Supplementary Material

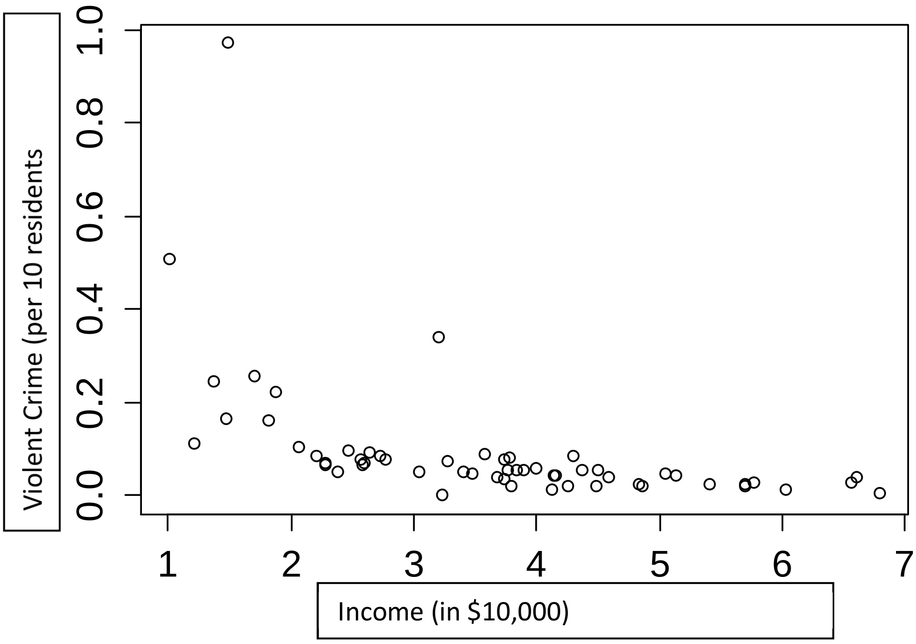
Additional Detail of Statistical Methods

Scaling of covariates

In order to facilitate interpretability of odds ratio estimates, continuous variables were rescaled (age divided by 10, NIHSS divided by 5). Using age in decades more closely links typical epidemiological considerations regarding changes in stroke risk (as the difference between age 60 and 70 is easier to conceptualize than age 60 and 61). Similarly, the NIHSS is often considered from comparing mild, medium and severe strokes, and a 5-point change (from 5 to 10) confers a clinically relevant change in severity, whereas a change from 5 to 6 does not. In addition, census level variables were rescaled as follows. The proportions (proportion over 65 and proportion Hispanic) were multiplied by 10 (therefore the parameter estimates represent a 10% increase in each of those covariates in the fitted models (e.g., 0.5 became 5). Median household income was divided by $10,000. Violent crimes per 10 residents of each census tract were calculated by dividing the number of violent crimes in each tract by the 2000 population of the census tract and then multiplying by 10. In addition, the models were computationally intense. The scaling allowed the Markov-Chain Monte Carlo fitting procedures to converge much more efficiently.

Hierarchical logistic regression and correlation between neighborhood income and crime

For the hierarchical logistic regression models (with exception of intercept only models), posterior estimates for the covariate parameters were calculated and converted into odds ratios. All continuous covariates were mean centered to facilitate the interpretability of the intercepts. Because of small numbers, stroke patients who were not Mexican-American or non-Hispanic white were excluded from the models. The Spearman correlation between neighborhood income and crime was 0.5 and is depicted via scatterplot in this figure.



As crime and neighborhood income were not perfectly correlated and variation in crime existed across moderate income neighborhoods we elected to include both in the models.

Predictive imputation model for education

Because of a high frequency of missing data for education, which was only collected in a subset of cases who had interviews, a non-parsimonious general linear model to predict individual level education when missing was used within the overall Bayesian hierarchical model. It is important to note, that in subjects without an interview, the other covariates (stroke severity, ethnicity, etc) were available from the medical record. Education data could be missing if the patient was not selected for interview or the interview was refused or unable to be conducted. Individual level of education was not available from 1931 cases. Of these 1070 were not approached based on the sampling techniques of BASIC at the time, 469 patients or proxies refused the interview, 333 could not be located, and 59 interviews were incomplete for other reasons. For the prediction model to impute education when missing, the predictors were census tract proportion with high school diploma, census tract proportion living in poverty, along with individual age, sex, and ethnicity. We fit models excluding subjects with missing data for education as well, the point estimates for all covariates were not different, but the confidence intervals were much wider given the large loss in sample size.

Bayesian model fitting

The models were fitted using WinBugs. Non-informative, diffuse priors were used for all parameters.

Below is model code for the full hierarchical model which imputes individual level of education when missing.

```
### model


model {



for (i in 1:n){
  y[i] ~ dbin(p.bound[i],1)
  p.bound[i] <- max(0,min(1,p[i]))

 logit(p[i]) <- alpha[hood[i]] +
              a[1]*(nih[i] - mean(nih[])) +
                        a[2]*(age[i]- mean(age[])) +
                        a[3]*nhw[i] +
                        a[4]*ich[i] +
                        a[5]*female[i] +
                        a[6]*hxstroke[i]+
           a[7]*(educ.cut[i] - mean(educ.cut[]))

 educ.cut[i] <- cut(educ[i]) #forces information to flow only from the imputation model to the model of interest

educ[i] ~ dnorm(mu.imp[i],tau.imp)
  mu.imp[i] <- b[1] + b[2]*HS[i] + b[3]*pov[i] +b[4]*age[i] + b[5]*female[i] + b[6]*nhw[i]


}

for (j in 1:n.hood){
  alpha[j] ~ dnorm(mu[j],tau)

   mu[j] <-          g[1] +
                        g[2]*(l2inc[j] - mean(l2inc[]))+
                        g[3]*(l2hisp[j]- mean(l2hisp[])) +
                        g[4]*(l2crime[j]- mean(l2crime[])) +
                        g[5]*(l2old[j] - mean(l2old[]))

  p.hood[j] <- 1/(1+exp(-alpha[j]))
  }



#tau ~ dunif(0,1000)
#tau.imp ~ dunif(0,1000)

tau ~ dgamma(0.001, 0.001)
tau.imp ~ dgamma(0.001, 0.001)
```

```
sigma2 <- 1 / tau

# use pi squared divided by 3 as the level 1 variance for a logistic regression model
vpc <- sigma2/(sigma2 + 3.289868)

for (i in 1:7){
        a[i] ~ dnorm(0, 1.0E-6)
   }

for (i in 1:5){
        g[i] ~ dnorm(0, 1.0E-6)
   }

for (i in 1:6){
    b[i] ~ dnorm(0, 1.0E-6)
}


}
```

Model convergence and intra-class correlation

The models were checked for convergence by inspection of graphical plots for all parameters; two chains were run with distinct initial values. For each model, the first 1000 iterations were discarded and then the following 9000 were retained. Posterior estimates for the parameters and variance from each model were obtained by sampling from the 18000 saved draws from the posterior. The intra-class correlation coefficient was estimated by dividing the tract level variance by the tract level variance plus $\pi^2/3$. (Reference for variance equation: Rodrıguez, G. and Elo, I., Intra-class correlation in random-effects models for binary data, The Stata Journal, 2003, 3(1):32-46.)

Model fit

Model fit was checked by comparing the Bayesian Information Criterion (analogous to the Akaike Information Criterion from frequentist logistic regression) for appropriately nested models. Lower values are better. As we moved from the empty model (intercept only), to the level 2 variables only model the BIC lowered indicating improved model fit. The addition of the level 1 variables created models not directly comparable (level 2 versus level 1 and 2) using the BIC and we did not sequentially check the model for fit after the addition of each level 1 variable.

Ranking of census tracts by EMS use: observed versus model based estimates

A secondary goal of this investigation was to determine whether we could identify neighborhoods that had less than expected EMS use for acute stroke. To determine this, the observed proportion arriving by EMS within each census tract was calculated and ranked. This ranking was repeated for the model-based estimated proportion arriving by EMS for each census tract (derived from intercept of the hierarchical logistic models) from the three models described above. The ten census tracts with the lowest proportion arriving by EMS based on observed use were compared to the ten census tracts with the lowest predicted EMS use from the final model. This allowed us to compare the observed variability between neighborhoods without adjustment and after controlling for individual and neighborhood level factors. This method of examining how rankings change before and after model based adjustment has previously been described in hospital rankings (Reference: Health Services Research Volume 45, Issue 6p1, pages 1614–1629, December 2010)