Combining Information from Survey and Non-survey data sources: Analysis of Health Care Costs

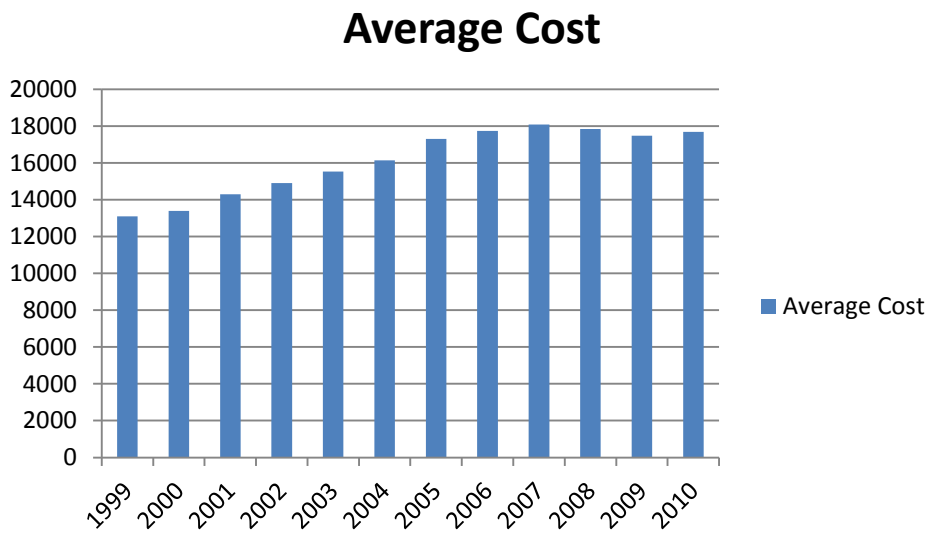
Trivellore Raghunathan (Raghu)

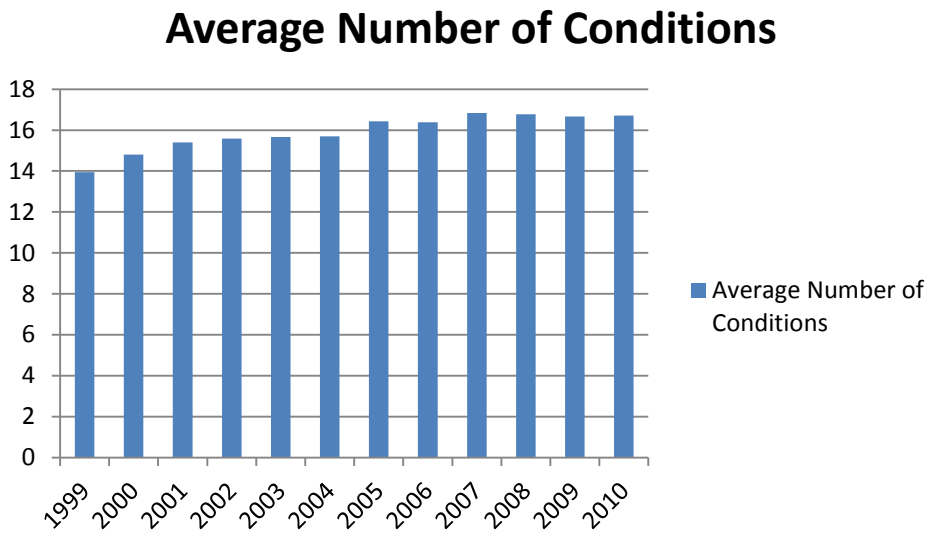
Survey Research Center

University of Michigan

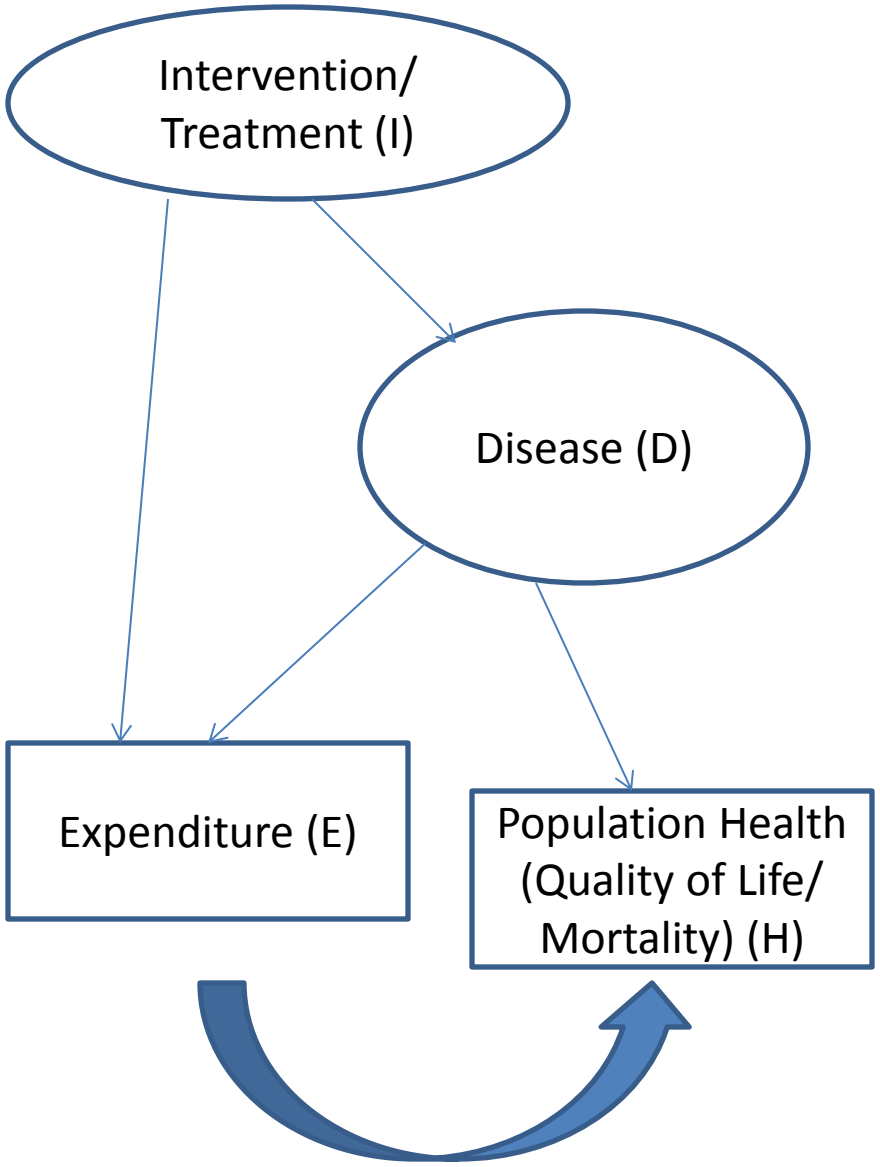
Joint work with David Cutler (Harvard), Susan Stewart (NBER), Kaushik Ghosh (NBER), Kassandra Messer (UM), Irina Bondarenko (UM), Pat Berglund (UM)

Trends in Average Cost and Number of Health Conditions (65 years of age or older)





Goals



- Each arrow denotes the relationship to be studied
- Thereby study the relationship between expenditure and population health
- Bayesian network modeling to be used to study the impact of the intervening on each arrowed relationship
- Simulation to represent alternative s “what if?” scenarios
- Essentially, we need information on $\{E_t, D_t, H_t, I_t, X_t; t = 1, 2, \ldots, T\}$

Four Broad Objectives

1. Estimate prevalence rates and assess trends for various diseases/screening
2. Estimate costs attributable to each disease and assess trends in these costs
3. Estimate population health (measuring Quality of Life Expectancy) attributable to each disease and assess trends in health
4. Use 3 sets of estimations to develop models for spending and health

Population and Data Sources

- Four age groups: >=65, 45-64, 25-44 and <=24
- Survey Data: MCBS, MEPS, NHIS, NHANES, HRS, PSID, NCS
- Non Survey Data: Medicare Claims, Provider data, IMS, HMO, Prescription prices
- Information from Clinical Studies
- Identified about 120 disease/screening conditions

Primary Data Source (Age 65 and older)

- Medicare Current Beneficiary Survey (MCBS)
  - Age 65 and older
  - Years 1999-2010 (2012)
- 80 diseases and screening dummy variables
- Community dwelling and Institutionalized (nursing home, assisted living) populations
- Purely covered on Medicare
- Adjustments
  - Propensity score weighting to compensate for excluding HMO enrollees
  - Multiplier to cost so that weighted estimated population total agrees with published national health expenditure
  - All costs are in 2010 dollars

Objective 1: Estimation of Prevalence

- 80 health conditions: Ever having this condition; some during the specific time period
  - Option 1: Use the Medicare claims (any claim) indicating particular ICD-9 codes
    - The prevalence rates based on this definition:
    - Reasonable for some chronic diseases
    - Low rate for acute conditions and some chronic diseases
  - Option 2: Calibrate the claims using benchmark data
    - Self-report from the National Health and Nutritional Examination Survey (NHANES)
    - All calibrated claims can be thought of as "Ever Having Disease"

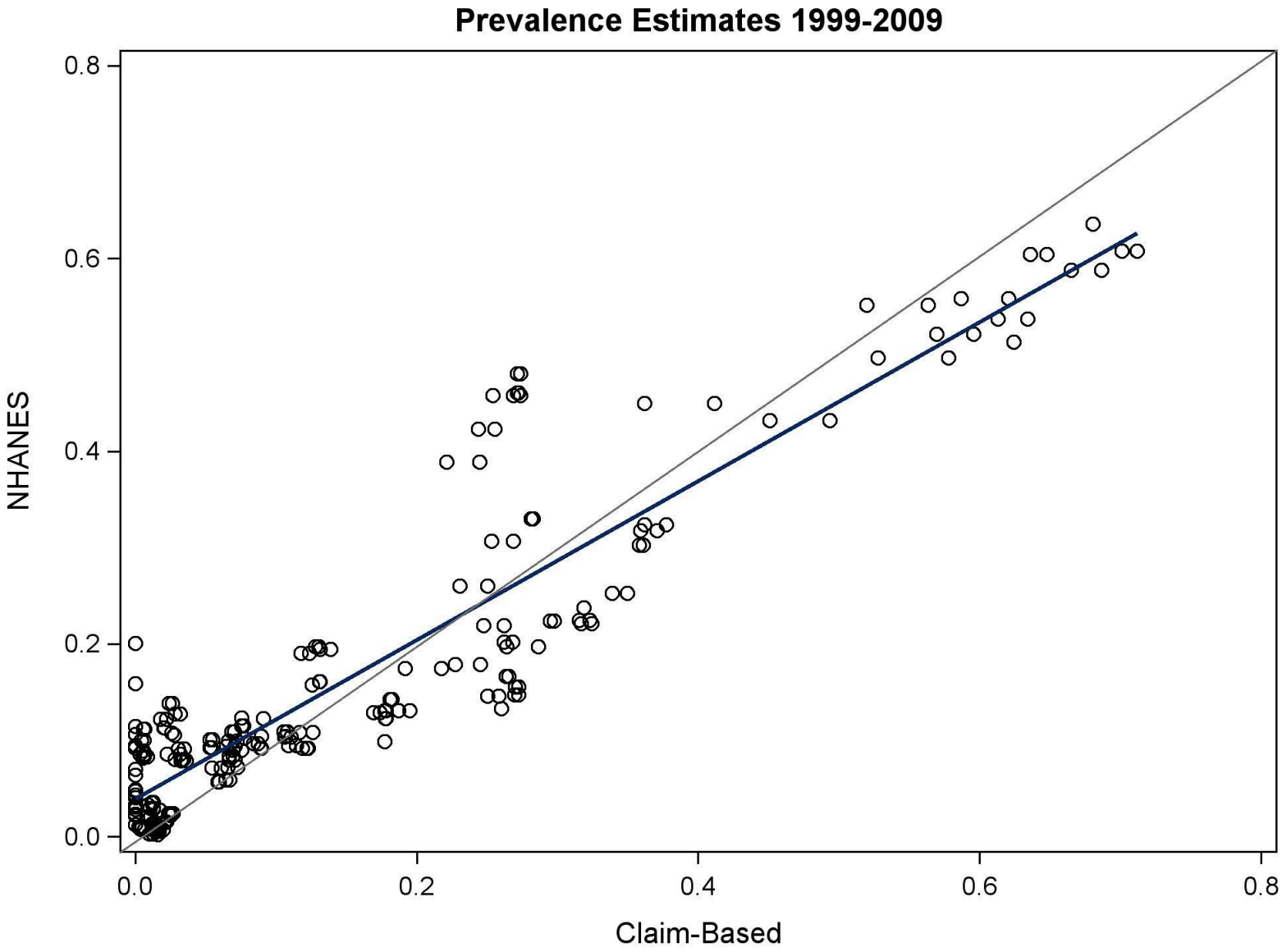
Prevalence of Health Conditions

Age group=65+, Year=2001

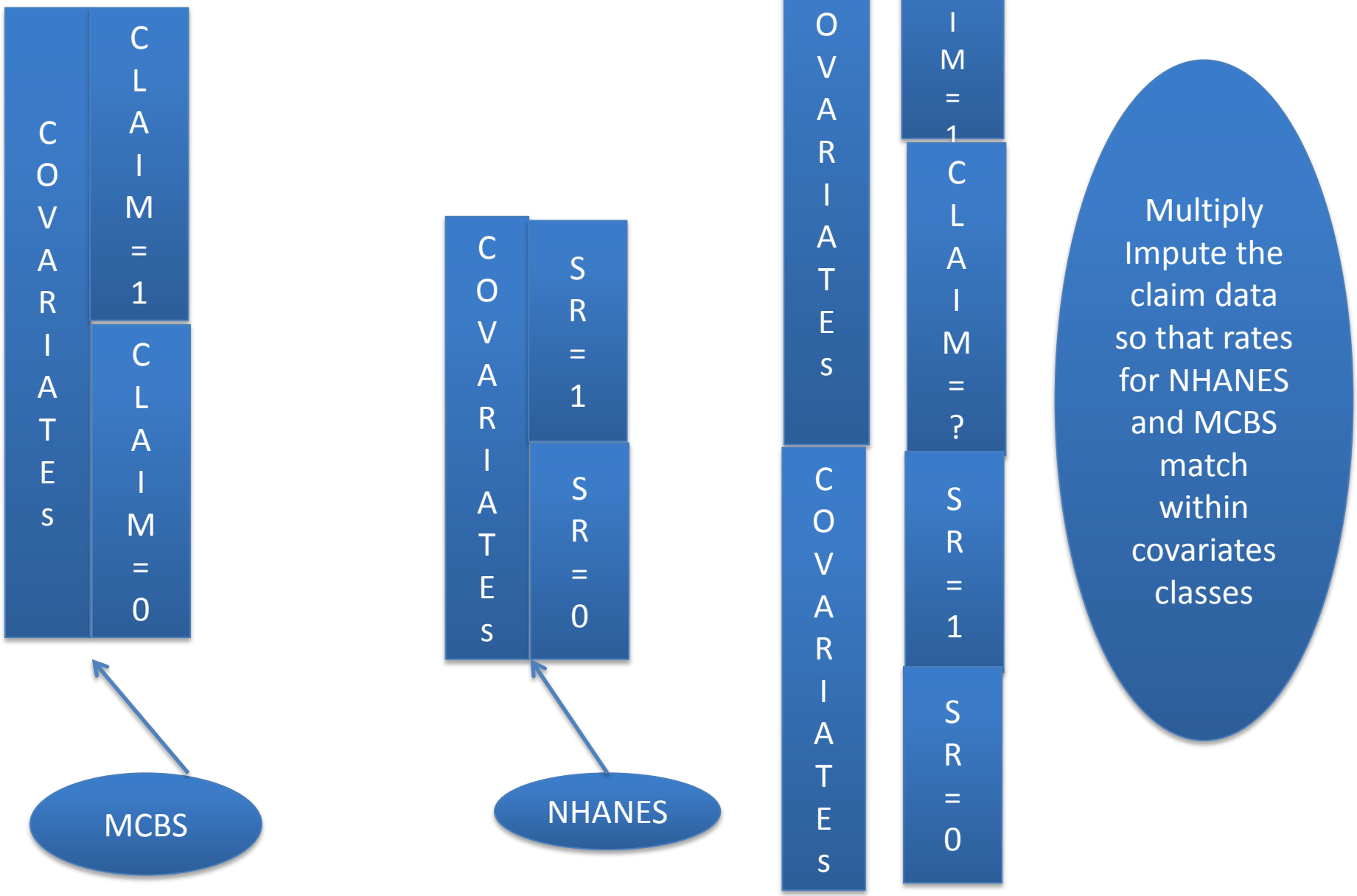
Disease	SR NHANES	SR MCBS	Claims MCBS
Hyperlipidemia	43.81 (2.15)		35.97 (0.68)
Hip Fracture	3.51 (0.82)	3.71 (0.19)	1.12 (0.12)
Asthma	9.29 (1.27)		4.3 (0.2)
Diabetes	17.9 (1.1)	18.9 (0.5)	18.6 (0.6)
Hypertension	55.9 (1.6)	59.6 (0.6)	47.9 (0.8)
Thyroid Disorders			13.90 (0.5)
Depression			4.69 (0.3)
Dermatologic Diseases			26.66 (0.63)

Claim-based disease definitions utilized AHRQ, CCS, and ICD-9-CM

A Scatter plot of self-report and Claim-based prevalence rates



Schematic Display



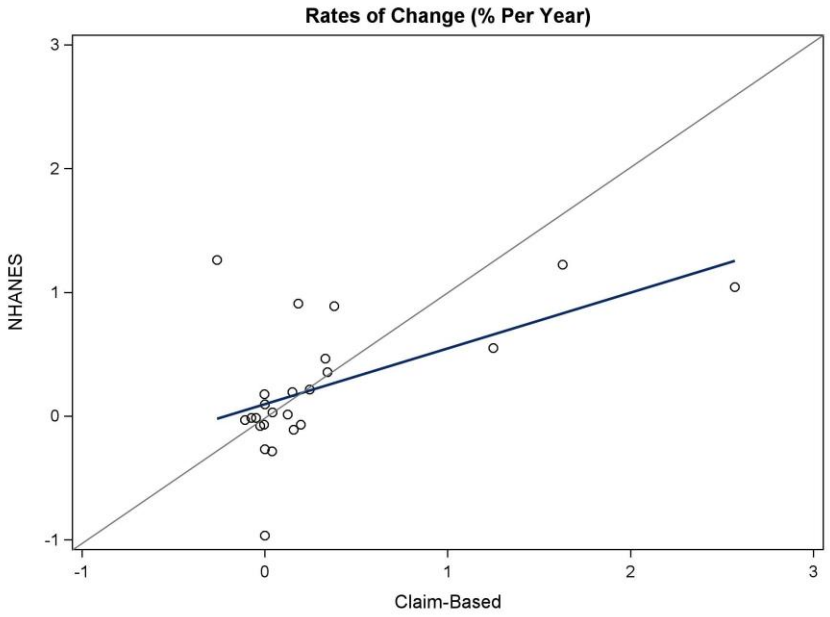
Calibration and Analysis

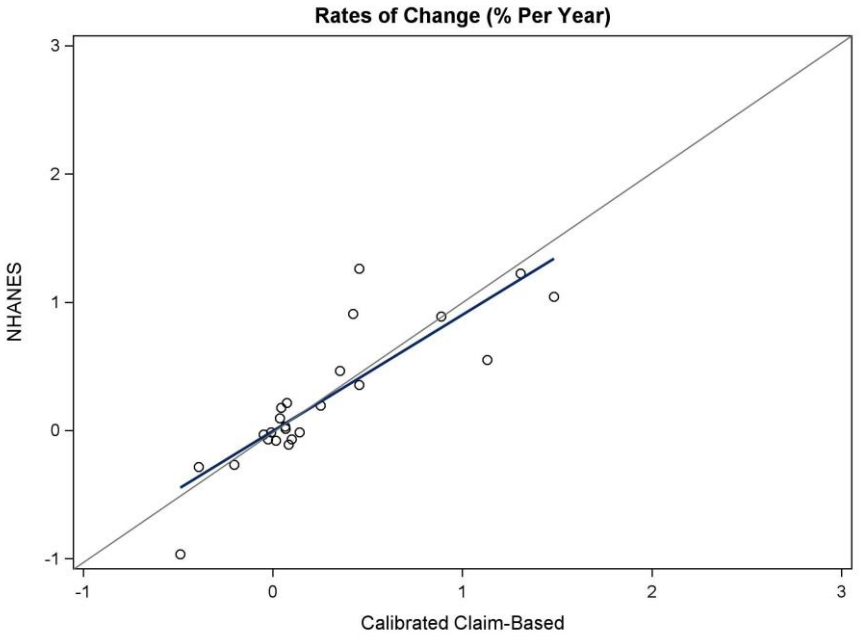
- For diseases with no self-report
  - Construct a measurement error model relating claim and calibrated claim
  - Impute calibrated claims based on this model
- Calibration carried out for each year, separately for Community and Institutionalized populations
- Five imputed data sets with calibrated claims
- All other missing covariates were also imputed
- Obtained the prevalence rates for each disease and year
- Performed a trend analysis using a hierarchical model (random intercepts and slope)
- Performed numerous model diagnostics

Estimated Prevalence of Select Cardiovascular Diseases and Risk Factors for Participants 65 Years and Older: NHANES 2009-2010 and MCBS 2009

Medical Condition	SR NHANES	SR MCBS: Community	MCBS: Claims	Calibrated Claims
Diabetes Mellitus	23.72 (1.38)	23.90 (0.64)	32.15 (0.67)	32.15 (0.67)
Undiagnosed Diabetes Mellitus	2.34 (0.58)			2.04 (0.27)
Hyperlipidemia	51.35 (2.33)	52.31 (1.15)	61.36 (0.89)	62.43 (1.44)
Undiagnosed Hyperlipidemia	1.71 (0.84)			1.64 (0.35)
Hypertension	63.59 (1.75)	69.21 (0.85)	68.50 (0.87)	71.41 (1.49)
Undiagnosed Hypertension	3.17 (1.35)			2.55 (0.57)
Acute myocardial infarction (AMI)	8.58 (0.73)	13.58 (0.55)	2.30 (0.19)	11.19 (1.08)
Acute hemorrhagic stroke			0.64 (0.09)	1.17 (0.21)
Ischemic stroke			5.11 (0.35)	8.12 (0.54)
Any stroke	8.18 (1.03)	11.40 (0.51)	5.40 (0.36)	8.62 (0.58)

Scatter plots of Trend Estimates from Self-report, Claim-Based and Calibrated Claim Based Prevalence Rates





Objective 2: Cost Attribution

- Attributable cost estimated as the difference between those with and without a particular disease other things (covariates and all other diseases) being equal

$$D_j = Disease$$

$$D_{(-\,j)} = Other\ Diseases$$

$$X = Covariates$$

$$Y = Total\ Cost$$

$$A_j = E(Y \mid D = 1, X, D_{(-\,j)}) - E(Y \mid D = 0, X, D_{(-\,j)})$$

Outline of Methods

- A logistic regression model to predict disease dummy variable with covariates and other disease dummy variables as predictors
- Propensity score used to create strata
- Mean difference in the cost for those with and without the disease was computed in each strata
- The weighted average of these differences was defined as the attributable cost for the disease
- Computed attributable cost for all 80 diseases and for all 12 years 1999-2010

Cost Model

- Aggregated individual level cost computed by adding attributable costs for individual level diseases

$$A_j = \textit{Attributed cost for Disease } j$$

$$D_{ij} = 1 \textit{ if subject } i \textit{ has disease } j \textit{ and } 0 \textit{ otherwise}$$

$$Ag.C_i = \sum_{j=1}^{80} A_j D_{ij}$$

- The Aggregated costs and the actual cost may not agree as the cost depends upon several other factors such as hospital stays, number of conditions etc

Cost Model (Contd.)

- Regression model adjustment to predict actual cost

$A_j = \textit{Attributed cost for Disease } j$

$D_{ij} = 1 \textit{ if subject } i \textit{ has disease } j \textit{ and } 0 \textit{ otherwise}$

$$Ag.C_i = \sum_{j=1}^{80} A_j D_{ij} \text{ (Aggregated Cost)}$$

$Ac.C_i = \textit{Actual Cost}$

$$Ac.C_i = Ag.C_i \left[\beta_0 + \sum_k \beta_k X_{ik} \right] + \varepsilon_i$$

$X_1 = \textit{Number of Health Conditions}$

$X_2 = \textit{Number of Health Conditions squared}$

$X_3 = \textit{Dummy variable for no inpatient stays}$

$X_4 = \textit{Number of inpatient stays}$

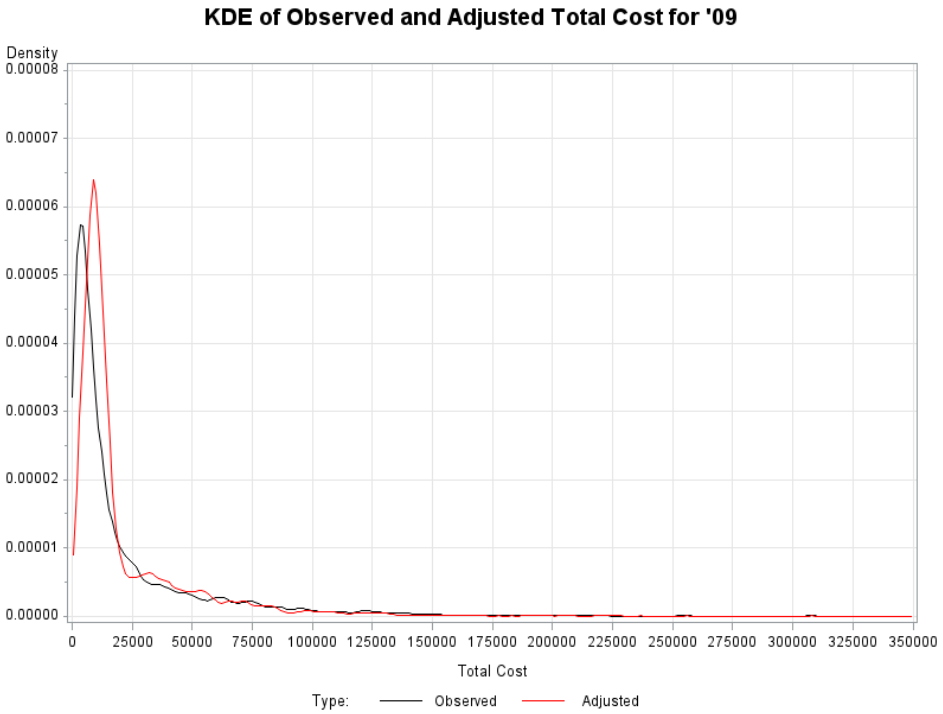
$X_5 = \textit{Number of inpatient nights}$

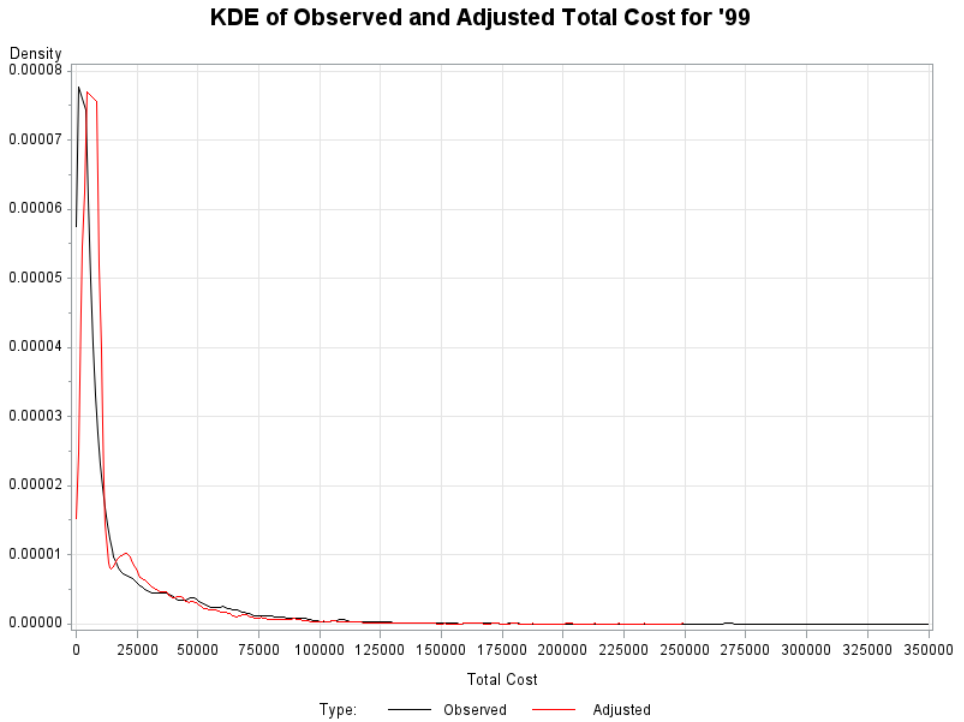
$X_6 = \textit{Dummy for Death during the year}$

$X_7 = \textit{Number of months alive during the year}$

$X_8 = \textit{Number of days institutionalized}$

Comparison of actual and adjusted cost

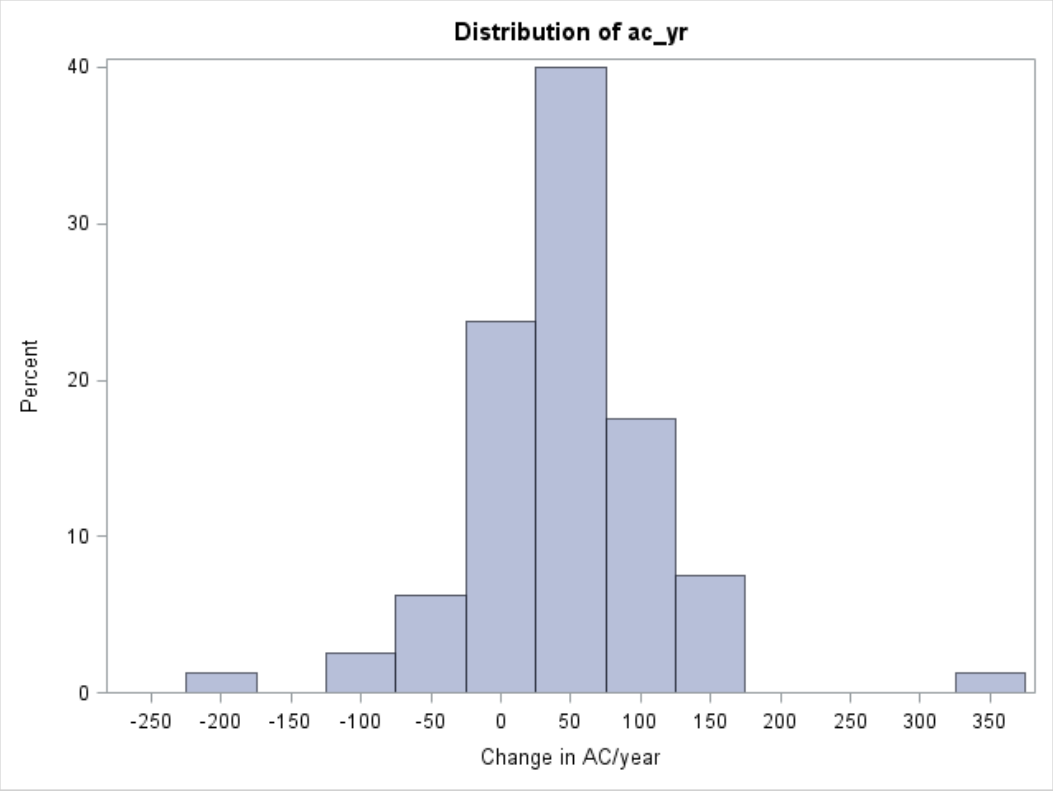




Attributable costs, Prevalence rate of diseases, and 6 covariates are the building blocks for predicting cost using the regression model

Changes in Attributable costs over the 11 year period

- Fitted a hierarchical random effect models for the attributable cost (random intercepts and slopes across the 80 diseases)



Cost-Disease Prevalence Dynamics

- Counter factual Cost per person were computed by applying the attributable cost for Year t to the Prevalence rate for Year s with all other covariates remaining the same.

Year	est99	est00	est01	est02	est03	est04	est05	est06	est07	est08	est09	est10
1999	13100 (227)	13663 (232)	14610 (268)	14819 (217)	14501 (233)	14398 (282)	14949 (293)	14758 (258)	15004 (309)	14733 (273)	14295 (288)	14204 (245)
2000	13553 (224)	13883 (230)	14795 (262)	15031 (206)	14657 (226)	14586 (279)	15201 (288)	15051 (252)	15311 (299)	15068 (267)	14630 (281)	14584 (250)
2001	13104 (223)	13408 (236)	14297 (273)	14531 (206)	14207 (227)	14176 (281)	14817 (292)	14618 (245)	14912 (308)	14708 (270)	14350 (295)	14312 (252)
2002	13426 (223)	13783 (235)	14614 (270)	14903 (214)	14631 (232)	14604 (297)	15225 (301)	15041 (257)	15364 (319)	15144 (282)	14743 (307)	14724 (262)
2003	14184 (238)	14585 (246)	15529 (280)	15820 (221)	15522 (244)	15521 (309)	16243 (314)	16069 (266)	16416 (331)	16145 (298)	15710 (317)	15731 (276)
2004	14761 (260)	15231 (272)	16346 (318)	16624 (240)	16231 (271)	16140 (330)	16905 (341)	16717 (284)	17039 (355)	16743 (314)	16192 (341)	16121 (286)
2005	15343 (258)	15774 (267)	16752 (312)	17039 (240)	16618 (266)	16580 (333)	17295 (346)	17095 (286)	17440 (359)	17182 (320)	16703 (343)	16677 (293)
2006	15879 (274)	16275 (278)	17344 (325)	17670 (250)	17273 (281)	17221 (343)	17969 (345)	17730 (291)	18143 (370)	17923 (332)	17444 (360)	17412 (307)
2007	15610 (264)	16038 (275)	17167 (319)	17466 (249)	17195 (273)	17175 (332)	17927 (340)	17731 (290)	18088 (367)	17846 (324)	17411 (349)	17365 (300)
2008	16379 (300)	16847 (303)	18141 (356)	18398 (275)	17953 (315)	17815 (374)	18617 (366)	18205 (311)	18532 (400)	17834 (335)	17197 (367)	17193 (305)
2009	15905 (268)	16325 (280)	17466 (307)	17825 (247)	17512 (273)	17492 (342)	18261 (345)	18015 (297)	18356 (368)	17981 (330)	17479 (352)	17477 (309)
2010	16325 (286)	16826 (298)	18114 (331)	18426 (268)	18031 (299)	17987 (370)	18764 (368)	18473 (320)	18812 (401)	18218 (344)	17595 (367)	17674 (322)

Analysis

Cost Year	Prevalence year				
	1999	2002	2005	2008	2010
1999	$13,100	$14,819	$14,949	$14,733	$14,204
2002	$13,426	$14,903	$15,225	$15,144	$14,724
2005	$15,343	$17,039	$17,295	$17,182	$16,677
2008	$16,379	$18,398	$18,617	$17,846	$17,193
2010	$16,325	$18,426	$18,764	$18,218	$17,674

Average yearly Change	Dollar (SE)	Percent (SE)	
Due to Prevalence	$94.00 ($12.00)	0.6% (0.1%)	
Due to Cost	$364.00 ($9.70)	2.3% (0.1%)	
Total	$458.00	2.9%	

Analysis of cost by disease group

Disease Group	Average Cost	Average yearly Change (Total)	Due to Prevalence	Due to Cost
Cancer	$623	$14.60	$5.80	$8.90
Chronic and Disabling Conditions	$1,670	$35.30	$9.80	$25.50
Recoverable Acute Conditions	$3,336	$52.50	$5.50	$47.00
Non-Fatal Chronic Conditions	$2,709	$139.00	$36.80	$103.00
Non-Fatal Acute Conditions	$3,730	$83.70	$3.50	$80.20
Other Ill-Defined Conditions	$3,920	$103.00	$51.60	$51.60
Screening	$174	$29.90	$(19.0)	$48.70

Objective 3: QOL and Disease

- Same approach used as in attributable cost estimation

$$D_j = Disease$$

$$D_{(-j)} = Other\ Diseases$$

$$X = Covariates$$

$$Q = QOL$$

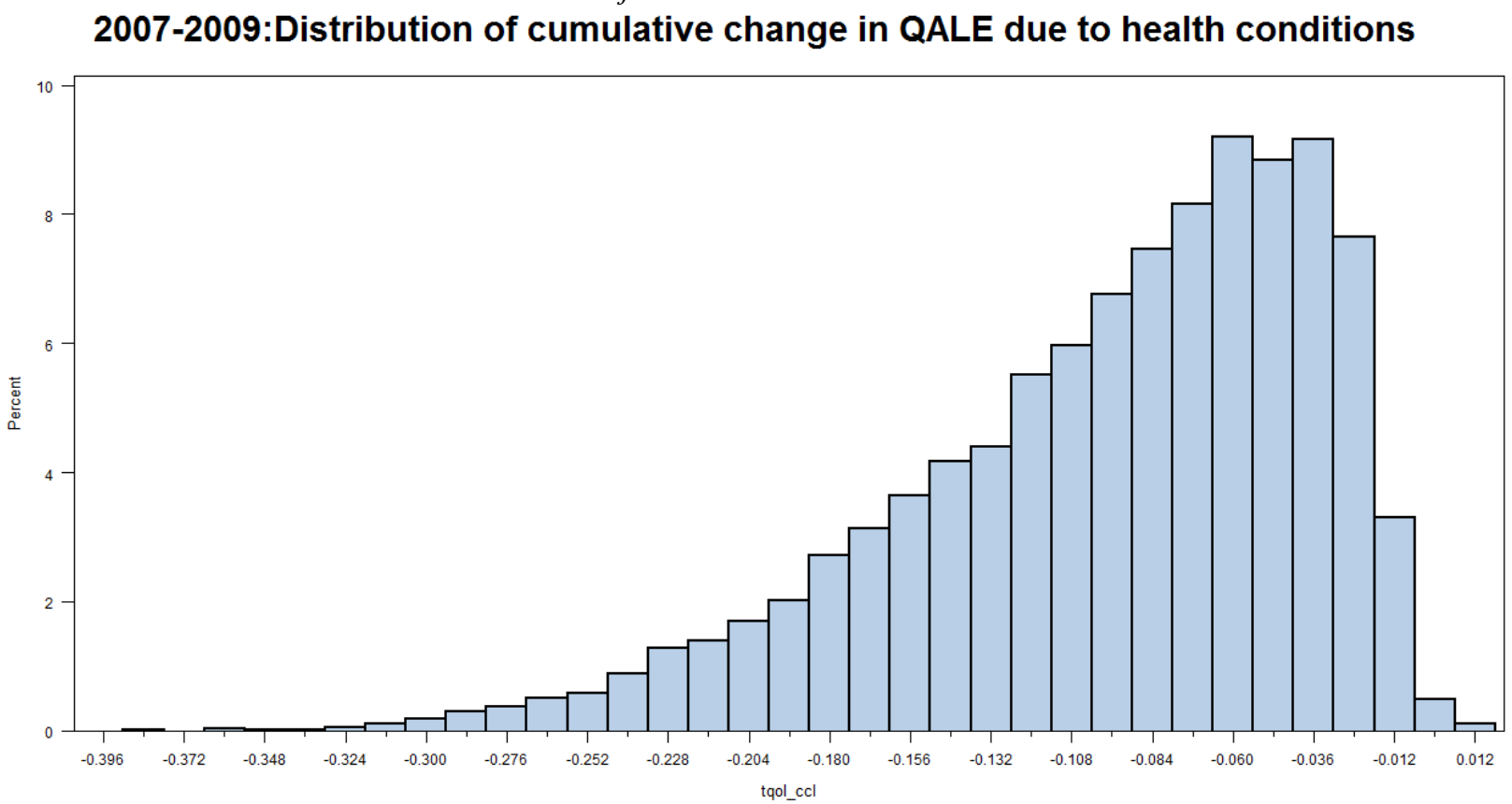
$$Q_j = E(Q \mid D = 1, X, D_{(-j)}) - E(Q \mid D = 0, X, D_{(-j)})$$

- Propensity score stratification method to estimate the attributable QOL to disease j
- Used the combined data for years 2007-2009

Aggregated Reduction in QOL due to diseases

- Computed the reduction by aggregating attributable QOL for the individual level diseases

$$DQ_i = \sum_{j=1}^{80} Q_j D_{ij}$$



Future Work

- Develop models to understand the dynamics of cost, health and diseases
- We have building blocks and now need to do in-depth analysis of all three dimensions, disease, health and cost
- Work on other age groups; negotiating access to HMO data
- Work at Census RDC
- Sparse data on Children
- Electronic medical records, linked to billing on a sample of hospitals might be needed