Opportunities and Challenges in Supply-Side Simulation: Physician-Based Models

Carole Roan Gresenz, David I. Auerbach, and Fabian Duarte

Objective. To provide a conceptual framework and to assess the availability of empirical data for supply-side microsimulation modeling in the context of health care.

Data Sources. Multiple secondary data sources, including the American Community Survey, Health Tracking Physician Survey, and SK&A physician database.

Study Design. We apply our conceptual framework to one entity in the health care market—physicians—and identify, assess, and compare data available for physician-based simulation models.

Principal Findings. Our conceptual framework describes three broad types of data required for supply-side microsimulation modeling. Our assessment of available data for modeling physician behavior suggests broad comparability across various sources on several dimensions and highlights the need for significant integration of data across multiple sources to provide a platform adequate for modeling. A growing literature provides potential estimates for use as behavioral parameters that could serve as the models’ engines. Sources of data for simulation modeling that account for the complex organizational and financial relationships among physicians and other supply-side entities are limited.

Conclusions. A key challenge for supply-side microsimulation modeling is optimally combining available data to harness their collective power. Several possibilities also exist for novel data collection. These have the potential to serve as catalysts for the next generation of supply-side-focused simulation models to inform health policy.

Key Words. Microsimulation, supply, physicians

Microsimulation modeling is one tool in the arsenal available to policy makers for understanding the effects of potential policy changes. Such models have been used for decades in a variety of contexts, such as understanding the effects of tax rate changes (e.g., the TAXSIM model) and income support programs (e.g., the TRIM2 model). A distinguishing feature of microsimulation models is that the objects of the simulation are individual units—as opposed
to aggregations of units—and their advantages include that they allow for examination of distributional effects of a policy change or other exogenous event and for understanding the effects of multiple, interacting policy changes (Chollet 1990; Giannarelli 1992; Mitton 2000).

Recent years have witnessed the development and increased use of a number of microsimulation models designed to understand the effect of health care reform efforts, including the RAND COMPARE (Comprehensive Assessment of Reform Efforts) model, CBO’s Health Insurance Simulation Model (HISim), and the Urban Institute’s Health Insurance Reform Simulation Model (HIRSM), among others. These modeling efforts have largely been focused on understanding consumer demand for health insurance coverage and health care under a range of policy scenarios, along with the associated costs of policies in terms of government and societal spending.

But simulation modeling efforts aimed at understanding the supply-side effects of health policies have been few and ad hoc—despite the fact that the Patient Protection and Affordable Care Act (ACA) includes a wide range of supply-oriented policy changes, and supply-side effects of major health reforms have the potential to dwarf demand-side effects under some circumstances (Stewart and Enterline 1961; Enterline, McDonald, and McDonald 1973; Finkelstein 2007). One factor underlying the relative paucity of supply side models is a lack of good data. As Citro (1991) emphasize:

Good data are a critical ingredient for models and other analysis tools to produce good estimates. Data that are of poor quality, scope and relevance will increase the uncertainty and decrease the validity of model outputs. Poor data also make it harder for models to respond to changing policy analysis needs in a timely and cost-effective manner.

Chollet (1990) notes further that:

Much of the effort involved in building and maintaining microsimulation models relates to enhancing and imputing the input data necessary to support even relatively simple analysis.
There are many supply-side entities in health care that may be useful to analyze in the context of a microsimulation model—including insurance companies, health plans, hospitals, urgent care centers, retail clinics, community health centers, physicians, nurse practitioners, psychologists, social workers, and others—to gauge the potential effects of the ACA. In this article, we focus for tractability on one key such entity in the health care market: physicians. The adequacy of the current supply of physicians to accommodate what is expected to be an increase in the number of insured individuals of nearly 30 million after full ACA implementation (Eibner and Price 2012) has been of particular concern (AcademyHealth 2012). We provide a framework for understanding the types of data required for supply-side microsimulation modeling, identify key data sources available to fulfill these needs in the context of a physician model, and assess and compare the strengths and limitations of data from various sources.

**CONCEPTUAL FRAMEWORK**

Microsimulation modeling requires a range of data that often must be carefully synthesized from multiple sources. In microsimulation models that focus on the demand side of health care, a typical approach is to draw on a core set of data representing a nationally representative sample of individuals and households and to supplement that data with information from other sources through imputation, statistical matching, or direct merging (Chollet 1990). For example, the RAND COMPARE microsimulation model uses the Survey of Income and Program Participation (SIPP) as the principal base dataset. Data from the Medical Expenditure Panel Survey (MEPS) regarding health care utilization and medical expenditures are statistically matched to SIPP respondents based on age, insurance status, health status region, and income, and individuals are assigned to synthetic firms with particular characteristics and health benefit offerings using data from the Kaiser/HRET survey and based on region, firm size, and industry (Eibner and Price 2012). Additional data regarding expected birth and death rates and rates of immigration may be used to project the population forward over time. Other data—for example, characteristics of geographic areas—may be merged with the core dataset to supplement the range of information available. Behavioral parameters, such as individuals’ decisions to enroll in Medicaid coverage when eligible, are sometimes derived from still other data and/or they may be drawn from estimates available in the existing literature.
Parallel data requirements apply to supply-side microsimulation modeling and can be grouped into three broad types. We describe each data type first in general and then as it applies to physician-based models.

Type 1
Unit-level data on the current supply of providers, including characteristics of the providers. These data are equivalent to the base data on the consumer side, such as the SIPP. Type 1 data include not just information sufficient for counting the number of providers of different types (e.g., primary care doctors, specialists, hospitals, community health centers) but also characteristics of the population of patients each provider serves as well as measures of capacity. In addition, Type 1 data include information about the organizational and financial relationships between each provider and other providers. These data are similar to demand-side information on, for example, the relationship between individuals and employers.

With respect to physicians in particular, Type 1 data include the number of physicians by location; characteristics of physicians, such as medical specialty, age, gender, race/ethnicity, and years since graduation from medical school; characteristics of their practice, including organizational features such as office or hospital-based practice and single- or multiphysician group; and measures of capacity such as hours worked, time spent in various activities, and number and payor mix of patients served. Type 1 data also include information about the insurance plans accepted by the physician, which insurance networks they participate in as a preferred provider, and hospitals with which they are associated.

Type 2
Data to project forward provider supply. These data correspond to death rates, birth rates, and immigration rates on the demand side that may be used to age the U.S. population forward to create a synthetic population for some future date in time.

Type 2 data for projecting physician supply require information on flows of individuals into and out of the profession, such as medical school enrollment, number of residency slots, retirement rates, and an understanding of how factors like the availability of federal or state loan repayment packages or tort reform influence individuals’ decisions to enter or exit the physician market.
Type 3

Data to understand provider behavior, such as the type of care provided and prices charged. Behavioral parameters are the models’ engines, typically interacting with policy changes to produce outcomes of interest. Type 3 includes not only primary data sources that can be used to estimate behavioral parameters but also secondary data sources that include published estimates of relevant parameters. These data parallel estimates of behavioral parameters on the demand side such as how the availability of public insurance affects firms’ offers of employer sponsored health insurance coverage. Type 3 also includes information about how various factors influence the structure of relationships across providers.

With respect to physicians, Type 3 data are required to inform questions about how various factors influence the decisions of physicians to accept certain kinds of patients, provide various types of services, or locate in various geographic areas as well as decisions about the price and quality of care provided.

**KEY DATA SOURCES FOR MODELING PHYSICIAN SUPPLY**

In this section, we identify and assess key sources available to populate the three types of data that together define the architecture for simulation modeling. Notably, some data sources may fall into multiple types. Our goal is not to compile an exhaustive list, but rather to characterize those sources likely to provide the greatest utility for microsimulation modeling.

*Data to Characterize Physician Supply (Type 1)*

The most comprehensive source of data on the number of physicians by location is the American Medical Association’s (AMA) Masterfile. All allopathic and most osteopathic medical students in the United States are entered into the Masterfile’s records and are contacted every several years throughout their lifetimes. The AMA also obtains information on International Medical Graduates who enter the United States for required residency positions and, thus, has records for nearly all MDs in the United States, not just those who are AMA members and not just those who are actively practicing medicine or who maintain their licensure. The data are further supplemented with data...
from state licensing boards and specialty associations. Core elements include information about the physician’s age, location, contact information, detailed specialty, some practice and employer characteristics, and major function (e.g., patient care, research, education, administration). A single variable captures work effort – ‘active’ physicians are defined as those working at least 20 hours per week.

A primary shortcoming of the AMA data for supply-side modeling is that variables such as hours worked, income, detailed practice setting, characteristics of patients seen, or time spent in various activities are not collected. Furthermore, the fact that the data are only updated every few years creates an inherent bias in the data, particularly in that physician work effort is substantially overestimated for older physicians, and likely underestimated for newly entering physicians (Staiger, Auerbach, and Buerhaus 2009). Researchers using the Master file as a basis for supply modeling typically make ad hoc adjustments to attempt to reduce those inherent biases (Colwill and Cultice 2003; Dill and Salsberg 2008).

Another source for information on the population of physicians is licensing data that state regulatory agencies collect. However, such data vary substantially across states in breadth and quality and there is little or no coordination across to standardize data elements or to aggregate data across states (although the Health Resources Services Administration—HRSA—is undertaking efforts toward such standardization). Individual states, such as North Carolina, provide a good example of efforts to capture better physician information, at least within state boundaries (see http://www.shepscenter.unc.edu/hp/presentations.htm).

One seldom, but increasingly used data source in physician workforce analysis is the Current Population Survey (CPS). The CPS contains more labor supply-relevant data than the AMA Masterfile, such as hours worked, family characteristics, and industry. Physicians are identified in survey if they report their primary occupation as a physician (or, if out of the labor force, they would remain in the sample if they identified physician as their predominant occupation over the past 5 years). However, the sample size is relatively small (on the order of 1,000 physicians per year are obtained) and information by physician specialty is not available. These data have been used previously as a benchmark against which to compare supply estimates derived from the AMA Masterfile.

In addition, since 2001, the American Community Survey (ACS) has collected information about physicians similar to that in the CPS, but with a larger sample. The ACS sample was enhanced in 2005 and, as a consequence,
the survey has been collecting information on roughly 10,000 physicians per year since that time. Geographic location is available at the MSA level. To date, few researchers have used the ACS for estimating physician supply, but the data hold significant promise going forward.

In addition, the Community Tracking Study, administered by the Center for Studying Health System Change, includes a physician survey which has been administered every 4 years since 1996 (although it may not be continued in the future). The physician survey component was named the Health Tracking Physician Survey (HTPS) in 2008. In earlier years, the HTPS sample was clustered within 60 metropolitan areas in the United States, but the 2008 sampling frame was changed to obtain a representative sample of physicians across the United States. Physicians are asked detailed questions about their practice size, type and organization, the financial incentives they face, the payer mix of the patients they see, and the percentage of their (or their organization’s) revenues derived from different payers. The HTPS also includes some information on physician employment by hospitals (Casalino et al. 2008). Although the sample size is not large enough to obtain representative supply estimates at geographic levels at the size of a state or smaller, the data are often used on their own to track national trends and in more complex supply analyses (Tu and Ginsburg 2006; Cunningham 2011).

Similar practice and organizational detail can be obtained from the National Ambulatory Medical Care Survey (NAMCS) but for office-based, patient care physicians only (also excluding anesthesiologists, radiologists, and pathologists). The NAMCS, which contains data going as far back as 1973, draws samples of such physicians from the AMA and then samples patient visits associated with those physicians. The NAMCS contain considerable detail on patient characteristics associated with the sampled physicians. Like the CTS, the NAMCS contains geographic identifiers at the county level that are available in a restricted version of the data. Researchers must submit a proposal to the National Center for Health Statistics (NCHS) to analyze the restricted-use version of NAMCS data on-site at a secure data facility (such as the NCHS Data Center or a Census Research Data Center) or through a secure remote access system.

Several other sources exist but have been less widely used in research for a variety of reasons. Private organizations such as IMS Health, SK&A, and the Medical Group Management Association (MGMA) assemble extensive databases on physicians primarily for marketing purposes. For example, data from SK&A, a Cegedim company, represent a census of practicing, primarily office-based physicians in the United States. Increasingly, private
organizations who have collected such data have made it available to researchers for a negotiated fee, although one factor limiting their broader use in research purposes has been the need for data validation and comparison to other well-known and widely used physician databases. These private sources of physician data, importantly, allow for characterizing physician relationships with other providers such as hospitals or health systems. The SK&A data, for instance, contain linkages for each physician in the database to hospitals or health systems with which they are affiliated; the AMGA and other organizations maintain information on large integrated physician organizations.

In addition, the Bureau of Labor Statistics Occupational Employment Statistics data obtain useful information such as earnings and geographic location by physician specialty. However, the data do not cover self-employed individuals, which in the case of physicians, leaves out a substantial portion of all physicians. Finally, Medicare claims data also contain useful information on physicians, who can be identified by their provider numbers and associated with characteristics of Medicare service utilization. However, physicians who do not bill Medicare are not included.

Data to Characterize Physician Entry and Exit Behavior (Type 2)

There are several data sources of data that can be used to understand entry into and exit from the physician workforce (Type 2). Data on medical school applicants, enrollments, and graduates are collected by the American Association of Medical Colleges, and summary data are generally made available on their website. Counts of physician residents by year and residency type can be obtained from data published annually by the National Residency Matching Program or in the September issue of the *Journal of the American Medical Association*. Physician retirement rates are generally inferred from analysis of physician labor supply activity by age, although, as noted above, the Masterfile underestimates physician retirement.

Data to Characterize Physician Behavior (Type 3)

As described earlier, Type 3 includes data to understand physician behavior as well as factors influencing physicians’ organization and financial relationships with other supply-side entities. For example, how do physicians respond to payment rate changes or insurance coverage increases? What factors affect the decisions of physicians to participate in certain insurance networks or to become a subsidiary of a larger hospital or health system?
There is a growing literature available to help characterize physician behavioral responses to changes such as increasing medical malpractice premiums, tort reform, modifications to public insurance program reimbursement rates or policies, and expansions of eligibility for public insurance programs. With many following the conceptual framework laid out in McGuire and Pauly (1991), numerous studies have explored labor supply, treatment decisions, and other responses by physicians to policy changes. For example, Enterline, McDonald, and McDonald (1973) find that universal coverage in Quebec led to a reduction in physicians’ hours worked, whereas a study by Garthwaite (2011) shows that implementation of the Children’s Health Insurance Program (CHIP) led to expanded physician participation in Medicaid but also to a decreased number of hours spent on patient care. White (2012) finds no effect of CHIP implementation on children’s use of physician services but improved access among children in response to increases in physician fees. Others studies have focused on physician responses to medical malpractice policies and premiums: Kessler, Sage, and Becker (2005) find that medical malpractice reform increased physician supply, whereas Polsky, Marcus, and Werner (2010) find that high malpractice premiums are associated with a reduced supply of obstetricians. Finally, Hadley et al. (2009) and Clemens and Gottlieb (2012) find that physicians increase the quantity of care provided to Medicare patients in response to Medicare payment rate increases.

**Empirical Assessment of Key Physician Data Sources**

While multiple sources are potential candidates for contributing to a physician input file for microsimulation, little is known about their relative consistency. To that end, we developed comparisons of four key data sources—the ACS, AMA Masterfile (as compiled by HRSA in the Area Resource File or ARF), HTPS, and SK&A physician database—along several dimensions. Specifically, we compared the estimated total number of physicians from each source; the number of physicians by specialty; the number of physicians by region; and the estimated number of physicians who accepted Medicaid. All data were from 2010 with the exception of the HTPS, for which the most recent available data are from 2008.

We first compared the number of active, office-based, nonresident physicians using the ARF—which compiles data from the AMA Masterfile; the SK&A database, which also includes practicing, primarily office-based physicians in the United States; and the ACS. We excluded physicians in the ACS
who were not working or who were under the age of 35 and employed in the hospital (likely residents); however, the ACS may include hospital-based and other physicians working as administrators, researchers or in other capacities outside of patient care. We were not able to otherwise further restrict the data because of limited practice information available from the ACS.

The ARF data put the total number of practicing, office-based physicians at approximately 553,000, whereas the SK&A figures suggest a total of 552,000, although the close-to-exact similitude of the figures is somewhat coincidental (Table 1), given that the ARF may retain some older physicians who have retired and would not be included in the SK&A counts (Staiger, Auerbach, and Buerhaus 2009) and the exclusion of hospital-based and other physicians from both samples is surely not identical. The ACS total is, not surprisingly, higher, at 673,000 because it includes some non-office-based physicians.

We also compared the percentage of the total supply of physicians estimated to be in different specialties using the ARF and SK&A data. The data sources suggest remarkably similar fractions of physicians specializing in family medicine and general practice (14 percent in both); internal medicine (14 percent in the ARF and 12 percent in the SK&A); medical specialties (47 and 43 percent); obstetrics/gynecology (6 and 5 percent); pediatrics (8 and 7 percent); psychiatry (6 and 4 percent); and surgical specialties (11 percent in both).

We compared physician supply by region using the same three data sources and also included information from the HTPS (although the HTPS

Table 1: Physician Supply in Total and by Specialty (2010): Comparison across Three Data Sources

<table>
<thead>
<tr>
<th>Total Number of Physicians</th>
<th>ARF 553,000 (%)</th>
<th>SK&amp;A 552,000 (%)</th>
<th>ACS 673,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family/general practice</td>
<td>14</td>
<td>14</td>
<td>–</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>14</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td>Medical specialties</td>
<td>43</td>
<td>47</td>
<td>–</td>
</tr>
<tr>
<td>Obstetrics/gynecology</td>
<td>6</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>8</td>
<td>7</td>
<td>–</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>6</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>Surgical specialties</td>
<td>11</td>
<td>11</td>
<td>–</td>
</tr>
</tbody>
</table>

Note. ARF is Area Resource File. SK&A is a private vendor of physician data. ACS is American Community Survey. ARF and SK&A counts represent practicing, office-based physicians only. The ACS count includes other types of physicians. Physician supply estimates by specialty are not available for the ACS.
sample is drawn from the AMA Masterfile data and weights are derived using these data). We find broad consistency across data sources in the estimated distribution of physicians across nine areas within the United States (Figure 1).

Finally, we used two data sources—SK&A and HTPS—to compare a key characteristic of physician behavior, namely, the percentage of primary care physicians who accept Medicaid. Our comparison is based on a dichotomous (0/1) variable indicating whether the physician accepts any Medicaid patients (which is different from, and higher than, the rate of acceptance of new Medicaid patients, which was recently reported by Cunningham [2011] based on the HTPS as 42 percent). Both data sources suggest that roughly two-thirds of primary care physicians accept patients with Medicaid coverage in the United States (67 percent in the HTPS vs. 64 percent in SK&A). We also explored the consistency across data sources in variation in the Medicaid acceptance rate across regions within the United States (specifically, U.S. Census Divisions). The variation in rates is highly correlated across data sources (correlation coefficient of .79), and Figure 2 shows the patterns of acceptance by region.

Figure 1: Estimated Distribution of Physicians by Region (2010): Comparison across Three Data Sources

Notes: ARF is Area Resource File. SK&A is a private vendor of physician data. ACS is American Community Survey. ARF and SK&A counts represent practicing, office-based physicians only. The ACS count includes other types of physicians.
Building a Microsimulation Input File

The physician data sources described above are often rich in certain dimensions but lacking in others. For example, AMA data contain a census of physicians and information on their specialty and demographic characteristics, but little about capacity, labor supply or organizational characteristics of the practice (e.g., hours worked, income, patient mix by insurance type). SK&A data also include a census of physicians but also no data on patient mix by insurance type, practice environment, or number of hours worked. HTPS contain rich information on patient mix and practice type but for a relatively small sample of surveyed physicians. The ACS obtains a representative, annual sample of roughly 10,000 physicians each year with rich labor supply information, but it lacks information unique to physicians such as specialty.

As a consequence and similar to demand-side models, building an input file for a microsimulation model of physician behavior is likely to require synthesis of data across several sources. Our empirical comparisons of several of the key data sources suggest a broad-level consistency among them. Statistically matching or imputing key variables, such as revenue by insurance type

Figure 2: Percentage of Primary Care Physicians Accepting Any Medicaid Patients: By Census Division

Notes: HTPS is Health Tracking Physician Survey. SK&A is a private vendor of physician data.
DISCUSSION AND CONCLUSION

Our conceptual framework describes three broad types of data required for supply-side simulation modeling, including base data to broadly characterize the population of supply-side entities; data to project forward provider supply to allow for longer term projections about the effects of various healthy policies on supply-side entities; and data to provide estimates of behavioral parameters among supply side entities.

Our assessment of available data for microsimulation modeling of physician behavior suggests first that a key challenge is optimally combining available data to harness their collective power. Our empirical analyses of several physician data sources provide evidence of broad-based consistency in at least the key variables analyzed in the selected datasets.

Several rich promising data sources exist for better understanding physician practice patterns, such as data collected and aggregated by private companies from pharmaceutical suppliers that offer a window into physician prescribing behavior and aggregated claims data synthesized by companies who act as intermediaries between providers and insurers in the claims process that can be used to characterize services delivered by providers to the full panel of patients they see, including those enrolled in private and public health insurance plans. However, access to such data can be limited by fees associated with their use. The federal government might play a role in negotiating regular, standardized abstracts of these types of privately collected data for a set price and ensuring their availability for health services research. State licensing boards regularly update and gather information on all providers through the licensing process—federal efforts to expand and harmonize the data collected through these processes, already under way via HRSA, could also prove a valuable source of data for modeling, particularly if it were made available to the public for minimal fees (Trude 2011).

In addition, with changes in provider relationships, such as through the formation of accountable care organizations, poised to take center stage among changes resulting from the ACA, data for tracking and monitoring these changes in a systematic way are essential but, at present, limited. Available physician databases contain some information on physician
relationships with hospitals, as described, and the American Hospital Association (AHA) collects some data on physician–hospital organization through its Annual Survey (such as about whether hospitals operate an independent physicians association, open physician–hospital organization, closed physician–hospital organization, management service organization, or fully integrated organization) (Cuellar and Gertler 2005). Researchers have also turned to qualitative data and/or state- or market-specific data. For example, Robinson (1998) uses a case study approach to study the growth structure and strategy in California and New Jersey of physician organizations that coordinate medical groups in multiple markets and contract with HMOs. More recently, Berenson, Ginsburg, and Kemper (2010) use qualitative data gathered through site visits to six California markets and interviews with representatives from hospitals, large employers, benefit consultants, insurance brokers, and other stakeholders to assess how provider bargaining power affects health insurers’ payment rates. Business filings, some of which are captured in proprietary legal databases (e.g., Lexis-Nexis), represent one potential but unexploited source of additional data. Pilot studies that focus on using such data to describe the dynamics of physician–hospital–insurer relationships in a particular market or for a particular hospital system or insurer are a first step in determining whether such data might also be useful for capturing these dynamics on a larger, national scale using novel and developing data mining techniques.

A rich literature around individual decisions regarding health insurance enrollment and purchasing behavior and health care utilization has been key to the development of demand-side microsimulation models. Continued development of behavioral studies that investigate the relationship between structural factors such as payment rates or practice environment changes and physician supply and treatment behavior is likewise crucial to the success of supply-side microsimulation models focused on physicians.

Data collection and analysis along these lines have the potential to serve as catalysts for the next generation of simulation models to inform health policy; namely, those that are focused on understanding the effects of health policy changes on the supply side of the health care market.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Support for this research was provided by RAND’s Investment in People and Ideas program and the Bing
Center for Health Economics at RAND. We extend our thanks to George Hart for his expert assistance with data analyses.

Disclosures: None.

Disclaimers: None.

NOTES

1. The AMA Masterfile incorporates most osteopathic physicians, by incorporating them into the Masterfile if they enter ACGME-accredited residency programs or when they enter the workforce in some cases. However, some osteopathic physicians never enter the AMA Masterfile, particularly those who enter residencies accredited by the AOA.

2. Supplemental surveys that captured some of this information, along with earnings and other data were terminated in 2001.

REFERENCES


Clemens, J., and J. Gottlieb. 2012 “Do Physicians’ Financial Incentives Affect Medical Treatment and Patient Health?” SIEPR Discussion Paper No. 11-017


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.