Identifying Chronic Conditions in Medicare Claims Data: Evaluating the Chronic Condition Data Warehouse Algorithm

Yelena Gorina and Ellen A. Kramarow

Objective. To examine the strengths and limitations of the Center for Medicare and Medicaid Services' Chronic Condition Data Warehouse (CCW) algorithm for identifying chronic conditions in older persons from Medicare beneficiary data.

Data Sources. Records from participants of the NHANES I Epidemiologic Follow-up Study (NHEFS 1971–1992) linked to Medicare claims data from 1991 to 2000.

Study Design. We estimated the percent of preexisting cases of chronic conditions correctly identified by the CCW algorithm during its reference period and the number of years of claims data necessary to find a preexisting condition.

Principal Findings. The CCW algorithm identified 69 percent of preexisting diabetes cases but only 17 percent of preexisting arthritis cases. Cases identified by the CCW are a mix of preexisting and newly diagnosed conditions.

Conclusions. The prevalence of conditions needing less frequent health care utilization (e.g., arthritis) may be underestimated by the CCW algorithm. The CCW reference periods may not be sufficient for all analytic purposes.

Key Words. CCW, NHEFS, Medicare claims, chronic conditions

As the population ages and the treatment and management of chronic conditions such as heart disease, cancer, and diabetes has improved, the number of older people with one or more chronic conditions has increased (Vogeli et al. 2007). In 2005, among persons 65 and older, 91.5 percent had at least one chronic condition, and 76.6 percent had at least two chronic conditions. About 59 percent of all medical care expenses for persons age 65 and older were for treatment of chronic conditions (Machlin, Cohen, and Beauregard 2008).

A significant body of research has used administrative databases to assess chronic conditions, but there are limitations in how well these data can identify a range of conditions, especially comorbid conditions (Taylor,

Fillenbaum, and Ezell 2002; Rector et al. 2004; Kern et al. 2006; Klabunde, Harlan, and Warren 2006; Harrold et al. 2007; Østbye et al. 2008). In support of the goals of Section 723 of the 2003 Medicare Prescription Drug, Improvement, and Modernization Act, the Center for Medicare and Medicaid Services (CMS) created the Chronic Condition Data Warehouse (CCW), consisting of CMS Medicare beneficiary data linked by a unique ID across multiple Medicare data sources. As part of the CCW, beneficiaries with chronic conditions are identified through a predefined algorithm based on particular diagnosis and procedure codes found on certain types of claims within a specified reference period. The ability to easily identify beneficiaries with particular chronic conditions in Medicare claims data has great potential to facilitate and expand research. The CCW algorithms were developed based on prior research using Medicare claims data to identify various chronic conditions (Katz et al. 1997; Herbert et al. 1999; Taylor, Fillenbaum, and Ezell 2002; Losina et al. 2003; Foley et al. 2005). Yet, to date, it is unknown how well the CCW algorithm identifies preexisting chronic conditions, which may have been first diagnosed years before the date of the claims records.

The aim of this paper is to examine the strengths and limitations of using CMS's CCW algorithm with Medicare claims data to identify chronic conditions in older persons. Records from the NHANES I Epidemiologic Follow-up Study (NHEFS), including data from questionnaires, physical examinations, medical facility records, and death certificates, have been linked to Medicare claims records. We selected five conditions common among older persons: diabetes, ischemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), dementia, and arthritis. We compared diagnoses for these five conditions derived from the two data sources (NHEFS and Medicare claims using the CCW algorithm). Specifically, we explored (a) the number of years of Medicare claims history necessary to find a preexisting condition and (b) the proportion of preexisting versus newly diagnosed conditions identified by the Medicare claims using the CCW algorithm.

Address correspondence to Ellen A. Kramarow, Ph.D., Centers for Disease Control and Prevention, National Center for Health Statistics, Office of Analysis and Epidemiology, 3311 Toledo Road, Room 6332, Hyattsville, MD 20782; e-mail: EKramarow@cdc.gov. Yelena Gorina, M.S., M.P.H., is with the Centers for Disease Control and Prevention, National Center for Health Statistics, Office of Analysis and Epidemiology, Hyattsville, MD.

DATA SOURCES

The NHEFS baseline interview and examination were conducted in 1971–1975 based on a national probability sample of 14,407 persons 25–74 years of age. Four follow-up interviews were conducted during 1982–1992 (Cohen et al. 1987; Finucane et al. 1990; Cox et al. 1992; Cox et al. 1997). We included in our analytic sample 4,846 NHEFS participants who were born in 1935 or earlier; who survived to January 1, 1991 and had at least one follow-up record from the time of the baseline interview to death or 1992, including a phone or face-to-face interview, hospital record or nursing home stay; and who could be linked to fee-for-service Medicare A and B records at some point between 1991 and 2000. These participants ranged in age from 56 to 95 in 1991, but were age 65 or older and eligible for Medicare at some point between 1991 and 2000. Although more recent NHANES data have recently been linked to CMS records (National Center for Health Statistics, Office of Analysis and Epidemiology 2010), the NHEFS data based on NHANES I give us the opportunity to examine chronic conditions over a 25-year period and to include information from facility records. Figure 1 gives a schematic view of the study sample size and inclusion criteria for our analysis. Of the original 14,407 NHEFS participants, 9,923 were born in 1935 or earlier so they would have been 65 years or older between 1991 and 2000 and thus could potentially be Medicare beneficiaries and generate claims. Of these 9,923 age-eligible participants, 5,814 survived to January 1, 1991, the beginning date of Medicare claims availability for this study. From the 5,814 survivors, 84 percent (4,846) could be linked to Medicare records at some point between 1991 and 2000. NHEFS participants were considered ineligible for linking to Medicare records if they refused to provide their Social Security number or Health Insurance Claim number. Participants were also considered ineligible if they refused to provide or had missing or incomplete information on last name and date of birth.

Detailed descriptions of the NHEFS methods have been published elsewhere (Madans et al. 1986; Cox et al. 1997) and are briefly summarized here.

Baseline NHEFS Information: 1971–1975

Basic demographic, behavioral, and medical history data were collected from all NHEFS participants at the time of their NHANES I interview during 1971–1975. A subset of participants was asked more detailed health and socioeconomic status questions. All NHEFS participants underwent a medical examination during 1971–1975.

Figure 1: Study Sample Size and Inclusion Criteria

CCW, Chronic Condition Data Warehouse; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; NHEFS, NHANES I Epidemiologic Follow-up Study.

NHEFS Follow-Up Interviews: 1982–1992

The NHEFS follow-up interviews with the study participant or his/her proxy were conducted in-person in 1982–1984, and via the telephone in 1986 (only

for persons age 55–74 at baseline), 1987, and 1992. Among our study sample of 4,846 NHEFS participants born in 1935 or earlier who survived to 1992 and were linked to Medicare records, 34 percent had all four follow-up interviews, 57 percent had three interviews, 7 percent had two, and 2 percent had only one follow-up interview. These interviews include decedent interviews with the next-of-kin of the decedent.

NHEFS Health Care Facility Stay Data: 1971–1992

The NHEFS health care facility stay data contain information about study participants' overnight stays in a hospital and/or nursing home. The medical records (exclusive of physician records) were collected for facility stays from baseline until the last conducted follow-up interview in 1992 or until death if death occurred before the interview. Up to 10 diagnoses at the time of a hospital discharge or at the time of the admission to a nursing home were available.

Death Records: 1971–2000

Death records are available for each of the study participants who died by the end of the study period (December 31, 2000). Although up to 20 causes of death can be coded on the death certificate, in 50 percent of death certificates three or fewer causes were reported, and only 2 percent of death certificates had five or more causes of death reported. Our study follow-up ends with 2000; however, the NHEFS linked mortality file has recently been updated to provide mortality follow-up through December 31, 2006. These data are described at the NCHS website (http://www.cdc.gov/nchs/data_access/data_linkage/mortality/nhefs_linkage.htm).

Medicare Claims Records: 1991–2000

Of the 5,814 NHEFS participants born in 1935 or earlier and who were alive on January 1, 1991, 4,846 were linked to Medicare A and B records at some point between January 1, 1991 and December 31, 2000. These Medicare records include claims from institutions (skilled nursing facility, home health agency, hospice); hospitals (inpatient and outpatient); physicians and other related providers (carrier files); as well as durable medical equipment claims. The linkage methodology is explained in detail elsewhere (National Center for Health Statistics, Office of Analysis and Epidemiology 2010). The mean number of years of claims records available for these participants was 5.7. Seventy-five percent of the cases had at least 2.5 years of claims data.

Table 1 presents the sociodemographic and health profile of the study participants with linked Medicare records compared with the original sample of NHEFS participants born in 1935 or earlier and with those participants who were not linked to Medicare records. The demographic composition of participants who were alive and enrolled in Medicare 20 years after the baseline differs from the original sample of NHEFS participants born in 1935 or earlier: the sample linked with Medicare records was younger and a higher percent was female. Baseline health characteristics of the linked sample, however, were similar to the full sample. The unlinked sample was somewhat younger than the linked sample, but it had similar sex and education distributions.

IDENTIFYING CHRONIC CONDITIONS

A study participant was identified as having a chronic condition in the NHEFS data if either of the following two conditions were met:

1. He/she responded positively to the question(s) related to this chronic condition during a baseline or follow-up interview or examination. In addition to the data from the baseline and all available follow-up interviews, the results of the baseline physical examination and X-rays were used to identify arthritis. Dementia was identified based on responses to the Mental Status Questionnaire administered during 1982–1984 and dementia-related questions asked during other follow-up interviews. The data from the baseline and all available follow-up interviews were used to identify diabetes and IHD. Finally, COPD was identified using the data from the baseline and 1982–1984 follow-up interviews. The date of the diagnosis was recorded as supplied by the participant; otherwise it was assigned to be the date of the interview.
2. An appropriate International Classification of Diseases (ICD) code, matching the CCW algorithm codes, was found in the medical records in the NHEFS for an overnight facility stay or a death record (ICD-9-CM for diagnoses and ICD-9 or ICD-10 for deaths). The date of the diagnosis was assigned as follows: for a nursing home stay—the date of admission to the facility; for a hospital stay—the date of discharge; for death records—1 day before the date of death.

The five conditions studied were reported, on average, about 1.5 times over all the sources (e.g., baseline, follow-up interviews, facility records),

Table 1: Baseline characteristics among older NHEFS participants

	NHEFS Participants Born in 1935 or Earlier		NHEFS Participants Born in 1935 or Earlier Who Survived to January 1, 1991 and Who Were Linked with 1991–2000 Medicare Claim Records*		NHEFS Participants Born in 1935 or Earlier Who Survived to January 1, 1991 and Who Were NOT Linked with 1991–2000 Medicare Claim Records*	
	N	Percent of All Participants (Unweighted)	N	Percent of All Participants (Unweighted)	N	Percent of All Participants (Unweighted)
All	9,923	100.0	4,846	100.0	968	100.0
Birth cohort						
1896–1905	2,845	28.7	626	12.9	104	10.7
1906–1915	2,268	22.9	1,009	20.8	151	15.6
1916–1925	2,165	21.8	1,394	28.8	281	29.0
1926–1935	2,644	26.6	1,817	37.5	432	44.6
Sex						
Men	4,354	43.9	1,757	36.3	350	36.2
Women	5,569	56.1	3,089	63.7	618	63.8
Race [†]						
White	8,318	83.8	4,218	87.0	788	81.4
Black	1,521	15.3	587	12.1	162	16.7

<i>Education at baseline[†]</i>					
Less than 12 years	5,211	52.5	2,068	42.7	417
High school	2,782	28.0	1,675	34.6	319
More than 12 years	1,848	18.6	1,069	22.1	224
<i>BMI status at baseline[‡]</i>					
Underweight	305	3.1	106	2.2	24
Healthy weight	4,243	42.8	2145	44.3	437
Overweight	3,555	35.8	1756	36.2	345
Obese	1,816	18.3	838	17.3	161
<i>Smoking status at baseline[§]</i>					
Never	1,874	37.8	1,077	41.7	187
Former	1,098	22.1	565	21.9	104
Current	1,991	40.1	940	36.4	223

*Participants had Medicare A and B at some point between 1991 and 2000.

[†]Percents may not sum to 100% because of missing data.

[‡]Underweight is defined as BMI < 18.5. Healthy weight is 18.5 ≤ BMI < 25. Overweight is 25 ≤ BMI < 30. Obesity is defined as BMI ≥ 30.

[§]Only a subsample of the NHEFS baseline cohort were asked questions on smoking. Percents are calculated from the subsample. BMI, body mass index; NHEFS, NHANES I Epidemiologic Follow-up Study.

excluding the death records. Arthritis was most likely to be mentioned first at the baseline (49 percent of identified cases), with COPD being mentioned at baseline by 37 percent of identified cases and diabetes by 21 percent. In contrast, IHD and dementia were more likely to be first mentioned later, as the participants aged: IHD was first reported at the 1992 interview by 33 percent of identified cases, and dementia was first mentioned in 1992 by 63 percent of the identified cases.

A study participant was identified as having a chronic condition in the linked Medicare data if the Medicare claims records during the study years met the criteria specified in the CCW algorithm. The CCW algorithm is based on claim diagnosis and procedure codes, specific criteria for reference time periods, and the number and type of qualifying claims and other criteria as defined in CMS's CCW. The reference period is the number of years before and including a year of interest during which the CCW algorithm criteria must be met to identify a chronic condition. For example, the CCW identifies a case of Alzheimer's disease when there is a ICD-9-CM diagnosis code of 331.0 anywhere on the beneficiary's submitted claims, including inpatient, skilled nursing facility and other claims during a 3-year period (i.e., the requested year of interest and the previous 2 years). As described in the CCW manual, a beneficiary can be identified with a chronic condition in 1 year but not in the next, depending on the reference period for the particular condition and whether the relevant claims occur in consecutive years. These definitions were developed by CMS in collaboration with the Research Data Access Center and the contractor who developed the CCW (CCW 2007).

In this analysis, we consider a "CCW chronic condition" to be identified by the date of the first relevant claim, according to the CCW algorithm, in the linked NHEFS-Medicare data, beginning with the earliest claims available from 1991. The claims history for study participants begins at different points from 1991 to 2000, depending on the age of the participant. Thus, we are using the same definitions as the CCW algorithm, but instead of "looking back" from an arbitrary date (e.g., the requested year of CCW data) we are looking forward from the first available claim. The CCW algorithm definitions used for the five selected conditions are shown in Appendix SA2.

STUDY DESIGN

Analyses were based on the 4,846 participants who were enrolled in Medicare claims at some point during the period 1991–2000. For those study

participants who were identified with a chronic condition based on the NHEFS data and who also had Medicare claims records, we calculated the percent of cases identified by claims during the CCW reference period and also the time needed by the CCW algorithm to identify the preexisting condition. Only beneficiaries identified with the chronic condition by NHEFS before the start of the claims history were eligible for this analysis. We determined the date of the first Medicare claim of any kind and the date of the first Medicare claim relevant to the chronic condition according to the CCW algorithm. The time period between these two dates was considered as the time needed for the CCW algorithm to identify a preexisting condition. In a second analysis, we identified cases with chronic conditions within the CCW reference period and calculated the percent of cases that could be considered preexisting because they were previously identified by NHEFS data. The remaining cases (CCW-identified cases which were not previously found in NHEFS) can be considered newly diagnosed cases.

RESULTS

How Many Years of Claims History Are Necessary to Find Preexisting Conditions?

To estimate the number of years of claims data needed to identify preexisting conditions, we selected those study participants with linked Medicare claims records at some point during 1991–2000 who were identified in NHEFS with selected chronic conditions before the first available Medicare claim. Table 2 presents the percent of participants with preexisting NHEFS conditions who were identified as having a condition using Medicare claims with the CCW algorithm, the average number of years of claims data needed for identification, and the percent of preexisting cases identified during the CCW reference period and during the entire study period.

Using the CCW algorithm with all the available years of claims data, ranging from an average of 3.9 years for dementia cases to 5.8 years for arthritis cases, we found 77 percent of the preexisting diabetes cases and 37 percent of the preexisting cases of arthritis. However, when the duration of the claims history is limited to the CCW reference period (2 years for diabetes and arthritis), 69 percent of preexisting cases of diabetes and only 17 percent of preexisting cases for arthritis were identified. For each condition, the cases that were not identified by the CCW algorithm had on average fewer years of claims available than those that were identified; however, the number of years of available claims was more than the CCW reference period (see Appendix

Table 2: Preexisting NHEFS cases identified by Medicare claims using CCW algorithm, 1991–2000

<i>Condition</i>	<i>Number of Preexisting NHEFS Cases*</i>	<i>Average Years of Claims History Available for Those with Condition</i>	<i>Percent of Preexisting NHEFS Cases Identified by Claims during Entire Study Period (1991–2000)</i>	<i>Percent of Preexisting NHEFS Cases Identified during CCW Reference Period</i>	<i>Average Time (years) between Start of Claims History and First Related Claim</i>	<i>CCW Reference Period (years)</i>
Diabetes	690	5.2	76.5	68.6	0.7	2
IHD	1,071	5.3	79.9	62.9	1.3	2
Dementia	262	3.9	52.3	39.3	1.8	3
COPD	831	5.5	48.4	24.4	2.1	1
Arthritis	2,874	5.8	36.6	17.0	3.0	2

Notes. Estimates are based on 4,846 study participants from the baseline NHEFS who were born in 1935 or earlier and who could be linked to Medicare A & B records at some point from 1991 to 2000.

*Cases identified by NHEFS before beginning of the Medicare claims history.

CCW, Chronic Condition Data Warehouse; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; NHEFS, NHANES I Epidemiologic Follow-up Study.

SA3). Overall, using a longer claims history than the CCW reference period enabled us to identify a higher percent of preexisting cases. Even so, only about half of dementia and COPD cases and slightly more than one-third of arthritis cases were found.

Distinguishing Preexisting Conditions from Newly Diagnosed Conditions When Using Claims Data

Depending on the nature of the research, it might be important to distinguish preexisting conditions from newly diagnosed conditions in Medicare claims data. For this analysis we chose study participants with linked Medicare claims records who were identified with selected chronic conditions by the CCW algorithm. A condition was considered preexisting if it was identified by NHEFS before the participant's first available claim; otherwise it was considered newly diagnosed by CCW.

Table 3 shows that the majority of cases of diabetes and arthritis identified by the CCW in the CCW reference period were preexisting cases (67 percent and 84 percent, respectively), while about one-half of IHD and COPD cases were preexisting. Only about 30 percent of dementia cases identified by the CCW were previously diagnosed in NHEFS.

Table 3: Newly diagnosed and preexisting cases identified by the Medicare claims using the CCW algorithm, 1991–2000

<i>Condition</i>	<i>Total Number of Cases Identified during CCW Reference Period</i>	<i>Number of Claims-Identified Cases Previously Identified by NHEFS</i>	<i>Percent of Claims- Identified Cases That Were Preexisting</i>	<i>CCW Reference Period (years)</i>
Diabetes	710	473	66.6	2
IHD	1336	674	50.4	2
Dementia	347	103	29.7	3
COPD	400	203	50.8	1
Arthritis	585	489	83.6	2

Notes. Estimates are based on 4,846 study participants from the baseline NHEFS who were born in 1935 or earlier and who could be linked to Medicare A & B records at some point from 1991 to 2000.

CCW, Chronic Condition Data Warehouse; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; NHEFS, NHANES I Epidemiologic Follow-up Study.

Looking beyond the CCW reference period, using all the available data from 1991 to 2000, the pattern is the same, although, as would be expected, the percent of new cases found increases. On average, 3.5 years of claims data were needed to find the first claim for new diabetes cases and 4.0 years were needed for arthritis (data not shown).

DISCUSSION

The NHEFS-Medicare linked dataset used in this paper allows us to evaluate the identification of chronic conditions by the CCW algorithm. We examined the CCW algorithm from two perspectives: (1) ability to identify preexisting conditions; and (2) ability to distinguish between preexisting and newly diagnosed conditions. The CCW has the potential to facilitate health services and epidemiological research on chronic conditions. However, researchers should be aware of its limitations to avoid drawing incorrect conclusions about the population of beneficiaries being analyzed. In particular, users of Medicare data with chronic conditions identified by the CCW algorithm should consider how the definitions (specifically the reference or look-back period) may affect the detection of chronic conditions.

The survey data linked to Medicare claims give us an opportunity to estimate the length of claims history needed for the CCW algorithm to

identify a preexisting condition. For example, if the study subject was diagnosed with a chronic condition in 1985 and his/her available Medicare claims history starts in January 1992, what is the probability that the CCW algorithm will identify a claim for this condition in the first year of enrollment? In the second year? The answers to these questions depend on the severity of the condition and the need to utilize health care services. Thus, a person with arthritis may not have claims related to arthritis for a long time, while a person with IHD is more likely to visit a doctor regularly and therefore will more quickly generate a claim with the corresponding ICD code. For three of the conditions (diabetes, IHD, and dementia), a relevant claim was identified on average within the CCW reference period. Yet there was variation in the proportion of preexisting cases that were identified. The CCW algorithm and reference period identified a higher proportion of preexisting cases of diabetes (69 percent) and IHD (63 percent) compared with the three other conditions. (See Table 2.)

The intersection of disease etiology and health care utilization influences the interpretation of the chronic conditions identified by the CCW algorithm. If the CCW reference period is strictly followed, we have shown that approximately 84 percent of arthritis cases identified were preexisting compared with approximately 50 percent of IHD and COPD cases. (See Table 3.) Thus, while most of the arthritis cases identified by the CCW algorithm in the reference period were preexisting cases, only a small proportion of the beneficiaries who actually had previously diagnosed arthritis (about 17 percent shown in Table 2) were captured by the algorithm.

Conclusions drawn from our study should keep in mind several limitations in data and methods, which may affect the ability of the CCW algorithm to identify preexisting chronic conditions. The sample sizes in the linked dataset were not sufficient to do a detailed analysis by age, sex, or race. It is possible that the identification of chronic conditions varies by demographic characteristics. Because only about half of the original NHEFS participants born in 1935 or earlier survived until 1991 and could be linked to Medicare claims, our analysis was conducted on a subsample that is not fully representative of the original NHEFS participants. In addition, the NHEFS data are from a longitudinal survey and include information from different points in time; not all respondents have the same amount of information.

Even though NHEFS covers approximately 20 years, the data may not be complete. NHEFS diagnoses are based on several sources: interview responses, examination results, medical records (hospital discharges and nursing

home admission records), and death records. Each source has its own limitations. For example, the self-report of conditions could contain errors. In addition, the study definition for some chronic conditions changed between waves of the follow-up study, in some cases because the medical definition of the condition changed, and in other cases because of changes in the questionnaire. For example, during the study period, the definition of diabetes changed and the previously recommended oral glucose tolerance test was replaced with a recommendation that the diagnosis of diabetes mellitus be based on two fasting plasma glucose levels of 126 mg/dl or higher (Mayfield 1998). New tests for dementia continue to be developed leading to advances in understanding and detection (Mani et al. 1999; Cummings 2000). The follow-up interviews did not ask about conditions such as COPD or IHD, and in these cases, the study data rely only on medical records for overnight facility stays. In addition, the hospital and nursing home records provided ICD-9-CM codes for the various conditions; finding a chronic condition using these records depends on the accuracy of the clinical coding and may be less precise for comorbid conditions than for primary diagnoses (Kern et al. 2006). Similarly, mortality records contain on average one to two comorbid conditions, in addition to the underlying cause of death, and they may not include all contributing chronic conditions among older deceased participants (Gorina and Lentzner 2008).

In contrast to NHEFS data, claims data (with the CCW algorithm) identify chronic conditions based only on utilization of health care for that particular condition, such as a doctor visit or hospitalization. Therefore, patients in remission or with conditions that do not require health care for some period of time may be missed by the claims data analysis (Joyce et al. 2005). Beneficiaries may underutilize health care due to limited access to care (e.g., in rural areas) and thus not generate claims for a particular chronic condition or may utilize services not covered by Medicare. Previously uninsured or under-insured new Medicare enrollees may utilize health care at a greater rate in the first few years after enrollment than before; thus, some conditions may be identified by the CCW algorithm at a greater rate in the first years after enrollment than in later years. Enrollees in this analysis with a claims history that starts at the time of enrollment in Medicare may have a greater chance of being diagnosed with chronic conditions than those whose early claims history is not available in the analytic sample. In addition, the accuracy of Medicare claims has to be taken into consideration (Taylor, Fillenbaum, and Ezell 2002; Losina et al. 2003; Handlon and Cleverley 2006). Finally, the CCW algorithm is applicable only to the claims of beneficiaries who were enrolled in fee-for-service Medicare Part

A and Part B and not enrolled in a Medicare HMO, and therefore may be subject to selection bias (Mello et al. 2003).

Our findings differ from the results of Katz et al. (1997), who find high levels of agreement for certain arthritic conditions between medical records and Medicare physician claims. The medical records in that study, however, were taken from visits to rheumatology specialists, where one would expect arthritis to be reported more accurately than it might be over the full range of providers, services, and facilities.

We also do not find the same high levels of agreement with claims data for dementia as found by Taylor, Fillenbaum, and Ezell (2002), who compared Medicare claims with an Alzheimer's disease registry. The identification of dementia in NHEFS is confounded by a variety of issues. Mortality selection plays a major role in who survives to be linked to Medicare records. Some cases of dementia may have been missed because the participants did not survive to the time period of the Medicare claims linkage. Because only 13 percent of our analytic sample was older than 65 years at baseline, it is not surprising that conditions especially associated with aging (e.g., dementia) may be harder to detect in longitudinal data. Dementia is difficult to diagnose in its early stages and may not be accurately reported in survey data or be the primary reason for health care utilization. Taylor, Fillenbaum, and Ezell (2002) do find that less severe cases of dementia are less likely to be captured by claims data.

The size of the Medicare beneficiary population and the high prevalence of chronic conditions among this population necessitates its continuing study to improve health and health care policy. The creation of the CCW offers potential for expanded use of Medicare claims data for analysis. Comparing conditions identified by the CCW with outside data, we have shown wide variation in the ability of the CCW algorithm to identify chronic conditions. The reference periods embedded in the CCW algorithm may not be sufficient for all analyses. Depending on the research question, users need to consider how the intersection of time since enrollment, specific condition, age of enrollee, number of years of claims available, and whether a condition is pre-existing or newly diagnosed, may affect a study's results and interpretations. For example, analyses using the CCW algorithm's identification of arthritis cases may underestimate the overall cost of treatment since the population captured in a given year misses many beneficiaries with previously diagnosed arthritis. Analyses using beneficiaries recently enrolled may miss cases where more years of data are needed to identify conditions. We encourage continued research and development on the CCW to refine this valuable resource.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors would like to thank James Lubitz for his help with the original design of this study and The National Center for Health Statistics Data Linkage Team for making the data available.

Disclosure: The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Disclaimer: None.

REFERENCES

- CCW. 2007. *Chronic Condition Data Warehouse User Manual, Version 2.0* [accessed on April 10, 2008]. CMS Research Data Assistant Center (ResDAC). Des Moines, IA: Iowa Foundation for Medical Care. Available at <http://www.resdac.umn.edu/Index.asp>
- Cohen, B. B., H. E. Barbano, C. S. Cox, et al. 1987. *Plan and Operation of the NHANES I Epidemiologic Follow-up Study, 1982–84. Vital and Health Statistics Report Series 1, No. 22.* Washington, DC: National Center for Health Statistics.
- Cox, C. S., M. E. Mussolino, S. T. Rothwell, M. A. Lane, C. D. Golden, J. H. Madans, and J. J. Feldman. 1997. *Plan and Operation of the NHANES I Epidemiologic Follow-up Study, 1992. Vital and Health Statistics Report Series 1, No. 35.* Washington, DC: National Center for Health Statistics.
- Cox, C. S., S. T. Rothwell, J. H. Madans, et al. 1992. *Plan and Operation of the NHANES I Epidemiologic Follow-up Study, 1987. Vital and Health Statistics Report Series 1, No. 27.* Washington, DC: National Center for Health Statistics.
- Cummings, J. L. 2000. "New Tests for Dementia." *Neurology* 55: 1601–2.
- Finucane, F. F., V. M. Freid, J. H. Madans, et al. 1990. *Plan and Operation of the NHANES I Epidemiologic Follow-up Study, 1986. Vital and Health Statistics Report Series 1, No. 25.* Hyattsville, MD: National Center for Health Statistics.
- Foley, R. N., A. M. Murray, S. Li, C. A. Herzog, A. M. McBean, P. W. Eggers, and A. J. Collins. 2005. "Chronic Kidney Disease and the Risk for Cardiovascular Disease, Renal Replacement, and Death in the United States Medicare Population, 1998 to 1999." *Journal of the American Society of Nephrology* 16: 489–95.
- Gorina, Y., and H. Lentzner. 2008. "Multiple Causes of Death in Old Age." *NCHS Aging Trends* 9.
- Handlon, L., and W. Cleverley. 2006. "Is Your Claim Editor Really Working?" *Healthcare Financial Management* 60: 80–4.
- Harrold, L., K. G. Saag, R. A. Yood, T. R. Mikuls, S. E. Andrade, H. Fouayzi, J. Davis, K. A. Chang, M. A. Raebel, A. V. Worley, and R. Platt. 2007. "Validity of Gout Diagnoses in Administrative Data." *Arthritis & Rheumatism (Arthritis Care & Research)* 57 (1): 103–8.

- Herbert, P. L., L. S. Geiss, E. F. Tierney, M. M. Engelgau, B. P. Yawn, and A. M. McBean. 1999. "Identifying Persons with Diabetes Using Medicare Claims Data." *American Journal of Medical Quality* 14 (6): 270-7.
- Joyce, G. F., E. B. Keeler, B. Shang, and D. P. Goldman. 2005. "The Lifetime Burden of Chronic Disease among the Elderly." *Health Affairs* 24 (suppl 2): W5R18-29.
- Katz, J. N., J. Barrett, M. H. Liang, A. M. Bacon, H. Kaplan, R. I. Kieval, S. M. Lindsey, W. N. Roberts, D. M. Sheff, R. T. Spencer, A. L. Weaver, and J. A. Baron. 1997. "Sensitivity and Positive Predictive Value of Medicare Part B Physician Claims for Rheumatologic Diagnoses and Procedures." *Arthritis and Rheumatism* 40 (9): 1594-600.
- Kern, E. F., M. Maney, D. R. Miller, T. Chin-Lin, A. Tiwari, M. Rajan, D. Aron, and L. Pogach. 2006. "Failure of ICD-9-CM Codes to Identify Patients with Comorbid Chronic Kidney Disease in Diabetes." *Health Services Research* 41 (2): 564-80.
- Klabunde, C. N., L. C. Harlan, and J. L. Warren. 2006. "Data Sources for Measuring Comorbidity: A Comparison of Hospital Records and Medicare Claims for Cancer Patients." *Medical Care* 44 (10): 921-8.
- Losina, E., J. Barrett, J. A. Baron, and J. N. Katz. 2003. "Accuracy of Medicare Claims Data for Rheumatologic Diagnoses in Total Hip Replacement Recipients." *Journal of Clinical Epidemiology* 56: 515-9.
- Machlin, S., J. Cohen, and K. Beauregard. 2008. *Health Care Expenses for Adults with Chronic Conditions, 2005. Statistical Brief #203*. Rockville, MD: Agency for Healthcare Research and Quality. Available at http://www.meps.ahrq.gov/mepsweb/data_files/publications/st203/stat203.pdf
- Madans, J. H., J. C. Kleinman, C. S. Cox, H. E. Barbano, J. J. Feldman, B. Cohen, F. F. Finucane, and J. Cornoni-Huntley. 1986. "Ten Years after NHANES I: Report of Initial Follow-up, 1982-84." *Public Health Reports* 101 (5): 465-73.
- Mani, S., M. B. Dick, M. J. Pazzani, E. L. Teng, D. Kempler, and I. M. Taussig. 1999. "Refinement of Neuro-Psychological Tests for Dementia Screening in a Cross Cultural Population Using Machine Learning." In *Lecture Notes in Artificial Intelligence: Artificial Intelligence in Medicine*, edited by Horn, W., Y. Shahar: G. Lindberg: S. Andreassen, and J. Wyatt. pp. 326-35. Aalborg, Denmark: AIMDM.
- Mayfield, J. 1998. "Diagnosis and Classification of Diabetes Mellitus: New Criteria." *American Family Physician* 58 (6): 1355-62, 1369-70.
- Mello, M. M., S. C. Stearns, E. C. Norton, and T. C. Ricketts. 2003. "Understanding Biased Selection in Medicare HMOs." *Health Services Research* 38 (3): 961-92.
- National Center for Health Statistics, Office of Analysis and Epidemiology. 2010. "Linkages between Survey Data from the National Center for Health Statistics and Medicare Program Data from the Centers for Medicare and Medicaid Services" [accessed on October 29, 2010]. Available at http://www.cdc.gov/nchs/data/datalinkage/cms_medicare_methods_report_final.pdf
- Østbye, T., D. H. Taylor, E. C. Clipp, L. Van Scoyoc, and B. L. Plassman. 2008. "Identification of Dementia: Agreement among National Survey Data,

- Medicare Claims, and Death Certificates." *Health Services Research* 43 (1, Part 1): 313–26.
- Rector, T. S., S. L. Wickstrom, M. Shah, N. T. Greenlee, P. Rheault, J. Rogowski, V. Freedman, J. Adams, and J. J. Escarce. 2004. "Specificity and Sensitivity of Claims-Based Algorithms for Identifying Members of Medicare+Choice Health Plans That Have Chronic Medical Conditions." *Health Services Research* 39 (6, Part I): 1839–57.
- Taylor, D. H., G. G. Fillenbaum, and M. E. Ezell. 2002. "The Accuracy of Medicare Claims Data in Identifying Alzheimer's Disease." *Journal of Clinical Epidemiology* 55 (9): 929–37.
- Vogeli, C., A. E. Shields, T. A. Lee, T. B. Gibson, W. D. Marder, K. B. Weiss, and D. Blumenthal. 2007. "Multiple Chronic Conditions: Prevalence, Health Consequences, and Implications for Quality, Care Management, and Costs." *Journal of General Internal Medicine* 22 (suppl 3): 391–5.

SUPPORTING INFORMATION

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

- Appendix SA1: Author Matrix.
- Appendix SA2: CCW Algorithm for Five Conditions.
- Appendix SA3: Means Years of Claims Available for Preexisting NHEFS Cases from 1991 to 2000.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.