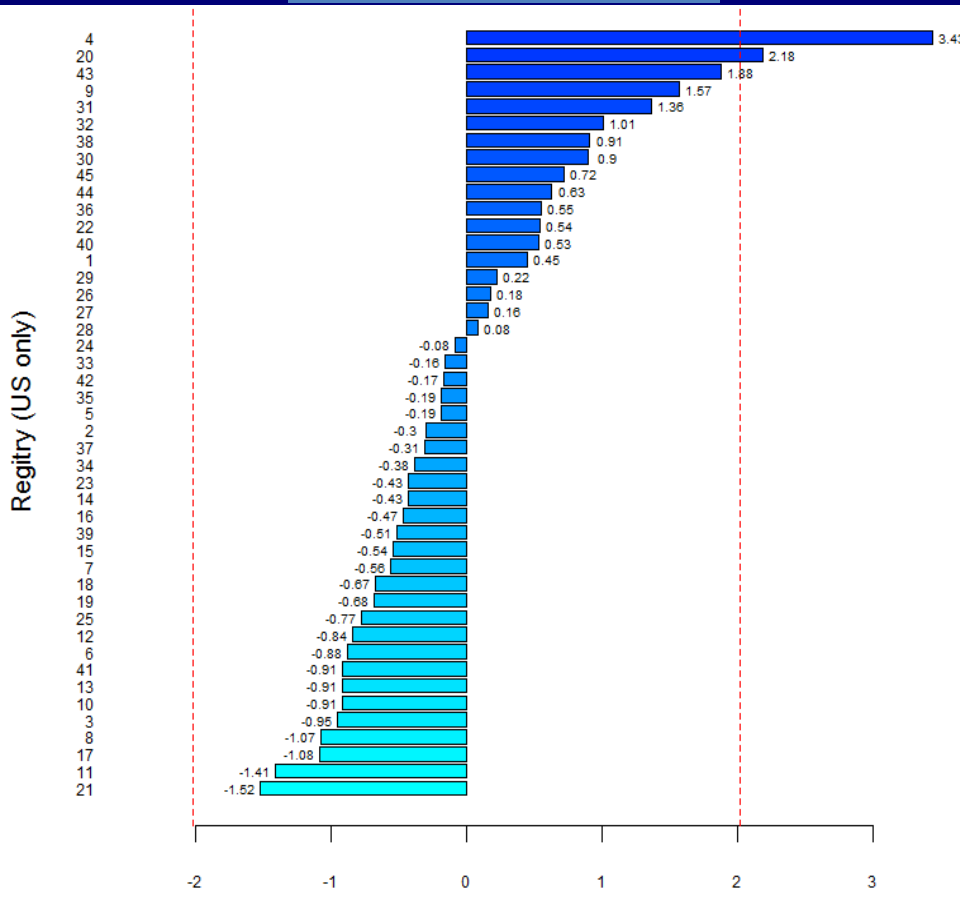


Lung Cancer: Model selection

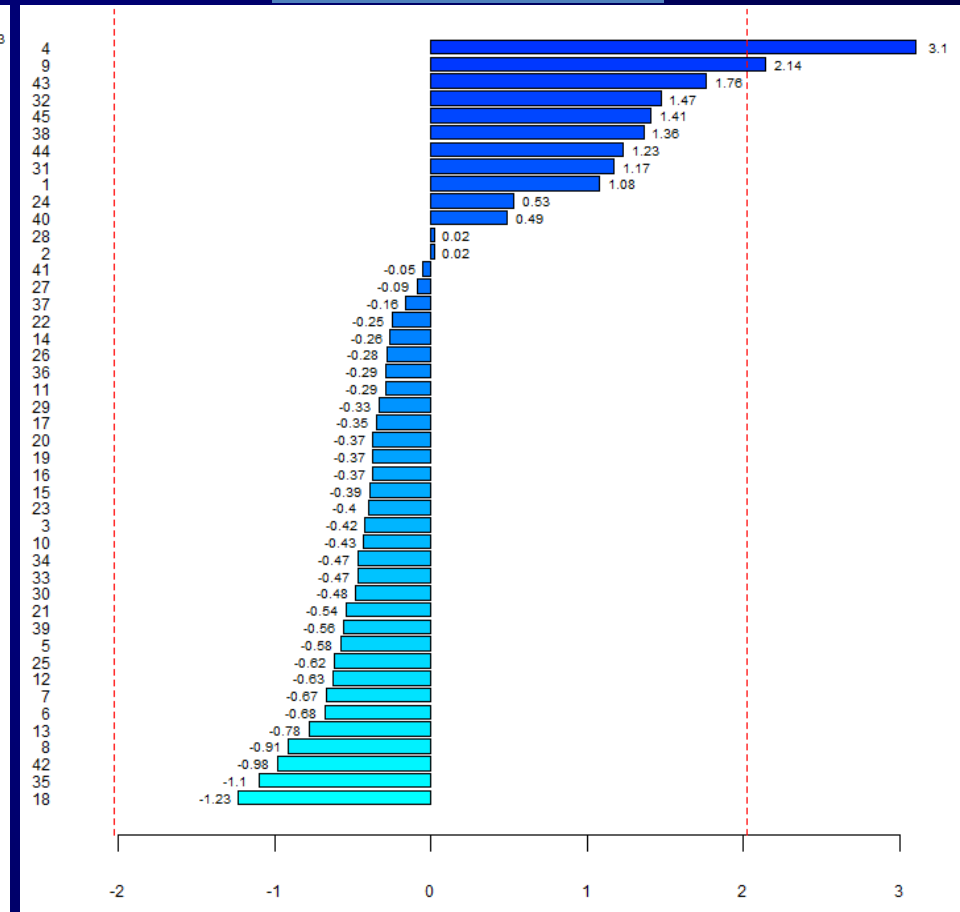
	Full Model		Reduced Model	
Predictor	Estimate(SE)	P value	Estimate(SE)	P value
Intercept	0.022 (0.63)	0.97	0.067 (0.02)	0.007
Male	0.149 (0.22)	0.50		
Y2005	-0.370 (0.45)	0.42		
Y2006	0.222 (0.76)	0.77		
Y2007	-0.959 (0.78)	0.23		
Age ≤ 39	0.256 (3.91)	0.95		
Age 60-74	0.576 (0.45)	0.21		
Age 75+	0.034 (0.37)	0.93		
Nonmetro	-0.068 (0.03)	0.06	-0.028 (0.02)	0.22
Nonmicr_confirm	-0.280 (0.20)	0.18	-0.085 (0.19)	0.65
Unknown_confirm	0.351 (0.41)	0.40	0.687 (0.39)	0.08
Black	-0.183 (0.13)	0.18	-0.049(0.07)	0.51
Other_race	-0.082 (0.05)	0.09	-0.062(0.04)	0.15
nonhospital	0.355 (0.10)	0.002	0.262 (0.097)	0.010

Lung Cancer: Plot of Standardized Residuals of Unknown Stage

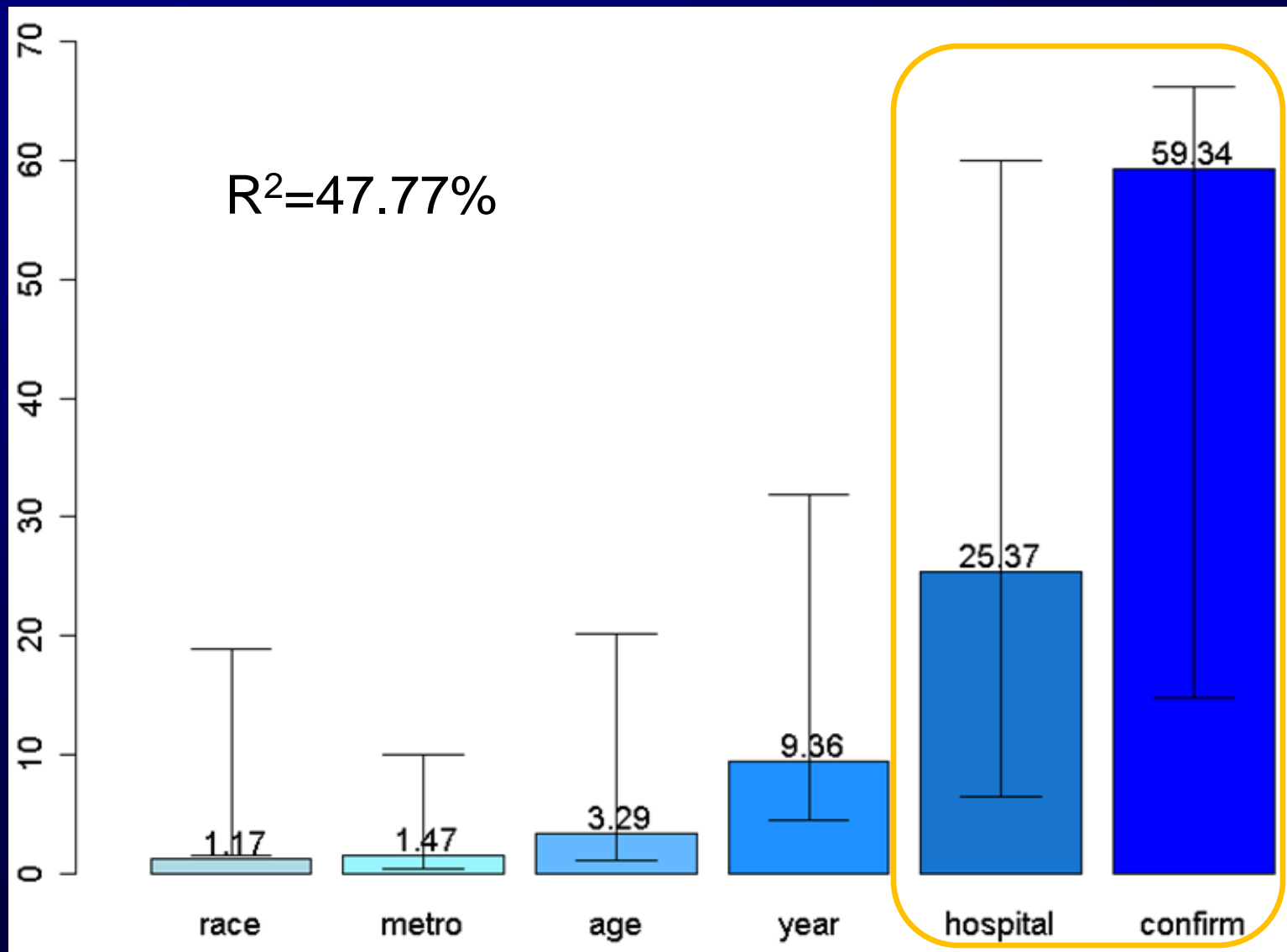
Before adjustment



After adjustment



Breast Cancer: Variable Relative Importance (95 % CI) in Full Main Effects Model

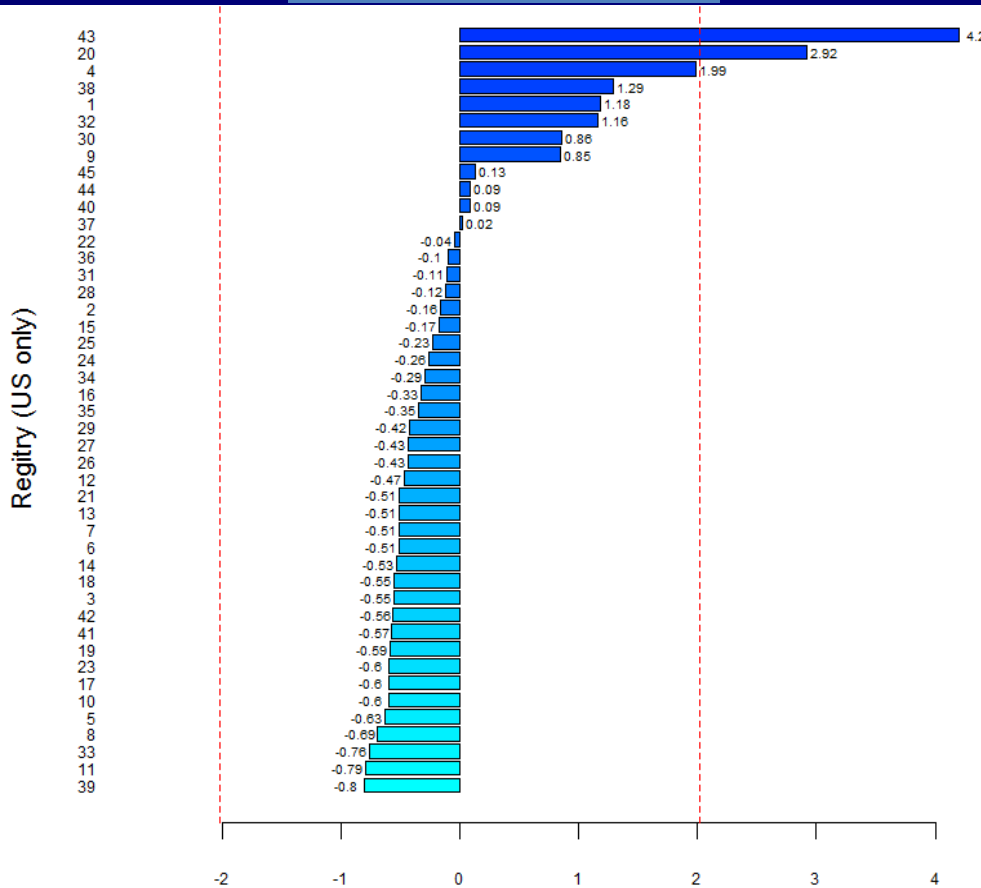


Breast Cancer: Model Selection

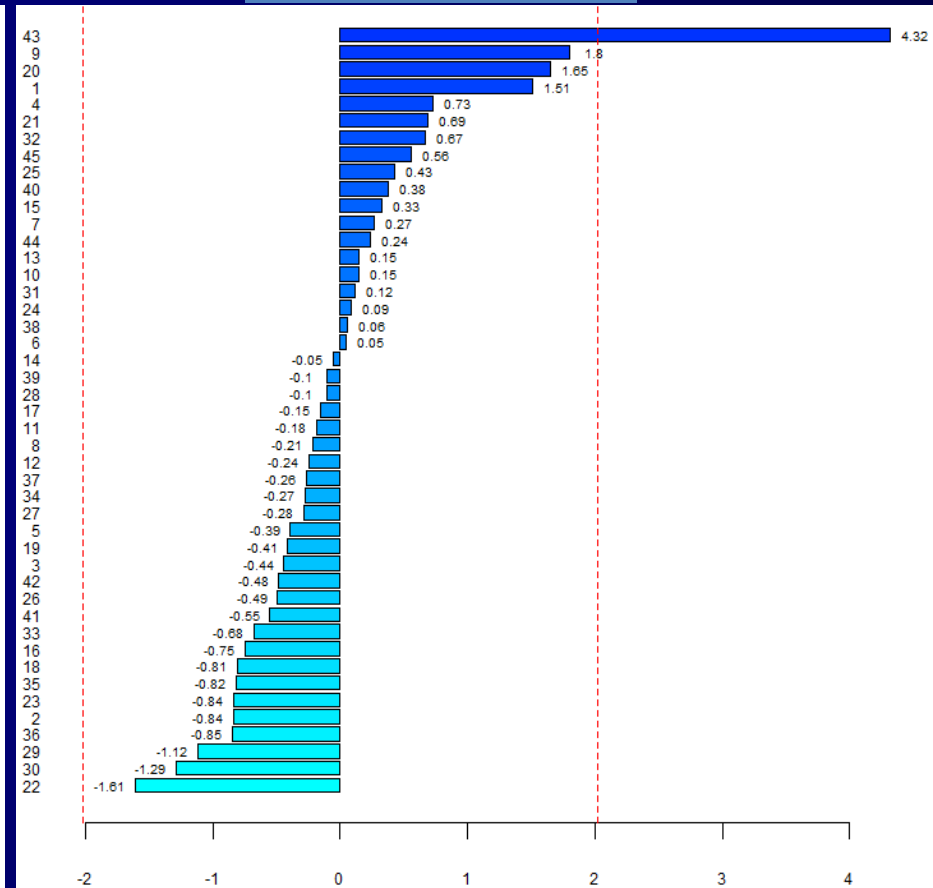
	Full Model		Reduced Model	
Predictor	Estimate(SE)	P value	Estimate(SE)	P value
Intercept	0.040(0.25)	0.87	0.0026(0.005)	0.64
Y2005	0.029(0.27)	0.91		
Y2006	0.197(0.55)	0.72		
Y2007	-0.370(0.38)	0.34		
Age ≤39	0.252(0.67)	0.71		
Age 80+	-0.093(0.17)	0.59		
Nonmetro	0.007(0.02)	0.68		
Nonmicr_confirm	0.870(0.92)	0.35	0.774(0.74)	0.302
Unknown_confirm	5.298(1.45)	0.0009	5.265(1.24)	0.00012
Black	-0.002(0.05)	0.97		
Other_race	0.0063(0.03)	0.84		
nonhospital	0.060(0.03)	0.046	0.065(0.03)	0.014

Breast Cancer: Plot of Standardized Residuals of Unknown Stage

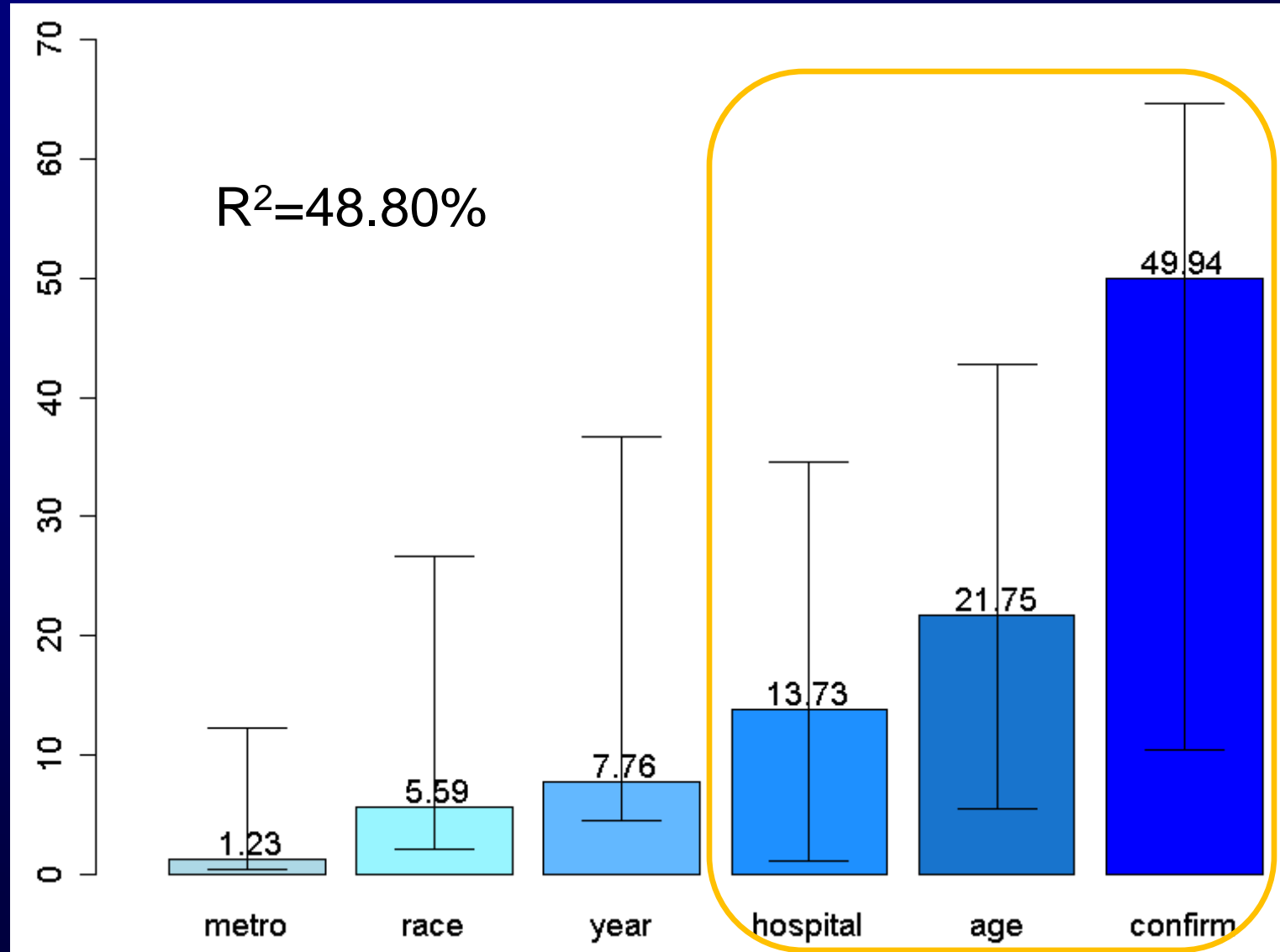
Before adjustment



After adjustment



Cervix Cancer: Variable Relative Importance (95 % CI) in Full Main Effects Model

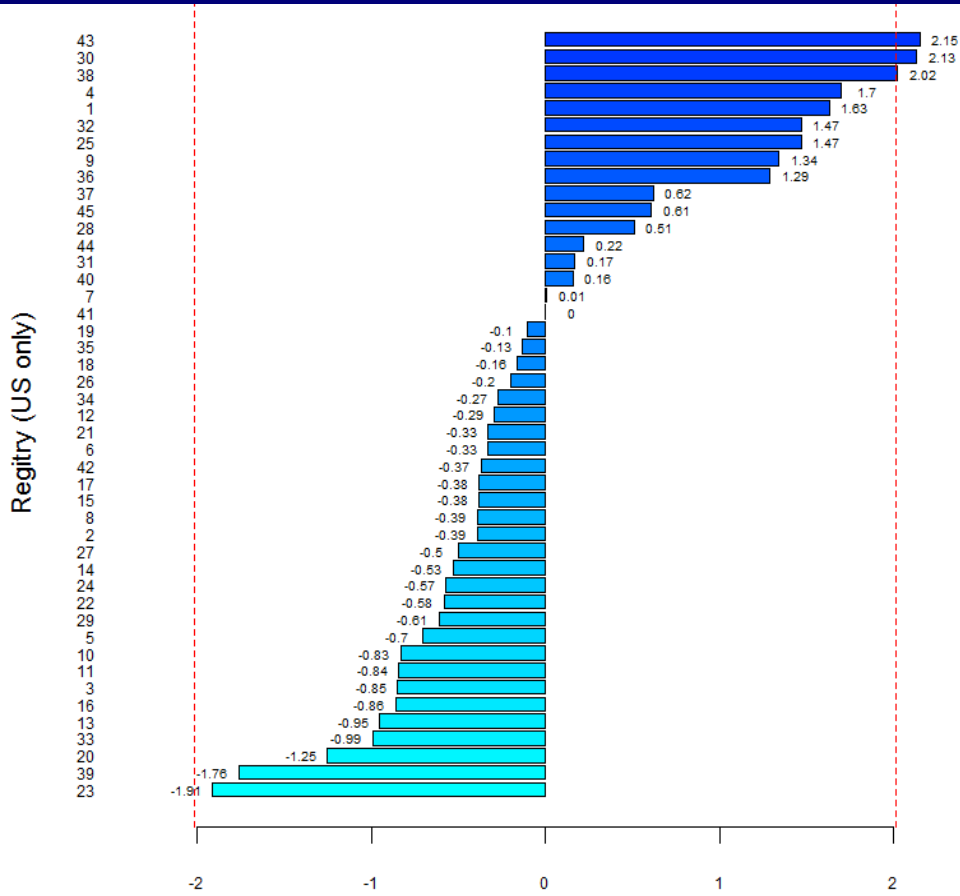


Cervix Cancer: Model Selection

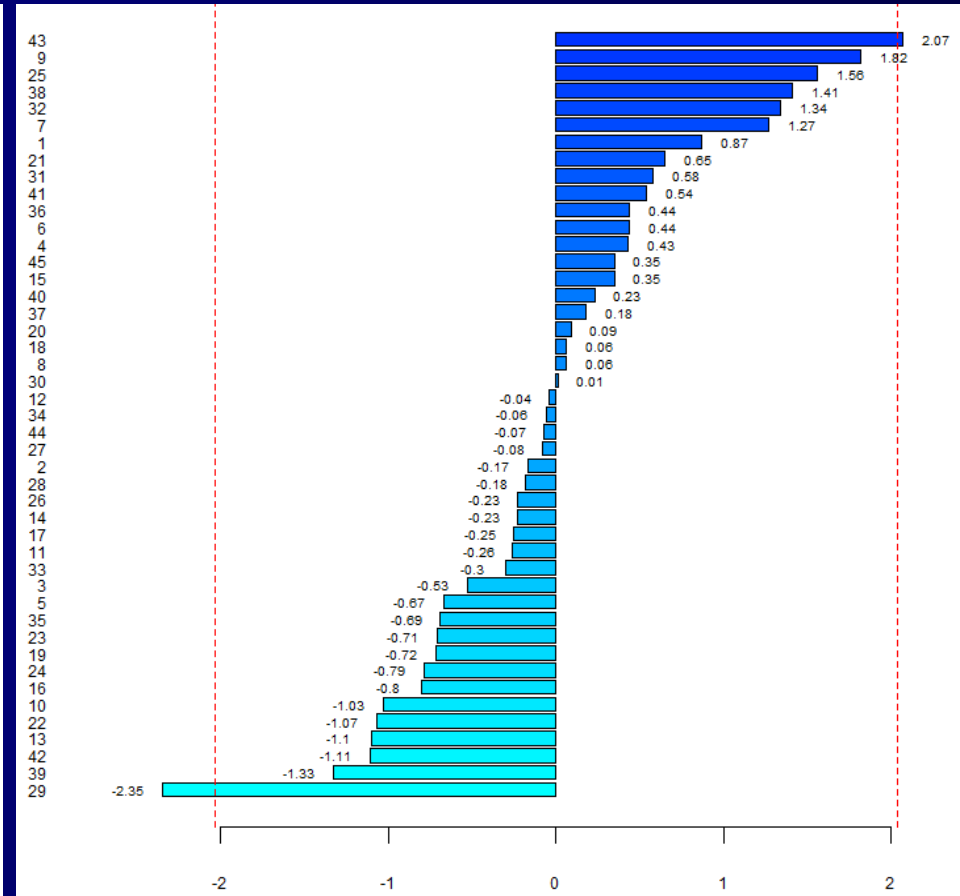
	Full Model		Reduced Model	
Predictor	Estimate(SE)	P value	Estimate(SE)	P value
Intercept	0.233(0.125)	0.07	0.234(0.12)	0.064
Y2005	-0.043(0.16)	0.79	-0.044(0.16)	0.79
Y2006	-0.167(0.18)	0.36	-0.167(0.18)	0.35
Y2007	-0.321(0.19)	0.09	-0.322(0.18)	0.086
Age ≤39	-0.195(0.07)	0.01	-0.195(0.07)	0.0096
Age 75+	-0.355(0.26)	0.17	-0.353(0.25)	0.17
Nonmetro	0.0019(0.02)	0.92		
Nonmicr_confirm	0.445(0.64)	0.49	0.442(0.63)	0.49
Unknown_confirm	2.501(0.64)	0.00042	2.490(0.62)	0.0003
Black	0.087(0.05)	0.069	0.085(0.04)	0.05
Other_race	0.016(0.04)	0.68	0.015(0.04)	0.69
nonhospital	0.162(0.07)	0.04	0.162(0.073)	0.03

Cervix Cancer: Plot of Standardized Residuals of Unknown Stage

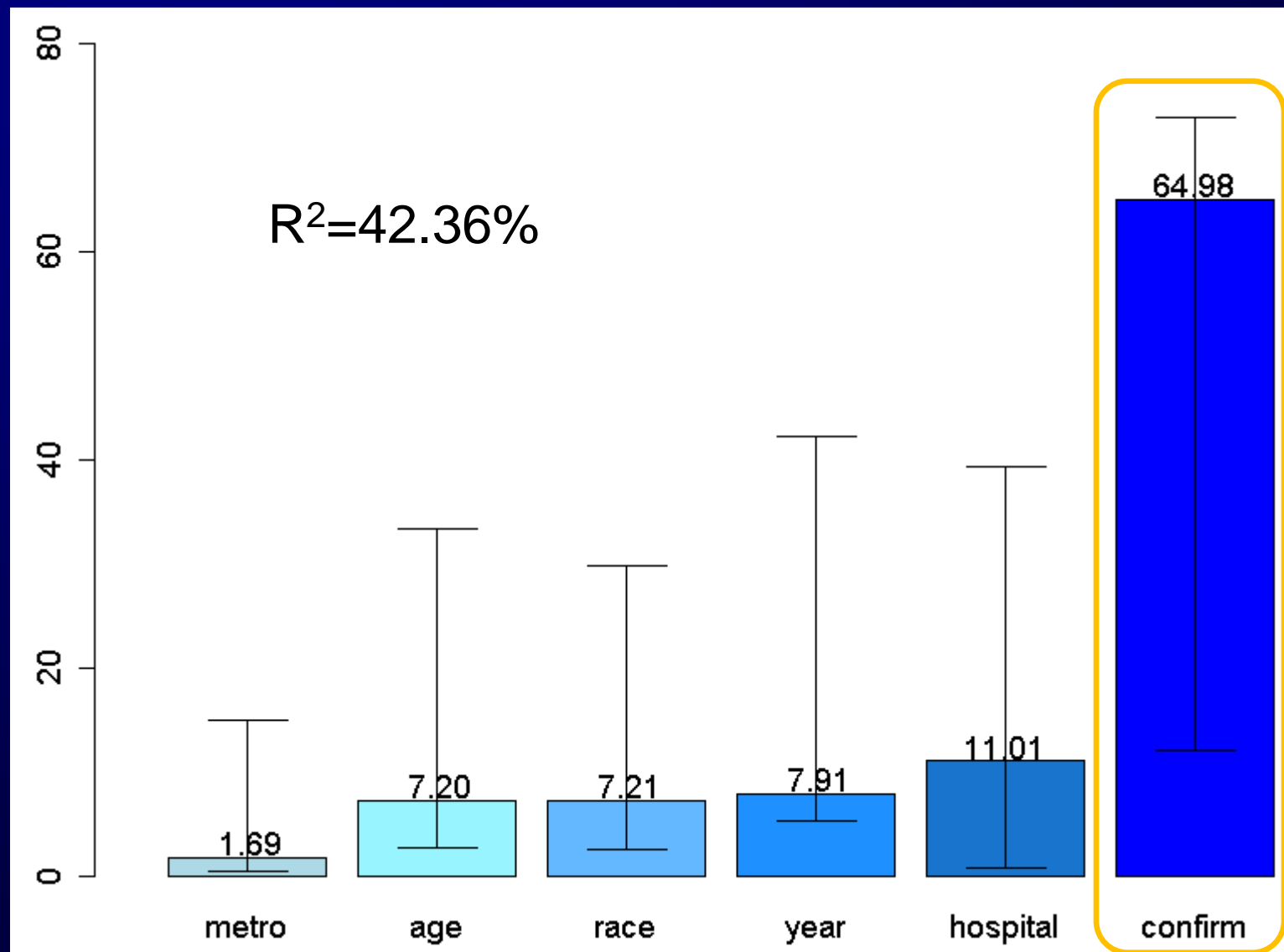
Before adjustment



After adjustment



Prostate Cancer: Variable Relative Importance (95 % CI) in Full Main Effects Model

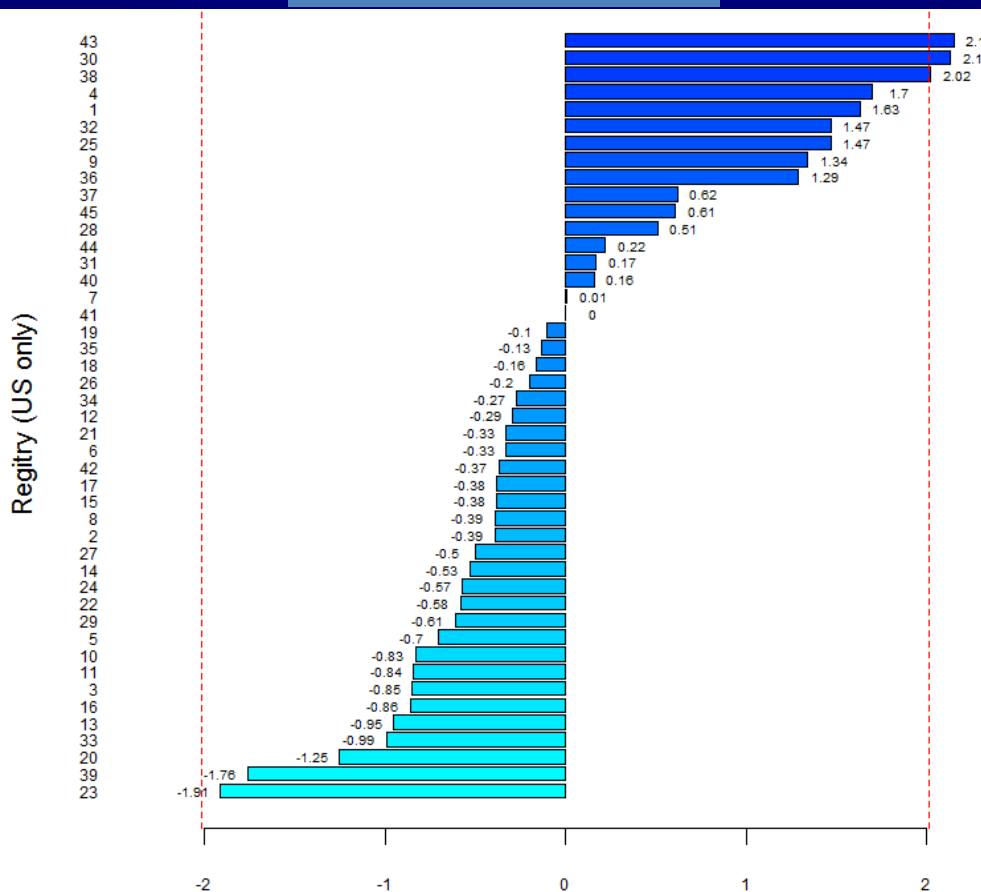


Prostate Cancer: Model Selection

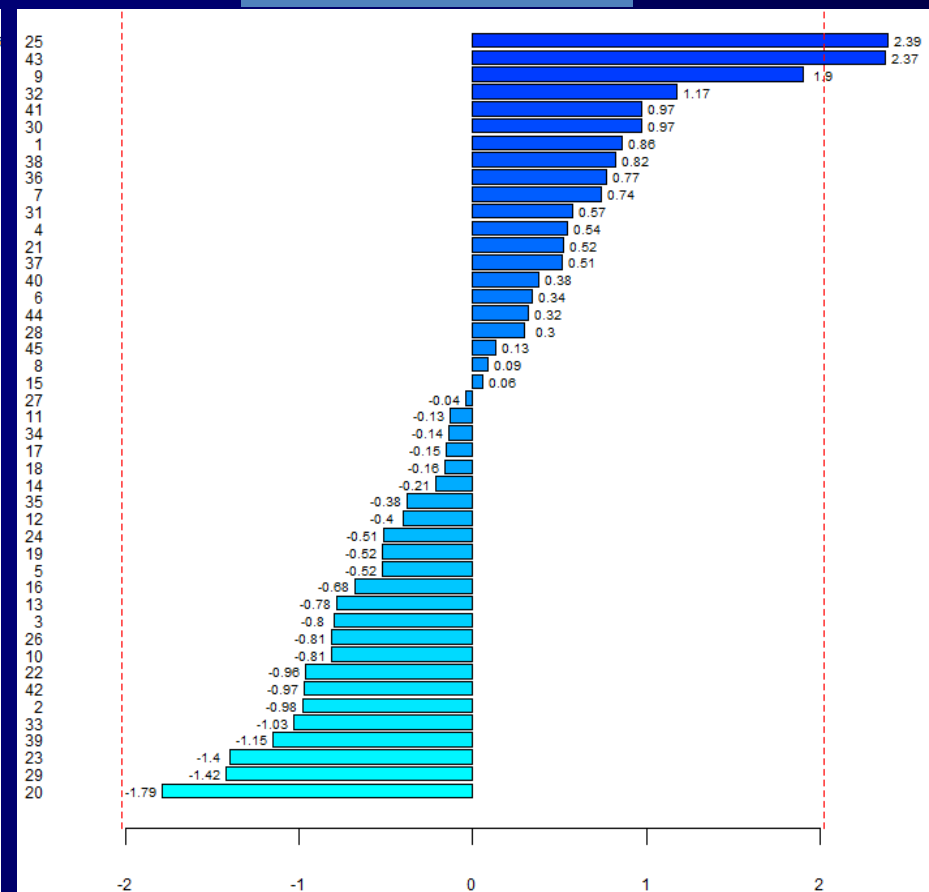
	Full Model		Reduced Model	
Predictor	Estimate(SE)	P value	Estimate(SE)	P value
Intercept	0.189(0.15)	0.22	0.029(0.01)	0.0005
Y2005	0.038(0.18)	0.83		
Y2006	-0.164(0.20)	0.41		
Y2007	-0.260(0.21)	0.22		
Age ≤49	-0.135(0.08)	0.11		
Age 80+	-0.145(0.34)	0.67		
Nonmetro	0.0012(0.02)	0.96		
Nonmicr_confirm	0.718(0.70)	0.31	0.078(0.59)	0.89
Unknown_confirm	2.730(0.66)	0.0002	2.214(0.60)	0.0007
Black	0.098(0.05)	0.06	0.040(0.038)	0.3
Other_race	0.026(0.04)	0.55	-0.010(0.04)	0.78
nonhospital	0.110(0.074)	0.15		

Prostate Cancer: Plot of Standardized Residuals of Unknown Stage

Before adjustment



After adjustment



Conclusions

- Overall, only a few registries have higher % of unknown stage and are varied by cancer site
- Unknown diagnostic confirmation was significantly associated with high proportion of unknown stage for all five types of cancers
- Non-hospital reporting source was positively related to high proportion of unknown stage for all cancer sites except for prostate cancer

Conclusions cont'd

- Year of diagnosis, age and race were significantly associated with the proportion of unknown stage for cervical cancer
 - Recently diagnosed cervical cases have lower proportion of unknown stage
 - Compared with age 40-74, younger patients (≤ 39) have lower proportion of unknown stage
 - Registries with higher proportion of black patients were most likely to have higher proportion of unknown stage