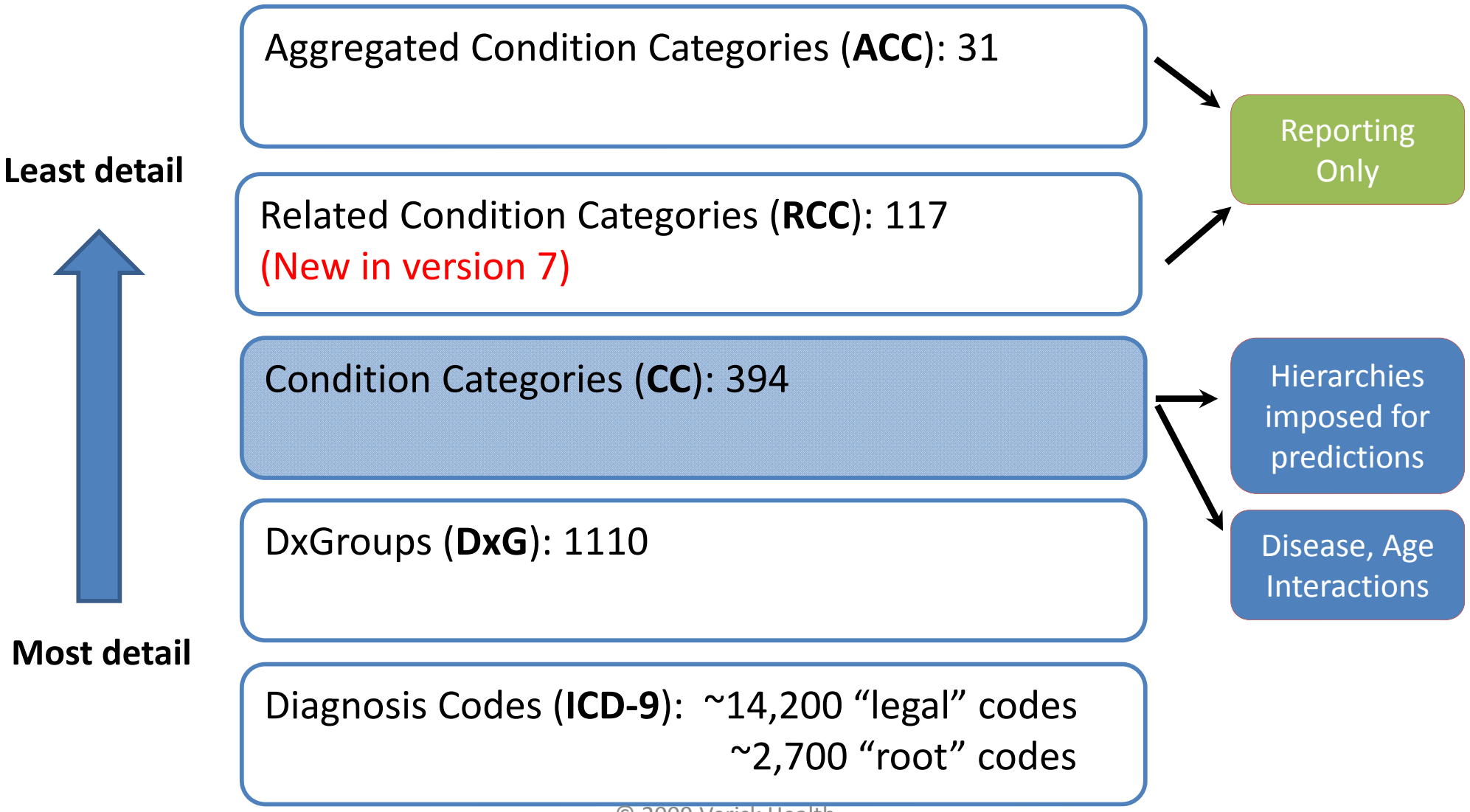
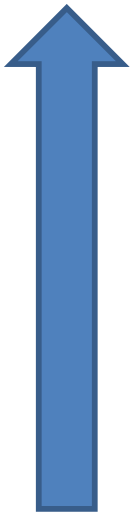
Appendix B
Medical Classification System Basics

(Formerly DxCG, Inc.)

Overview: Medical Classification System (version 7)



Medical Classification Example



Least detail

Most detail

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

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 w acute complications
- Hospitalized for Arrhythmia

8.25
Risk Score

Full Structure of "All Medical" Risk Model

Richer models include interactions by age and by disease groups

$$Risk = \beta_{0agecat} + \beta_1 HCC_1 + \beta_2 HCC_2 + \beta_3 HCC_3 + \ldots + \beta_n HCC_n$$

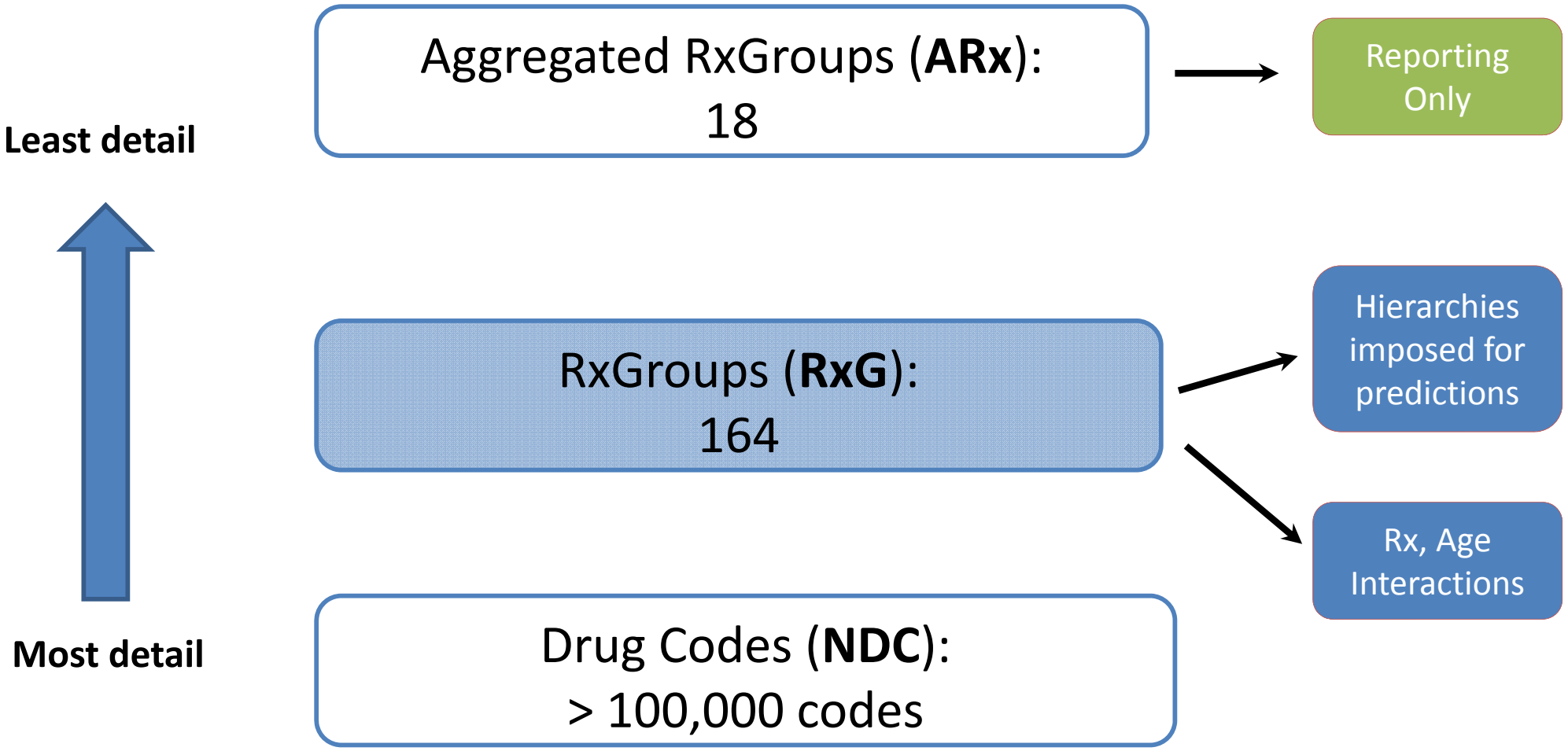
$$+(Age < 18) * (\beta_{k0} + \beta_{k1} HCC_1 + \ldots + \beta_{kn} HCC_n)$$

$$+(Age > 65) * (\beta_{e0} + \beta_{e1} HCC_1 + \ldots + \beta_{en} HCC_n)$$

+(Selected interactions of HCC groups)

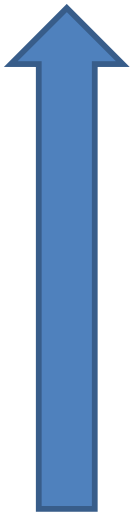
Appendix C
Pharmacy Classification System Basics

Pharmacy Classification System (version 3)



Pharmacy Classification Example

Least detail

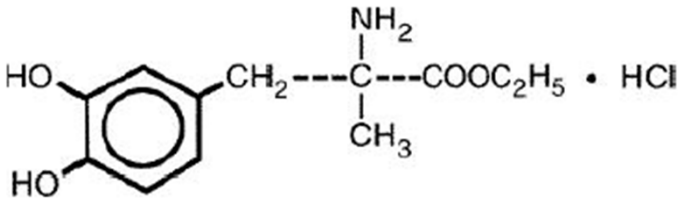


Most detail

ARx 05:
Cardiovascular

RxG 40:
Antiadrenergic agents, centrally acting

NDC 00003290710:
Methyldopate



Misc. Statistical Issues with modeling health care spending (e.g. risk adjustment models)

Econometric problems with modeling total health care costs

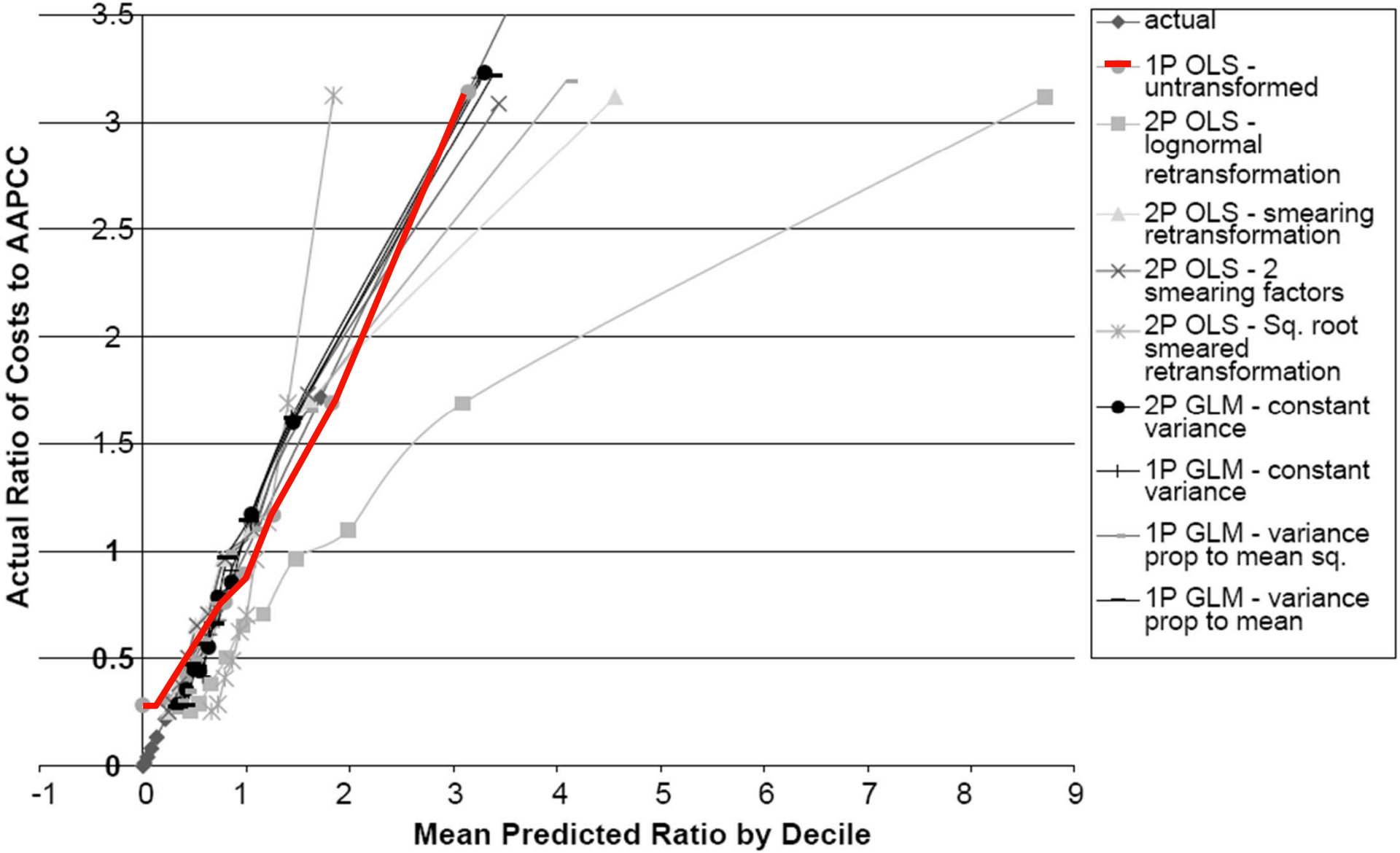
Buntin and Zaslavsky, JHE 2004 "Too much ado about two part models…"

Key problems:

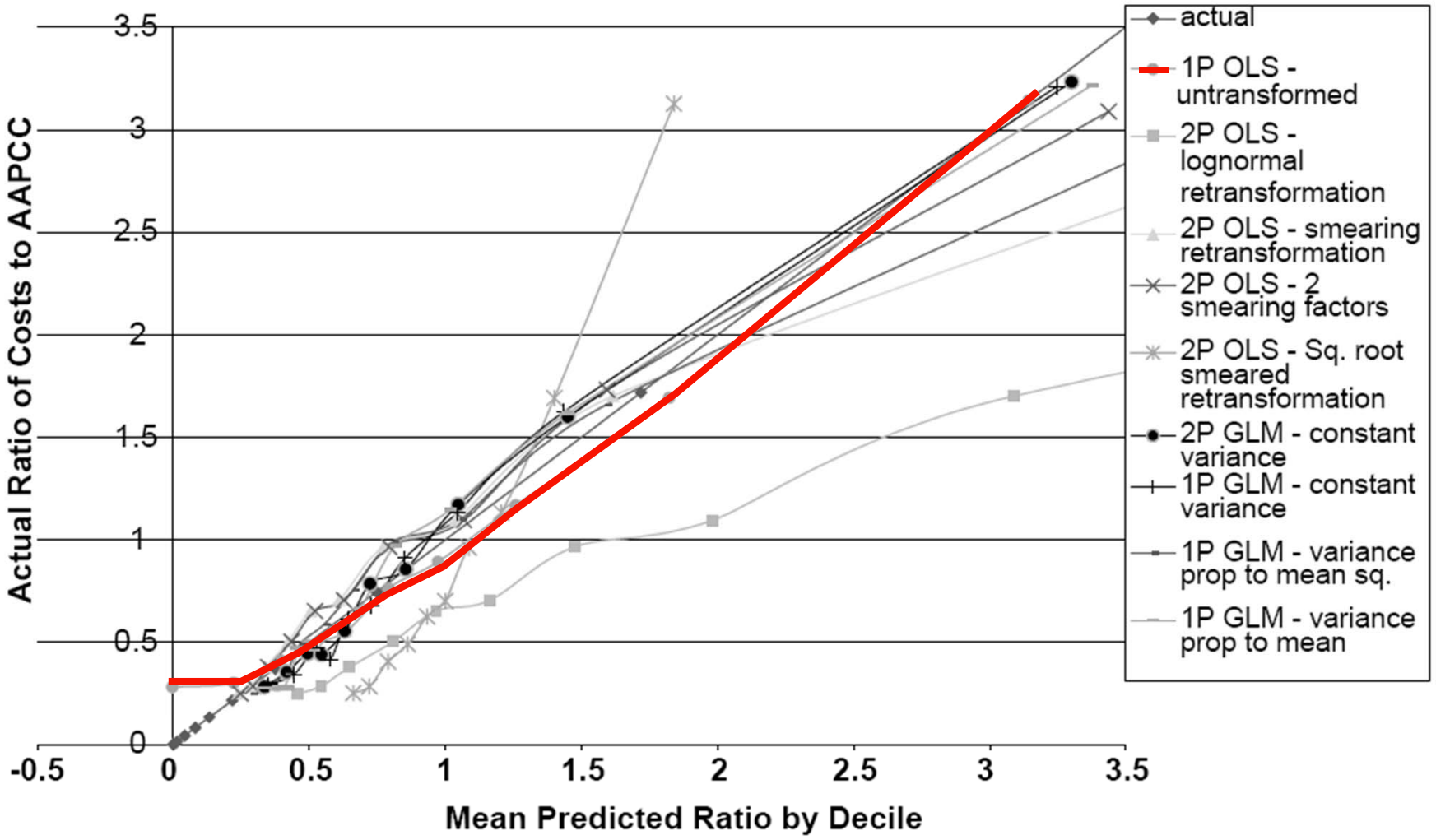
- Restricted range (nonnegative)
- Spike at zero
- Skewness (thick right hand tail)
- Serious heteroskedasticity
- OLS biased and inefficient

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

Estimator	Mean of sample	Beneficiaries with chronic conditions	Beneficiaries without chronic conditions	Beneficiaries with ADL limitations	Beneficiaries without ADL limitations	Beneficiaries in poor health
1P OLS—untransformed	0.97	1.30	0.40	2.70	0.78	3.18
2P OLS—lognormal retransformation	1.98	2.72	0.72	6.95	1.46	8.26
2P OLS—smearing retransformation	1.03	1.43	0.38	3.63	0.76	4.32
2P OLS—2 smearing factors	0.93	1.26	0.38	2.84	0.73	3.32
2P OLS—square root smeared retransformation	1.05	1.20	0.79	1.68	0.98	1.79
2P GLM—constant variance	0.97	1.23	0.56	2.57	0.81	3.18
1P GLM—constant variance	0.98	1.23	0.56	2.57	0.81	3.18
1P GLM—variance proportional to mean squared	1.02	1.37	0.44	3.19	0.79	3.66
1P GLM—variance proportional to mean	0.97	1.24	0.50	2.70	0.78	3.18
Actual	0.97	1.28	0.43	2.70	0.78	3.18

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

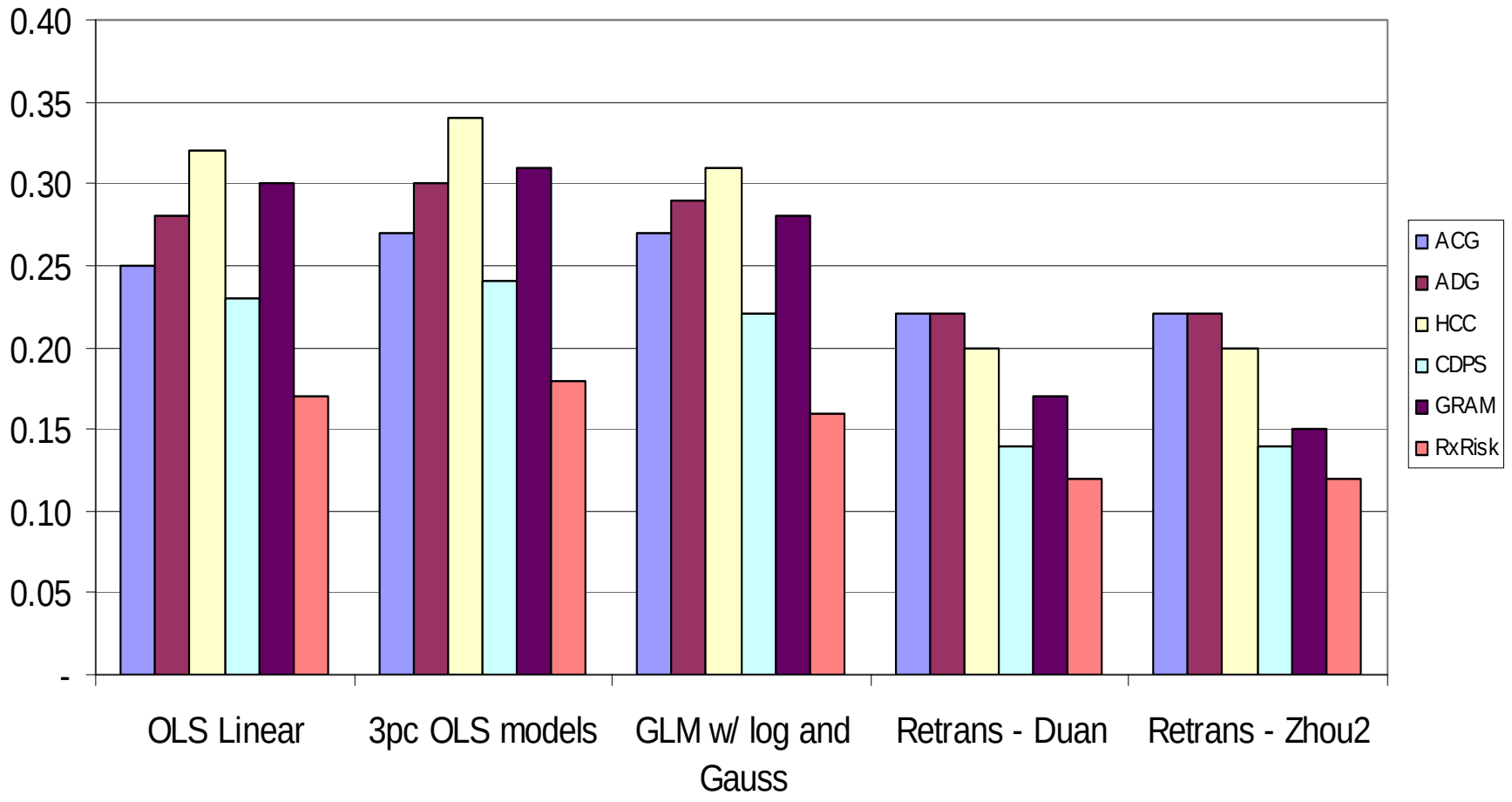
Table 2
Mean square error (MSE), mean absolute prediction error (MAPE), and mean square forecast errors (MSFE) of alternative estimators

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

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

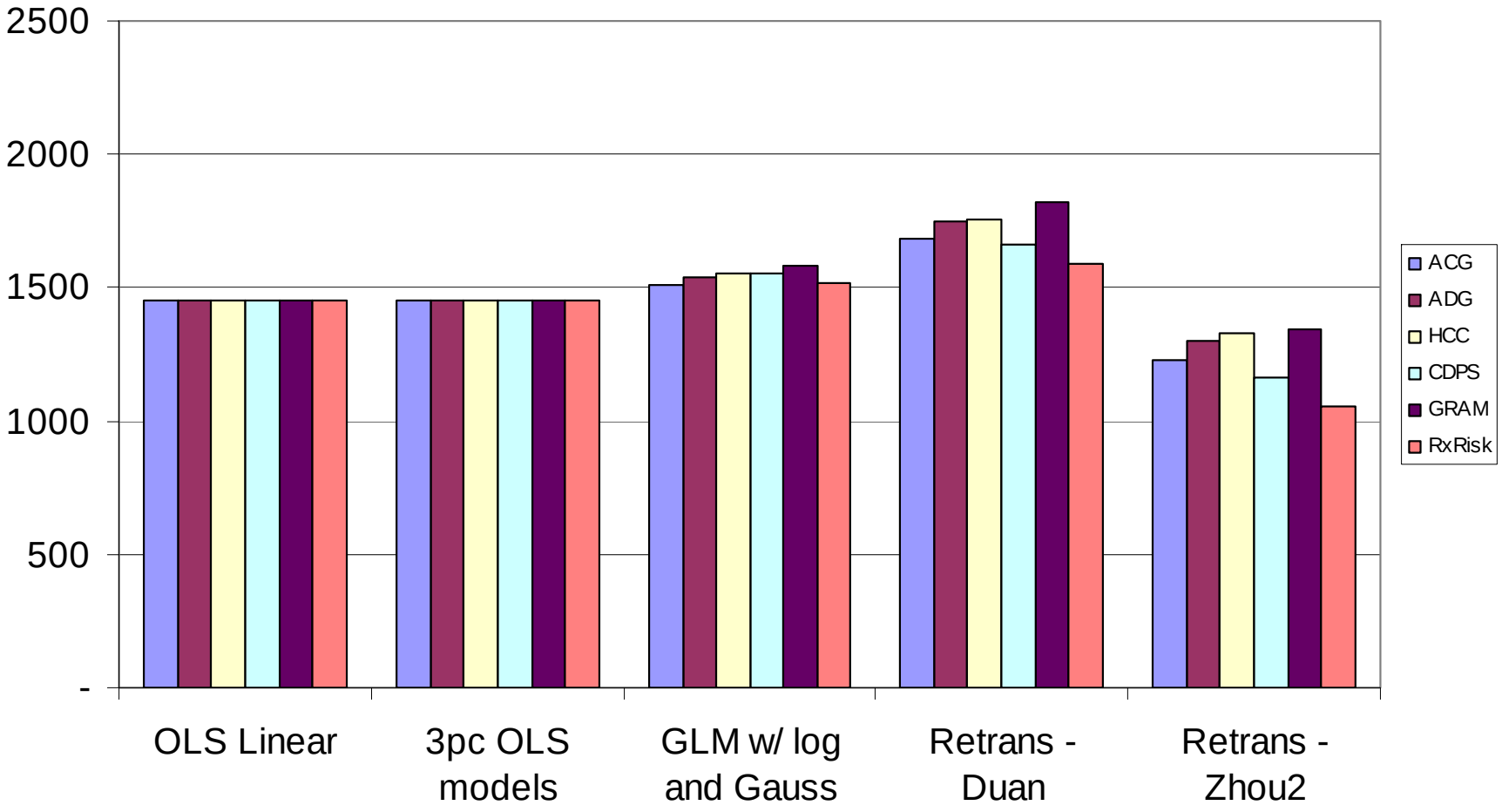
R^2 from five estimation methods using six different classification systems

Fishman et al, 2005, Issues in Evaluating Alternative Risk Assessment Models: Evidence from the US Veteran Population, working paper presented at iHEA, July 2005.



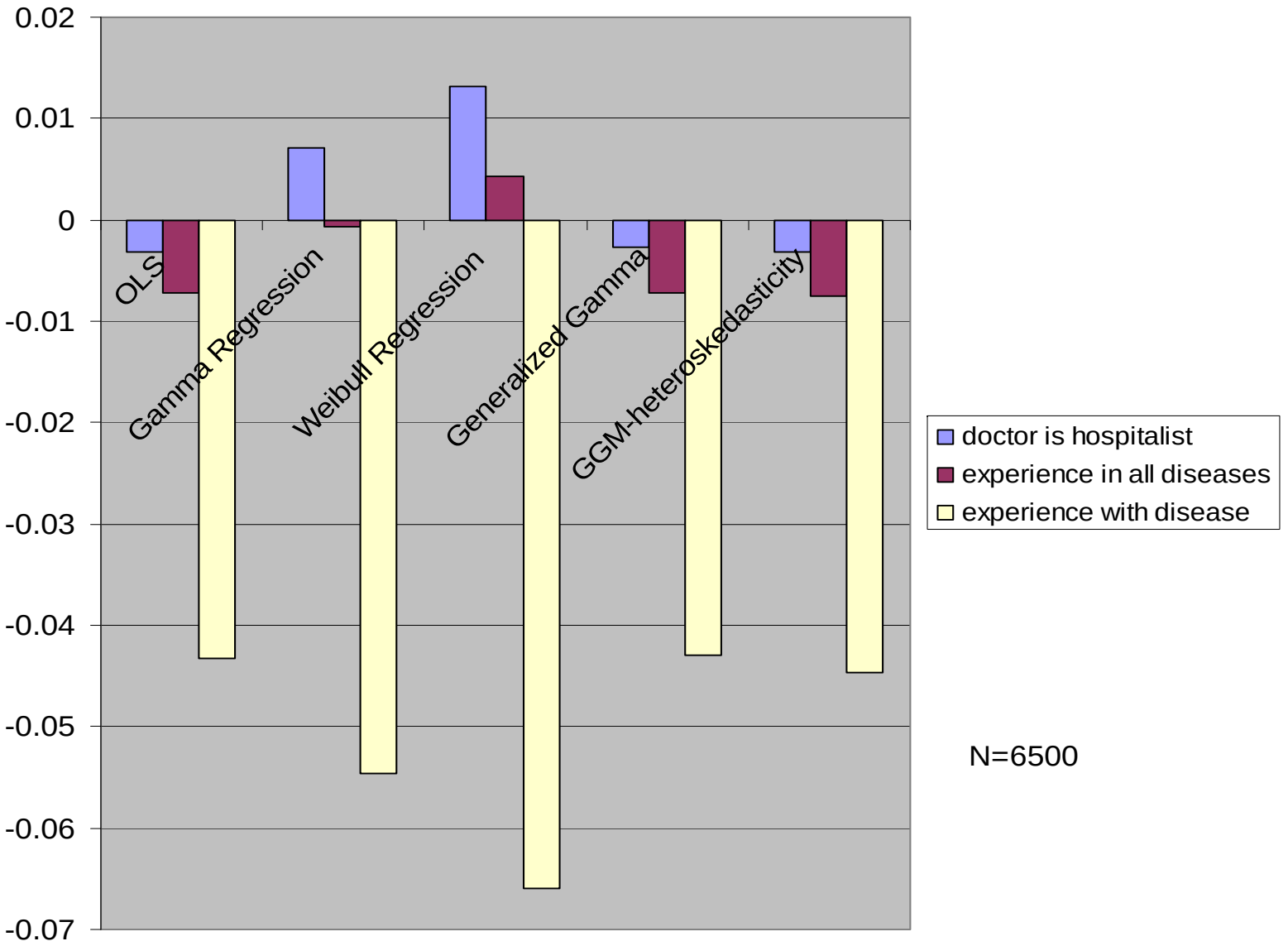
Mean predictions from five estimation methods, six different RA systems on US Veterans data

Fishman et al, 2005, Issues in Evaluating Alternative Risk Assessment Models: Evidence from the US Veteran Population



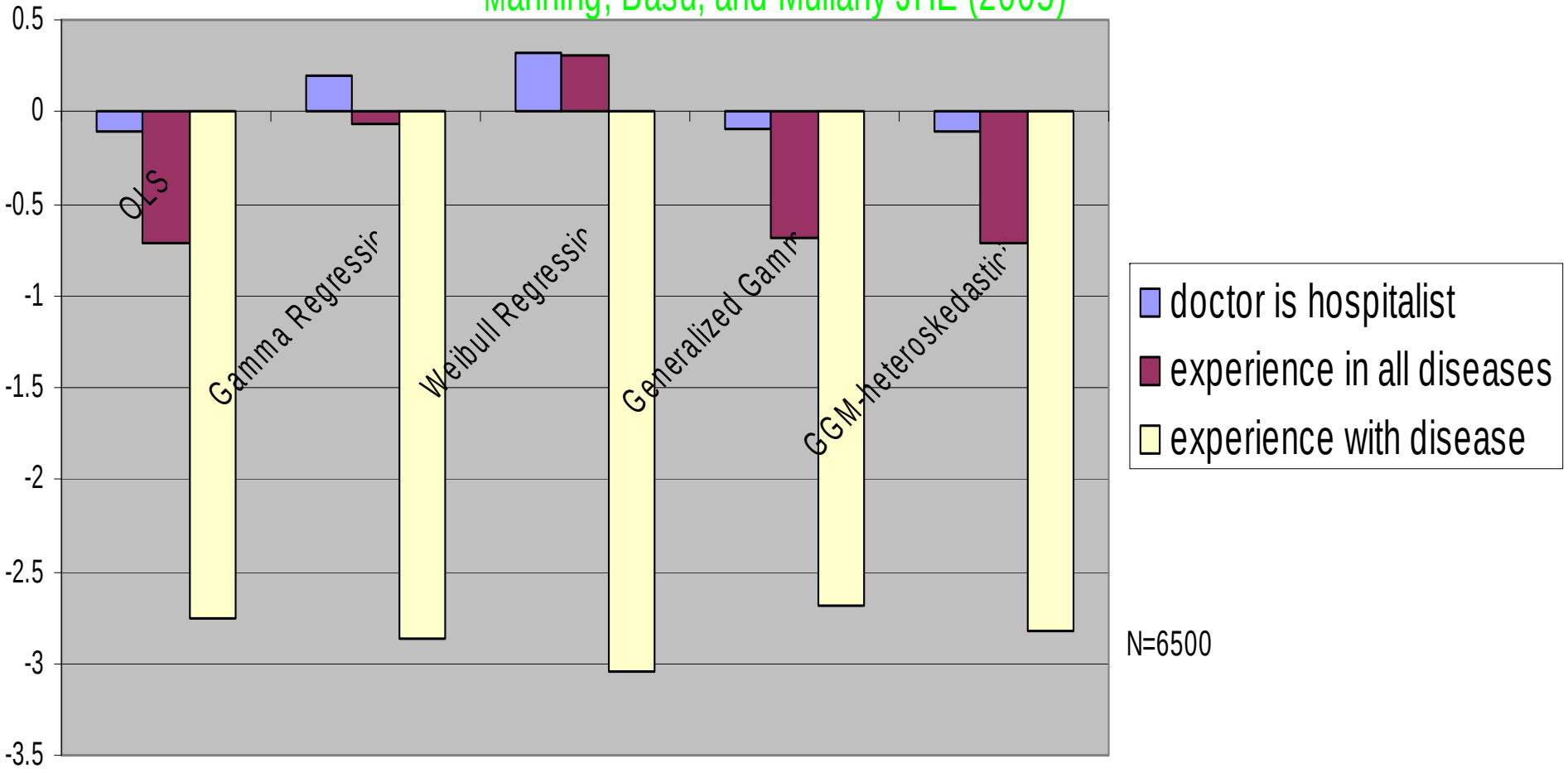
Comparision of three slope coefficients from five empirical distributions (dependent variable: ln(inpatient spending per hospitalization))

Manning, Basu, and Mullahy JHE (2005)



Comparision of z scores of slope coeficients from five empirical distributions, dependent variable: ln(inpatient spending per hospitalization),

Manning, Basu, and Mullahy JHE (2005)



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

	Weighted OLS	OLS	Square Root model (hetero-skedasticity-corrected)	Two part linear model	GLM with link = log, dist = normal
Partial Year Eligibles included?	Yes	No	No	No	No
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
	R^2	R^2	R^2	R^2	R^2
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

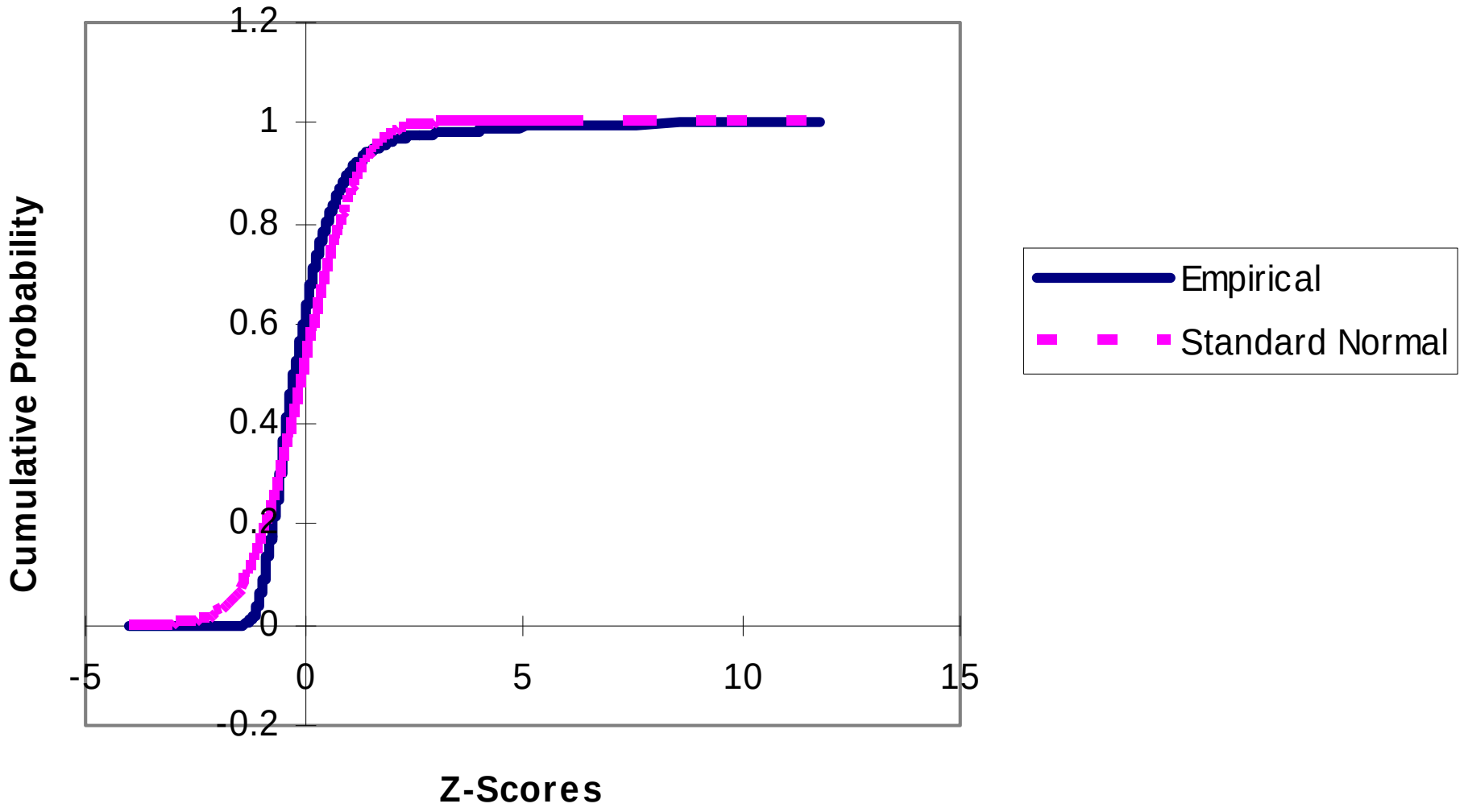
	Estimation Method		
	Weighted LS	**OLS**	**Two-Part Linear Model**
Partial Year Eligibles included?	Yes	No	No
Sample Mean	3560	3463	3463
Number of Observations	5,298,819	4,688,097	4,688,097
Information Set	**Rsquare**	**Rsquare**	**Rsquare**
Age and Gender only	0.0266	0.0293	0.0277
Prior Year total covered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

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?

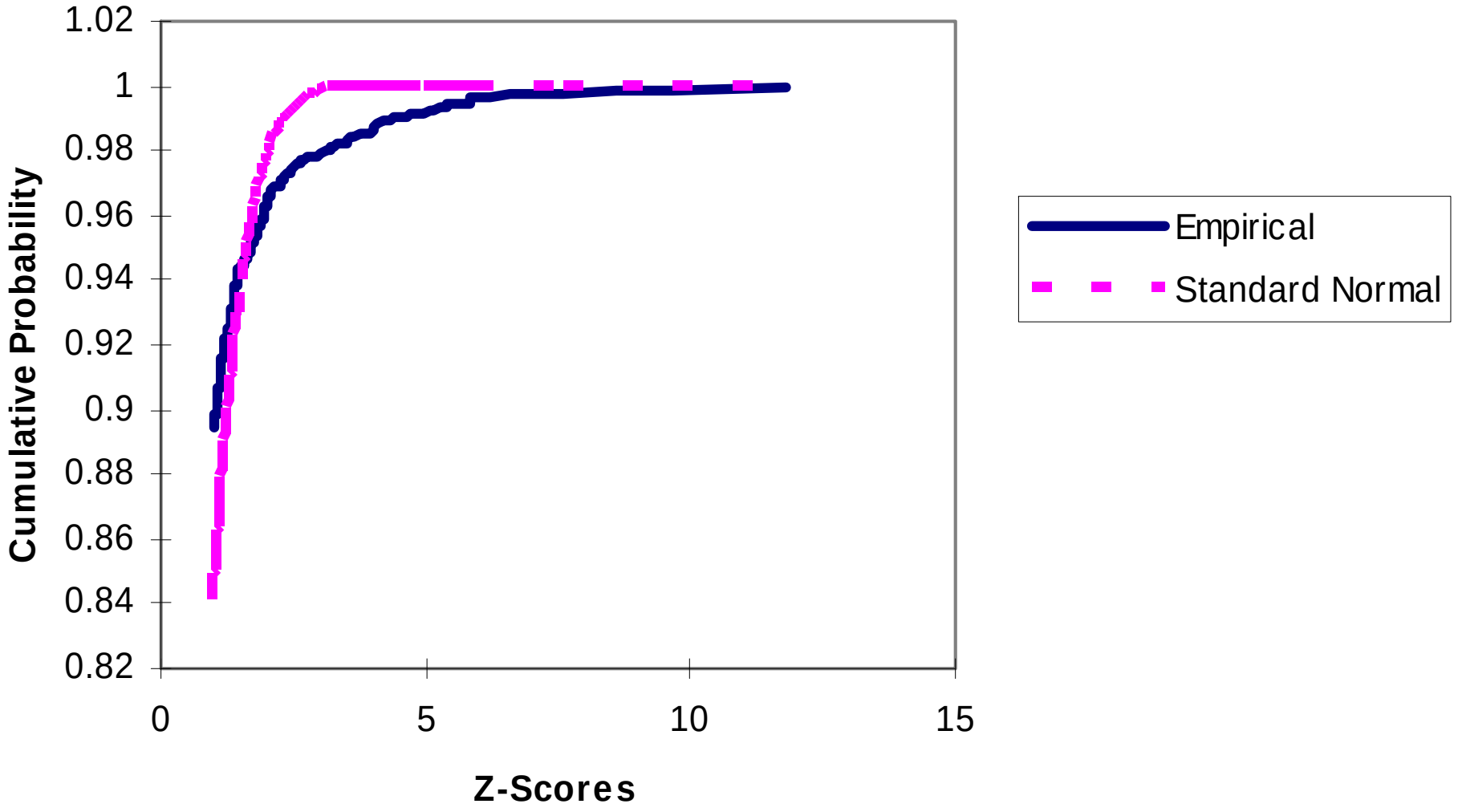
Figure 1
Comparison of Full Empirical and Standard Normal Distributions (1000 draws of N=100)



Unpublished research, Ellis

Figure 2

Comparison of Upper Tails of Empirical Distribution and Standard Normal Distribution (1000 draws of N=100)



Unpublished research, Ellis

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.
- R2 values are also biased in small samples
  - Conventional R^2 biased upward
  - Validated R^2 biased downward

Ellis and Mookim (2008) Working paper

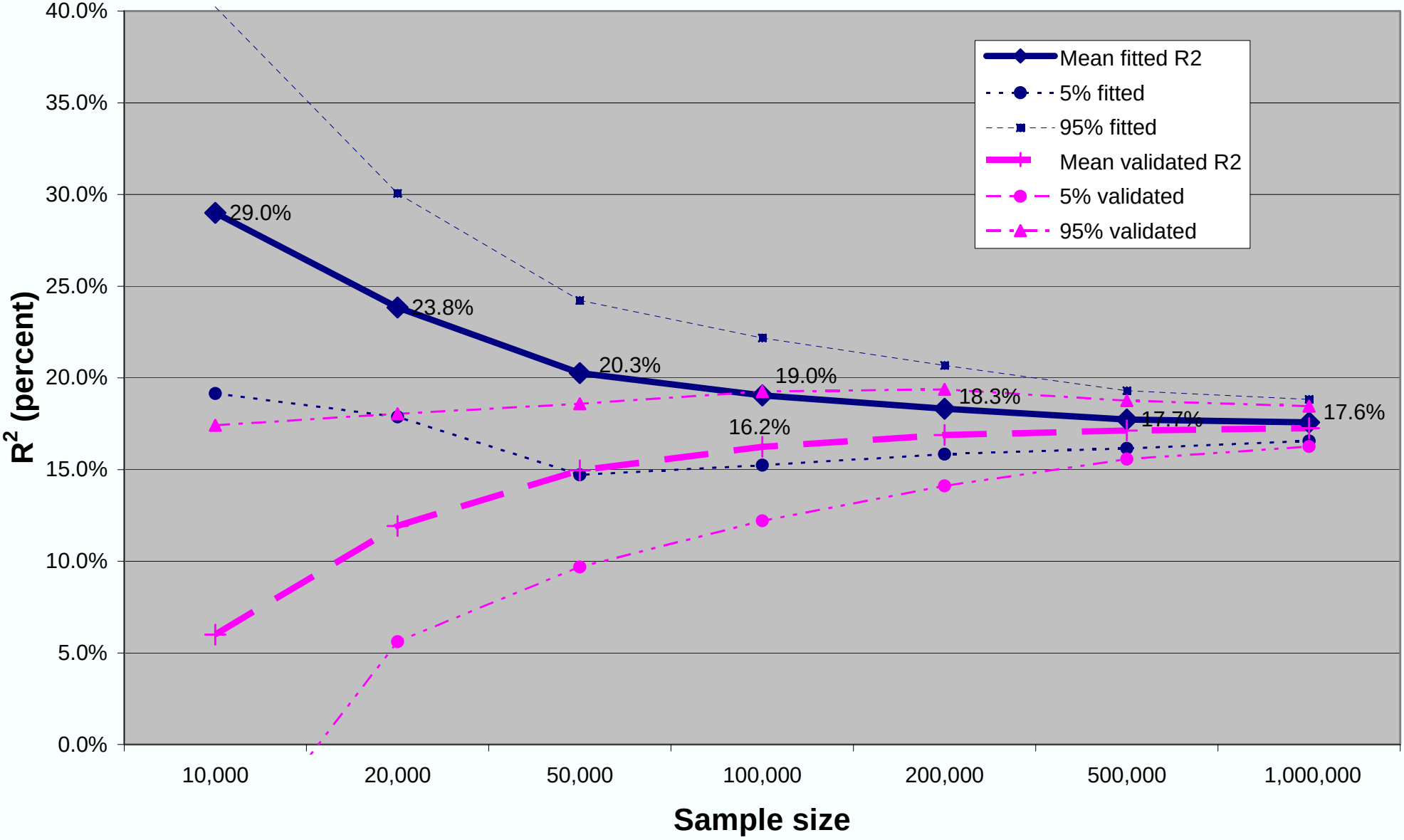
A within-sample method of validating predictive power with special application to risk adjustment models

---

- 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