

6-1-2010

## Comparison of GE Centricity Electronic Medical Record database and National Ambulatory Medical Care Survey findings on the prevalence of major conditions in the United States.

Albert G. Crawford

*Department of Health Policy, Jefferson Medical College*

Christine Cote

*Johnson and Johnson, Corporate Office of Science and Technology*


Joseph Couto

*Thomas Jefferson University School of Population Health*

Mehmet Daskiran

*Johnson and Johnson, Corporate Office of Science and Technology*

Follow this and additional works at: <https://jdc.jefferson.edu/healthpolicyfaculty>

 Place of Gunnarsson, Community Health and Preventive Medicine Commons, and the Health Services Research Commons

### Let us know how access to this document benefits you

*See next page for additional authors*

---

#### Recommended Citation

Crawford, Albert G.; Cote, Christine; Couto, Joseph; Daskiran, Mehmet; Gunnarsson, Candace; Haas, Kara; Haas, Sara; Nigam, Somesh C; Schuette, Rob; and Yaskin, Joseph, "Comparison of GE Centricity Electronic Medical Record database and National Ambulatory Medical Care Survey findings on the prevalence of major conditions in the United States." (2010). *College of Population Health Faculty Papers*. Paper 55. <https://jdc.jefferson.edu/healthpolicyfaculty/55>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in College of Population Health Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

---

## Authors

Albert G. Crawford, Christine Cote, Joseph Couto, Mehmet Daskiran, Candace Gunnarsson, Kara Haas, Sara Haas, Somesh C Nigam, Rob Schuette, and Joseph Yaskin

# Comparison of GE Centricity Electronic Medical Record Database and National Ambulatory Medical Care Survey Findings on the Prevalence of Major Conditions in the United States

Albert G. Crawford, PhD,<sup>1</sup> Christine Cote, MD,<sup>2</sup> Joseph Couto, PharmD,<sup>1</sup> Mehmet Daskiran, PhD,<sup>2</sup>  
Candace Gunnarsson, PhD,<sup>3</sup> Kara Haas, MD,<sup>4</sup> Sara Haas, MS,<sup>3</sup> Somesh C. Nigam, PhD,<sup>5</sup>  
Rob Schuette, PhD,<sup>3</sup> and Joseph Yaskin, MSS<sup>1</sup>

## Abstract

The study objective was to facilitate investigations by assessing the external validity and generalizability of the Centricity Electronic Medical Record (EMR) database and analytical results to the US population using the National Ambulatory Medical Care Survey (NAMCS) data and results as an appropriate validation resource. Demographic and diagnostic data from the NAMCS were compared to similar data from the Centricity EMR database, and the impact of the different methods of data collection was analyzed.

Compared to NAMCS survey data on visits, Centricity EMR data shows higher proportions of visits by younger patients and by females. Other comparisons suggest more acute visits in Centricity and more chronic visits in NAMCS. The key finding from the Centricity EMR is more visits for the 13 chronic conditions highlighted in the NAMCS survey, with virtually all comparisons showing higher proportions in Centricity.

Although data and results from Centricity and NAMCS are not perfectly comparable, once techniques are employed to deal with limitations, Centricity data appear more sensitive in capturing diagnoses, especially chronic diagnoses. Likely explanations include differences in data collection using the EMR versus the survey, particularly more comprehensive medical documentation requirements for the Centricity EMR and its inclusion of laboratory results and medication data collected over time, compared to the survey, which focused on the primary reason for that visit. It is likely that Centricity data reflect medical problems more accurately and provide a more accurate estimate of the distribution of diagnoses in ambulatory visits in the United States. Further research should address potential methodological approaches to maximize the validity and utility of EMR databases. (*Population Health Management* 2010;13:139–150)

**A**GROWING PROPORTION of US ambulatory care physician practices are adopting electronic medical record (EMR) systems to support various clinical processes including documentation of patient encounters and secure exchange of data with other providers. In 2007, one leading commercial EMR system, GE Centricity,<sup>a</sup> was used by more than 20,000 clinicians to manage the medical records of about 30 million patients in 49 states. These electronic records contain a wide range of clinical and demographic variables. A structured

user interface supports accurate and consistent documentation by physicians and supports internal data validity, both individually and in the aggregate. A significant attribute of Centricity, like other EMRs, is its capacity to track patients' clinical conditions over time. Taken together, these attributes of Centricity and other EMRs make them rich resources for epidemiologic, population health, and outcomes research. Specifically, the Centricity EMR data structure supports multivariate analytic designs including retrospective cohort

<sup>1</sup>Thomas Jefferson University School of Population Health, Philadelphia, Pennsylvania.

<sup>2</sup>Johnson & Johnson, Corporate Office of Science and Technology, New Brunswick, New Jersey.

<sup>3</sup>S<sup>2</sup> Statistical Solutions, Inc., Cincinnati, Ohio.

<sup>4</sup>Ethicon Endo-Surgery, Inc., Scientific Affairs/Medical Affairs, New Brunswick

<sup>5</sup>Johnson & Johnson, Office of Evidence Based Medicine, Medical Device and Diagnostics, New Brunswick, New Jersey.

<sup>a</sup>Centricity Physician Office is a registered trademark of GE Medical Systems Information Technologies.

studies. Researchers' interest in Centricity is already evident in the growing medical research literature that is reviewed in this paper.

The purpose of this investigation is to facilitate future investigations by assessing the external validity and generalizability of Centricity data and analytic results to the US population as a whole. Our literature review identified the National Ambulatory Medical Care Survey (NAMCS) as an appropriate validation resource for this study. Supported by the Federal Centers for Disease Control and Prevention (CDC), NAMCS collects data regarding patient visits from a sample of physicians. Like Centricity, the NAMCS provides a de-identified data set that includes patients' demographic characteristics, physicians' diagnoses, and medications ordered and/or provided. NAMCS also publishes reports with descriptive statistics on patients' demographic characteristics and conditions including diagnoses and treatments. However, the 2 data sets employ different methods of data collection: NAMCS collects data from physicians using a survey tool at the time of the patient visit, posing a modest threat to internal validity, while the Centricity data set is based on an EMR, in which high data quality stem from professional responsibility, the neglect of which could compromise quality care and endanger patient safety. NAMCS data include sample weights that allow making statistical inferences from its national probability sample, providing a basis for interpreting possible deviations between nationally representative NAMCS results and corresponding results derived from the Centricity EMR database. We evaluated a number of demographic, diagnostic, and disease state variables for external validity and generalizability.

## Background

A review of the peer-reviewed literature provides a basis for the present effort to assess the external validity and generalizability of the GE Centricity EMR data set and its capacity to support population-level research on chronic conditions in the United States. Although the methods employed in this literature review were not formal, we believe that the results are valid, especially given our ability to query Centricity staff about all publications to date based on Centricity EMR data.

Our review includes several examples of recently published research that employed secondary analyses of federal data sets and/or data derived from widely used EMR systems.

In a study by Burt and associates,<sup>1</sup> data from NAMCS and its counterpart in hospitals, the National Hospital Ambulatory Medical Care Survey (NHAMCS), were used to investigate physician treatment and prescribing patterns. Specifically, using a multiplicity estimator, NAMCS and NHAMCS data on the number of past visits the patient made to the sample provider during the 1-year period prior to the sampled visit were used to estimate the number of patients. The resulting distribution of patients by annual number of visits is similar to the estimated distribution of persons in the United States making ambulatory care visits based on a population-based survey.

Another recent report, by Hing and associates,<sup>2</sup> describes ambulatory visits to physician offices in the United States

and presents statistics on selected characteristics of physician practices, patients, office visits, and trends in visits. The data were collected in the 2004 NAMCS and were weighted to produce annual national estimates by employing an estimator using revised non-response adjustment. The authors found that in 2004 an estimated 910.9 million visits were made to physician offices in the United States, an overall rate of 315.9 visits per 100 persons, with 58.9% of visits to physicians in the specialties of general and family practice, internal medicine, pediatrics, and obstetrics and gynecology. Essential hypertension, malignant neoplasms, acute upper respiratory infection, and diabetes mellitus were the leading illness-related primary diagnoses.

An earlier paper by Machlin and colleagues<sup>3</sup> compared 1996 survey estimates using NAMCS, NHAMCS, the Medical Expenditure Panel Survey, and the National Health Interview Survey and described methodological issues to consider when using these data sets to measure ambulatory utilization.

There has been considerable research in the last few years based on analyses of Centricity EMR data. In fact, focusing on just US national research using Medical Quality Improvement Consortium (MQIC) data in particular, there were 12 journal publications during 2006–2007, and 31 poster/podium presentations during 2004–2008.<sup>4</sup> For example, Brixner and colleagues<sup>5</sup> used clinical (biometric), diagnosis (ICD-9-CM codes), and treatment (prescription) information in the Centricity EMR database to examine the prevalence of cardiometabolic risk (CMR) factors that contribute to metabolic syndrome in the primary care setting. In the study population of 475,651 patients with information on indicators of CMR, 15.3% and 11.8% were found to have metabolic syndrome according to National Cholesterol Education Program and International Diabetes Federation criteria, respectively; 34.2% had a body mass index (BMI)  $\geq 27$  kg/m<sup>2</sup> as a risk factor, 56.0% had high blood pressure, 10.7% had high triglycerides, 16.0% had low high-density lipoprotein cholesterol (HDL-C), 8.8% had impaired fasting glucose, and 7.2% had diabetes. Brixner and colleagues conclude that "the distribution of CMR factors in a primary care database is similar to that established by prospective national health surveys such as NAMCS. A key source of identification of risk factors are clinical outcomes including BMI and lab values. Future studies on metabolic syndrome need to link clinically based information with more readily available treatment and diagnosis information."

In a 2008 article, Brixner and colleagues<sup>6</sup> evaluated the relationships between CMR factors and BMI as recorded in the Centricity EMR. Patients with a BMI  $\geq 18$  kg/m<sup>2</sup> in the EMR at any time during the 10-year period from 1996–2005 were stratified into groups by number of CMR factors, and individual risk factor for those with only 1. The authors identified a total of 499,593 patients with a BMI  $\geq 18$  kg/m<sup>2</sup>; 56.4% had a BMI  $> 27$  kg/m<sup>2</sup>, while 43.6% had a BMI of 18–27 kg/m<sup>2</sup>. Compared with patients with no risk factors, patients with 1–4 risk factors were significantly more likely to have a BMI  $> 27$  kg/m<sup>2</sup>; 48.4% without CMR factors had a BMI  $> 27$  kg/m<sup>2</sup>, compared with 63.3%, 79.8%, 84.6%, and 88.5% for patients with 1–4 cardiometabolic risk factors, respectively. Adjusted odds ratios (ORs) for having a BMI  $> 27$  kg/m<sup>2</sup> were 2.64 for type 2 diabetes, 2.21 for elevated triglycerides, 1.91 for hypertension, and 1.45 for low

HDL-C. Adjusted ORs for having a BMI > 27 kg/m<sup>2</sup> were 3.58 for patients with any 2 risk factors, 4.24 for patients with any 3 risk factors, and 5.07 for patients with any 4 risk factors, relative to patients with no CMR factors. Brixner and colleagues conclude that "For patients with cardiometabolic risk factors, compared with patients with no risk factors, the odds of having a BMI > 27 kg/m<sup>2</sup> were multiplied by 1.45–5.07, depending on the type and number of risk factors. Diagnoses and treatment indicators for cardiometabolic risk factors are potential indicators of obesity."

In another recent use of secondary data from the Centricity EMR database, Wang and colleagues<sup>7</sup> compared the risk of incident hypertension associated with the use of celecoxib and nonselective non-steroidal anti-inflammatory drugs (NSAIDs).

One of the most relevant and useful analyses of Centricity EMR data, especially for the present analyses, is Gill and Chen's 2008 evaluation of lipid management,<sup>8</sup> which includes adequate lipid testing, achievement of lipid goals, and appropriate use of lipid-lowering medication. Lipid testing was adequate for 62% of high-risk, 67% of moderate-risk, and 36% of low-risk patients. Lipid goals were achieved in 65% of high-risk, 66% of moderate-risk, and 90% of low-risk patients; 35% of high-risk, 45% of moderate-risk, and 32% of low-risk patients achieved adequate testing and optimal goals; and medications were appropriately prescribed for 70% of high-risk, 47% of moderate-risk, and 48% of low-risk patients. Gill and Chen note that "National EHR networks are excellent vehicles for large outpatient quality of care studies, particularly for measuring clinical outcomes such as lipid levels."

In summary, these studies show that national EMR databases such as Centricity are valuable tools for health services research including epidemiologic and outcomes research and studies of provider behavior. EMR data sets have several attributes that provide significant opportunities for such research. First, identification of a patient's primary reason for a visit not only provides deeper information than that available in *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) diagnosis codes in claims databases, but it also allows cross-tabulation with disease state and prescribing information. Second, a data set derived from algorithm-based extraction of data from EMRs is inherently more comprehensive and accurate than a data set derived from patient- or visit-level survey records, the validity of which depends on the cooperation and attention of responding physicians and office staff. The factorial analysis of metabolic syndrome by Brixner and colleagues<sup>5</sup> is an early indicator of the potential for multivariate analyses of EMR data, specifically investigations of comorbidities and disease staging. Currently, results of such studies may be of limited generalizability to the degree that patient demographic and clinical characteristics represented in EMR data sets have not been validated as representative of the US population. Comparing patient attributes represented in the NAMCS data to those in the MQIC data set derived from the GE Centricity EMR, the present study provides such results to address the generalizability of this EMR-based data set. Specifically, the analysis explores various factors (ie, differences in data collection methods) that pose potential analytical limitations. The results are intended to provide insight into the strengths and weaknesses of the GE Centricity EMR

as a tool for population health research, informing future studies and stimulating promising lines of research.

## Methods

### *National Ambulatory Medical Care Survey*

As described on the CDC Web site,<sup>9</sup> "The National Ambulatory Medical Care Survey (NAMCS) is a national survey designed to meet the need for objective, reliable information about the provision and use of ambulatory medical care services in the United States. Findings are based on a sample of visits to non-federally employed office-based physicians who are primarily engaged in direct patient care. Physicians in the specialties of anesthesiology, pathology, and radiology are excluded from the survey. The survey [has been] conducted ... annually since 1989."

"Specially trained interviewers visit the physicians prior to their participation in the survey in order to provide them with survey materials and instruct them on how to complete the forms. Data collection from the physician, rather than from the patient, provides an analytic base that expands information on ambulatory care collected through other NCHS surveys. Each physician is randomly assigned to a 1-week reporting period. During this period, data for a systematic random sample of visits are recorded by the physician or office staff on an encounter form provided for that purpose. Data are obtained on patients' symptoms, physicians' diagnoses, and medications ordered or provided. The survey also provides statistics on the demographic characteristics of patients and services provided, including information on diagnostic procedures, patient management, and planned future treatment."

An enhancement implemented in the 2005 NAMCS allowed chronic condition-specific comparisons: "The emphasis for the 2005 survey year was chronic conditions. Additions to the routine encounter data that related to chronic conditions included: a chronic disease checklist, including conditions affecting the respiratory, cardiovascular, renal, and endocrine systems; arthritis; cancer; depression; obesity; and osteoporosis."

### *GE Centricity database*

The GE Centricity EMR database captures patient-level clinical data elements obtained from the Centricity Physician Office EMR (formerly Logician) for Clinical Data Services (CDS) reporting. The Centricity ambulatory care EMR and its predecessors have been used for over 20 years, are certified by the Certification Commission for Healthcare Information Technology (CCHIT), and are currently used by over 30,000 clinicians in the United States. Centricity CDS includes data provided by 7259 clinicians (including approximately 60% primary care providers and 40% specialists) at 98 installations with 133 unique provider members. CDS includes de-identified, standardized data on more than 8,900,000 patients; and the data on at least half of these patients spans more than 985 days, for a median of approximately 2.7 years of continuous care.

While the present study uses all data in CDS, some research cited in our literature review analyzes data generated by MQIC, a national network of outpatient practices. All practices that use the Centricity EMR are invited to join MQIC, but membership in MQIC is completely voluntary. In

2008 MQIC included over 4,000,000 patients cared for by over 5000 physicians and other providers from over 90 institutions in 35 states throughout the United States. These practices ranged from solo practices to large multipractice institutions with over 1000 providers; roughly 63% of MQIC providers are primary care physicians (including family medicine, general internal medicine, general pediatrics, and geriatrics).

#### *Exclusion criteria applied to the Centricity EMR data*

Exclusion criteria and algorithms were developed to eliminate potential distortions in aggregated Centricity EMR data resulting from user interface design, database structure, the de-identification process, or methods of recording procedures. Problem/complaint text extracted from records was filtered to ensure that patients merely evaluated for a condition (without confirmation of diagnosis) were not counted as "positives" along with patients explicitly diagnosed with that condition (included in this process of elimination are diagnoses containing "family history of," "rule out," "risk of," "screening of," "symptoms of," and "question of"). Exclusion criteria were also employed to reduce the likelihood of overestimation resulting from "backfilling" of retrospective visit data for physician practices when they first adopt the Centricity EMR.

#### *Office visits and study population*

This analysis uses 2005 GE Centricity EMR data on "activities" and patient demographics. Several sets of inclusion/exclusion criteria were applied prior to analysis (Figure 1). First, all 2005 activity data were extracted from the database

(Step 1). The index date range was from January 1, 2005 to December 31, 2005. Then, patients who had an activity record in 2005 and at least 1 documented activity in both 2004 and 2006 were selected to assure that they were active and still in the same health care system throughout 2005 (Step 2). According to the NAMCS 2005 summary report,<sup>10</sup> NAMCS excludes office visits to physicians in the specialties of anesthesiology, pathology, and radiology. In the GE Centricity EMR database, 56% of the patients' specialty information was listed as "unknown." Among the 3.9 million patients with specialty information, specialty data for only 872 patients indicated the 3 specialties of anesthesiology, pathology, or radiology. Specialty information provided by the responsible service provider was patient based, and activity data did not include specialty information. Therefore, this selection criterion was not applied to the data. In NAMCS, certain types of contacts were also excluded, including those made by telephone, those made outside the physician's office (eg, house calls), visits made in hospital settings (unless the physician has a private office in a hospital and that office meets the NAMCS definition of an "office"), visits made in institutional settings by patients for whom the institution has primary responsibility over time (eg, nursing homes), and visits to doctors' offices that are made for administrative purposes only (eg, to leave a specimen, pay a bill, pick up insurance forms). Therefore, among the various activity types, only activities meeting these "office visit" criteria were included in the study (Step 3). In addition to the criteria mentioned earlier, a small number of patients had missing sex and/or age information and were excluded from the final analysis data set (Step 4).

In the 2005 NAMCS survey data, only a single "primary" ICD-9-CM diagnosis code was recorded by physicians (see Tables 2 and 3), and additional chronic conditions were recorded through checkbox responses regardless of visit diagnosis. In the GE Centricity EMR database, however, activity data were available from multiple sources: problem, complaint, medication, prescription, observation, and order data. To maximize the validity of the comparison of data in the 2 data sets, only office visits associated with the problem table (only visits which resulted in 1 or more ICD-9-CM diagnosis codes) were included in the analyses reported in Table 2 (Step 5). In situations where more than 1 ICD-9-CM diagnosis code is listed during a single office visit, a fraction of each ICD-9-CM code was calculated by dividing 1 into the total number of ICD-9-CM diagnosis codes, and this fraction was used to calculate the annual number of visits with specified ICD-9-CM codes. A small number of ICD-9-CM "E" (external injury) codes were present in the database; these were included in the "other" category in Table 2.

For the chronic disease comparisons shown in Table 3, we first identified the corresponding ICD-9-CM diagnosis codes and searched the problem data set for these codes through the end of 2005. The diagnosis codes for these chronic diseases are listed in the Notes for Table 3. All chronic diseases were considered separately, and the fractional approach described above was not applied in this part of the analysis.

#### *Analytic methods*

Standard errors of percentages published for the NAMCS data were used to generate 99% confidence intervals, to

GE Centricity Database (Jan. 1 <sup>st</sup> 1996 – Mar. 31 <sup>st</sup> 2008) Patient population: 8,901,057 Activity count: 412,175,606
<b>Step 1</b>
GE Centricity Database (Jan. 1 <sup>st</sup> – Dec. 31 <sup>st</sup> 2005) Patient population: 3,683,868 Activity count: 57,634,290
<b>Step 2</b>
GE Centricity Database (Jan. 1 <sup>st</sup> – Dec. 31 <sup>st</sup> 2005) Inclusion criteria: At least 1 documented activity in both 2004 and 2006 Patient population: 1,816,664 Activity count: 39,374,063
<b>Step 3</b>
GE Centricity Database (Jan. 1 <sup>st</sup> – Dec. 31 <sup>st</sup> 2005) Inclusion criteria: Activity type needs to be an "office visit" Patient population: 1,462,126 Activity count: 6,202,445
<b>Step 4</b>
GE Centricity Database (Jan. 1 <sup>st</sup> – Dec. 31 <sup>st</sup> 2005) Inclusion criteria: non-missing sex and age Patient population: 1,462,002 Activity count: 6,201,616 (Final 2005 office visit data for Table 1)
<b>Step 5</b>
GE Centricity Database (Jan. 1 <sup>st</sup> – Dec. 31 <sup>st</sup> 2005) Inclusion criteria: Visits with an ICD-9-CM diagnosis Patient population: 950,431 Activity count: 2,210,587 (Final 2005 office visit data for Tables 2-3)

**FIG. 1.** Steps in the selection of office visits.

which point estimates generated for the Centricity EMR data were compared to yield conclusions about statistically significant differences at the  $P < .01$  level between the 2 sets of results. Given the very large numbers involved, almost all of the differences examined were statistically significant. The actual percentage differences were then reexamined to assess the substantive significance of each finding, with a criterion of a difference of at least 2% to indicate substantive significance.

## Results

Table 1 compares NAMCS and GE Centricity EMR results on the number and percent distribution of office visits, with corresponding standard errors, by patient age and sex for the United States for 2005. The GE Centricity EMR data include higher proportions of patient visits for ages up to age 64 and lower proportions for ages 65 and older, as follows: 18.0% vs. 16.7% for younger than age 15 years, 8.6% vs. 7.3% for ages 15–24 years, 22.0% vs. 20.9% for ages 25–44 years, 29.9% vs. 29.4% for ages 45–64 years, 10.8% vs. 12.4% for ages 65–74 years, and 10.7% vs. 13.3% for ages 75 years and older. While all 6 of these differences are statistically significant ( $P < .01$ ), they appear relatively modest. On the other hand, the GE Centricity EMR data include a significantly and substantially higher percentage of female patient visits (63.3% vs. 58.2%,  $P < .01$ ).

Table 2 compares NAMCS and GE Centricity EMR results on the number and percent distribution of office visits by physician's primary diagnosis. The comparison is problematic in that the GE Centricity EMR data include a significantly higher proportion of visits with diagnoses in the symptoms, signs, and ill-defined conditions category (16.0% vs. 6.3%,  $P < .01$ ) and significantly lower proportions in the supplementary classification, all other diagnoses, and unknown categories (11.9% vs. 18.6%, 1.5% vs. 2.7%, and 0.0% vs. 0.9%, all  $P < .01$ ), with a combined difference across these 3 categories of 13.3% vs. 22.2%. Given a close correspondence between the NAMCS and GE Centricity EMR in the overall proportions in all 4 residual categories combined (28.5% vs. 29.3%), we will compare differences in the more substantive diagnostic categories, but we will do so with the required caution. All 11 such comparisons are statistically significant ( $P < .01$ ); however, only 3 seem noteworthy - the lower proportions of neoplasm and circulatory condition diagnoses (1.2% vs. 4.1% and 3.6% vs. 8.5%, respectively), and the higher proportion of respiratory condition diagnoses (17.9% vs. 11.5%).

Table 3 compares NAMCS and GE Centricity results for the number and percent distribution of office visits by selected chronic conditions for the United States for 2005. The Centricity database is substantially more likely than the NAMCS data to show at least 1 condition (63.4% vs. 52.7%) and substantially less likely to show none (36.6% vs. 43.8%) or blank (0.0% vs. 3.5%). Of 13 possible comparisons regarding selected chronic diagnoses, all showed statistically significant differences between the 2 data sets, and all but 1 showed a higher proportion in the Centricity data: hypertension (24.4% vs. 22.8%), hyperlipidemia (25.7% vs. 13.5%), diabetes (10.2% vs. 9.8%), depression (16.0% vs. 8.8%), obesity (8.9% vs. 7.1%), cancer (14.4% vs. 5.9%), asthma (10.7% vs. 5.7%), chronic obstructive pulmonary disease (COPD);

10.3% vs. 4.2%), ischemic heart disease (6.1% vs. 4.1%), cerebrovascular disease (4.2% vs. 1.9%), congestive heart failure (3.5% vs. 1.6%), and chronic renal failure (2.9% vs. 1.2%). The one exception to the pattern detailed above was the higher percentage of visits with a diagnosis of arthritis in NAMCS vs. Centricity (14.3% vs. 13.2%). Five of the differences are especially sharp (ie, exceeding 5%): those for hyperlipidemia, depression, cancer, asthma, and COPD.

## Discussion

Regarding demographic data, comparisons of NAMCS and Centricity EMR patient visits by age and sex show that Centricity visits are somewhat more likely to involve younger patients (the 6 percentage differences by age group ranged from 0.9%–2.6%), and substantially more likely to involve females (5.1%). Although all of these differences are statistically significant, given the large sample size, we believe that sex may have more potential for confounding than age. This issue should be addressed by additional research.

Comparisons of NAMCS and Centricity EMR results for physician's primary diagnosis show lower proportions of neoplasm and circulatory condition diagnoses in Centricity and a higher proportion of respiratory condition diagnoses, suggesting relatively more visits for chronic conditions in NAMCS and relatively more visits for acute conditions in Centricity.

Turning to the selected chronic conditions highlighted in the NAMCS results, the Centricity data include a substantially higher proportion of visits with at least 1 condition and a substantially lower proportion with none of the selected chronic conditions. Of the 13 possible chronic diagnosis comparisons, all but 1 (ie, arthritis) showed a significantly higher proportion in the Centricity data.

Among the potential explanations for the apparent greater sensitivity of Centricity data to chronic condition diagnoses are (1) methodological differences between the NAMCS and Centricity data collection approaches and data sets, (2) differential characteristics of providers who use EMRs like Centricity, and (3) in particular, differential characteristics of providers who not only use an EMR but also participate in a data collection and analysis program like MQIC.

Regarding the first potential explanation, the key methodological difference in data collection is that the NAMCS focuses on the "primary reason for the visit," while Centricity accumulates all problems and diagnoses and thus has greater capacity to identify all relevant diagnoses. Moreover, data extracted from the Centricity EMR using algorithms to maximize validity may be inherently more comprehensive and accurate than patient- or visit-level survey data, the validity of which depends on the motivation of patients, physicians, other clinicians, and office staff. Additional strengths of the Centricity EMR database are (1) its incorporation of documentation of a wide range of diagnostic and therapeutic services, specifically laboratory test results and medications ordered, and (2) its incorporation of all such data over time, allowing greater sensitivity in capturing diagnoses. As a result, it is likely that the Centricity data reflect more of the medical problems that motivated patient visits than the NAMCS data.

At least 2 specific interpretations of our major findings relate to the differential characteristics of EMRs and the

TABLE 1. NUMBER, PERCENT DISTRIBUTION, AND ANNUAL RATE OF OFFICE VISITS WITH CORRESPONDING STANDARD ERRORS, BY PATIENT AGE AND SEX: UNITED STATES, 2005:  
COMPARISON OF NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS) AND GE CENTRICITY RESULTS

Patient characteristics	NAMCS					GE Centricity					
	Number of visits	Standard error	Percent distribution	Standard error of percent	Confidence Interval 99%	Number of visits per 100 persons per year <sup>1</sup>	Standard error of rate	Number of visits	Percent distribution	Number of visits per 100 persons per year <sup>1</sup>	Significance based on NAMCS CI
All visits	963,617	40,611	100.0	...	...	100	...	6202	100.0		
Sex and age											
Female	560,355	24,487	58.2	0.6	56.7–59.7	376.8	16.5	3836	61.8	483.3	***
<15 years	72,804	5166	7.6	0.5	6.3–8.9	245.7	17.4	382	6.2	380.3	
15–24 years	44,654	2787	4.6	0.3	3.8–5.4	220.0	13.7	340	5.5	392.9	***
25–44 years	129,588	7056	13.4	0.6	11.9–14.9	313.0	17.0	882	14.2	401.1	
45–64 years	166,275	8738	17.3	0.4	16.3–18.3	446.9	23.5	1153	18.6	426.3	***
65–74 years	66,993	4419	7.0	0.3	6.2–7.8	671.3	44.3	490	7.9	519.5	
≥75 years	80,042	5958	8.3	0.4	7.3–9.3	784.6	58.4	589	9.5	569.5	***
Male	403,262	18,050	41.8	0.6	40.3–43.3	283.1	12.7	2366	38.2	403.1	***
<15 years	87,748	6819	9.1	0.6	7.6–10.6	282.9	22.0	411	6.6	386.0	
15–24 years	25,930	2065	2.7	0.2	2.2–3.2	125.1	10.0	121	2.0	279.1	
25–44 years	72,031	4732	7.5	0.4	6.5–8.5	177.5	11.7	375	6.0	343.8	***
45–64 years	116,905	6384	12.1	0.4	11.1–13.1	332.7	18.2	746	12.0	396.3	
65–74 years	52,068	3600	5.4	0.3	4.6–6.2	618.6	42.8	359	5.8	481.0	
≥75 years	48,581	3377	5.0	0.3	4.2–5.8	741.4	51.5	354	5.7	543.2	

<sup>1</sup>Visit rates for age, sex, race, and ethnicity are based on the July 1, 2005, set of estimates of the civilian noninstitutional population of the United States as developed by the Population Division, US Census Bureau.

Numbers may not add to totals because of rounding.

14 Advance DataNo. 387 + June 29, 2007.

..., category not applicable.

\*\*\* $p < .01$ .



TABLE 2. NUMBER AND PERCENT DISTRIBUTION OF OFFICE VISITS WITH STANDARD ERRORS, BY PHYSICIAN'S PRIMARY DIAGNOSIS: UNITED STATES, 2005:  
COMPARISON OF NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS) AND GE CENTRICITY RESULTS

Major disease category	ICD-9-CM code range <sup>1</sup>	NAMCS					GE Centricity			Significance based on NAMCS CI
		Number of visits	Standard error	Percent distribution	Standard error of percent	Confidence Interval 99%		Number of visits in thousands	Percent distribution	
						Lower Bound	Upper Bound			
All visits	...	963,617	40,611	100	...			2211	100.0	
Infectious and parasitic diseases	001-139	26,720	2062	2.8	0.2	2.3	3.3	90.5	4.1	***
Neoplasms	140-239	39,200	5285	4.1	0.5	2.8	5.4	26.1	1.2	***
Endocrine, nutritional and metabolic diseases, and immunity disorders	240-279	56,408	4655	5.9	0.4	4.9	6.9	93.0	4.2	***
Mental disorders	290-319	47,094	3777	4.9	0.4	3.9	5.9	76.2	3.4	***
Diseases of the nervous system and sense organs	320-389	86,128	6644	8.9	0.5	7.6	10.2	167.0	7.6	***
Diseases of the circulatory system	390-459	81,836	6683	8.5	0.6	7.0	10.1	79.6	3.6	***
Diseases of the respiratory system	460-519	110,999	7670	11.5	0.7	9.7	13.3	395.5	17.9	***
Diseases of the digestive system	520-579	28,678	2225	3	0.2	2.5	3.5	82.2	3.7	***
Diseases of the genitourinary system	580-629	42,256	3131	4.4	0.3	3.6	5.2	114.6	5.2	***
Diseases of the skin and subcutaneous tissue	680-709	44,443	3189	4.6	0.3	3.8	5.4	125.0	5.7	***
Diseases of the musculoskeletal system and connective tissue	710-739	80,601	6320	8.4	0.5	7.1	9.7	224.4	10.1	***
Symptoms, signs, and ill-defined conditions	780-799	60,536	3916	6.3	0.3	5.5	7.1	352.7	16.0	***
Injury and poisoning	800-999	45,137	3087	4.7	0.3	3.9	5.5	88.8	4.0	
Supplementary classification	V01-V82	179,276	11,124	18.6	0.9	16.3	20.9	262.3	11.9	***
All other diagnoses <sup>2</sup>	...	25,609	3533	2.7	0.3	1.9	3.5	32.7	1.5	***
Unknown <sup>3</sup>	...	8697	1645	0.9	0.2	0.4	1.4	6.8	0.0	***

... category not applicable.

<sup>1</sup>Based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).

<sup>2</sup>Includes diseases of the blood and blood-forming organs (280-289); complications of pregnancy, childbirth, and the puerperium (630-677); congenital anomalies (740-759); certain conditions originating in the perinatal period (760-779); and entries not codable to the ICD-9-CM (eg, illegible entries, left against medical advice, transferred, entries of "none," or "no diagnoses").

<sup>3</sup>Includes blank diagnoses.

Numbers may not add to totals because of rounding.

Advance DataNo. 387 + June 29, 2007 23.

\*\*\* $p < .01$ .

TABLE 3. NUMBER AND PERCENT DISTRIBUTION OF OFFICE VISITS BY SELECTED CHRONIC CONDITIONS, ACCORDING TO PATIENT AGE AND SEX: UNITED STATES, 2005:  
COMPARISON OF NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS) AND GE CENTRICITY RESULTS

Chronic conditions <sup>1</sup> (number of visits in thousands)	NAMCS						GE Centricity							
	Total NAMCS	By patient age			By patient sex			Total GE Centricity	By patient age			By patient sex		
		<45 years	45–64 years	65–74 years	75 years and older	Female	Male		<45 years	45–64 years	65–74 years	75 years and older	Female	Male
All visits	963,617	432,754	283,180	119,061	128,623	560,355	403,262	2,210.6	1,074.4	660.2	238.6	237.4	1,399.2	811.4
At least 1 condition	507,390	121,891	187,209	94,472	103,818	298,366	209,024	1,402.0	469.9	518.5	206.4	207.2	897.1	504.9
None	422,515	295,386	85,466	21,112	20,551	242,189	180,327	808.6	604.6	141.7	32.2	30.2	502.1	306.5
Blank	33,712	15,476	10,505	3477	4254	19,800	13,911	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hypertension	219,476	23,060	84,640	52,131	59,646	127,072	92,404	538.7	72.5	233.5	114.2	118.5	337.0	201.7
Arthritis	137,942	20,861	51,744	27,168	38,169	91,610	46,332	290.7	24.2	113.2	70.0	83.3	209.1	81.7
Hyperlipidemia	130,053	12,806	55,561	31,658	30,028	68,208	61,845	567.9	84.3	263.6	118.9	101.1	340.8	227.1
Diabetes	94,359	10,387	38,628	23,810	21,534	50,625	43,733	226.3	30.3	99.3	52.3	44.4	133.3	93.0
Depression	84,610	30,935	35,059	8538	10,078	59,418	25,192	352.6	146.4	134.9	35.6	35.7	271.8	80.8
Obesity	68,201	24,275	29,574	9298	5054	45,028	23,173	197.4	84.8	81.1	20.9	10.5	140.7	56.7
Cancer	57,025	4817	18,809	15,014	18,384	32,384	24,640	319.1	76.3	121.1	59.0	62.7	203.6	115.6
Asthma	54,545	28,913	14,491	5764	5376	33,276	21,269	235.9	146.6	58.3	17.3	13.7	152.8	83.1
COPD	40,754	7746	11,425	10,818	10,764	22,458	18,295	227.0	67.3	80.9	39.0	39.9	144.5	82.5
Ischemic heart disease	39,559	1472	10,179	12,092	15,815	16,245	23,314	135.1	4.1	43.4	38.1	49.5	63.7	71.4
Osteoporosis	27,977	*	6466	8050	12,096	24,842	3136	92.8	2.0	23.7	25.8	41.2	86.5	6.4
Cerebrovascular disease	18,319	731	4322	5489	7777	9000	9319	77.0	3.2	19.0	20.7	34.0	47.1	29.9
CHF	15,692	*	3589	3409	8392	8950	6742	65.2	2.3	15.7	16.7	30.5	38.6	26.6
Chronic renal failure	11,486	*	2741	3039	4969	5633	5852	32.8	1.9	8.8	8.4	13.7	17.1	15.7

\*Figure does not meet standards of reliability or precision.

<sup>1</sup>Presence of chronic conditions, regardless of visit diagnosis, were based on checkbox responses. ICD-9-CM diagnosis codes analyzed in the Centricity data are as follows: hypertension (401–405), arthritis (710–716, excluding 711.0), hyperlipidemia (272, 272.0, 272.1, 272.2, 272.3, 272.4), diabetes (250), depression (300.4, 311, 296.2, 296.3), obesity (278.0), cancer (140–239), asthma (493), chronic obstructive pulmonary disease (490–492, 494–496), ischemic heart disease (410–414), osteoporosis (733.0), cerebrovascular disease (430–438), congestive heart failure (429.1, 429.2, 428 excluding 428.21, 428.31, and 428.31), and chronic renal failure (585).

COPD chronic obstructive pulmonary disease; CHF, congestive heart failure.

Numbers may not add to totals because more than 1 condition may be reported per visit.

28 Advance DataNo. 387 + June 29, 2007.

TABLE 3 (CONTINUED). NUMBER AND PERCENT DISTRIBUTION OF OFFICE VISITS BY SELECTED CHRONIC CONDITIONS, ACCORDING TO PATIENT AGE AND SEX: UNITED STATES, 2005: COMPARISON OF NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS) AND GE CENTRICITY RESULTS

Chronic Condition <sup>1</sup>	NAMCS										GE Centricity									
	By patient age					By patient sex					Total GE Centricity					By patient age				
	<45 years		45–64 years		65–74 years		≥75 years		Total		<45 years		45–64 years		65–74 years		≥75 years		Female	
	Total NAMCS CI 99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%
All visits	49.4–56.0	25.6–30.8	62.8–69.4	76.0–82.6	76.8–84.6	49.6–56.8	48.2–55.4	100	63.4	***	100	43.7	***	100	78.5	***	100	86.5	***	100
At least 1 condition	40.7–46.9	65.5–71.1	26.9–33.5	14.6–20.8	12.4–19.6	39.9–46.5	41.1–48.3	36.6	0.0	***	56.3	***	21.5	***	13.5	***	12.7	***	35.9	***
None	2.5–4.5	2.6–4.6	2.4–5.0	1.4–4.4	2.0–4.6	2.5–4.5	2.4–4.4	0.0	24.4	***	0.0	***	0.0	***	0.0	***	0.0	***	0.0	***
Hypertension	20.2–25.4	4.3–6.3	26.6–33.2	38.9–48.7	42.0–50.8	19.9–25.5	20.1–25.7	24.4	13.2	***	6.7	***	35.4	***	47.9	***	49.9	***	24.1	***
Arthritis	12.5–16.1	3.8–5.8	15.5–21.1	19.5–26.1	25.3–34.1	14.0–18.6	10.0–13.0	10.7	10.3	***	2.3	***	17.1	***	29.3	***	35.1	***	14.9	***
Hyperlipidemia	11.4–15.6	2.2–3.8	16.5–22.7	22.2–31.0	19.2–27.4	10.1–14.3	12.7–17.9	25.7	10.2	***	7.8	***	39.9	***	49.8	***	42.6	***	24.4	***
Diabetes	8.5–11.1	1.6–3.2	11.8–15.4	17.2–22.8	14.1–19.3	7.7–10.3	9.3–12.3	16.0	8.9	***	2.8	***	15.0	***	21.9	***	18.7	***	9.5	***
Depression	7.3–10.3	5.8–8.4	9.8–15.0	4.9–9.5	5.5–10.1	8.5–12.7	4.9–7.5	14.4	14.4	***	13.6	***	20.4	***	14.9	***	15.0	***	19.4	***
Obesity	5.8–8.4	4.3–6.9	8.1–12.7	5.5–10.1	2.4–5.4	6.5–9.5	4.4–7.0	10.7	10.3	***	7.9	***	12.3	***	8.8	***	4.4	***	10.1	***
Cancer	4.4–7.4	0.6–1.6	4.5–8.7	9.0–16.2	11.0–17.6	4.0–7.6	4.8–7.4	10.7	10.3	***	7.1	***	18.3	***	24.7	***	26.4	***	14.5	***
Asthma	4.9–6.5	5.4–8.0	4.1–6.1	3.3–6.3	2.7–5.7	4.9–6.9	4.0–6.6	10.7	10.3	***	13.6	***	8.8	***	7.3	***	5.8	***	10.9	***
COPD	3.4–5.0	1.3–2.3	3.0–5.0	6.5–11.7	5.8–11.0	3.2–4.8	3.5–5.5	10.3	6.1	***	6.3	***	12.2	***	16.3	***	16.8	***	10.3	***
Ischemic heart disease	3.3–4.9	0.0–0.6	2.8–4.4	8.1–12.3	9.7–14.9	2.1–3.7	4.5–7.1	4.2	4.2	***	0.4	***	6.6	***	16.0	***	20.8	***	4.6	***
Cerebrovascular disease	1.4–2.4	0.2–0.2	1.0–2.0	3.1–6.1	4.2–7.8	1.1–2.1	1.5–3.1	3.5	2.9	***	0.2	***	3.6	***	10.8	***	17.4	***	6.2	***
CHF	1.1–2.1	NA	0.5–2.1	1.9–3.9	4.2–8.8	1.1–2.1	0.9–2.5	2.9	2.4	***	0.3	NA	2.9	***	8.7	***	14.3	***	3.4	***
Chronic renal failure	0.7–1.7	NA	0.2–1.8	1.3–3.9	2.4–5.4	0.7–1.3	0.7–2.3	2.9	2.9	***	0.2	NA	2.4	***	7.0	***	12.9	***	2.8	***

<sup>1</sup>Presence of chronic conditions, regardless of visit diagnosis, were based on checkbox responses.

COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure.

Numbers may not add to totals because more than 1 condition may be reported per visit.

28 Advance DataNo. 387 + June 29, 2007.

\*\*\* $p < .01$ .

TABLE 3 (CONTINUED). NUMBER AND PERCENT DISTRIBUTION OF OFFICE VISITS BY SELECTED CHRONIC CONDITIONS<sup>1</sup>, ACCORDING TO PATIENT AGE AND SEX: UNITED STATES, 2005: COMPARISON OF NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS) AND GE CENTRICITY RESULTS

Chronic Condition <sup>1</sup>	NAMCS													
	Total NAMCS							By patient sex						
								Female			Male			
	Percent Distribution	SE of % Dist.	SE of % Dist.	Lower Bound	Upper Bound	Percent Distribution	SE of % Dist.	Lower Bound	Upper Bound	Percent Distribution	SE of % Dist.	Lower Bound	Upper Bound	Percent Distribution
All visits	100	...	...	...	...	100.0	...	...	...	100	...	...	...	100
At least 1 condition	52.7	1.3	1.4	49.4	56.0	53.2	1.4	49.6	56.8	51.8	1.4	48.2	55.4	63.4
None	43.8	1.2	1.3	40.7	46.9	43.2	1.3	39.9	46.5	44.7	1.4	41.1	48.3	36.6
Blank	3.5	0.4	0.4	2.5	4.5	3.5	0.4	2.5	4.5	3.4	0.4	2.4	4.4	0.0
Hypertension	22.8	1	1.1	20.2	25.4	22.7	1.1	19.9	25.5	22.9	1.1	20.1	25.7	24.4
Arthritis	14.3	0.7	0.9	12.5	16.1	16.3	0.9	14.0	18.6	11.5	0.6	10.0	13.0	13.2
Hyperlipidemia	13.5	0.8	0.8	11.4	15.6	12.2	0.8	10.1	14.3	15.3	1	12.7	17.9	25.7
Diabetes	9.8	0.5	0.5	8.5	11.1	9.0	0.5	7.7	10.3	10.8	0.6	9.3	12.3	10.2
Depression	8.8	0.6	0.6	7.3	10.3	10.6	0.8	8.5	12.7	6.2	0.5	4.9	7.5	16.0
Obesity	7.1	0.5	0.6	5.8	8.4	8.0	0.6	6.5	9.5	5.7	0.5	4.4	7.0	8.9
Cancer	5.9	0.6	0.7	4.4	7.4	5.8	0.7	4.0	7.6	6.1	0.5	4.8	7.4	14.4
Asthma	5.7	0.3	0.4	4.9	6.5	5.9	0.4	4.9	6.9	5.3	0.5	4.0	6.6	10.7
COPD	4.2	0.3	0.3	3.4	5.0	4.0	0.3	3.2	4.8	4.5	0.4	3.5	5.5	10.3
Ischemic heart disease	4.1	0.3	0.3	3.3	4.9	2.9	0.3	2.1	3.7	5.8	0.5	4.5	7.1	6.1
Cerebrovascular disease	1.9	0.2	0.2	1.4	2.4	1.6	0.2	1.1	2.1	2.3	0.3	1.5	3.1	4.2
CHF	1.6	0.2	0.2	1.1	2.1	1.6	0.2	1.1	2.1	1.7	0.3	0.9	2.5	3.5
Chronic renal failure	1.2	0.2	0.1	0.7	1.7	1.0	0.1	0.7	1.3	1.5	0.3	0.7	2.3	2.9
														2.8
														3.3

... Category not applicable.

<sup>1</sup>Presence of chronic conditions, regardless of visit diagnosis, were based on checkbox responses.

COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure.

NOTE: Numbers may not add to totals because more than 1 condition may be reported per visit.

28 Advance DataNo. 387 + June 29, 2007.

providers who adopt them. First, EMRs facilitate more complete and accurate documentation of diagnoses as well as laboratory results and medication orders; specifically, when diagnoses are listed in progress notes, the EMR automatically uses them to populate the problem list. Second, as noted by Gill and Chen<sup>8</sup> regarding MQIC in particular, "there may be differences between EHR users in general and those who participate in MQIC... while MQIC includes a large and diverse group of providers, participation is voluntary, so there may be some self-selection. For example, one reason that practices join MQIC is to have access to quality reporting. It may be that practices that are more interested in measuring (and improving) quality of care are more likely to join MQIC." It is a reasonable inference that providers who use the Centricity EMR and participate in MQIC document diagnoses more completely and thoroughly than other providers. If this is the case, the key findings of the present analysis – substantially higher proportions of visits for key chronic conditions in Centricity compared to NAMCS – represent not an overestimate but an accurate estimate of the distribution of diagnoses in ambulatory visits in the United States.

### *Limitations*

The limitations of this study generally stem from the methodological differences between the NAMCS and Centricity EMR data sets. Although NAMCS data collection is based on probability sampling techniques that support estimates for the United States as a whole, there is no such methodology and capability in the Centricity database. Thus, differences in results derived from the 2 data sets may reflect a variety of factors including, among others, the age, sex, and other demographic characteristics of the patients whose visits were captured; the clinical characteristics of those patients; and the characteristics of the physicians who use the Centricity EMR, which may, in turn, be associated with patients' demographic and clinical characteristics.

As already described, several approaches were employed to reduce the effects of such methodological artifacts. Specifically, exclusion criteria were employed to reduce potential overestimation of patients' diagnoses in Centricity: (1) filtering problem/complaint text extracted from records to guarantee that patients merely evaluated for a condition (without confirmation of diagnosis) were not counted as "positives" along with patients explicitly diagnosed with that condition; and (2) excluding retrospective visit data that was "backfilled" when providers first adopted the Centricity EMR. However, available data do not allow assessment of the success of these strategies at eliminating all duplication and overestimation.

More generally, analyses of EMR data are subject to the same limitations posed by any use of paper or electronic medical records; data will not be available in the medical record if they were not available to the physician. Such gaps result whenever patients receive care from other providers but documentation is not provided to the original physician. Moreover, gaps may result to the degree that paper-based documentation is not transferred to the EMR and to the degree that there are free-text data in the EMR that are not available for analysis because of the complexities and costs of such text mining.

Among the specific potential limitations of the Centricity EMR database are that data on a single patient may be included in the database with more than 1 patient identifier, allowing for duplication and overestimation when aggregate data are analyzed. Additionally, data reflect problems or complaints, rather than ICD-9-CM diagnoses, as are used in NAMCS; as a result, analyses must begin by searching the entire patient record, including multiple codes and text strings, to guarantee a comprehensive search for and valid identification of diagnoses. A final limitation is the rate of missing data on race/ethnicity in the Centricity data, which was too high to allow comparisons with the NAMCS data.

On the other hand, among the strengths of the Centricity EMR database are its incorporation of documentation of a wide range of diagnostic and therapeutic services, specifically laboratory test results (with exact values and units of measurement) and medications ordered, and, perhaps most importantly, all of these types of data at multiple points in time, allowing longitudinal analyses not possible with most federal data sets, including NAMCS.

It is expected that use of ambulatory EMR systems like GE Centricity will expand substantially during the next decade. Many general trends and, specifically, increased federal incentives will combine to increase EMR utilization, increase the number of patients whose records will be incorporated in EMRs and EMR-based research databases such as GE CDS and MQIC, and increase the number of years of continuous patient data available for longitudinal research including outcomes research, quality measurement, and research on quality.

A broad range of clinical research will be facilitated by increasing EMR utilization and the increasing size, scope, and span of EMR-based databases. One way of projecting the potential for such studies is to examine the range of studies conducted to date using GE Centricity data. A recent list of such publications shows 12 articles published in peer-reviewed journals from 2006-2008 and 31 presentations at professional society meetings from 2004-2008. Among the topics of the journal articles were quality of care for patients with diabetes; oral antidiabetic medication use and outcomes (2 articles); cardiometabolic risk factors; the effectiveness of statins to lower low-density lipoprotein cholesterol (2 articles); the effects of a randomized controlled trial on antihypertensive prescribing; gastrointestinal complications of over-the-counter NSAIDs; effects of second-generation antipsychotics on weight gain; diagnosis, treatment, and/or outcomes for COPD (2 articles); and antibiotic use for adult upper respiratory infections. Undoubtedly, the range of conditions and outcomes studied will grow. In particular, studies of rare conditions and outcomes will become more common, given the huge numbers of cases available for longitudinal analyses as well as the ever-widening time span.

### **Conclusions**

This study has compared key demographic characteristics (ie, age, sex) and clinical characteristics (ie, diagnoses) in the GE Centricity EMR database and the federal NAMCS. Compared to NAMCS patient visits, Centricity EMR patient visits are somewhat more likely to involve younger patients and considerably more likely to involve females. Centricity visits include lower percentages of the more severe chronic conditions of neoplasms and circulatory conditions and

higher percentages of respiratory conditions, suggesting that there is a higher proportion of visits for acute conditions in the Centricity database and a higher proportion of visits for chronic conditions in NAMCS. However, regarding the 13 chronic conditions highlighted by NAMCS, Centricity data show a substantially higher proportion of visits for such conditions in general, and for 12 of the 13 specific conditions.

While Centricity EMR data and results are not perfectly comparable to NAMCS data and results, Centricity data may actually be more sensitive in capturing diagnoses, especially chronic condition diagnoses. More research is needed to assess the validity and utility of Centricity and other EMR databases, and, specifically, to assess the potential for evolving methodologies to maximize such validity and utility.

### Author Disclosure Statement

Ethicon Endo-Surgery, Inc., sponsored this research. Ethicon Endo-surgery, Inc. employs Dr. Haas and contracted with all other authors of this manuscript. Drs. Crawford, Cote, Couto, Daskiran, Gunnarsson, Haas, Nigam, and Schuette, and Mr. Yaskin and Ms. Haas disclosed no other financial ties or conflicts of interest.

### References

1. Burt CW, Hing E, Woodwell D. Electronic medical record use by office-based physicians: United States, 2005. Available at: <http://www.cdc.gov/nchs/products/pubs/pubd/hestats/electronic/electronic.htm>. Accessed July 31, 2006.
2. Hing E, Cherry DK, Woodwell DA. National ambulatory medical care survey: 2004 summary. *Advance Data*. 2006; 374:1–33.
3. Machlin SR, Valluzzi JL, Chevarley FM, Thorpe JM. Measuring ambulatory health care use in the United States: a comparison of 1996 estimates across four federal surveys. *J Econ Social Measurement*. 2001;27:57–69.
4. GE Centricity Data Services. Journal articles and presentations. 2009.
5. Brixner D, Said Q, Kirkness C, Oberg B, Ben-Joseph R, Oderda G. Assessment of cardiometabolic risk factors in a national primary care electronic health record database. *Value in Health*. 2007;10:S29–S36.
6. Brixner D, Gbate SR, McAdam-Marx C, Ben-Joseph R, Said Q. Association between cardiometabolic risk factors and body mass index based on diagnosis and treatment codes in an electronic medical record database. *J Manag Care Pharm*. 2008;14:757–767.
7. Wang J, Mullins CD, Mamdani M, Rublee DA, Shaya FT. New diagnosis of hypertension among celecoxib and non-selective NSAID users: a population-based cohort study. *Ann Pharmacother*. 2007;41(6):937–943.
8. Gill J, Chen Y. Quality of lipid management in outpatient care: A national study using electronic health records. *Am J Med Qual*. 2008;23:375–381.
9. National Center for Health Statistics. NAMCS description. Available at: <http://www.cdc.gov/nchs/about/major/ahcd/namcsdes.htm>. Accessed November 3, 2008.
10. Cherry DK, Woodwell DA, Rechtsteiner EA. National Ambulatory Medical Care Survey: 2005 summary. *Adv Data*. 2007;387:1–39.

**This article has been cited by:**

1. Claudio Marelli, Candace Gunnarsson, Susan Ross, Sara Haas, Donna F. Stroup, Paul Cload, Paul Clopton, Anthony N. DeMaria. 2011. Statins and Risk of Cancer. *Journal of the American College of Cardiology* **58**:5, 530-537. [[CrossRef](#)]