Appendix B
Medical Classification System Basics

(Formally DxCG, Inc.)
Overview: Medical Classification System (version 7)

- **Aggregated Condition Categories (ACC):** 31
- **Related Condition Categories (RCC):** 117  
  *(New in version 7)*
- **Condition Categories (CC):** 394
- **DxGroups (DxG):** 1110
- **Diagnosis Codes (ICD-9):** ~14,200 “legal” codes
  ~2,700 “root” codes

- Reporting Only
- Hierarchies imposed for predictions
- Disease, Age Interactions

© 2009 Verisk Health
Medical Classification Example

ACC MU: Musculoskeletal

RCC HIP: Hip

CC HIP.15: Hip fracture/dislocation

DxG HIP.15.35: Traumatic dislocation of hip

ICD-9 835.01: Closed dislocation of hip – posterior dislocation

© 2009 Verisk Health
DxCG Medical Model Basics

Demographic
- Member ID: 00001
- Name: John Smith
- Gender: Male
- Age: 50

Medical Profile
- Hypertension
- Type I Diabetes
- Congestive Heart Failure
- Alcohol Dependence

6.35 Risk Score
Inpatient and Pharmaceutical Profile-based Risk Score

Demographic

- Member ID: 00001
- Name: John Smith
- Gender: Male
- Age: 50

Clinical Profile

- Antiadrenergic agents, centrally acting
- Calcium channel blocking agents
- Insulin
- Hospitalized for diabetes with acute complications
- Hospitalized for Arrhythmia

© 2009 Verisk Health

8.25 Risk Score
Full Structure of “All Medical” Risk Model

Richer models include interactions by age and by disease groups

\[ Risk = \beta_{0_{agecat}} + \beta_1 HCC_1 + \beta_2 HCC_2 + \beta_3 HCC_3 + ... + \beta_n HCC_n \]

\[(Age < 18) \times (\beta_{k0} + \beta_{k1} HCC_1 + ... + \beta_{kn} HCC_n)\]

\[(Age > 65) \times (\beta_{e0} + \beta_{e1} HCC_1 + ... + \beta_{en} HCC_n)\]

\+(Selected interactions of HCC groups)
Appendix C
Pharmacy Classification System Basics
Pharmacy Classification System (version 3)

Aggregated RxGroups (ARx): 18

RxGroups (RxG): 164

Drug Codes (NDC): > 100,000 codes

Least detail

Most detail

Reporting Only

Hierarchies imposed for predictions

Rx, Age Interactions
Pharmacy Classification Example

Least detail

ARx 05:
Cardiovascular

RxG 40:
Antiadrenergic agents, centrally acting

Most detail

NDC 00003290710:
Methyldopate

© 2009 Verisk Health
Misc. Statistical Issues with modeling health care spending (e.g. risk adjustment models)
Econometric problems with modeling total health care costs

Buntin and Zaslavsky, JHE 2004 “Too much ado about two part models…”

Key problems:
• Restricted range (nonnegative)
• Spike at zero
• Skewness (thick right hand tail)
• Serious heteroskedasticity
• OLS biased and inefficient

Further challenges:
• Large number of possible predictors
• Extremely large samples make nonlinear models challenging, and force a tradeoff between complexity and specification richness
• Explanatory variables highly skewed, often binary
• Models used for policy implementation reward simplicity
• Not everyone is present for the entire year
• Every model is biased and inefficient if it is misspecified
Scheme 1. Mean predictions of alternative estimators plotted against actual cost ratios, by decile.
Scheme 2. Detail, lower end of the distribution, mean predictions of alternative estimators plotted against actual cost ratios, by decile.

Buntin and Zaslavsky, JHE 2004 “Too much ado about two part models...”
## Table 1
Predictions of the mean cost ratio for the entire sample and relevant subgroups, from alternative estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Mean of sample</th>
<th>Beneficiaries with chronic conditions</th>
<th>Beneficiaries without chronic conditions</th>
<th>Beneficiaries with ADL limitations</th>
<th>Beneficiaries without ADL limitations</th>
<th>Beneficiaries in poor health</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P OLS—untransformed</td>
<td>0.97</td>
<td>1.30</td>
<td>0.40</td>
<td>2.70</td>
<td>0.78</td>
<td>3.18</td>
</tr>
<tr>
<td>2P OLS—lognormal retransformation</td>
<td>1.98</td>
<td>2.72</td>
<td>0.72</td>
<td>6.95</td>
<td>1.46</td>
<td>8.26</td>
</tr>
<tr>
<td>2P OLS—smearing retransformation</td>
<td>1.03</td>
<td>1.43</td>
<td>0.38</td>
<td>3.63</td>
<td>0.76</td>
<td>4.32</td>
</tr>
<tr>
<td>2P OLS—2 smearing factors</td>
<td>0.93</td>
<td>1.26</td>
<td>0.38</td>
<td>2.84</td>
<td>0.73</td>
<td>3.32</td>
</tr>
<tr>
<td>2P OLS—square root smeared retransformation</td>
<td>1.05</td>
<td>1.20</td>
<td>0.79</td>
<td>1.68</td>
<td>0.98</td>
<td>1.79</td>
</tr>
<tr>
<td>2P GLM—constant variance</td>
<td>0.97</td>
<td>1.23</td>
<td>0.56</td>
<td>2.57</td>
<td>0.81</td>
<td>3.18</td>
</tr>
<tr>
<td>1P GLM—constant variance</td>
<td>0.98</td>
<td>1.23</td>
<td>0.56</td>
<td>2.57</td>
<td>0.81</td>
<td>3.18</td>
</tr>
<tr>
<td>1P GLM—variance proportional to mean squared</td>
<td>1.02</td>
<td>1.37</td>
<td>0.44</td>
<td>3.19</td>
<td>0.79</td>
<td>3.66</td>
</tr>
<tr>
<td>1P GLM—variance proportional to mean</td>
<td>0.97</td>
<td>1.24</td>
<td>0.50</td>
<td>2.70</td>
<td>0.78</td>
<td>3.18</td>
</tr>
<tr>
<td>Actual</td>
<td>0.97</td>
<td>1.28</td>
<td>0.43</td>
<td>2.70</td>
<td>0.78</td>
<td>3.18</td>
</tr>
</tbody>
</table>
Table 2
Mean square error (MSE), mean absolute prediction error (MAPE), and mean square forecast errors (MSFE) of alternative estimators

<table>
<thead>
<tr>
<th>Method</th>
<th>(1) MSE whole sample</th>
<th>(2) MAPE whole sample</th>
<th>(3) Average MSFE over 100 split-samples</th>
<th>(4) Average MAPE over 100 split-samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P OLS—untransformed</td>
<td>4.792</td>
<td>1.073</td>
<td>4.849</td>
<td>1.081</td>
</tr>
<tr>
<td>2P OLS—lognormal retransformation</td>
<td>10.081</td>
<td>1.704</td>
<td>10.478</td>
<td>1.718</td>
</tr>
<tr>
<td>2P OLS—smearing retransformation</td>
<td>5.241</td>
<td>1.080</td>
<td>5.367</td>
<td>1.097</td>
</tr>
<tr>
<td>2P OLS—2 smearing factors</td>
<td>4.853</td>
<td>1.032</td>
<td>4.903</td>
<td>1.042</td>
</tr>
<tr>
<td>2P OLS—square root smeared retransformation</td>
<td>5.133</td>
<td>1.180</td>
<td>5.158</td>
<td>1.185</td>
</tr>
<tr>
<td>2P GLM—constant variance</td>
<td>4.687</td>
<td>1.064</td>
<td>4.924</td>
<td>1.067</td>
</tr>
<tr>
<td>1P GLM—constant variance</td>
<td>4.689</td>
<td>1.070</td>
<td>4.923</td>
<td>1.072</td>
</tr>
<tr>
<td>1P GLM—variance proportional to mean squared</td>
<td>4.945</td>
<td>1.075</td>
<td>5.038</td>
<td>1.085</td>
</tr>
<tr>
<td>1P GLM—variance proportional to mean</td>
<td>4.718</td>
<td>1.052</td>
<td>4.815</td>
<td>1.060</td>
</tr>
</tbody>
</table>

Buntin and Zaslavsky, JHE 2004 “Too much ado about two part models...”
$R^2$ from five estimation methods using six different classification systems

Mean predictions from five estimation methods, six different RA systems on US Veterans data

Comparision of three slope coefficients from five empirical distributions (dependent variable: ln(inpatient spending per hospitalization))

Manning, Basu, and Mullahy JHE (2005)

N=6500
Comparision of z scores of slope coefficients from five empirical distributions, dependent variable: ln(inpatient spending per hospitalization),
Manning, Basu, and Mullahy JHE (2005)

N=6500

- doctor is hospitalist
- experience in all diseases
- experience with disease
Bottom line on estimation of nonlinear models versus OLS

• Several very careful studies using health expenditure data have shown that even though OLS is “biased and inefficient” it still does better on measures that we commonly care about: $R^2$ and means of subsamples. It also does well for hypothesis testing.
• OLS is much easier to explain to policy makers, and more transparent.
• Very large sample sizes mean that OLS is more efficient than nonlinear models run on smaller samples
• Rather than worrying about error specification, it may be more productive to worry about omitted variable bias from having to simplify how diseases are captured in the model.
Table 3: Predictive power of various information sets and various models
Dependent variable is 1997 US Medicare total covered charges

<table>
<thead>
<tr>
<th>Partial Year Eligibles included?</th>
<th>Weighted OLS</th>
<th>OLS</th>
<th>Square Root model (heteroskedasticity-corrected)</th>
<th>Two part linear model</th>
<th>GLM with link = log, dist = normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Sample Mean                     | 6,886        | 5,063| 5,063                                        | 5,063                 | 5,063                             |
| Number of Observations          | 1,380,863    | 1,273,471| 1,273,471                                   | 1,273,471             | 1,273,471                         |

| Age and gender only             | 0.011         | 0.010| 0.009                                        | 0.010                 | 0.010                             |
| Prior year total covered charges*| 0.089         | 0.096| 0.113                                        | 0.120                 | 0.105                             |
| Diagnoses organized by DCG/HCC* | 0.104         | 0.108| 0.103                                        | 0.107                 | 0.105                             |

| Covered charges by DCG/HCC*     | 0.099         | 0.107| 0.103                                        | 0.105                 | 0.095                             |
| Covered charges by Place of Service* | 0.140     | 0.145| 0.136                                        | 0.145                 | 0.126                             |
| Covered charges by Physician Specialty* | 0.142     | 0.152| 0.143                                        | 0.152                 | 0.131                             |
| Covered charges by Type of Service* | 0.150     | 0.155| 0.146                                        | 0.154                 | 0.134                             |
| All of the above except diagnoses* | 0.154     | 0.160| 0.151                                        | 0.160                 | 0.138                             |

"Kitchen sink": All of the above* | 0.169        | 0.171| 0.161                                        | 0.169                 | 0.147                             |

*All Regressions included a constant and 21 age-gender dummy variables

Source: Ellis and McGuire, 2006, Table 1.
# Predictive Power of Various Information Sets

**US Commercially insured sample, 2004-2005, prospective model**

<table>
<thead>
<tr>
<th>Information Set</th>
<th>Weighted LS</th>
<th>OLS</th>
<th>Two-Part Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Year Eligibles included?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sample Mean</td>
<td>3560</td>
<td>3463</td>
<td>3463</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>5,298,819</td>
<td>4,688,097</td>
<td>4,688,097</td>
</tr>
<tr>
<td><strong>Rsquare</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age and Gender only</td>
<td>0.0266</td>
<td>0.0293</td>
<td>0.0277</td>
</tr>
<tr>
<td>Prior Year total covered charges</td>
<td>0.0982</td>
<td>0.1027</td>
<td>0.0992</td>
</tr>
<tr>
<td>Simple HCC</td>
<td>0.1692</td>
<td>0.1746</td>
<td>0.1749</td>
</tr>
<tr>
<td>Covered charges by Place of Service</td>
<td>0.1894</td>
<td>0.2042</td>
<td>0.2055</td>
</tr>
<tr>
<td>Covered charges by Physician Specialty</td>
<td>0.1779</td>
<td>0.1924</td>
<td>0.1938</td>
</tr>
<tr>
<td>Covered charges by Type of Services</td>
<td>0.1977</td>
<td>0.2107</td>
<td>0.2036</td>
</tr>
</tbody>
</table>

Jiang, Ellis, and Kuo 2008
Caveate on test statistics with OLS

• OLS does have biased standard errors, which will tend to overstate significance

• test statistics such as z scores are biased

• Experiment: take 1000 random draws of 100 people from a large sample, then calculate z scores using each sample’s standard deviation and means. Law of large number suggests this should be approximately normally distributed. Is it?
Figure 1
Comparison of Full Empirical and Standard Normal Distributions (1000 draws of N=100)

Cumulative Probability

Z-Scores

Unpublished research, Ellis
Figure 2

Comparison of Upper Tails of Empirical Distribution and Standard Normal Distribution (1000 draws of N=100)
Therefore be careful with OLS

• Use a very high z-score (t-statistic) for hypothesis testing in deciding what variables to include or not.
• Suggest using $z > 4$.
• Easy to get $z$ scores of 10 or 50 with millions of observations. Don’t believe them.
• $R^2$ values are also biased in small samples
  – Conventional $R^2$ biased upward
  – Validated $R^2$ biased downward
A within-sample method of validating predictive power with special application to risk adjustment models

• Traditional approach is to use split sample methods to evaluate overfitting.
• Inefficient in that only one split is traditionally considered.
• Predictive power is understated in validated measures of goodness of fit, for the same reason that fitted measures overstate power.
• Ellis and Mookim use systematic within-sample fitting and validation to generate more powerful measures.
Prospective HCC model, fitted and validated $R^2$, by sample sizes

Based on 100+ Monte Carlo draws of each size from 4.7 million US lives

Ellis and Mookim, 2008