

The benefit of using both claims data and electronic medical record data in health care analysis



By John Wilson, MD, Vice President of Clinical Analytics, OptumInsight,
and Adam Bock, MD

Those interested in understanding and improving the health care system have long used information in medical claims—bills submitted by physicians and hospitals for payment by commercial and government health plans. This information helps health administrators, researchers, and policy makers to understand the cost and quality of health care and identify patients at risk of developing chronic conditions, pinpoint billing fraud, and improve patient care (e.g., by finding patients who are overdue for mammography or other recommended care).

So what is claims data? Claims data consists of the billing codes that physicians, pharmacies, hospitals, and other health care providers submit to payers (e.g., insurance companies, Medicare). This data has the benefit of following a relatively consistent format and of using a standard set of pre-established codes that describe specific diagnoses, procedures, and drugs. Additionally, since all health care providers want to be paid for their services, nearly every encounter that a patient has with the medical system leads to the generation of a claim, creating an abundant and standardized source of patient information.

More recently, however, a new world of data has become increasingly available for analysis: data from the electronic medical record (EMR). Because the EMR is the software which is accessed directly by physicians to record the details of their encounters with patients, it contains a rich array of data not available elsewhere.

This paper makes the case that neither claims data nor EMR data alone allow for optimal analysis of patient health status, but rather that the best practice of health analysis depends on using *both* of these data types together.

The advantages of claims data

Before extolling the virtues of EMR data, it should be said that claims data has a lot going for it. Perhaps the main advantage is that it is only through claims data that a holistic view of the patient's interactions with the health care system can be seen.

Consider, for example, Steve, a patient with diabetes. Steve might have an appointment in March with his primary doctor. Before the appointment, he might stop at a lab affiliated with his doctor's office to have his cholesterol checked. Then in April, Steve might visit his eye doctor to make sure that diabetes hasn't affected his vision. Finally, in May he might go to his local pharmacy to refill his insulin prescription.

So how will all of this activity be documented in the EMR? In many cases, only a subset of this activity is recorded within an EMR. This is true because, in spite of federal incentives encouraging EMR adoption, in 2011 only 57 percent of office-based physicians used an EMR.¹ So, if Steve's primary doctor uses an EMR but his eye doctor does not, data from

his primary doctor's EMR may contain no evidence whatsoever that an eye exam was performed. Additionally, while Steve's doctor may use the EMR to write and print a prescription for Steve's medications, the EMR system may not be connected to the pharmacy and therefore may not record whether Steve ever filled (or refilled) his prescription. And even when physician offices are electronically connected to pharmacies (36 percent of offices in 2010),² that connection often allows only the electronic transmission of the prescription to the pharmacy without sending any information back to the physician about whether or not the prescription was ever filled or refilled.

From this simple example, we can see several advantages of claims data:

- **Assessing medication compliance:** Claims data include important details about medications. Every fill/refill of a prescription, complete with date of that event, shows up. In addition, there is information about the actual drug that was dispensed, the amount of drug that was dispensed, and the number of days—if the medication is taken as prescribed—that the medication should last.

With all of this data, it is possible to assess whether or not a patient is taking a medication as directed or not. This evaluation cannot be done using EMR data, given the lack of information for prescription refills.

- **Good reflection of tests, procedures, and services provided:** From the above discussion, we can see that EMR data might not capture the fact that Steve had an eye exam. Similarly, any services provided by a provider not using an EMR will fail to be reflected in EMR data. The claims data, however, will contain evidence of them because all of these services need to be reimbursed.

What does EMR data add?

More complete condition identification

There are a variety of reasons that physicians may fail to completely record on a claim all the diagnoses from a visit. For one thing, physicians are constantly pressed for time, and every second spent recording billing codes is a second that takes them away from direct patient care. Additionally, in a fee-for-service setting, the payment that a physician receives for an office visit is not directly related to the number or type of conditions for which the physician codes (see Appendix A for more detail). Hence, in many cases, the EMR will have a more complete set of diagnoses for a given patient than claims data.

Because of this, claims data is often an imperfect reflection of the actual status of a patient. Several studies support this.

One study examined how often people with the condition of chronic kidney disease (CKD) had a claims-based diagnosis code

¹ NCHS Data Brief, Number 79, November 2011: Electronic Health Record Systems and Intent to Apply for Meaningful Use Incentives Among Office-based Physician Practices: United States, 2001–2011

² <http://www.surescripts.com/pdfs/national-progress-report.pdf>

for this condition over a one-year period. In this study,³ results of a blood test (the estimated glomerular filtration rate or eGFR) were used to determine whether or not kidney disease existed. If this test was abnormal on at least two separate occasions over a year, the patients met the definition CKD. The authors then examined all claims data for the patients who had CKD diagnosed by virtue of lab testing. They found that only 20–42 percent of these patients had a diagnosis code for CKD on a claim over the one-year period. Put another way, if one year of claims data was all that had been present, 58–80 percent of people with CKD would not have been identified.

Another study published in the *Journal of the American Medical Association (JAMA)*⁴ showed that of children with EMR blood pressure values that were high on at least three separate doctor visits, only 26 percent of them had a claim with a diagnosis of hypertension on it.

In addition to showing the shortcomings of claims data in identifying conditions, these data suggest another powerful conclusion: Use of clinical data from the EMR can significantly improve condition identification. The use of lab result data elements can support identification of people with CKD even without a coded diagnosis. In addition, the use of vital sign data can allow identification of people with hypertension despite the lack of a claim-based diagnosis of this condition.

In fact, there are a variety of data elements that might be available in the EMR which, when analyzed, can allow the identification of a condition that was either not recognized or not coded for by the physician. So, one way in which EMR data enables better condition identification is by providing access to data elements (e.g., lab results and vital signs) that allow one to *impute* a diagnosis—even if that diagnosis was never made.

Allowing for imputed diagnosis is just one way in which EMR data improves condition identification over and above the use of claims data. The EMR also has something which claims data do not: the concept of a *'problem list'*.

Claims data is, by its nature, temporally limited. Meaning the claim reflects only the diagnoses and services that occurred on the date when the claim was submitted. It is not designed to convey information about what happened in the past. So for a patient who had heart surgery or an appendectomy two years ago, there is no reason that those items will appear on a claim today. Similarly, if a patient had a diagnosis of heart failure two years ago, that diagnosis may not appear on a claim during a subsequent time period, even if the condition persists. The EMR, however, has a way to transcend the concept of time by which claims data are constrained: the problem list.

3 Failure of ICD-9-CM codes to identify patients with comorbid chronic kidney disease in diabetes. Kern EF, Maney M, Miller DR, Tseng CL, Tiwari A, Rajan M, Aron D, Pogach L. *Health Serv Res.* 2006 Apr;41(2):564-80.

4 Underdiagnosis of hypertension in children and adolescents. Hansen ML, Gunn PW, Kaelber DC *JAMA.* 2007 Aug 22;298(8):874-9.

The problem list is an area in the electronic record where providers can keep track of the list of medical problems affecting a patient. The EMR maintains this list independently from any particular medical visit/encounter. Hence, use of the EMR problem list allows identification of conditions which may not be identified via claims data.

More timely

In addition to the ability of EMR data to enhance condition identification as discussed above, EMR data has another advantage: timeliness. Sometimes we would like to understand that a patient has experienced a certain event as soon as that event has occurred. For example, we might like a nurse to provide a follow-up call to a patient the day after an emergency room visit. If we are dependent on claims data alone to identify the emergency room visit, there may be a delay (sometimes of months) until a claim for this visit is received and processed. However, providers interact with the EMR during (or soon after) the patient encounter. Hence, EMR data is generated in real time, and a system which evaluates data from the EMR can allow a much more rapid response.

Information in addition to diagnosis

It is also important to note that the EMR is a much richer dataset than is claims data. At its heart, claims data is designed to hold only those pieces of information that are required to facilitate payment by an insurance company: what service was provided; the diagnosis; who was the service provider; how much money is owed for that service. However, the EMR contains a lot more detail:

- Vital signs and lab results (as noted in the CKD and high blood pressure examples)
- Information from the patient:
 - Habits: smoking and alcohol use
 - List of non-prescription drugs (which don't appear in claims data) taken by the patient, such as aspirin
 - Results of surveys given to the patient:
 - i.e. PHQ-9 is often used for depression monitoring or screening
 - i.e. Asthma Control Test (ACT) which is used to monitor asthma severity
 - i.e. SF-36 which measures patient quality of life
- Information recorded by nurses, pharmacists, and other team members involved in case or disease management

All of the information above is of vital importance in understanding the full health status of a patient and yet is not available within claims data. If we have a patient with a history of a heart attack, we need to know whether or not the patient is taking aspirin—but claims data will not contain this information. If we are trying to identify smokers, we cannot depend on claims data to provide this information. If we want to track a patient's response to depression treatment, we need to see the results of a survey tool like PHQ-9 over time: again, claims data does not afford us the opportunity to evaluate this.

It is true that sometimes these non-claims data types may be missing from a given EMR and, in some cases, this data will need to be imported from other sources. However, the main point being made here is that there are myriad non-claims data types that have value and that, whether via EMR or another route, having this data is useful.

Table 1

Comparison: Claims Data vs. EMR Data		
	Claims Data	EMR Data
Scope of Data	Broad: Captures information from all doctors/providers caring for a patient	Limited: Captures only the portion of care provided by doctors using the EMR
Scope of Patients	Insured patients only	All patients (including uninsured)
Prescription Data	An accurate record of all prescriptions that were filled including dates of refills	Contains only that a physician prescribed a drug but not whether or not it was filled/refilled
Non Prescription Drugs	Not present	Present
Data Richness	Limited: diagnosis, procedures	Rich: lab results, vital signs, patient surveys, habits (smoking, etc), problem list, etc.

Scope of patients

An additional benefit of EMR data is that it contains data on everyone a provider group cares for. Claims data, by its nature, is owned by the insurance company which covers each patient. It is nearly unheard of for a provider to have access to claims from all commercial insurance carriers used by all patients for whom they care. In addition, government payers like Medicare and Medicaid do not release patient identified claim data, and uninsured patients have no source of claim data. However, the EMR does have the clinical information for all of a provider's patients, and so any analysis that is done to enhance the health status of a provider group's full patient population needs to draw upon EMR data as its source.

Conclusion

The goal of using data to understand the health status and health care activities of a patient is a bit like looking at a panoramic vista through the lens of a camera. The camera brings certain items into focus but it is not capable of capturing the entire scene. Similarly, any single source of health data will have limitations (see Table 1). Hence, neither claims data nor EMR data alone can provide a complete, accurate, and timely view of a person's health status. The future of health data analysis will depend on integrating both data sources together and leveraging the strengths of each.

Appendix A: office visit reimbursement

When a physician bills for an office visit, that bill is submitted with a CPT® code which explains the intensity of the visit. It is the CPT® code that determines the payment for the office visit rather than either the type or number of diagnosis codes (ICD-9 codes). As seen by the descriptions below, there are requirements for billing a higher versus lower cost CPT® code, but those requirements are not directly related to the number/type of diagnoses.

Less expensive office visit:

name	code	description
CPT®	99211	Office or other outpatient visit for the evaluation and management of an established patient, that may not require the presence of a physician. Usually, the presenting problem(s) are minimal. Typically, 5 minutes are spent performing or supervising these services.

More expensive office visit:

Name	code	description
CPT®	99215	Office or other outpatient visit for the evaluation and management of an established patient, which requires at least 2 of these 3 key components: a comprehensive history; a comprehensive examination; medical decision making of high complexity. Counseling and/or coordination of care with other providers or agencies are provided consistent with the nature of the problem(s) and the patient's and/or family's needs. Usually, the presenting problem(s) are of moderate to high severity. Physicians typically spend 40 minutes face-to-face with the patient and/or family.

CPT is a registered trademark of the American Medical Association.

The views and opinions expressed herein are not necessarily those of Optum, its employees or others. Optum provides this white paper for informational purposes only. It is not intended as advice for a particular situation, nor is it intended to be legal or professional advice. Consult with an appropriate professional for your situation.

The information in this document is subject to change without notice. This documentation contains proprietary information, which is protected by U.S. and international copyright.

All rights are reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, without the express written permission of Optum. Copyright 2012 Optum.



13625 Technology Drive, Eden Prairie, MN 55344

Optum and its respective marks are trademarks of Optum, Inc. All other brand or product names are trademarks or registered marks of their respective owners. Because we are continuously improving our products and services, Optum reserves the right to change specifications without prior notice. Optum is an equal opportunity employer.

12-27912 02/12 © 2012 Optum. All Rights Reserved.