Hospital and Physician Quality Assessment: Case Study 2

Sun, Muyao

11/11/2017

Introduction

Choosing a hospital is a complex decision reflecting personal needs and preference. However, a clear ranking of hospitals provide a starting point to learn their quality and compare hospitals locally. The Centers for Medicare and Medicaid Services plans to have a systematic way to evaluate hospitals' performance. Also, assessing the performance for physicians in hospitals can be helpful for both patients and health care providers. In order to rank hosiptals and physicians, I utilize the data on outcomes of cardiac surgical procedures in New York State from 2009-2011 and build a binomial generalized linear mixed model to measure their performance, mainly relied on the mortality rate within 30 days.

Data Exploration

The dataset on outcomes of cardiac surgical procedures includes 361 records and each record contains physician name, hospital name, regions in NY state, cardiac surgical procedure type (Valve or Valve/CABG involving valve replacement and CABG not involving valve replacement), number of cases taken by the physician, number of deaths, observed mortality rate per 100 cases, and expected mortality rate per 100 cases. The expected mortality rate takes account for the health status of patients undergoing each procedure, avoiding penalize physicians who have received many severe and complex cases.

No missing values exist in this data set. Physician name, hospital name, regions in NY state and procedure type are coded into categorical variables, and each contains 140, 40, 10 and 2 levels, respectively. Manhattan is the region occupying the largest proportion of records (32%) with total 8 hospitals and 47 physicians. Cardiac surgical procedures are reported as Valve or Valve/CABG with valve replacement (51%) and CABG without valve replacement (49%). The rest four variables related to outcomes are treated as numeric values. Table 1 summarizes these four variables separately for each region. Manhattan still has the highest total number of cases and deaths, but Kings has the highest average observed mortality rate and Long Island has the highest average expected mortality rate.

Table 1: Outcomes of cardiac surgical procedures

Region	Cases(total)	Deaths(total)	Obs MR(avg)	Exp MR(avg)
Bronx	3022	77	1.8713043	2.347826
Capital District	6519	119	1.4248571	1.765429
Central NY	8015	193	2.3800000	2.276250
Kings	3047	72	2.5403571	2.322500

Region	Cases(total)	Deaths(total)	Obs MR(avg)	Exp MR(avg)
Manhattan	17627	473	2.2902609	1.831478
NY Metro - Long Island	14913	374	2.3128261	2.496522
NY Metro - New Rochelle	5053	66	1.0741667	1.815417
Queens	2394	29	0.9230769	1.891539
Staten Island	2250	46	1.4112500	1.662500
Western NY - Buffalo	7147	128	1.3121429	1.554286
Western NY - Rochester	6532	187	2.3417647	1.904706

Among 361 records, 140 unique physicians and 40 unique hospitals are included, which indicate that one physician may have several records with different surgical procedure or in different hospital, equivalently, the data is correlated. Also, the values in number of cases and number of deaths cell varies dramatically, even for a same physician. A physician can take many cardiac surgeries in one hospital but rare cases in another one. Thus, a model can deal with correlated data and be able to borrow information within group is required in order to gain insight from this data set.

Method

The outcome of interest is whether or not a patient taken by a physician in a hospital after a cardic procedure die. A specific interest is to evaluate each physician's performance and each hospital's performance based on the mortality rate. In this analysis, I use Bayesian binomial generalized linear mixed model for the outcome Y_{ijk} of death for patient taken by physician i in hospital j with cardiac surgical procedure k. Mixed effect model is a robust analytical approach to model the sources of correlation in hierarchical data and allow me to determine which additional random components should be included in my model. The Bayesian version can be modeled with prior belief and return a posterior distribution of fixed and random effects instead of just point estimates, which allows more analysis.

Three explanatory variables (region, surgical procedure and interaction between these two) and two random intercepts (hospital and physician) are included in the model. Random effects on different levels of the model (physician and hospital) are assumed independent. The model is defined as following:

$$logit(Pr(Y_{ijk} = 1)) = \beta_0 + \beta_1 Region_j + \beta_2 Procedure_{ijk} + \beta_3 Region_j \times Procedure_{ijk} + b_{0_i} + b_{0_j}$$

where b_{0_i} and b_{0_j} follow $N(0, \sigma^2)$. Before fitting the model, the priors for $(\beta_1, \beta_2, \beta_3)$ are set to $N(0, 2^2)$ with default uniform prior on β_0 and half-t prior with degree of freedom 3 and scale parameter 10 on σ.

Rhat in model summary is considered as an indicator to posterior convergence information. With the default setting of Baysian mixed model in R (package brms), the Rhat for several parameters is larger than 1, which requires a strong prior or more iterations to converge. Thus the model (bm1) in this analysis is fitted using 4 chains, each with 4000 iterations (the first 2000 are warmup), leading to a total of 8000 posterior samples for each parameter.

Result

Two random effects are included in this model. The standard deviations of the random effects reflect how much variability there are between individuals across all records. The standard deviation for random intercept physician is 0.11 (0.006,0.268) and random intercept hospital is 0.37 (0.287,0.467), telling that $\frac{0.11^2}{0.11^2+0.37^2+\sigma_\mu^2} \times 100\%$ of total variance is explained by physician difference and $\frac{0.37^2}{0.11^2+0.37^2+\sigma_\mu^2} \times 100\%$ of total variance is explained by hospital difference where σ_μ is the standard deviation of between groups.

Three fixed effects are included in this model. Since explanatory variable region has 10 levels (using Queens as referent), total 22 parameter estimates are given by the model. For example, the estimate on procedure Valve/CABG is 0.38. For a specific physician in a specific hospital with use of Valve/CABG in Queens, having higher odd of being dead (exp(0.38)) comparing to (s)he in same hospital with use of CABG. The detailed infomation about odds ratio (posterior median and 95% credible intervals) comparing the odds of death in each region to that in Queens, separately for each procedure type is shown in Fig 1.
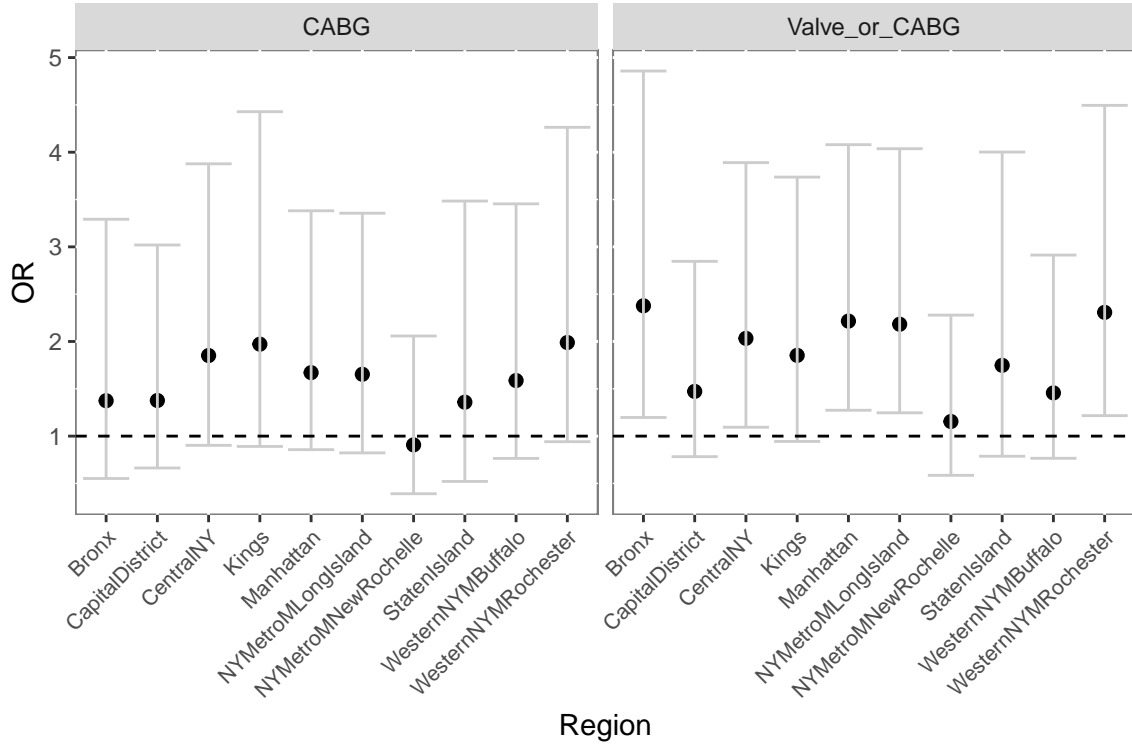

Figure 1: Odds ratio comparing the odds of death in each region

According to the above plot, Rochester has the highest odds ratio of death with use of CABG relative to Queens; but with the use of Valve or CABG, Bronx has the highest odds ratio. Since each physician and hospital have their own favor on procedures, a same region can have different odds ratio of death comparing to Queens with different use of procedures. Bronx is one of wealthiest counties in NY. People living in this county are more likely to attend hospitals in this area and they may have more tendency or be more financial capable to receive procedure involving valve replacement (1898 Valve cases and 1124 CABG cases). This may one reason to explain why Bronx has the highest odds ratio of death with use of Valve but a relative low odds ratio with use of

CABG. I still need to go through the region and explore hospitals and physicians to have a better idea of their performance.

Comparing the model-predicted mortality rate to already given expected mortality rate can give me an understanding of how the physician in a hospital performs. If the model-predicted mortality rate is higher than the expected mortality rate (or the ratio of the model-predicted mortality rate to the expected mortality rate is greater than 1), the physician performs worse than the expected given the illness of his or her patients. All the ratio of unique physician and hospital combination in each region is shown in Fig 3 (in last page), separately for two different procedures.

According to the posterior probability that the ratio of model-predicted to the expected is greater than 1, top 5% physician and hospital combinations and bottom 5% combinations are listed in below tables.

Table 2: Bottom 5% the posterior probability that the ratio is greater than 1

names	posterior_median	posterior_conf	prob_excess
Mosca.R.S (NYU.Hospitals.Center)	3.7501	(1.66,7.99)	0.9990
Knight.P (Strong.Memorial.Hosp)	1.5960	(1.19,2.11)	0.9990
Stelzer.P (Mount.Sinai.Hospital)	2.2296	(1.34,3.57)	0.9991
Schwartz.C.F (Bellevue.Hospital.Ctr)	2.6439	(1.55,4.39)	0.9994
Ciuffo.G.B (Mount.Sinai.Hospital)	2.1753	(1.41,3.27)	0.9994
Tortolani.A (NYP-.Weill.Cornell)	3.2886	(1.76,5.99)	0.9998
Reddy.R.C (Mount.Sinai.Hospital)	2.0233	(1.42,2.84)	0.9999
Bell-Thomson.J (Erie.County.Med.Ctr)	3.4967	(2.06,5.79)	1.0000
Bell-Thomson.J (Erie.County.Med.Ctr)	4.6719	(2.89,7.36)	1.0000
Loulmet.D.F (Lenox.Hill.Hospital)	4.7085	(2.63,8.06)	1.0000
Weinstein.S (Montefiore.-.Moses)	7.1988	(3.04,16.12)	1.0000
Nguyen.K (Mount.Sinai.Hospital)	6.3394	(2.9,14.13)	1.0000
Reddy.R.C (Mount.Sinai.Hospital)	2.4059	(1.6,3.54)	1.0000
Nguyen.K (Mount.Sinai.Hospital)	4.3669	(2.06,9.21)	1.0000
Chen.J.M (NYP-.Columbia.Presby.)	9.7789	(4.6,20.73)	1.0000
Naka.Y (NYP-.Weill.Cornell)	4.9835	(2.84,7.65)	1.0000
Naka.Y (NYP-.Weill.Cornell)	9.5869	(5.61,13.96)	1.0000
Tortolani.A (NYP-.Weill.Cornell)	4.8226	(2.64,8.35)	1.0000

Some physicians' model-predicted poor performance is consistent among all the hospitials (s)he is working. For example, Dr. Tortolani is working for both NY Methodist Hospital and NYP- Weill Cornell, and had posterior ratio of the model-predicted mortality rate to the expected mortality rate larger than 1 in both hospitals with probability near 1, which indicates Dr. Tortolani has a worse performance than expected. Since Dr. Tortolani only takes 4 cases of CABG procedure and 5 cases of Valve in NYP- Weill Cornell hospital, and has one case died for each procedure, I am wondering if his poor performance is related to the quality of NYP- Weill Cornell hospital. I checked the random effects of Dr. Tortolani and both hospitals (s)he works for 2009-2011. The random effect for Dr. Tortolani is positive (0.3) but the random effects for both hospitals are negative. And also the random effect for Dr. Ciaburri who is also working for NY Methodist Hospital is negative.

4

Although the effect of hospitals tend to draw down the odds of death, Dr. Tortolani's performance tends to lead to higher of mortality rate. No matter based on the random effect of Dr. Tortolani and ratio compared to expected mortality rate, Dr. Tortolani's performance cannot be evaluated as good.

Some physicians' unsatisfied model-predicted performance may not reflect their real quality. Dr. Chen works for NYP-Columbia Presbyterian Hospital and NYP- Weill Cornell Hospital, and receives total 7 cases of Valve replacement without any death case. But his posterior ratio in NYP-Columbia Presbyterian Hospital is extremely high (9.8) which may not due to his performance is not good but records on this particular physician are few. Not enough data cannot accurately predict a physician's performance. The high predicted odds of death may be caused by high random effect of the hospital. The random effect for NYP-Columbia Presbyterian Hospital is positive (0.6) and random effect for Dr. Chen is negative but close to 0, which demonstrates the high model predicted mortality rate does not correctly and completely reflect the physician's performance. Dr. Argenziano and Dr. Naka are two physicians I consider that need to pay more attention to by the governors of NYP-Columbia Presbyterian Hospital. They both have posterior ratio larger than 1 with a high probability and positive random effect towards higher odds of death. Their most surgical procedures are done in this hospital though Dr. Naka also works in NYP- Weill Cornell receiving only 2 cases. Thus the NYP-Columbia Presbyterian Hospital cannot gather more information about their performance out of their hospital.

Table 3: Top 5% the posterior probability that the ratio
is greater than 1

names	posterior_median	posterior_conf	prob_excess
Lang.S (NYP-.Weill.Cornell)	0.2485	(0.12,0.47)	0.0000
Michler.R.E (Montefiore.-.Weiler)	0.2653	(0.12,0.52)	0.0001
Spielvogel.D (Westchester.Med.Ctr)	0.4311	(0.25,0.69)	0.0001
Kalimi.R (North.Shore.Univ.Hosp)	0.4685	(0.29,0.72)	0.0002
Palazzo.R (LIJ.Medical.Center)	0.3341	(0.15,0.65)	0.0004
Scheinerman.S.J (LIJ.Medical.Center)	0.3681	(0.18,0.69)	0.0009
Sarabu.M (Vassar.Bros..Med.Ctr)	0.4361	(0.24,0.74)	0.0009
Girardi.L (NYP-.Weill.Cornell)	0.5612	(0.37,0.82)	0.0016
Spielvogel.D (Westchester.Med.Ctr)	0.4088	(0.21,0.74)	0.0019
Kalimi.R (Southside.Hospital)	0.5065	(0.29,0.82)	0.0026
Jacobowitz.I (Maimonides.Medical.Ctr)	0.5185	(0.3,0.84)	0.0034
Malekan.R (Westchester.Med.Ctr)	0.3748	(0.17,0.76)	0.0034
Graver.L (LIJ.Medical.Center)	0.5463	(0.32,0.87)	0.0036
Fernandez.H.A (St..Francis.Hospital)	0.5710	(0.36,0.86)	0.0038
Lansman.S (Westchester.Med.Ctr)	0.4320	(0.21,0.82)	0.0038
Taylor.J (St..Francis.Hospital)	0.6416	(0.44,0.91)	0.0062
Malekan.R (Westchester.Med.Ctr)	0.4663	(0.24,0.85)	0.0066
Kalimi.R (Southside.Hospital)	0.5794	(0.34,0.9)	0.0070

The above table only includes the top 5% (total 18) physician and hospital combinations. If I extend this table, I can see that Dr. Ciaburri ranks 41th by the posterior probability of ratio greater

than 1. Dr. Ciaburri is a physician who is working for three hospitals and his most of work is done in NY Methodist Hospital. (S)he does not have any death cases among all three hospitals in 2009-2011. According to the fitted model, (s)he has a negative random effect towards odds of death and a posterior median ratio 0.53 (0.24,1.07) that indicates that (s)he performs better than his/her expected performance at most of time. I consider that Dr. Ciaburri is a good physician with satisfying performance beyond expectation, but I am not sure if I could conclude that (s)he has the best mortality record in the state.

According to the top 5% table, I have five candidates to be the best physician in the state: Dr. Spielvogel, Dr. Kalimi, Dr. Scheinerman, Dr. Sarabu and Dr. Malekan. They all have relative low posterior median of the ratio and low probability that ratio exceeds 1. Most of them only work for one hospital but have enough number of cases to analyze. Their expected mortality rate is moderately high indicating that they have reputation among patients and health care providers and be able to take patients in severe condition. Since their model-predicted mortality rate may be affected by the random effect of the hospital they are working for, I take a look for random effects on themselves. All physicians have a negative random effect except Dr. Malekan. Among my five candidates, Dr. Kalimi has the smallest value of random effect, which makes him/her able to get lower odds of death if all the five physician work for a same hospital. Also, Dr. Kalimi is the only one physician who is working for two hospitals and have good performance in both hospitals. Moreover, the average expected mortality rate for Dr. Kalimi is the highest among my five candidates, which shows (s)he always take patients with severe and complex cardiac diseases. I consider Dr. Kalimi is the best physician in NY state.

Having discussed the performance of individual physicians, I am interested in the ranking of hospital since people usually go to the best hospital in their area. The criterion I used to compare physicians' performance is based on their posterior ratio of model-predicted mortality rate to the expected mortality rate. However, the expected mortality rate given in data set is on physician level not on the hospital level. I cannot use it directly for eveluating the performance of hospitals thus I calculated the weighted posterior median ratio for each hospital. I use the ratio of cases taken by a specific physician to total cases taken in the hospital as weight to evaluate the importance of a specific physician in a given hospital. The individual physician's posterior median ratio weighted by his/her important is the posterior median ratio for the hospital. The formula is as following:

$$Expected\ MR_{Hospital_j} = \sum_i \frac{cases_i}{\sum_i cases_i} \frac{model\ predicted\ MR_i}{Expected\ MR_i}$$

Fig 2 shows the ranking of hospitals based on weighted posterior median ratio of model-predicted mortality rate to expected mortality rate. As we can seen, the top five hospitals are LIJ Medical Center, Westchester Medical Center, NY Hospital Queens, Good Sam Suffern and Vassar Bros. Medical Center; among these five hospitals, three hospitals are in region NY Metro - New Rochelle and the rest two are in region Queens. I've also look these hospitals' random effect but they are not all negative. However, I do not think evaluating a hospital based on their random effects is accurate and complete. Firstly, it is impossible to have a same physician to work in all hospitals. Moreover, a physician's random effect can be strong to offset the random effect of the hospital (s)he is working for thus a relative higher random effect of hospitals does not mean the odds of death in this hospital will be high.

The five low ranking hosiptals are Strong Memorial Hospital, Mount Sinai Hospital, St. Lukes at St. Lukes, Beth Israel Medical Center and Arnot Ogden Medical Center; among these five

hospitals, three hospitals are in region Manhattan and the rest two are in Western NY Rochester. St. Lukes at St. Lukes and Arnot Ogden Medical Center are hospitals with small size and each only has two physicians working on cardiac surgeries. They may be penalized by one physician whose performance is really unsatisfying. Also, I notice Mount Sinai Hospital has the largest positive value random effect though the rest four have both positive and negative random effects. The physicians working in this hospital may have higher odds of death due to the high random effect of this hospital.
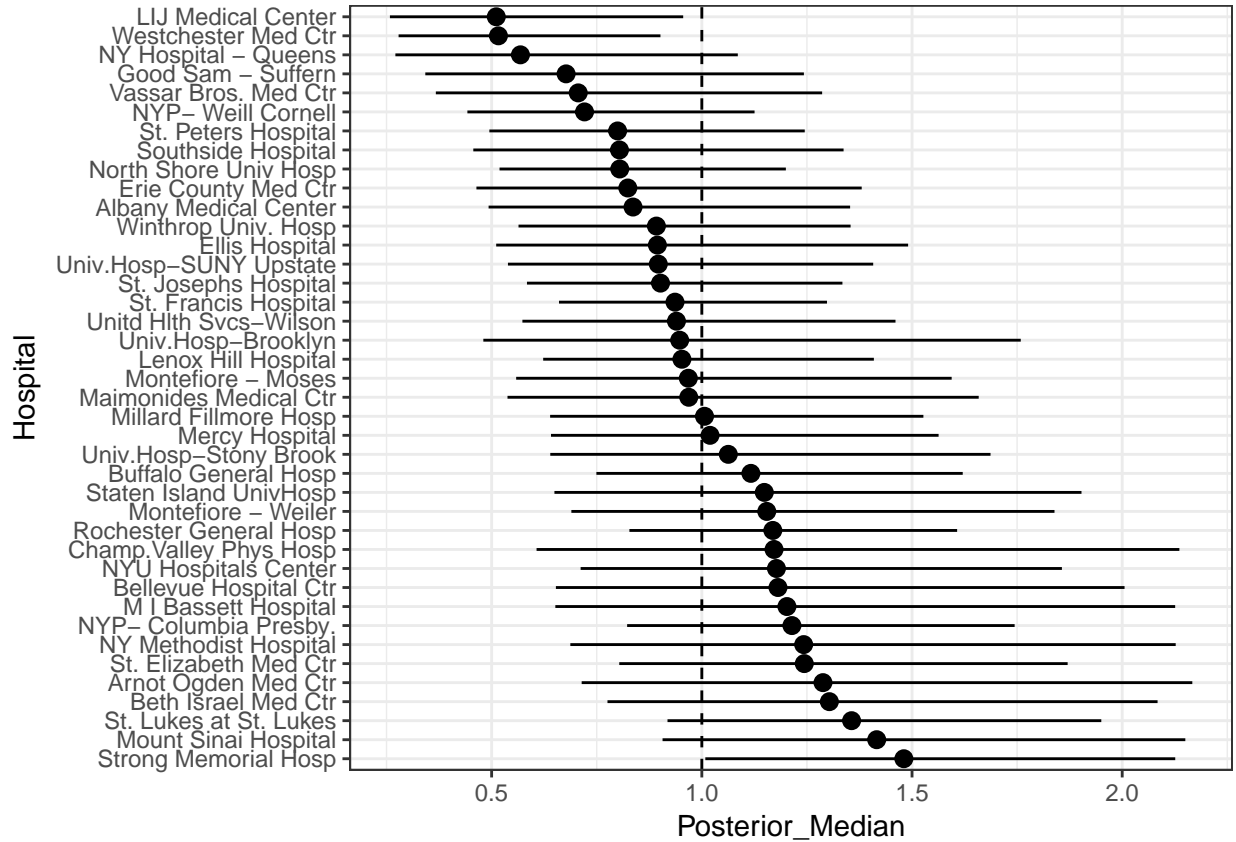

Figure 2: Ranking for hospitals based on weighted posterior median ratio

Conclusion

In this analysis, I evaluate the performance of hospitals in NY state using data on outcomes of cardiac surgical procedures from 2009 to 2011 and also assess the performance of physicians in hospitals. The top five achieving hospitals are LIJ Medical Center, Westchester Medical Center, NY Hospital Queens, Good Sam Suffern and Vassar Bros. Medical Center; the bottom five achieving hospitals are Strong Memorial Hospital, Mount Sinai Hospital, St. Lukes at St. Lukes, Beth Israel Medical Center and Arnot Ogden Medical Center. The performace of hospital is evaluated based on the weighted posterior ratio of model-predicted mortality rate to expected mortality rate. Since this posterior median ratio is calculated from the information of individual physicians, the high ranking hosiptals and high ranking physicians are highly correlated. If the hospital has one top ranking physician, the hospital is more likely to have a good ranking. The weighting system needs to be improved to address the problem of this highly correlated ranking in order to get more precise evaluating method on hospitals.

Some physicians may have taken many cases in one hospital but few cases in another hospital. Mixed effect model has an advantage to borrow information from the records of such physician in one hospital to the other and estimates the overall random effect for such physician. If I only use the ratio of observed mortality rate to expected mortality rate, I will lose information that the physician is a same person who only works in multiple hospitals and have not precise estimate on the performance of physician.

Reference

1. Burkner, Paul-Christian. "brms: An R Package for Bayesian Multilevel Models using Stan."

2. Gelman A, Rubin DB (1992). "Inference from Iterative Simulation Using Multiple Sequences." Statistical Science, pp. 457–472.

3. Duke Stat. "Bayesian Random Effect Models."

4. Wiki. "New York counties ranked by per capita income."

Figure 3: Ratio of the model-predicted mortality rate to the expected