Application of Hierarchical Bayesian Models with Poststratification for Small Area Estimation from Complex Survey Data

Vladislav Beresovsky[1], Catharine W. Burt[1], Van Parsons[1], Nathaniel Schenker[1], Ryan Mutter[2]

[1]National Center for Health Statistics, 3311 Toledo Rd, Hyattsville, MD 20782

[2]Agency for Healthcare Research and Quality, 540 Gaither Rd, Rockville, MD 20850

Abstract

Small area estimation from stratified multilevel surveys is well known to be challenging because of extreme variability of survey weights and the high level of data clustering. These challenges complicate county- and state- level estimates of healthcare indicators such as proportions of visits with asthma and injury diagnoses at emergency departments (ED) from the National Hospital Ambulatory Medical Care Survey (NHAMCS). In this study, proportions of visits with asthma and injury diagnoses to hospital EDs were predicted by various multilevel logistic regression models and then aggregated to state level estimates. County level population covariates from the Area Resource File, hospital level covariates from Verispan Hospital Database and survey design information were used for modeling fixed effects. Aggregation of predicted hospital proportions to state level estimates utilizing the available number of ED visits to each hospital amounts to poststratification with cells defined at the state level. We evaluated models by comparing predictions with estimates based on administrative data from the Healthcare Cost and Utilization Project (HCUP) databases.

Key Words: health care utilization, small area estimation, hierarchical Bayesian model, multilevel logistic regression, poststratification

Introduction

The National Hospital Ambulatory Medical Care Survey (NHAMCS) is an annual national multilevel probability sample of visits to the emergency, outpatient, and ambulatory surgery departments of noninstitutional general and short-stay hospitals. It was designed to provide high-level healthcare utilization estimates for national, regional, and Metropolitan Statistical Area (MSA) versus non-MSA designated areas [1].

In addition to these traditional estimates there is also growing interest in estimates for smaller geographical localities such as states and counties. However, the possibility of producing reliable direct estimates in small areas using NHAMCS ED data is an open question because of limited coverage in small areas, small sample sizes within many covered small areas, and the high level of data clustering used in the sample design. For instance, 75% of all EDs are located in counties that were not sampled in NHAMCS 2007 and another 15% of EDs are located in counties with only one or two sampled EDs. Coverage for states is somewhat better, yet estimates for many states are still unreliable and some states have no sampled hospitals in the NHAMCS.

Apparent limitations of direct estimates in small areas suggest a need to consider model-based methods. There are numerous applications of hierarchical linear and logistic models for estimation of small areas [2]-[5]. It has been demonstrated that models with random effects better account for data variability in small areas than fixed effects models and avoid inefficiencies of direct estimates in small areas with a small number of observations.

In this paper we use hierarchical logistic regression models to estimate state-level proportions of visits with injury and asthma diagnoses. We use hierarchical Bayesian methods to obtain posterior distributions of model parameters and predictions of proportions in hospitals. When variability between small areas is small compared to variability within small areas, the hierarchical Bayesian estimator seems to have the smallest mean square error (MSE) compared to empirical Bayesian and direct estimators [6]. Predicted hospital level proportions are aggregated to the state levels using the available number of ED visits to each hospital from administrative data files. These aggregation procedures are similar to poststratification described in detail elsewhere [3]-[5].

In order to evaluate the validity of models for the small area estimation (SAE), estimated state level proportions are compared with proportions calculated from the administrative data available from the HCUP databases [7] which represent true population estimates from 26 states participating in HCUP.

The hierarchical Bayesian estimation routine used to produce a posterior distribution of state level predictions makes possible the estimation of posterior means, standard deviations and credible intervals. Comparisons with the true population proportions from the HCUP databases facilitates obtaining measures of model fit, such as bias, root mean square error (RMSE) and coverage by credible intervals for each state.

We obtain estimates and compare performances of various models which differ by application of fixed and random effects and compare model-based estimates with direct estimates of proportions and standard errors. Results of these comparisons are summarized to obtain systematic conclusions about the reliability and feasibility of model-based and direct estimates in small areas.

1. Methods

1.1. Data

To achieve greater precision in estimating model parameters we combined NHAMCS ED data for 2006 and 2007. Combined data included 362 hospitals from the 2006 sample and 337 hospitals from the 2007 sample. Having a larger number of sampled hospitals is expected to improve the efficiency of estimation. Some of the hospitals (250) were sampled in both years, but studied proportions may still vary between years due to the seasonality of the data collection process [1]. In such cases proportions from both years were modeled independently.

For each sampled ED visit, NHAMCS classifies and codes up to three provider diagnoses DIAG1-DIAG3 according to the International Classification of Diseases, version 9, Clinical Modification (ICD-9-CM, www.cdc.gov/nchs/icd/icd9cm.htm); whereas the HCUP state databases can have many more visit diagnoses. There were challenges in finding ways to identify visits with injury and asthma diagnoses in NHAMCS and HCUP so estimates from both sources would be comparable. We compared direct estimates of proportions from NHAMCS data with HCUP at national and regional levels and found that the closest match existed between asthma proportions estimated from the principal diagnoses in both NHAMCS and HCUP. For visits with injury diagnoses the closest match was found between proportions estimated from all listed diagnoses in NHAMCS and the first two recorded diagnoses in HCUP. This approximate compatibility of direct estimates for larger domains provides the justification for using HCUP state level estimates for validation of predicted proportions of asthma and injury related visits from NHAMCS data using model-based methods. Comparative data are available from the first author.

Covariates for proposed models included county level variables available from the 2007 Area Resource File distributed by Health Resources and Services Administration (HRSA). We also used hospital-level covariates from the 2006 and 2007 Verispan Hospital Databases. This additional information helps to explain variability between hospitals and provides for developing models with higher sensitivity. Since census region and MSA status defined stratification of geographical PSUs, they were also used as model covariates to account for survey design.

All continuous covariates were standardized by centering and normalizing based on the arithmetic mean and standard deviation:

$$X^{STD} = \frac{\left(X - \overline{X}\right)}{StdDev\left(X\right)} \tag{1}$$

In addition to this linear transformation, outliers were truncated at the 99th percentile to improve robustness of the model. Standardizing covariates in many cases improved the convergence of numeric algorithms; model parameters were measured on the same scale, making them more easily interpretable.

The dependent variable in the left hand side of the modeling equations was weighted hospital-level proportion of visits with either asthma or injury diagnosis:

$$p_i = \frac{\sum_{l \in i} w_{il} d_{il}}{\sum_{l \in i} w_{il}} \tag{2}$$

where w_{il} are survey weights and $d_{il} = 1$ for asthma (injury), 0 otherwise.

1.2. Logistic and logistic-normal models

Logistic regression models with normally distributed random effects, simulating possible clustering in the data distribution, are commonly used to estimate small area proportions [2]-[5]. Since covariates are available for each hospital and county in the population, we modeled proportions at the hospital level and then aggregated them to the state level using the known number of ED visits to each hospital within states. Using fixed effects allowed modeling of the first moments of the outcome variables in the absence of correlations in the variance-covariance matrix. Using random effects in the model accounted for clustering effects from the sample design in outcome variables. We explored how fixed and random effects influenced predicted proportions in states as well as associated variances and credible intervals. In the course of this study we considered the following list of models which differ by the use of fixed and random effects.

Complete pooling model (1)
The simplest model to consider is a "complete pooling" model (Gelman [8]). This is a simple logistic regression model without random effects including only fixed effects specified for all hospitals in the population. Information is pooled from the complete sample, and small area predictions are not directly affected by data from an individual hospital or state.

$$\operatorname{logit}\left(p_i\right)=\alpha+\mathbf{X}_i\boldsymbol{\beta} \tag{3}$$

where p_i is the proportion of asthma or injury visits to hospital i and $\mathbf{X}_i$ are county and hospital level covariates.

No pooling model (2)
The name for this model was also borrowed from Gelman [8]. This model also does not include random effects, but in addition to covariates used in "complete pooling" model, it also includes identifiers of individual states α_j as intercepts.

$$\operatorname{logit}\left(p_{ij}\right)=\alpha_j+\mathbf{X}_i\boldsymbol{\beta} \tag{4}$$

where p_{ij} is proportion of asthma or injury for hospital i within state j.

Random effect on hospital level (3)
This is a two-level hierarchical model accounting for additional correlation of visits within hospitals by including in the model random effects at the hospital level $a_{(i)}$

$$\operatorname{logit}\left(p_i\right)=\alpha+\mathbf{X}_i\boldsymbol{\beta}+a_{(i)} \tag{5}$$

Random effect on state level (4)
This model is similar to the previous one, but it accounts for the correlation of visits within states by including state level random effects $a_{(j)}$

$$\operatorname{logit}\left(p_{ij}\right)=\alpha+\mathbf{X}_i\boldsymbol{\beta}+a_{(j)} \tag{6}$$

Random effect on state, county and hospital levels (5)
By including in the model random effects at the state $a_{(j)}$, county $a_{(k)}$ and hospital $a_{(i)}$ levels, we accounted for all possible clustering in the distribution of visits with injury or asthma diagnoses.

$$\operatorname{logit}\left(p_{ikj}\right)=\alpha+\mathbf{X}_i\boldsymbol{\beta}+a_{(j)}+a_{(k)}+a_{(i)} \tag{7}$$

Though "complete pooling" and "no pooling" models include only fixed effects, similar results can be obtained from the hierarchical models by enforcing extreme values of the variance of the random effect. In the case of the "complete pooling" model this variance must be infinitesimally small, restricting all deviations from predictions by fixed effects and completely ignoring all information from data in small areas. For the "no pooling" model, the variance must be extremely large, so model predictions do not deviate from the maximum likelihood estimates in small areas.
We examined how the SAE proportions predicted by different models compared to true state level proportions from HCUP data and classified various models according to their structure and performance.

1.3. Hierarchical Bayesian estimation

Although there are many ways to estimate parameters of mixed models and make predictions for small areas, in this study we used a hierarchical Bayesian approach. The main advantage of this method is the ability to compute posterior distributions of model parameters and predicted proportions for small areas. By knowing the posterior distribution of predicted proportions and the true state level proportions from HCUP, it is possible to estimate measures of model performance including bias, root mean square error (RMSE) and credible intervals for each state, and to determine average coverage of HCUP proportions by predicted credible intervals. The empirical Bayesian method also generates the posterior distribution of predicted proportions, but its results heavily depend on empirical estimation of variances in small areas. In addition, it does not account for errors of empirical variances. As a result, it has the well known problem of underestimating variances of estimated model parameters [9].
On the other hand, results of the hierarchical Bayesian method can be sensitive to the assumptions made about the prior distribution of model parameters. The usual way to minimize the influence of prior selection is to use noninformative priors, with variability far exceeding expected true values. Since covariates are normalized, values of model parameters are close to each other and are not much larger than 1. Normally distributed diffuse priors with variance 100 were used for intercepts and slope coefficients. For random effects normally distributed priors with mean 0 and variance uniformly distributed in interval $\left[0,100\right]$ were also selected. Being concerned about the possible effect of prior selection on the outcome, we compared the means of the posterior distribution of model parameters with estimates obtained by using likelihood-based methods implemented in the R procedure glmer. Both methods produced very similar results.
To obtain the posterior distribution we used the WinBugs software utilizing Bayesian inference using Gibbs sampling (BUGS) [10]. Good convergence of the algorithm was achieved by simulating three chains of 6000 iterations, half of which were discarded as "burn in", insuring independence of posterior distribution from arbitrary selection of the model parameters initiating each chain of iterations. To achieve independence of observations in the simulated posterior distributions, observations were thinned by a factor of 10. Resulting posterior distributions of hospital proportions used for estimation contained about 900 simulated values from three chains of iterations.

1.4. Aggregation of predicted hospital proportions to state level

Predicted hospital proportions need to be aggregated to state level for **c**omparison with proportions calculated from HCUP databases. From the Verispan Hospital Database we know the number of visits N_{ij} to every hospital i within small area j. Aggregation of hospital proportions to states can be viewed as a poststratification process in which population covariates N_{ij} are used as weights for projecting predicted proportions in hospitals $\hat{p}_{ij}$ to state proportions $\hat{P}_j$:

$$\hat{P}_j = \frac{\sum_{i\in j} N_{ij}\hat{p}_{ij}}{\sum_{i\in j} N_{ij}} = \sum_{i\in j} w_{ij}\hat{p}_{ij},\, w_{ij} = N_{ij} \Big/ \sum_{i\in j} N_{ij} \tag{8}$$

1.5. Comparing posterior distribution of state level proportions with HCUP

Having posterior distributions as the output for the modeling process is very convenient for comparison with perceived true results from administrative data available from HCUP. From state level posterior distributions of proportions of visits with asthma and injury diagnoses it is possible to calculate the mean, standard deviation and 50% and 95% credible intervals. Assuming HCUP proportions are true population values for 26 states, bias and RMSE for state j can be calculated as the expectation over the posterior distribution of predicted state proportions $P_j^{Posterior}$:

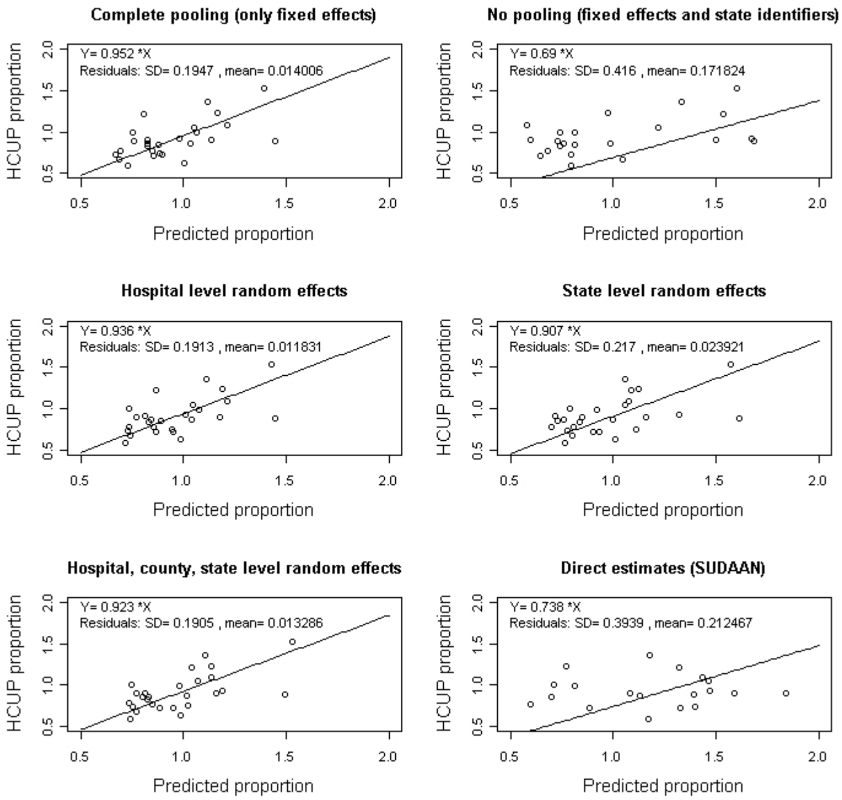
$$\begin{aligned} Bias_j &= E_{Posterior}\left(P_j^{Posterior} - P_j^{HCUP}\right) \\ RMSE_j &= \left(E_{Posterior}\left(P_j^{Posterior} - P_j^{HCUP}\right)^2\right)^{1/2} \end{aligned} \tag{9}$$

In addition to model based estimates we also present results of direct estimates obtained using SUDAAN software for analyzing complex survey data. Stochastic assumptions of randomization-based estimation allow estimation of means, standard deviation and confidence intervals, all of which make comparing various aspects of direct and model based estimations possible.

2. Results

Comparison of the goodness of fit of estimates by different SAE models is shown in Figures 1.a and 1.b, which present scatter plots of HCUP proportions of asthma and injury visits versus the means of posterior distributions of predicted proportions or direct estimates, with regression line drawn without intercept.

(a) ASTHMA


Complete pooling (only fixed effects)
Y= 0.952 *X
Residuals: SD= 0.1947 , mean= 0.014006
No pooling (fixed effects and state identifiers)
Y= 0.69 *X
Residuals: SD= 0.416 , mean= 0.171824
Hospital level random effects
Y= 0.936 *X
Residuals: SD= 0.1913 , mean= 0.011831
State level random effects
Y= 0.907 *X
Residuals: SD= 0.217 , mean= 0.023921
Hospital, county, state level random effects
Y= 0.923 *X
Residuals: SD= 0.1905 , mean= 0.013286
Direct estimates (SUDAAN)
Y= 0.738 *X
Residuals: SD= 0.3939 , mean= 0.212467
HCUP proportion
Predicted proportion

(b) INJURY

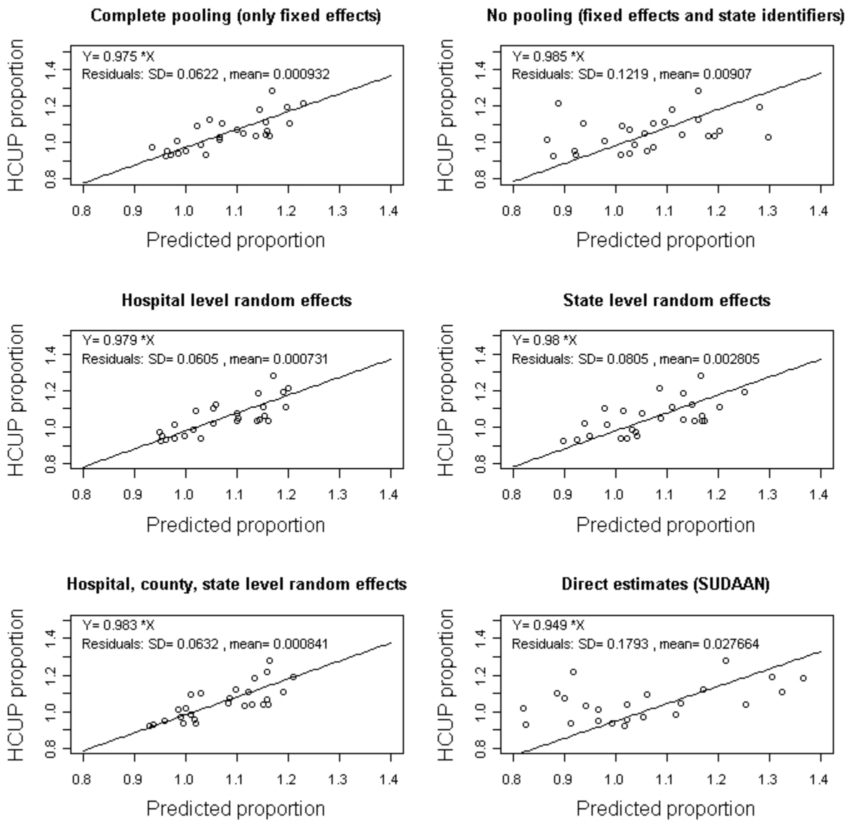
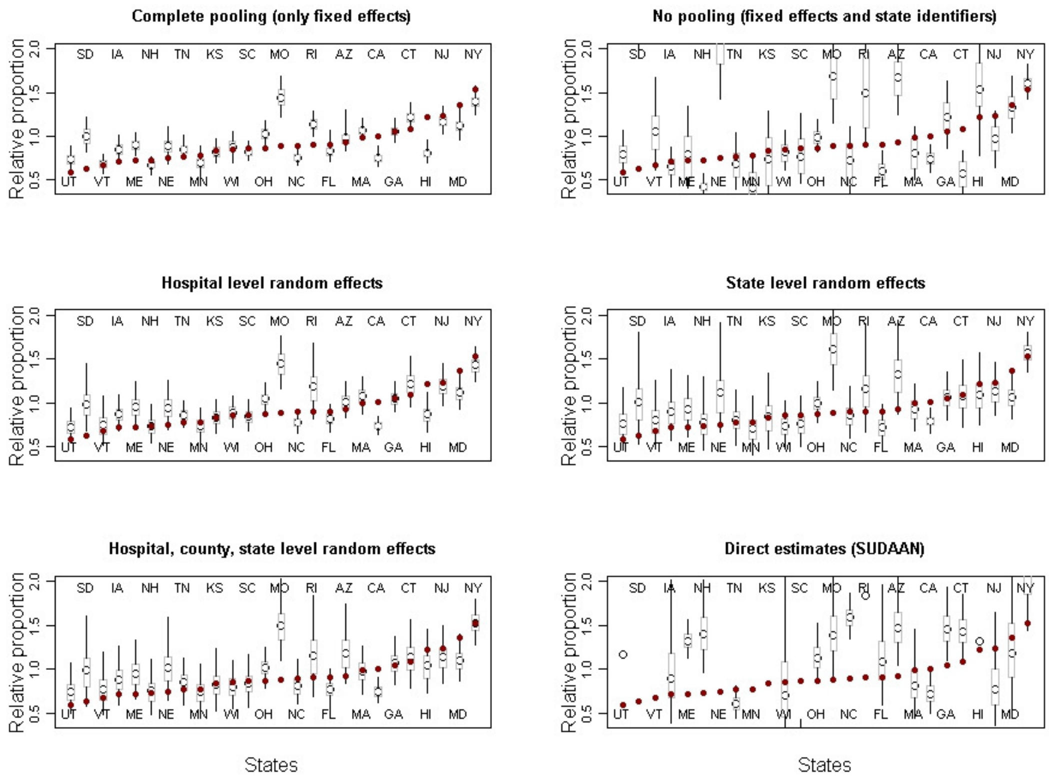


Figure 1: State level HCUP proportions of asthma (a) and injury (b) visits vs predictions using different methods. Asthma and injury proportions are relative to the corresponding national average.

Because the national average of proportion of visits with injury diagnoses (25.41%) is much higher than the proportion of visits with asthma (1.55%), state level predictions for injury fit HCUP proportions much better than predictions for asthma. Standard deviation of residuals for injury models is approximately 3 times lower than for asthma models. However, both sets of results show similar trends, which indicate that predictions by the complete pooling model and models with hospital and hospital/county/state level random effects fit HCUP proportions at similar levels. The model with random effects at the state level has a somewhat weaker fit. The poorest fit is observed for randomization-based estimates, and the "no pooling" model is just a little better. In general, estimates based on information pooled across all states had a better chance of fitting HCUP data than estimates relying more heavily on state data, while ignoring hospital-level clustering.

The next set of results (Figures 2.a and 2.b) display means, credible intervals and coverage of HCUP proportions for individual states.

(a) ASTHMA


Complete pooling (only fixed effects)
No pooling (fixed effects and state identifiers)
Hospital level random effects
State level random effects
Hospital, county, state level random effects
Direct estimates (SUDAAN)
Relative proportion
States

(b) INJURY

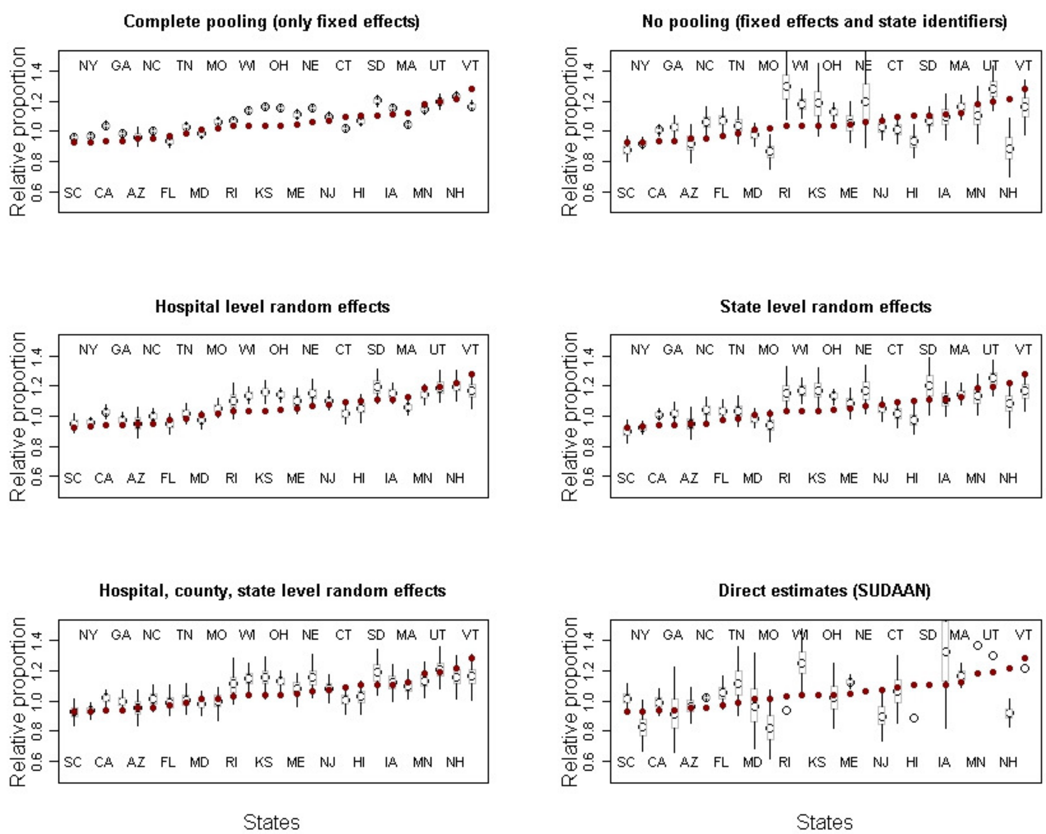

Complete pooling (only fixed effects)
No pooling (fixed effects and state identifiers)
Hospital level random effects
State level random effects
Hospital, county, state level random effects
Direct estimates (SUDAAN)
Relative proportion
States

Figure 2: Estimates of state level proportions of asthma (a) and injury (b) visits with 50% and 95 % credible intervals (filled circles show HCUP proportions). Asthma and injury proportions are relative to the corresponding national average.

Estimates of proportions of injury related visits are much more efficient than for asthma visits. Values of standard deviation and the width of credible intervals are approximately 3-4 times smaller for all estimation methods of injury proportions than for corresponding estimates of asthma proportions. As expected, in both cases credible intervals of the “complete pooling” model are the narrowest and direct estimates are the widest. Although means of posterior distributions of predicted proportions by models with hospital and hospital/county/state random effects are not much different from the “complete pooling” model, credible intervals of models with random effects are substantially wider and, consequently, coverage of HCUP proportions is much better. Because of the inefficiency of direct estimates and the “no pooling” model the coverage of HCUP proportions is not very good, despite the fact that confidence intervals are wider than of other methods. In conclusion, efficiency of estimates placing more weight on information coming from individual states rather than pooled across the whole sample and ignoring clustering at hospital level is inferior to other methods of estimation.
These conclusions are summarized in the Tables 1.a and 1.b comparing for all methods of estimation RMSE, standard deviation and coverage of HCUP proportions by credible intervals (CI) averaged over 26 states for which HCUP proportions of asthma and injury related visits are available.
For easier comparison, measures of the efficiency of various methods- RMSE and standard deviation - are expressed as ratios over the same values for the “complete pooling” model. The average coverage for each estimation method is determined as the percent of states for which HCUP proportions are within credible intervals from the predicted proportions. RMSE and standard deviation ratios are small, and the average coverage is high for the efficient estimation method. The presented average measures do not satisfy rigorous statistical standards since HCUP reports only 26 states, but they do allow a high level of comparison and categorization of estimation techniques.

Table 1: Visits with asthma (a) and injury (b) diagnosis: relative RMSE, standard deviation and coverage of HCUP proportions by CI , averaged over states with available HCUP proportions.
(a) ASTHMA

Model	Average RMSE	Average standard deviation	Average coverage by CI 50%	Average coverage by CI 95%
Complete pooling (only fixed effects)	100%	100%	23%	62%
Hospital level random effects	115%	148%	31%	77%
Hospital, county, state level random effects	138%	221%	46%	92%
State level random effects	154%	242%	38%	92%
No pooling (fixed effects and state identifiers)	266%	298%	31%	62%
Direct estimates (SUDAAN)		321%	19%	38%

(b) INJURY

Model	Average RMSE	Average standard deviation	Average coverage by CI 50%	Average coverage by CI 95%
Complete pooling (only fixed effects)	100%	100%	8%	27%
Hospital level random effects	114%	243%	12%	73%
Hospital, county, state level random effects	132%	350%	42%	88%
State level random effects	149%	329%	19%	77%
No pooling (fixed effects and state identifiers)	197%	412%	19%	62%
Direct estimates (SUDAAN)		557%	21%	54%

3. Discussion, conclusions and further research

The obtained results suggest that accounting for direct estimates in small areas versus pooling information from the whole sample will yield different predictions for small areas. Methods relying more heavily on pooling information can be considered "pessimistic" in a sense that they assume that local data cannot be used for small area estimation and only synthetic estimation has any value. As a result, these methods have artificially small standard deviation and narrow credible intervals. Methods giving higher importance to data available in small areas can be considered "optimistic". They have more realistic standard deviations and wider credible intervals, corresponding to the amount of data available in each small area. According to this classification, we ordered the methods from more "pessimistic" to more "optimistic" as follows:

Complete pooling (only fixed effects);
Hospital level random effects;
Hospital, county, state level random effects;
State level random effects;
No pooling (fixed effects and state identifiers);
Direct estimates (SUDAAN).

The relative efficiency of each method depends on the importance of available population covariates and the amount and quality of data in small areas. Since NHAMCS ED sample data are very sparse and highly clustered at the hospital level, available data in small areas do not provide good information for small area estimation. As a result, the most "optimistic" methods -"no pooling" model and direct estimates are very inefficient and demonstrate inferior coverage of HCUP proportions. On the other hand, the "complete pooling" model relies only on population covariates, ignores variability of data in small areas, and has the narrowest credible intervals and the lowest percent of coverage.
The numerical results of this study suggest that the model including random effects at the state, county and hospital levels performs better than other considered methods for estimating state-level proportions of injury and asthma diagnoses. Its credible intervals

are narrower than the more "optimistic" methods, but percent of coverage is the largest for both injury and asthma related visits. The fact that estimates from this model are noticeably better than estimates from the model having just state level random effects stresses the importance of accounting for hospital level clustering.
It will be interesting to conduct a simulation study to observe under which conditions the "optimistic" methods relying on data in small areas will become more efficient than "pessimistic" methods. The relevance of using small area data for small area predictions varies depending on sampling factors, the correlation within small areas, and the extent to which the sampling process makes sampled data in small areas different from the actual population.

References

[1] National Center for Health Statistics. Description of the National Hospital Ambulatory Medical Care Survey (NHAMCS) ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHAMCS/.
[2] Malec, D., Sedransk, J., Moriarity, C.L., and LeClere, F.B., 1997. Small Area Inference for Binary Variables in the National Health Interview Survey, *JASA* 92, 815-826.
[3] Gelman, A., Little , T.C., 1997. Poststratification Into Many Categories Using Hierarchical Logistic Regression, Survey Methodology 23, 127-135.
[4] Gelman, A., 2007. Struggles with Survey Weighting and Regression Modeling, Statistical Science 22, 153-164.
[5] Lazzeroni, L.C., Little, R.J.A., 1998. Random-effects Models for Smoothing Poststratification Weights, Journal of Official Statistics 14, 61-78.
[6] Casella, G., An Introduction to Empirical Bayes Analysis, 1985. The American Statistician 39, 83-87.
[7] Agency for Healthcare Quality and Research. National Emergency Department Sample (NEDS). http://www.hcup-us.ahrq.gov/nedsoverview.jsp
[8] Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models, *Cambridge University Press*.
[9] Hall, P. and Maiti, T., 2006. On parametric bootstrap methods for small area prediction, *J. R. Statist. Soc. B* 68, 221-238.
[10] Spiegelhalter, D.J., Thomas, A., Best, N., Gilks, W., and Lunn, D., (1994, 2002). BUGS: Bayesian inference using Gibbs sampling. *MRC Biostatistics Unit, Cambridge*, England.