The B.E. Journal of Economic Analysis & Policy

Contributions

Volume 11, *Issue* 2 2011 *Article* 8

INDUSTRIAL ORGANIZATION AND HEALTHCARE

Data Impediments to Empirical Work on Health Insurance Markets

Leemore Dafny* David Dranove†

Frank Limbrock‡ Fiona Scott Morton**

*Northwestern University, l-dafny@kellogg.northwestern.edu

†Northwestern University, d-dranove@kellogg.northwestern.edu

‡Northwestern University, f-limbrock@kellogg.northwestern.edu

**Yale School of Management, fiona.scottmorton@yale.edu

Recommended Citation

Leemore Dafny, David Dranove, Frank Limbrock, and Fiona Scott Morton (2011) "Data Impediments to Empirical Work on Health Insurance Markets," *The B.E. Journal of Economic Analysis & Policy*: Vol. 11: Iss. 2 (Contributions), Article 8.

Data Impediments to Empirical Work on Health Insurance Markets*

Leemore Dafny, David Dranove, Frank Limbrock, and Fiona Scott Morton

Abstract

We compare four datasets that researchers might use to study competition in the health insurance industry. We show that the two datasets most commonly used to estimate market concentration differ considerably from each other (both in levels and in changes over time), and reflect implausibly high volatility in market shares. By comparison, market share volatility is much lower in a private dataset gathered by a leading investment bank, and in state-level hospital discharge data. We also demonstrate that the outcome of regressions using these data vary considerably by the source used. We conclude that researchers should be cautious about using available data and recommend a new source be developed for public use.

KEYWORDS: market structure, concentration, health insurance

*Leemore Dafny, Kellogg School of Management, Northwestern University. David Dranove, Kellogg School of Management, Northwestern University. Frank Limbrock, Kellogg School of Management, Northwestern University. Fiona Scott Morton, Yale School of Management. We thank Cory Capps, Don Fullerton, Jose Guardado, and the anonymous referee for valuable suggestions.

1. Introduction

Rising healthcare costs in the United States are a persis7tent and significant concern. A rich literature documents the role of several factors underlying this trend, including the development of new technologies coupled with moral hazard generated by insurance, supplier-induced demand, and reimbursement systems that pay for quantity rather than quality. In this article, we focus on one explanation that has recently received attention in the media as well as the scholarly literature: insufficient competition in the health insurance industry. The vast majority of the nonelderly population in the U.S. purchases private health insurance, so that more than half of all spending is directly or indirectly controlled through this industry.

During the healthcare debate preceding the passage of the Patient Protection and Affordable Care Act of 2010 (PPACA), estimates of insurance industry concentration and market shares were frequently cited. For example, a February 25, 2010 article entitled "Health Insurance Competition Vanishing: Study" in US News and World Report states

> "Competition in the health insurance industry is vanishing, according to an American Medical Association report that looked at data from 43 states and 313 metropolitan markets. In 24 of the states, the two largest insurers had a combined market share of 70 percent or more. Last year, 18 of 42 states had that type of market situation. Among the other findings: In 54 percent of metropolitan markets, at least one insurer had a market share of 50 percent or more -- up from 40 percent of metropolitan markets the year before."

The shares are dramatic. By these measures, health insurance would be considered one of the most highly concentrated major industries in the United States. However, digging deeper into the data reveals a number of very serious problems, some of which were publicly highlighted by Capps (2009), in comments submitted to the FTC/DOJ as part of their review process for revising the Horizontal Merger Guidelines. In particular, Capps highlights some irregularities in the frequently-cited American Medical Association data, and contrasts it with data reported by the National Association of Insurance Commissioners.

This paper will take the reader on a more detailed tour of the available public and private data sources for health insurance market shares. We will demonstrate that estimates of state-level Herfindahls have low correlations across

data sources, and changes in shares are almost completely uncorrelated.[1] We also document significant differences across sources in both the magnitude of market shares for leading insurers, and the identity of market leaders. We further illustrate how these data inconsistencies can influence empirical analysis by using different data sources in the same regression specifications. Not only do the sources yield different estimates utilizing cross-sectional variation in concentration measures, but we see the same problem utilizing within-state variation in these measures. The NAIC, AMA, and others may have produced their data for different purposes, which could explain the lack of consistency. However, we document intertemporal volatility in Herfindahls and market shares that strongly suggests the presence of measurement error within each data set.

These findings highlight the need for high-quality, publicly-available data sources on private health insurance. Detailed public data is available for several other healthcare sectors, such as prescription drugs (MEPS), outpatient visits, diagnoses, and tests (NAMCS, NHAMCS), and hospitals (inpatient discharge data). Given that private insurers are poised to serve an *additional* 15 million enrollees as a result of PPACA, and many of these will use government subsidies to purchase coverage, the need to accurately assess market conditions has never been greater.[2] In the meantime, our findings suggest that researchers should exercise extreme caution when using the sources profiled here.

2. Data

We rely on four unique data sources to obtain information on market shares and concentration (as measured by the Herfindahl) at the state level:

1. The National Association of Insurance Commissioners (NAIC)
2. The American Medical Association (AMA)
3. Goldman Sachs (GS)
4. California's Office of Statewide Health Planning and Development (OSHPD)

[1] The Herfindahl index is defined as the sum of squared market shares, thus it ranges between zero and one.

[2] Table 4, http://www.cbo.gov/ftpdocs/113xx/doc11379/AmendReconProp.pdf (downloaded 7/7/2011)

Table 1 provides a brief summary of the differences across these datasets. The NAIC data are collected directly by state insurance commissioners. The AMA data are constructed using data from a private company called HealthLeaders-Interstudy, which combines state insurance commissioner data with its own research. The dataset created by Goldman Sachs (GS) is the result of original research by analysts at the firm. We include it because it represents the best efforts of financially-motivated researchers to create accurate data on market shares using all available information. Last, we also utilize data on hospital discharges in the state of California from the Office of Statewide Health Planning and Development (OSHPD). The OSHPD data capture the entire population of hospital inpatients, and include information on payer identities for HMO patients. We use these data to derive market share estimates for California, and compare these estimates to those reported by the AMA and NAIC. In the subsections that follow, we describe each source of data in greater detail.

Table 1: Data Sets

Dataset	NAIC	AMA	GS	OSHPD
Original Sources	State insurance commissioners	InterStudy databases, built using data from state insurance commissioners and insurer surveys	Private research and public sources	California hospital discharge data
Years Available	2001-2009	2000-2007	2006-2007	2004-2007 (used; more are available)
States Covered (2007)	50 (includes DC, not CA)	43	48	1 (CA)
Plan types included	All	HMO and PPO (HMO only in Section 3.2.1)	All	HMO
Self Insured in Data?	No	Varies (See Section 2)	No (in sample used)	Yes
# largest insurers identified	All	2	Varies (1-8)	14
Source of Herfindahl	Our Calculation	Reported	Our Imputation	Our Calculation

2.1 The National Data Sources

The NAIC data are available from 2001-2009 at the insurer-state-year level. Health insurance companies in each state report annual enrollment and premium data on a form known as the "NAIC Health Insurance Company Annual Statement blank," which mirrors the forms utilized by other lines of insurance business regulated by states, such as Property and Casualty and Auto. Only fully-insured, as opposed to self-insured, plans are regulated by the state, so the data should only reflect this market segment. However, independent audits are not performed.[3] The blanks are the source for annual publications on market shares issued by the NAIC. The variables include total state premiums written and enrollment totals in nine distinct categories, which are listed in Table 2. To maximize comparability with the other datasets, which focus on comprehensive insurance, we exclude the following categories: Vision Only, Dental Only, Medicare Supplement, Long-Term Care, Stop-Loss, and Other. All states are included in the NAIC data except California, where regulatory reporting requirements differ.

Table 2. NAIC Data Categories

Category	Included in Our Sample
Comprehensive – individual	Yes
Comprehensive – group	Yes
Medicare Supplement	No
Vision only	No
Dental only	No
Federal employees health benefit plan	Yes
Title XVIII – Medicare	Yes
Title XIX – Medicaid	Yes
Long-Term Care	No
Disability Income	No
Stop-Loss	No
Other	No

Source: Exhibit of Premiums, Enrollment, and Utilization (reported separately by insurer, state, and year).

The AMA data was first issued in a 2001 report entitled "Competition in Health Insurance" (American Medical Association 2001). This report has been

[3] NAIC analysts are unable to state with certainty that only fully-insured data are included in health blanks.

updated annually, with the exception of 2006, and we use the updates through 2009 in our analysis. The data years correspond to 2001-2007. As noted above, the AMA constructs its reports using data purchased from HealthLeaders/Interstudy, formerly known as Interstudy. The AMA staff adjusts the Interstudy data to enhance comparability of data across markets. The nature of these adjustments varies from year to year. For example, in 2007, the AMA data do not include self-insured HMOs, PPOs administered by non-risk-bearing organizations (e.g., "PPO rental networks"), all enrollment in areas outside an insurer's reported service areas (except for commuters who are not insured under Blue Cross Blue Shield), and states in which the HealthLeaders/Interstudy data did not account for at least 30 percent of the estimated privately-insured population. Market shares and Herfindahls are reported by state and by metropolitan areas, the exact identity and number of which varies from year to year. Data are also reported separately for three distinct product markets: HMO, PPO, and HMO and PPO combined. Unless otherwise noted, we use the data for the combined product market at the state level, as the other sources do not distinguish between the plan types and report state-level figures.

The final source of national data is Goldman Sachs, which provided us with their estimates of enrollment data for the leading insurers in 40 states during 2006 and 2007, separately by fully-insured and self-insured segments. To maximize comparability to our other data sources, we only use the fully-insured segment in our analysis. The leading insurer is always identified, however the number of additional insurers included appears to be arbitrary. Insurers not specifically identified are included in an "other" category. We use these data to compute an upper-bound estimate of state-level Herfindahls by allocating the market share in the "other" category to the smallest number of insurers needed to ensure that the largest "other" insurer has a market share no larger than the smallest identified insurer. For example, if Goldman Sachs identifies insurer 1 as having 50 percent share, and insurer 2 as having 30 percent share, then we allocate the remaining 20 percent to the "other" insurer. To take a slightly more complex example, consider a case in which insurer 1 has a market share of 50 percent and insurer 2 has a share of 15 percent and is the only other insurer identified by name. We assume the remaining 35 percent is distributed among 3 insurers with 15, 15 and 5 percent market share – the highest concentration (as measured by the Herfindahl) possible if none of these is larger than the smallest identified insurer.

Given that the national sources rely primarily (in the case of the AMA data) or exclusively (NAIC and GS) on fully-insured enrollees, it is worth emphasizing that less than half of the privately-insured are covered through fully-

insured plans.[4] Thus, the analysis below pertains only to this segment of the market, which is believed to be more concentrated than the self-insured segment. (This popular wisdom is corroborated in another proprietary data source used in Dafny, Duggan and Ramanarayanan (2010).) Fully insured plans are regulated by the states, whereas self-insured plans are regulated by the federal government.

Before proceeding, we note that another potential data source, the Medical Expenditure Panel Survey – Insurance Component (MEPS-IC), could not be utilized to estimate market share data because this national survey does not consistently gather insurer identities. The MEPS-IC surveys establishments in sufficient number to construct state-level estimates of employer insurance offerings, however plans are often recorded simply as "Plan A" and "Plan B."[5]

2.2 The California Data

The OSHPD hospital discharge dataset is widely used in health services research. Like most state discharge data sets, it provides a complete census of all inpatient admissions. Unlike most other states, California reports the identity of the insurer for a subset of discharges, specifically those financed by HMOs.[6] The data also identify the resident zip code of the patient and the location of the hospital, allowing us to measure insurer market shares based on the location of the enrollee or hospital.

3. Analysis

Our goal is to identify the consistency of information provided within one source over time as well as across different sources of insurance data. To that end, we present pairwise comparisons of the three national sources, focusing first on concentration levels, next on the short and long-run volatility of these levels, and last on changes in concentration levels over time. We use the OSHPD data primarily to highlight the much greater volatility in insurer market shares derived from insurance rather than patient data.

---

[4] According to the Employer Health Benefits 2010 Annual Survey performed by the Kaiser Family Foundation/Health Research Educational Trust, 59 percent of covered workers are in self-insured plans. The Employee Benefit Research Institute reports that over 90 percent of the privately-insured nonelderly obtain coverage through their employers (as opposed to the individual market), leaving fewer than 50 percent in fully-insured plans. (Employee Benefit Research Institute "Fast Facts", November 15, 2010.)

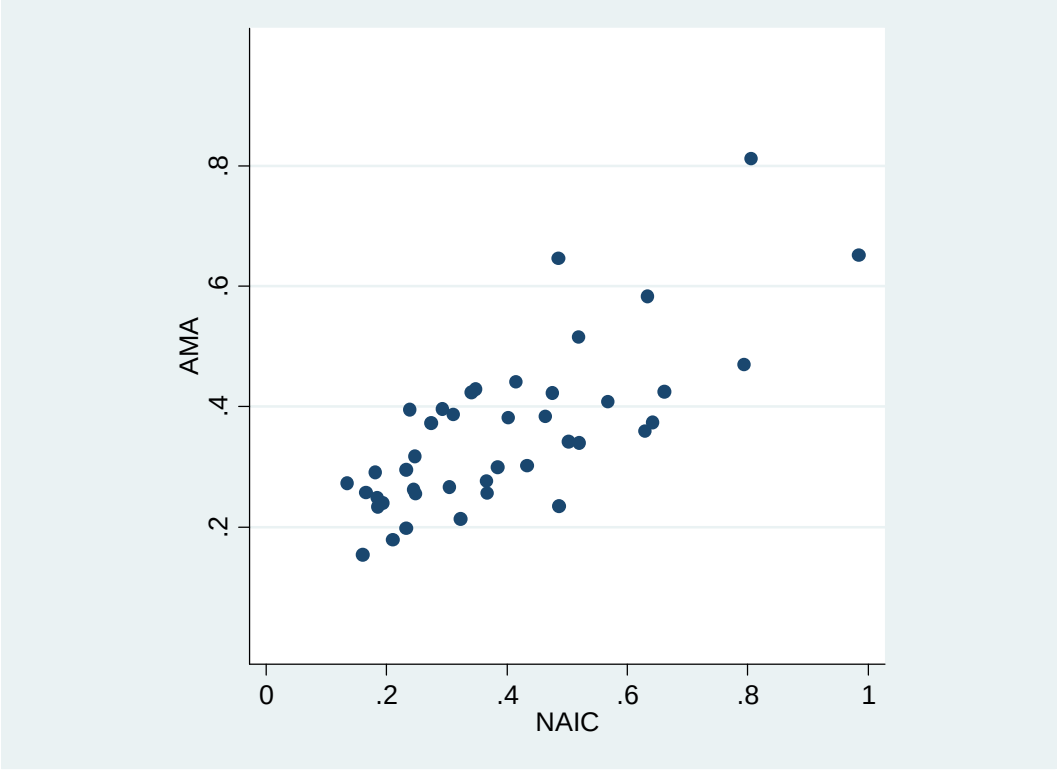
[5] Per personal correspondence with the Agency for Healthcare Research and Quality.

[6] Other states that identify specific payers include West Virginia, Massachusetts, and Maryland.

3.1 Comparison of State-level Herfindahls

We begin by comparing state-level Herfindahls computed from the three national data sources. The mean Herfindahl is approximately 0.35 using the AMA data, 0.40 using the NAIC data and .38 using the Goldman Sachs data. The correlation between the AMA and NAIC Herfindahls in 2007 is 0.74, implying an R^2 of just 0.55 from a univariate regression in which one measure is regressed on the other.[7] These data are graphed in Figure 1 below. The mean absolute value of the state-level differences between these datasets is 0.11. The correlation of the GS Herfindahl with the AMA Herfindahl is 0.79, and with the NAIC Herfindahl 0.83.

Figure 1: AMA vs. NAIC Herfindahls, 2007



Note: N=42. The following states are missing in the 2007 AMA data: Connecticut, Mississippi, Montana, North Dakota, Pennsylvania, Wisconsin, West Virgina. The NAIC data excludes California. Data for this figure is listed in Table 3.

[7] Across all state-years with both AMA and NAIC observations, this correlation is .68, implying an R2 of .46.

When one examines individual state Herfindahls, more striking differences emerge. For illustrative purposes, Table 3 presents the 2007 AMA and NAIC Herfindahls graphed in Figure 1. We lack permission to list the Goldman Sachs data, but we discuss aggregate facts below. Our qualitative findings are similar when we examine other years.

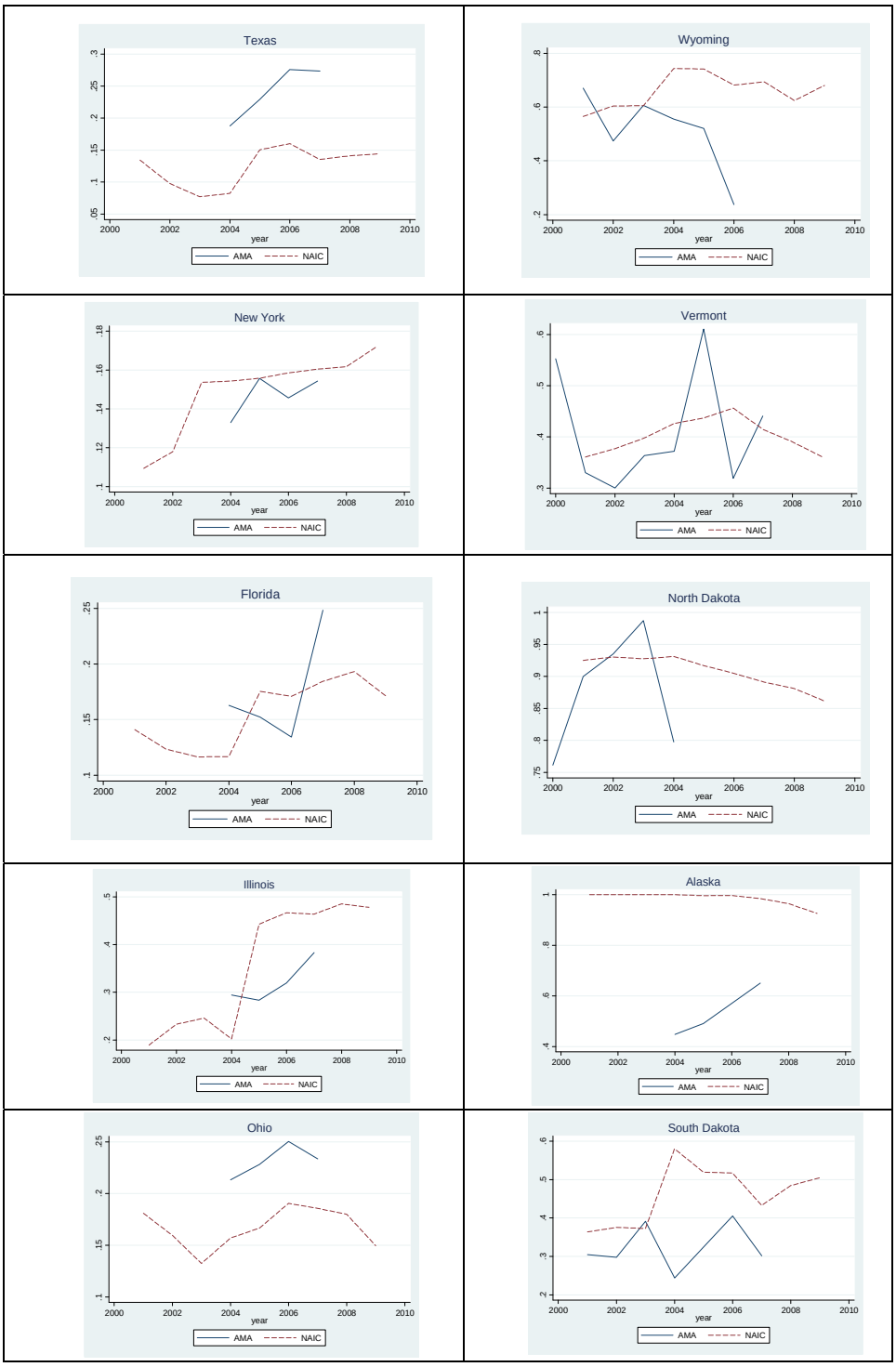
The AMA and NAIC Herfindahls differ by 0.10 or more in 17 out of 42 states. A difference of 0.10 is generated, for example, by a combination of two firms each with 22.5-percent market share into a single firm with a 45 percent share. The AMA and GS Herfindahls differ by 0.10 or more in 10 out of 40 states. The corresponding figure is 14 out of 47 states for the NAIC-GS comparison. Further visual evidence is presented in Figure 2, which consists of 10 individual graphs, one each for the 5 most and 5 least-populous states. The figure presents timeseries of Herfindahls for the AMA and NAIC data in each state. The graphs show larger discrepancies in the smallest states, as well as greater volatility in the AMA data series (which we discuss below).

These differences can profoundly affect the outcomes of antitrust cases. According to the 2010 Horizontal Merger Guidelines issued by the Department of Justice and the Federal Trade Commission, a market with a Herfindahl of 0.25 or higher is considered highly concentrated while a market with a Herfindahl of 0.15 or less is considered unconcentrated. Markets with Herfindahls between 0.15 and 0.25 are moderately concentrated. Antitrust agencies are more likely to challenge mergers in highly concentrated markets. Using the state as the market purely for illustrative purposes, we note that the AMA and NAIC agree that markets are highly concentrated in 27 out of 42 states. NAIC and GS agree that markets are highly concentrated in 32 out of 47 states, while AMA and GS agree that 27 out of 40 states are highly concentrated. The AMA reports no unconcentrated markets in 2007; there are 3 according to NAIC, one of which (WI) is missing in the AMA data; the other (TX) is highly concentrated according to AMA. According to the GS data, for which Herfindahls are upper bounds by construction, one state (WI) is unconcentrated. Finally, there are 7 states that are considered highly concentrated using either the AMA or NAIC data, moderately concentrated using the other, and the difference in Herfindahl is 0.09 points or higher. To summarize, our sources often paint different pictures of the extent of concentration. Given that researchers frequently compare behavior across markets designated as concentrated or unconcentrated, these discrepancies can lead to incorrect findings.

Table 3: Herfindahl Indices for 2007, AMA and NAIC data

State	AMA	NAIC
AK	0.651	0.985
AL	0.811	0.806
AR	0.469	0.794
AZ	0.296	0.233
CO	0.198	0.233
DC	0.318	0.248
DE	0.381	0.403
FL	0.249	0.184
GA	0.395	0.239
HI	0.646	0.486
IA	0.582	0.634
ID	0.342	0.502
IL	0.383	0.464
IN	0.373	0.274
KS	0.257	0.165
KY	0.299	0.384
LA	0.340	0.521
MA	0.424	0.341
MD	0.267	0.305
ME	0.359	0.630
MI	0.396	0.293
MN	0.387	0.311
MO	0.291	0.182
NC	0.374	0.643
NE	0.408	0.569
NH	0.235	0.487
NJ	0.263	0.246
NM	0.255	0.249
NV	0.213	0.322
NY	0.154	0.160
OH	0.234	0.186
OK	0.276	0.366
OR	0.179	0.210
RI	0.515	0.520
SC	0.424	0.662
SD	0.302	0.433
TN	0.422	0.476
TX	0.273	0.135
UT	0.256	0.367
VA	0.429	0.348
VT	0.441	0.415
WA	0.240	0.194

Figure 2: AMA and NAIC Herfindahls, Five Most and Least Populous States



3.2 Short-run Volatility

The NAIC, AMA, and GS may have produced their data for different purposes, which could explain the lack of consistency. However, a visual inspection of the time-series data for each state reveals substantial volatility in annual Herfindahls. Such volatility suggests the presence of measurement error. In this section, we document the volatility more thoroughly.

Table 4 lists the within-state standard deviations of the AMA and NAIC Herfindahls. For each state we restrict the analysis to those years when data is available from both sources. (Because the GS data includes only 2 years, we do not report standard deviations for this source.) The mean standard deviation in both data sets is 0.06. There are 49 instances in the AMA data where the AMA Herfindahl increases or decreases by 0.10 or more year-on-year, which is more than one-fifth of all observed year-on-year changes. In the NAIC data, we observe a change of 0.10 or more 24 times. (In the Goldman Sachs data, we observe only one change of such magnitude.) Once again, these findings have important antitrust implications. Given the role of the Herfindahl in generating antitrust scrutiny, it is critical for it to be accurately measured. Consider that in the AMA data there are 18 instances (out of 210 candidate state/years) where the Herfindahl in consecutive years falls on either side of the 0.25 "very concentrated" threshold and there was a year-on-year change in the Herfindahl of at least 0.05. In 7 instances, the 0.25 threshold is crossed and the year-on-year Herfindahl change is at least 0.10. In other words, concentration in these states appears to move dramatically over time in a policy-relevant way.

In the NAIC data, we observe 6 instances where the Herfindahl crosses the 0.25 threshold in either direction and the Herfindahl changes by at least 0.05. In 4 instances, the 0.25 threshold is crossed and the Herfindahl changes by at least 0.10. Interestingly, there are no such changes In the Goldman Sachs data.

In addition to the high degree of short-run volatility displayed *within* each source, there is disagreement *across* the sources with regard to the sign of year-on-year changes. There are 9 occasions when both the AMA Herfindahl changes by at least 0.05, and the NAIC Herfindahl for the same state and years moves in the opposite direction from the AMA Herfindahl. There are also 4 instances where the reverse happens – the NAIC Herfindahl crosses the threshold by at least 0.05 and the AMA Herfindahl moves in the opposite direction.

Finally, we also find that volatility is inversely correlated with state population: the correlation is -0.37 for the AMA data and -0.18 in the NAIC data. Although entirely expected, this finding suggests volatility will be even more problematic for analysts who use these data and adopt smaller geographic market definitions, as most courts and researchers have done.

Table 4: Standard Deviation of Herfindahls between 2001-2007, overlapping years only, for AMA and NAIC

State	AMA	NAIC
AK	0.107	0.008
AL	0.075	0.195
AR	0.079	0.080
AZ	0.027	0.012
CO	0.024	0.032
CT	0.041	0.018
DE	0.075	0.082
FL	0.051	0.031
GA	0.046	0.144
HI	0.030	0.092
IA	0.085	0.172
ID	0.044	0.144
IL	0.045	0.128
IN	0.088	0.066
KS	0.033	0.023
KY	0.039	0.059
LA	0.058	0.043
MA	0.062	0.038
MD	0.040	0.004
ME	0.082	0.061
MI	0.019	0.022
MN	0.053	0.014
MO	0.101	0.006
MT	0.082	0.068
NC	0.046	0.070
ND	0.080	0.003
NE	0.095	0.108
NH	0.101	0.053
NJ	0.053	0.015
NM	0.025	0.037
NV	0.031	0.064
NY	0.010	0.003
OH	0.015	0.016
OK	0.042	0.038
OR	0.015	0.010
RI	0.075	0.069
SC	0.074	0.049
SD	0.062	0.090
TN	0.075	0.038
TX	0.042	0.035
UT	0.029	0.017
VA	0.079	0.094
VT	0.107	0.034
WA	0.021	0.002
WI	0.116	0.004
WV	0.050	0.029
WY	0.150	0.076

3.2.1 Comparison of AMA and Hospital Inpatient Data

As suggested above, the volatility of the AMA and NAIC data casts doubt on their validity. To confirm our doubts, we examined one additional source: hospital discharge data for the state of California. As noted above, the OSHPD data includes the name of the health plan for HMO patients. We use these data to compute statewide HMO market shares and concentration and compare these estimates with statewide HMO market share and concentration figures provided by the AMA (in its data series pertaining to the HMO product market only). Before discussing the results, we note that the OSHPD market share *levels* may not align well with the AMA levels because those admitted to the hospital are not a random sample of HMO enrollees. However we would not expect this factor to influence yearly *volatility*.

Table 5 reports the market share leaders and Herfindahls for the years 2004-2007, separately for the AMA and the OSHPD inpatient data. It is readily apparent that the OSHPD time series is far more stable than the AMA series. The Herfindahl as measured using the OSHPD data remains within a .011 range over this time period, while the Herfindahl reported by AMA varies by .042. The raw market share data are more volatile, too: in the OSHPD data, the market share of the leading insurer (Kaiser Permanente) varies within a 2 percent window; in the AMA data, Kaiser's share varies by 6 percent during this four-year window.

Table 5: Market Shares and Herfindahls from AMA HMO Market Data and California Hospital Data

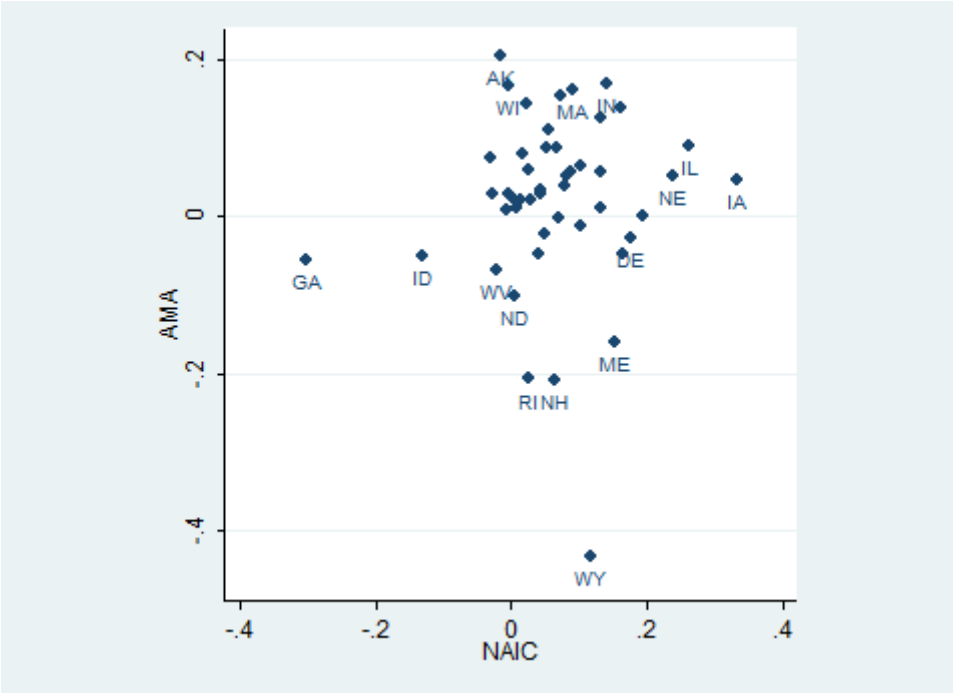
	AMA HMO Market Data				
	Herfindahl	**Insurer 1**	**Insurer 1 Share**	**Insurer 2**	**Insurer 2 Share**
2004	0.246	Kaiser	43%	WellPoint Inc.	14%
2005	0.238	Kaiser	42%	WellPoint Inc.	18%
2006	0.264	Kaiser	46%	UnitedHealth	12%
2007	0.289	Kaiser	49%	Health Net	12%
	California Hospital Discharge Data				
	Herfindahl	**Insurer 1**	**Insurer 1 Share**	**Insurer 2**	**Insurer 2 Share**
2004	0.169	Kaiser	35%	WellPoint Inc.	14%
2005	0.173	Kaiser	35%	WellPoint Inc.	14%
2006	0.179	Kaiser	36%	WellPoint Inc.	14%
2007	0.180	Kaiser	37%	WellPoint Inc.	14%

Note: In the California hospital discharge data, WellPoint Inc. is coded as BC of California.

3.3 Long-run volatility

The preceding section suggests that analyses relying on year-on-year changes in the Herfindahl, as reported by the AMA or NAIC, are unlikely to yield consistent (let alone convincing) results. Here we examine the approach often adopted by researchers who face data measurement issues: taking long differences. Long differences are presumed to be less noisy than short differences. To evaluate this approach, for every state we identified the earliest and the latest year for which we observe Herfindahls from both data sources; the average time span is 4.4 years. We then subtracted the earlier from the later value. Figure 3 shows these long differences in a scatterplot. We find that the two measures deliver very different estimates of long-term changes for several states, 14 of which are directly marked on the graph. For example, the NAIC data implies an increase in Herfindahl of 0.15 for Maine, while the AMA data reports a decrease of 0.16. Across all 47 states in this comparison, the correlation of long-term changes is 0.07.

Figure 3: Long differences in Herfindahls, 2002- 2007



3.4 Comparison of Market Share Data

A similar set of problems emerge when we examine individual health plan market shares. Tables 6 and 7 list the names and shares of the market leaders in 2007.

Table 6: Largest Insurers by State, AMA and NAIC data, 2007

State	AMA Insurer 1	Percent	NAIC Insurer 1	Percent
AK	Premera BC	79	Premera BC	99
AL	BCBS AL	90	BCBS AL	90
AR	BCBS AR	65	BCBS AR	89
AZ	BCBS AZ	45	BCBS AZ	39
CO	WellPoint Inc.	28	Kaiser	37
DC	UnitedHlthcare	42	Carefirst Inc Group	39
DE	BCBS of DE	56	Magellan Hlth Serv Inc	59
FL	BCBS FL	43	BCBS FL	37
GA	WellPoint Inc.	60	WellPoint Inc.	42
HI	BCBS HI	77	BCBS HI	66
IA	Wellmark	75	Wellmark	78
ID	BC of ID	49	BC of ID	64
IL	HCSC (BCBS)	59	HCSC (BCBS)	67
IN	WellPoint Inc.	57	WellPoint Inc.	46
KS	BCBS of KS	45	Preferred Hlth Systems	24
KY	WellPoint Inc.	44	WellPoint Inc.	56
LA	BCBS LA	53	BCBS LA	70
MA	BCBS MA	63	BCBS MA	54
MD	CareFirst BCBS	42	CareFirst BCBS	48
ME	WellPoint Inc.	54	WellPoint Inc.	78
MI	BCBS MI	59	BCBS MI	51
MN	BCBS MN	55	BCBS MN	43
MO	WellPoint Inc.	48	Coventry Corp Group	33
NC	BCBS NC	54	BCBS NC	79
NE	BCBS NE	58	BCBS NE	73
NH	WellPoint Inc.	38	WellPoint Inc.	64
NJ	Aetna	35	BCBS OF NJ GRP	41
NM	Presbyterian Hlth	38	FHC Hlth Systems Grp	32
NV	Sierra Hlth	34	Sierra Hlth	51
NY	WellPoint Inc.	26	UnitedHealth Group	25
OH	WellPoint Inc.	38	WellPoint Inc.	36
OK	HCSC (BCBS)	42	HCSC (BCBS)	55
OR	Regence BCBS	27	Regence BCBS	36
RI	BCBS RI	68	BCBS RI	68
SC	BCBS SC	63	BCBS SC	81
SD	Wellmark	49	Wellmark	63
TN	BCBS TN	63	BCBS TN	68
TX	HCSC (BCBS)	44	HCSC (BCBS)	31
UT	Intermountain Hlth	34	Intermountain Hlth	48
VA	WellPoint Inc.	64	WellPoint Inc.	57
VT	BCBS VT	64	BCBS VT	57
WA	Premera Blue Cross	38	Regence BlueShield	27

Table 7: Second-Largest Insurers by State, AMA and NAIC data, 2007

State	AMA Insurer 2	Percent	NAIC Insurer 2	Percent
AK	Aetna	15	Wellcare Group	0
AL	UnitedHlthCare	4	Viva Health Inc	5
AR	UnitedHlthCare	21	QCA Health Plan Inc	5
AZ	UnitedHlthCare	27	UnitedHlthCare	22
CO	UnitedHlthCare	26	Wellpoint Inc Grp	25
DC	CareFirst	33	UnitedHlthCare	19
DE	Aetna	19	BCBSD Inc	18
FL	UnitedHlthCare	19	Humana	12
GA	UnitedHlthCare	17	Wellcare Group	16
HI	Kaiser	21	Kaiser	22
IA	UnitedHlthCare	15	UnitedHlthCare	12
ID	Primary Hlth	29	Regence Group	31
IL	WellPoint Inc.	12	WellPoint Inc.	10
IN	HlthCare Group	19	Hoosier Motor Mut Ins Co	19
KS	UnitedHlthCare	15	Childrens Mercys Family Hlth Partner	19
KY	Humana	28	Humana	21
LA	UnitedHlthCare	19	Humana	13
MA	Tufts	13	Harvard Pilgrim	17
MD	UnitedHlthCare	22	UnitedHlthCare	22
ME	Aetna	23	Aetna	12
MI	Aetna	19	Spectrum Hlth Group	12
MN	Medica	25	Medica	31
MO	UnitedHlthCare	18	BCBS OF KC Group	21
NC	UnitedHlthCare	26	UnitedHlthCare	11
NE	UnitedHlthCare	24	Coventry Corp Group	15
NH	Harvard Pilgrim	23	Harvard Pilgrim	26
NJ	UnitedHlthCare	27	Magellan Hlth Serv Inc Grp	21
NM	Ardent Hlth Services	24	Presbyterian Hlthcare Serv Grp	31
NV	WellPoint Inc.	24	WellPoint Inc.	21
NY	GHI	21	WellPoint Inc.	24
OH	Medical Mutual	24	Medical Mutual	16
OK	CommunityCare	20	CommunityCare	21
OR	Providence Hlth	26	Kaiser	23
RI	UnitedHlthCare	23	UnitedHlthCare	19
SC	Kanawha HC Solutions	11	Select Health Of SC Inc	7
SD	Sioux Valley Hlth	18	Sioux Valley Hlth	12
TN	UnitedHlthCare	12	PHP Group	8
TX	UnitedHlthCare	22	Amerigroup	10
UT	Regence BCBS	32	Regence BCBS	33
VA	Aetna	9	Sentara Health Management Group	11
VT	MVP Hlth	14	Vermont Health Plan Llc	22
WA	Regence BlueShield	27	Premera Blue Cross	25

The average share of the top insurer is approximately 52 percent in the AMA data and 55 percent in the NAIC data. But that is where the similarity ends. The AMA and NAIC data disagree on the market share leader in 8 out of 42 states, and in 19 of the 33 states where the AMA and NAIC identify the same market leader, the leaders' market shares differ by 10 percent or more. The AMA and GS data disagree on the market share leader in 6 out of 40 states, and in 12 of the 34 states where the AMA and GS identify the same market leader, the leaders' market shares differ by 10 percent or more. Last, the NAIC and GS data disagree on the market share leader in 13 out of 47 states, and in 12 of the 34 states where the NAIC and GS data identify the same market leader, the leaders' market shares differ by 10 percent or more.

As with the Herfindahl data, the differences in market shares can profoundly affect antitrust analysis. For example, courts have never found a defendant to possess monopoly power with a market share below 50 percent (Barnett and Wellford 2008). Consider that there are 9 states in which one data source lists the leader's market share at over 50 percent while the other data source puts it at least ten points lower and below 50 percent. As with merger cases, the outcome of a monopolization claim could depend on which data source the parties choose to use. Table 7 lists the names and shares of the second largest insurer in each state. There is no longer any semblance of agreement between the data sets. The two sources agree on the identity of the second largest insurer in just 16 out of 42 states. The AMA and GS data agree on the identity of the second largest insurer in 18 out of 37 states, and the NAIC and GS data agree in 22 out of 43 states.[8]

3.4.1 Mergers are not detectable in the data

Following Capps (2009), we also explore our data sources to see whether they capture the most salient examples where market concentration is changing, namely, mergers. If the data do capture mergers, then it might be reasonable to use the merger events as instruments for market structure (so long as they are orthogonal to other determinants of the outcome of interest). This is the approach pursued by Dafny et al (2010), which utilizes a proprietary dataset on employer-sponsored health insurance offered by a sample of large firms to calculate market shares. We attempted to determine whether the AMA and NAIC data are amenable to such an approach. We obtained a list of recent insurance mergers from the American Medical Association. Close inspection of the data reveals that changes in Herfindahls that might have resulted from these mergers are frequently lost against background noise. When detectable, the sources often disagree on the

[8] The number of states is smaller than in the largest-insurer comparisons because the GS data identifies only one insurer by name in some states.

magnitude and even the direction of share shifts. For example, the market share of United Healthcare in Arizona rose from 22 percent in 2005 (the year United Healthcare acquired Pacificare) to 33 percent in 2006, and fell back to 27 percent in 2007, according to the AMA. In these same years, the NAIC data show United's share holding steady, at 26 percent in 2005, 25 percent in 2006, and 27 percent in 2007. Colorado shows an 11 percent change between 2005 and 2006 according to the AMA, while the NAIC data report a decrease of 3 percent.

4. Implications for Health Services Research

At a minimum, noisy data will cause attenuation bias in most empirical analyses. In this section we examine whether the results of regressions using AMA and NAIC market shares might otherwise be reliable.

4.1 Using the Herfindahls in Regression Studies

In this section we offer some evidence that the choice of data source can profoundly affect empirical studies of insurer market power. We specifically examine the relationship between insurance market concentration and the uninsured rate. We choose this relationship both because the data are readily available and there is a plausible story connecting concentration (and, implicitly, higher premiums) to reduced insurance coverage. We do not address identification in this simple analysis, as our intent is not to convincingly demonstrate that concentration affects coverage. Rather, we want to raise concerns about the reliability of any analysis based on a particular measure of concentration.

We obtained data on uninsured rates by state and year from the Annual Social and Economic Supplement (ASEC) portions of the 2002-2008 March Current Population Surveys, which pertain to the preceding year. We begin with a simple cross-sectional regression analysis, using data for 2007. The key predictor is the Herfindahl index calculated from each data source. We control for economic conditions using the unemployment rate (obtained from the U.S. Bureau of Labor Statistics) and median household income (U.S. Census Bureau). To ease the comparison across Herfindahl data sources, we perform the regressions on the 42 states that are in both the AMA and NAIC samples. We also present results using the GS estimates, which are available for 40 of these states. If concentration were associated with higher prices, and therefore more people dropping out of the insurance market, we might expect a positive coefficient on the Herfindahl in these tables.

Table 8 presents our regression findings. The coefficient on the Herfindahl is negative (counterintuitively) across all three models, however its

magnitude varies three-fold, from -3.48 (with a standard error of 3.40) to -9.91 (and a standard error of 4.75). Thus, both magnitudes and precision are impacted by the data source.

Table 8: OLS Regressions of Percent Uninsured on Herfindahls and Control Variables

	2007			2001-2007	
	(1)	(2)	(3)	(4)	(5)
Data Source	**AMA**	**NAIC**	**GS**	**AMA**	**NAIC**
Herfindahl	-9.905**	-3.482	-7.672	-0.730	0.211
	(4.751)	(3.396)	(4.601)	(1.388)	(1.289)
Median Income	-0.229**	-0.233**	-0.215**	-0.222***	-0.220***
	(0.090)	(0.094)	(0.101)	(0.053)	(0.054)
Unemployment	-0.139	-0.111	0.130	-0.241	-0.238
	(0.674)	(0.701)	(0.717)	(0.208)	(0.208)
State Fixed Effects	No	No	No	Yes	Yes
Year Fixed Effects	No	No	No	Yes	Yes
Constant	31.35	29.25	29.11	-	-
Sample size	42	42	40	248	248
Adjusted R2	0.158	0.086	0.086	0.912	0.912

Sources: Current Population Survey, Table HIA-6 (Uninsured Rate); U.S. Census Bureau (Median Household Income); Bureau of Labor Statistics (Unemployment Rate).

Analysts often rely on fixed effects models to control for unobserved time-invariant predictors of the dependent variable that might be correlated with a key predictor variable. Such fixed effects models effectively rely on intertemporal, within-market changes in the values of the key predictors. We estimate such models using all states and years that are jointly available in the AMA and NAIC data. The specification is analogous to the cross-sectional specification previously discussed, however state and year fixed effects are included. The coefficient estimates, displayed in columns 4 and 5 of Table 8, have opposite signs, although both are noisily estimated. The range of results is not particularly surprising given the correlation in first differences for the AMA and NAIC data (using all years) is 0.07. To check whether this might reflect different timing of data collection and reporting, we correlated one period lags and leads of the first differences; the correlations are never larger than 0.12. Thus, fixed effects regressions using the two data sources will frequently generate inconsistent findings. Interestingly, using all of the available data for the NAIC sample (rather than the subset of state-years also included in the AMA reports) yields a positive and statistically-significant coefficient on the Herfindahl (1.757 with a standard

error of 0.961). Thus, results are also impacted by differences in the sample utilized.

Table 9 reports results from a regression that exploits long-run changes in Herfindahls to identify their effect on uninsurance rates (a "long differences" model). Only the first and last years for which AMA and NAIC Herfindahls were available for a given state are included in the estimation. The corresponding state-year controls are included, as are year and state fixed effects. The coefficient estimates for the Herfindahl again vary widely.[9]

Table 9: OLS Regressions of Percent Uninsured on Herfindahls and Control Variables, First and Last Coinciding AMA and NAIC Data Years Only

	(1)	(2)
Data Source	**AMA**	**NAIC**
Herfindahl	-1.343	2.605
	(2.717)	(2.817)
Median Income	-1.343	-0.171
	(2.717)	(0.120)
Unemployment	0.308	0.302
	(0.327)	(0.317)
Indicator "Last year"	1.765**	1.448**
	(0.690)	(0.684)
State Fixed Effects	Yes	Yes
Sample size	96	96
Adjusted R^2	0.893	0.894

Sources: Current Population Survey, Table HIA-6 (Uninsured Rate); U.S. Census Bureau (Median Household Income); Bureau of Labor Statistics (Unemployment Rate).

5. Conclusions

We conclude that the publicly-available sources of data on health insurance market shares are unreliable. They show great variability across years relative to both a reasonable prior and to the variability exhibited in hospital discharge data. They do not reflect merger activity. In addition, they omit important components of the market such as self-insured healthplans.

The private insurance industry plays a tremendously important role in the U.S. healthcare sector. Over 65 percent of the nonelderly purchased private insurance plans in 2009, and the majority of the 17 percent covered through Medicaid are also enrolled in private plans.[10] Within the elderly population, 95 percent are enrolled in Medicare, and one quarter of these opt for a private

[9] AMA is aware of inconsistencies in the way the data are constructed before and after (data year) 2006 and has cautioned us against computing trends across this boundary.
[10] Employee Benefit Research Institute "Fast Facts", November 15, 2010.

"Medicare Advantage" insurance plan.[11] And these figures refer to comprehensive medical insurance only; millions more rely on the private insurance industry for prescription drug coverage through Medicare Part D, and for supplemental benefits (e.g. Medigap). Finally, as a result of PPACA, the private insurance industry is projected to cover an *additional* 15 million lives by 2019.

It is therefore of great policy interest to evaluate industry market structure and performance. Such facts can help policymakers decide whether to rely on the private sector when implementing expansions in coverage or benefits, and provide useful insights on the design of competition policy for this sector.

The results in this paper show that researchers and antitrust analysts cannot currently generate accurate empirical analyses of competition in the health insurance industry using readily-available market share data. The marked differences in shares and concentration reported across different data sources will, at a minimum, force researchers to choose among competing data sets. The implausibly high volatility within data sets suggests that such an endeavor may be futile; the results using any of the national data sets that we have considered, especially the NAIC and AMA data, might be far from convincing. We conclude that it is in the strong interest of the nation for a government entity to collect comprehensive and accurate data on health insurance markets and to make it available for research and policy analysis. In order to be maximally useful, such data should represent enrollees in both fully and self-insured plans across the range of buyer groups (individual, small group, large group), and include geographic identifiers.

References

American Medical Association (2001), "Competition in Health Insurance: A Comprehensive Study of U.S. Markets." Updates 2002-2005, 2007-2009

Barnett, T. O. and H.B. Wellford (2008). "The DOJ's Single-Firm Conduct Report: Promoting Consumer Welfare Through Clearer Standards for Section 2 of the Sherman Act," *Global Competition Policy*, October.

Capps, C. (2009). Comment, Project No. P092900, Submission of America's Health Insurance Plans to the Department of Justice and the Federal Trade Commission on the Horizontal Merger Guidelines Review Project. Available at http://www.ftc.gov/os/comments/horizontalmergerguides/545095-00009.pdf

[11] Kaiser Family Foundation (2010), "Medicare Advantage 2010 Data Spotlight: Benefits and Cost-Sharing," February 2010. Available at http://www.kff.org/medicare/upload/8047.pdf

Congressional Budget Office (2010). "Estimate of Direct Spending and Revenue Effects for the Amendment in the Nature of a Substitute Released on March 18, 2010." http://www.cbo.gov/ftpdocs/113xx/doc11355/hr4872.pdf

Dafny, L., Duggan, M. and S. Ramanarayanan (2010). "Paying a Premium on Your Premium? Consolidation in the U.S. Health Insurance Industry," *NBER Working Paper 15434.*

Employee Benefit Research Institute (2010), "How Many of the U.S. Nonelderly Population Have health Insurance?", Fast Facts", November 15. http:// www.ebri.org/pdf/FFE183.15Nov10.Uninsured.Final.pdf

Kaiser Family Foundation (2010), "Medicare Advantage 2010 Data Spotlight: Benefits and Cost-Sharing," February 2010. http://www.kff.org/medicare/upload/8047.pdf