



# HHS Public Access

Author manuscript

*Biol Blood Marrow Transplant*. Author manuscript; available in PMC 2017 October 01.

Published in final edited form as:

*Biol Blood Marrow Transplant*. 2016 October ; 22(10): 1738–1746. doi:10.1016/j.bbmt.2016.05.005.

## Administrative Claims Data for Economic Analyses in Hematopoietic Cell Transplantation: Challenges and Opportunities

Jaime M. Preussler, MS<sup>\*</sup>,<sup>1</sup>, Lih-Wen Mau, PhD, MPH<sup>\*</sup>,<sup>1</sup>, Navneet S Majhail, MD, MS<sup>2</sup>, Christa L. Meyer, MS<sup>1</sup>, Ellen Denzen, MS<sup>1</sup>, Kristen C. Edsall, RN, BSN, MSN<sup>1</sup>, Stephanie H. Farnia, MPH<sup>1</sup>, Alicia Silver, MPP<sup>1</sup>, Wael Saber, MD, MS<sup>3</sup>, Linda J. Burns, MD<sup>1</sup>, and David J. Vanness, PhD<sup>4</sup>

<sup>1</sup>National Marrow Donor Program/Be The Match, Minneapolis, MN

<sup>2</sup>Blood & Marrow Transplant Program, Cleveland Clinic, Cleveland, OH

<sup>3</sup>Center for International Blood and Marrow Transplant Research, Milwaukee, WI

<sup>4</sup>University of Wisconsin-Madison, Madison, WI

### Abstract

There is an increasing need for the development of approaches to measure quality, costs and resource utilization patterns among allogeneic hematopoietic cell transplant (HCT) patients. Administrative claims data provide an opportunity to examine service utilization and costs, particularly from the payer's perspective. However, because administrative claims data are primarily designed for reimbursement purposes, challenges arise when using it for research. We use a case study with data derived from the 2007–2011 Truven Health MarketScan Research database to discuss opportunities and challenges for the use of administrative claims data to examine the costs and service utilization of allogeneic HCT and chemotherapy alone for patients with acute myeloid leukemia (AML). Starting with a cohort of 29,915 potentially eligible patients with a diagnosis of AML, we were able to identify 211 patients treated with HCT and 774 treated with chemotherapy only where we were sufficiently confident of the diagnosis and treatment path to allow analysis.

Administrative claims data provide an avenue to meet the need for health care costs, resource utilization, and outcome information. However, when using these data, a balance between clinical knowledge and applied methods is critical to identifying a valid study cohort and accurate measures of costs and resource utilization.

---

**Address for Correspondence:** Lih-Wen Mau, PhD, MPH, National Marrow Donor Program/Be The Match, 500 N. 5<sup>th</sup> Street, Minneapolis, MN 55401, Phone: 763-406-8497, lmau@nmdp.org.

<sup>\*</sup>Co-first author

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Conflicts of Interest:** None of the authors has any financial conflicts of interest to report

## Keywords

Hematopoietic cell transplantation; administrative claims

---

## Introduction

As health care expenditures continue to rise, policy-makers and payers are paying closer attention to the so-called “value equation,” assessing whether the high costs of medical interventions such as hematopoietic cell transplantation (HCT) are justified by their outcomes.<sup>1,2</sup> Accurate and comprehensive analysis of costs requires the correct identification, measurement and valuation of resources used in the delivery of the intervention and management of sequelae events, including potential savings from avoiding other costly downstream events.<sup>3</sup>

While research on health outcomes presents its own set of well-understood challenges<sup>4–6</sup>, obtaining decision-relevant information about the costs associated with treatment in the United States (US) is especially problematic, in part because of the decentralized nature of the US health care finance system. Varying contractual arrangements among payers and providers mean that estimates of actual resource costs based on charges (gross or net) from any single provider or payer may lack generalizability. Administrative claims datasets representing the experiences of patients and providers are available (at a cost) from third party administrators. While such data can be useful for providing comprehensive and generalizable information for cost analysis, valid estimates of the effects of treatment on cost also require assessment of patient factors that might influence both treatment selection and outcomes, which may be difficult to obtain from claims data.

In this paper, we highlight several opportunities and challenges researchers can expect to encounter when using administrative claims datasets to assess medical costs related to HCT. To do so, we present a case study of how we constructed cohorts from administrative claims data representing the first year experience after diagnosis of acute myeloid leukemia (AML) for US patients aged 50–64 who received allogeneic HCT or chemotherapy only.

## Overview of Administrative Claims Data

Generally, when health care services are delivered in the US, providers receive reimbursement by submitting a claim to a payer. On each claim, providers must document the underlying medical reason for the services delivered during an encounter. To do so, they currently use one or more medical conditions coded under the International Statistical Classification of Diseases and Related Health Problems (ICD-9) lexicon.<sup>7</sup> The total number of diagnoses varies by administrative claims database. As of October 1, 2015, the US transitioned from the ICD-9 version of the system to ICD-10, a more granular classification<sup>8</sup>, and the number of codes increased from 17,000 for ICD-9 to 140,000 with ICD-10. Ideally, the list of ICD codes justifying each encounter would contain the primary reason for the encounter at the most specific level of detail available (e.g., ICD-9 205.01, Myeloid Leukemia Acute in Remission) in the first position, and in subsequent positions any other conditions or comorbidities present which may make the encounter more complex.

Each claim consists of one or more line items containing a code identifying the specific medical services delivered to the patient during an encounter. Services provided by a medical professional are identified by a procedure code according to the Current Procedural Terminology, Fourth Edition (CPT-4)<sup>9</sup> lexicon licensed by the American Medical Association, the Level II Healthcare Common Procedure Coding System (HCPCS);<sup>10</sup> or an ICD-9 or ICD-10 procedure code. Procedure codes may be “modified” by a sub-code identifying special circumstances, such as when multiple instances of the same service are provided at the same time, or when the service is split into professional and technical components and billed separately. In the hospital setting, each procedure code must be accompanied by a four-digit revenue code, which identifies the type and location of where services are delivered. Each dispensed pharmaceutical is identified by a National Drug Code (NDC),<sup>11</sup> a ten-digit number uniquely identifying the vendor, compound and packaging for the product, or a J code, part of the HCPCS Level II code set used to identify drugs.<sup>12</sup>

The use of administrative claims data to estimate patterns of utilization and cost is an example of the increasingly important use of “real world data” (i.e., data captured in the everyday course of delivery and reimbursement of health services, and not specifically in the context of research).<sup>13–15</sup> Administrative claims data frequently include both inpatient and outpatient claims, providing the opportunity to evaluate service utilization across care sites. Depending on the design of the study, administrative claims data may better reflect real-world populations and patterns of clinical practice compared to the highly selective and protocol-driven environment of clinical trials. Comprehensive data for the patient’s enrollment period allows for cross-sectional and longitudinal study designs, enabling the identification of cohorts and tracking of patients over time.

Decisions about provision, coverage, and utilization of health care interventions, particularly high-cost and resource intensive procedures like allogeneic HCT, are increasingly subject to scrutiny by payers.<sup>16</sup> From the payer’s perspective, cost is reflected in actual negotiated, paid claims, which differ substantially from billed charges due to negotiated discounts and denied claims.<sup>17–19</sup> Administrative claims data, therefore, may serve as the most relevant source of information on health care costs from the payer’s perspective.

## Administrative Claims Data Challenges in HCT

While administrative claims data may provide useful information for assessing health care costs, its use also presents several challenges, some of which are amplified in the context of HCT. Ideally, researchers would observe pre-disease health resource utilization patterns, disease onset and progression to a point in time where the treatment choice is in equipoise (a point in time at which the treatment could “go either way”; e.g., AML in first complete remission (CR1)). In contrast to prospective cohort studies, it is not possible to identify disease onset or even initial diagnosis with certainty in claims data, because those events may have occurred prior to the start date of available data and/or before the patient was enrolled in the captured insurance plan. Given the severity and progression of hematologic conditions such as AML, treatment occurs fairly quickly after diagnosis. Claims data may not allow us to know with complete certainty whether the condition being treated is *de novo* or arising as a secondary consequence of treatment for another disease. It can also be

difficult to identify the end point of an episode, whether it be remission, the end of a treatment course, change of insurance plan, or mortality. Survival in long-term remission may be right-censored, as patients may change insurance plans and therefore exit the data.

Furthermore, ICD codes used to identify the presence of a health condition are not recorded for research purposes but rather to justify payment for claims. Variations in contractual relationships between payers and providers may alter the incentives for complete coding of medical conditions. In the specific case of HCT, payers and providers have many different forms of contractual relationships for payment including fully or partially bundled services, and the specific details of those agreements are rarely reported, in order to preserve the proprietary nature of those agreements for competitive purposes.<sup>20</sup> Within bundled payment arrangements, the start and end time defining the covered episode of care may differ. Some payment options allow a percentage of charges for some components of the transplant; still others are discounted fee-for-service, or can be fee-for-service initially and then change to global payment at the time of transplant. Under bundled payment arrangements, incentives to identify and accurately code all service items may be less than under discounted or partial fee-for-service because reimbursement is no longer tied to individual services delivered. For studies attempting to assess the costs of HCT from the payer's perspective, this may not be a significant limitation – as long as the correct bundled payment is identified. Individual payers or providers fare better in determining the true amount paid, as they are privy to all payment arrangements made that are not easy to ascertain in claims databases. For studies attempting to measure cost reflecting the societal perspective by applying representative service average costs (for example by using Medicare-assigned RVUs or DRG relative weights) to identified service items, such costs may be underestimated for patients with bundled payment arrangements. Readers of and researchers performing cost studies using administrative claims must understand these imperfections, amplified in HCT due to the complex nature of reimbursement methodologies, and be careful when making decisions based on these data and studies.

Additional challenges stemming from the necessity of using codes designed for billing include: limitations on the number of conditions that can be coded for billing purposes; secondary diagnosis codes are not necessarily ordered by importance; codes may be based on the suspicion of disease in order to justify diagnostic testing; advances in technology and documentation could change or improve coding over time; physician documentation may not be comprehensive; and codes may be assigned by non-clinician coders, who rely on physician's documentation and who may lack training to resolve complex medical visits.

## **Six Criteria for Applying Administrative Claims Datasets to Cost Analysis**

In light of the potential limitations of using claims data, we begin by defining what would be optimal claims data research criteria – and then identify real world points of departure. To serve as a valid and generalizable source of data for cost analysis, administrative claims datasets ideally should satisfy six criteria (Figure 1): 1) the sample from which the cohort is derived should be representative of the population of interest; 2) each service provided should have an accurate attribution of which diagnosis or diagnoses led to the service; 3) claims should capture all services actually provided to a patient throughout the episode of

care related to the condition of interest; 4) the identification of individuals with the condition of interest receiving the treatments being compared/examined should be both accurate and yield large enough treatment groups to provide adequate statistical power to detect meaningful differences in costs; 5) estimates of cost are comprehensive with respect to type of service delivered and site of care and reflect actual outlays by payers; 6) any and all concurrent or previous diagnoses or other individual factors that make the delivery of the interventions being compared/examined more complex or costly or alter the likelihood that specific treatment be chosen (relative to other available treatments) should be available.

The relative degree to which these six criteria can be fulfilled essentially represents a tradeoff between “sensitivity” (the ability to identify correctly those who have the disease or receive the treatment of interest)<sup>21</sup> and “specificity” (the ability to identify correctly those who do not have the disease or not receive the treatment of interest).<sup>21</sup> Relaxing the various criteria to include more patients will result in a larger cohort, with increased statistical power and sensitivity. However, as the size of the cohort is increased, the probability of including individuals not in the target population of interest or whose available claims do not reflect the complete and accurate utilization of health care services related to treatment increases.

To illustrate the challenges in satisfying these six criteria, we present our experience using the Truven Health MarketScan Research Database (Truven) to assemble a valid cohort for assessing costs in the year after AML diagnosis for US patients age 50–64 receiving either allogeneic HCT or chemotherapy only.

### **Case study: Assessing Costs in the First Year after Diagnosis of AML for US Patients Age 50–64 Who Received Either Allogeneic HCT or Chemotherapy Alone**

AML is the most common indication for allogeneic HCT.<sup>22</sup> The incidence of AML is higher in older adults and the median age at AML diagnosis is 67 years, with more than 70% of individuals diagnosed at age 55 or greater.<sup>23</sup> Standard treatment for adult AML typically involves induction chemotherapy with a goal of inducing clinical remission, with the decision of consolidation with chemotherapy or HCT dependent upon cytogenetic and molecular risk features and other prognostic factors.

To date, studies assessing costs of HCT in the US have focused on the patient’s perspective<sup>24,25</sup> or on costs from single institutions.<sup>26–31</sup> Few studies have compared the costs of HCT to other treatments for AML, such as chemotherapy alone,<sup>32</sup> which limits the usefulness of their results for payers and policy makers who are interested in the utilization and cost impact of HCT for the treatment of AML. Administrative claims data are generally drawn from large populations, which is beneficial for the study of rare diseases or treatments. Nevertheless, the use of administrative claims data for assessing the costs and outcomes associated with HCT has been rather limited, and provides an opportunity for exploration.

The Truven dataset contains claims from approximately 100 payers and includes medical and prescription drug claims for more than 115 million unique patients with insurance

coverage provided by large employers, commercial health plans and government and public agencies, with broad geographic representation.<sup>33</sup> This represented about 25% of covered lives in the US in 2009.<sup>34</sup> Claims are linked to patient information across visits over time, allowing patients to be tracked longitudinally. The database contains data on diagnosis; service utilization, including specific service lines (inpatient, outpatient, and prescription drug); payments (including patient deductibles and co-payments); enrollment eligibility; and claims of individuals in fee-for-service plans and fully capitated or partially capitated plans.<sup>33</sup> The feasibility of using this database to identify patients who received HCT and their costs has been demonstrated.<sup>35</sup>

We used Truven data to construct a cohort of non-Medicare US individuals aged 50–64 years with a primary diagnosis of AML who received HCT or chemotherapy alone from January 2007 to December 2011. The goal was to examine the costs and resource utilization of all services a patient received during the first year after AML diagnosis. While establishing this cohort, we identified several challenges that are unique to the HCT population. Using this case study, we will highlight areas where administrative claims data are applied for studying costs of HCT and share issues faced in defining our study cohort and the approach we took to establish a cohort that best represents a population of HCT recipients and non-HCT patients with AML (Figure 2).

### Identification of the AML Cohort

To identify a cohort representative of the population of interest, we were faced with several challenges (Table 1). For example, we wanted to accurately attribute diagnoses to each service (criterion 2 above). The Truven inpatient admissions file lists up to 15 diagnoses per inpatient admission, and the outpatient file allows for two (2007–2008) to four (2009–2011) diagnoses to be listed per outpatient claim. Originally, to identify patients with AML, all diagnosis positions were used: primary (the first diagnosis position) and secondary (any diagnosis positions other than the first). However, this resulted in the inclusion of many patients who, upon closer investigation, did not appear to have received any of the expected therapies for AML – which led to suspicion of whether these were in fact patients with AML. Therefore, only patients with a primary AML diagnosis were included (N=29,915; Figure 2).

If a patient received a primary diagnosis of AML on an inpatient claim, we assumed that it was the AML diagnosis that was the main reason for the hospitalization. If a patient received a single outpatient visit of AML, a second outpatient claim with AML within 3 months was required, because a single outpatient claim may simply have represented evaluation for suspected AML which was later deemed negative. The criteria of one inpatient claim or two outpatient claims within three months identified the “diagnosis date” as the first date of service containing the primary AML diagnosis. This led to the exclusion of 13,044 patients with only one outpatient AML diagnosis within three months.

Another issue in assembling a representative cohort was to include only patients with de novo AML in order to prevent confounding cases of secondary AML related to prior diagnoses and therapies. This challenge also incorporated criterion 3 (complete capture of all claims throughout the episode of care). An episode of care may be defined as “a series of

temporally contiguous healthcare services related to the treatment of a given spell of illness or provided in response to a specific request by the patient or other relevant entity.”<sup>36</sup> A “clear window” of time preceding the diagnosis, in which the patient is eligible for insurance coverage but during which no AML-indicated services are provided, is required to identify the beginning of the episode. Increasing the period of time before which no claim of AML is present (such as 6 months or 1 year prior) increases the specificity of the cohort, but potentially excludes valid members of the cohort who were not eligible for the full observation period. Patients who were not enrolled 60 days prior to the diagnosis date were excluded because a minimum 2-month prior window was critical to determining both the likely onset of disease and comorbidities present at the time of diagnosis. Thus, patients had to have claims between 3/1/2007 and 12/31/2010 to allow for having claims 60 days prior to, and one year after diagnosis (n=12,088). This also allowed us to look at the age on the date of diagnosis, with the age of interest of the study being age 50–64 (n=4,365). Three patients who were not enrolled in an insurance plan on the diagnosis date were also excluded.

### Identifying Receipt of Treatments

Criterion 4 requires the accurate identification of treatments received by individuals in the cohort. Strict conditions for allocating patients to either HCT or chemotherapy alone increases the specificity of each group and improves the chances of an internally valid comparison of costs. However, strict treatment allocation conditions also results in a larger number of individuals who cannot be allocated to treatment with confidence and who are dropped from analysis, thus reducing sample size and statistical power.

Within our AML cohort, we found patients who did not have any claims for chemotherapy administration. For chemotherapy cohort eligibility, patients had to have received chemotherapy. For the HCT cohort, we determined that patients also had to have received chemotherapy for inclusion, as it would be very unusual for a patient with AML to proceed to transplantation without induction chemotherapy. (Cases like this for the HCT group likely represented patients who may have switched their health insurance prior to HCT.) Thus, patients without any chemotherapy administration (n=858) were excluded. Patients who received chemotherapy prior to AML diagnosis were also excluded (n=746), as it could mean that a patient received chemotherapy for another indication (such as another blood disease or solid tumor). If this were the case, the intent of chemotherapy treatment after AML diagnosis could not necessarily be attributed to AML, but rather to the prior indication or that it was secondary AML, which could contribute to bias.

When identifying HCT, there were questions about the timing of HCT in relation to the date of diagnosis. Three patients received HCT within 15 days of their AML diagnosis date. A transplant so close to diagnosis was considered clinically unlikely, and the decision was made to exclude these patients. Patients with HCT prior to AML diagnosis were also excluded (n=82).

Previous studies<sup>37</sup> varied in the way that chemotherapy was identified, including use of DRG codes, revenue codes, NDCs, and chemotherapy administration and diagnosis codes. Initially, we identified chemotherapy based on presence of chemotherapy administration or diagnosis codes on inpatient or outpatient claims. However, preliminary analysis identified

patients with relatively few outpatient chemotherapy claims and no inpatient chemotherapy claims, or they appeared to be receiving medication not related to AML treatment. J codes were used to identify medications on the date of the chemotherapy administration or diagnosis. The J code medications were then reviewed by HCT physicians on the research team to specify those used to treat AML. These medications were then used to classify chemotherapy expressly for AML. An algorithm was developed to determine receipt of AML chemotherapy using the codes identified or combinations of codes. The algorithm included: presence of an AML chemotherapy NDC or J code; DRGs specifying chemotherapy with acute leukemia; DRGs specifying acute leukemia coupled with diagnosis, procedure, or revenue codes (which identify the hospital cost center to which the service is allocated) for chemotherapy; and revenue codes coupled with an AML diagnosis (in the primary diagnosis position) (n=1,172).

Finally, eleven patients in the chemotherapy only group had ICD-9 codes for graft versus host disease (GVHD) (ICD-9 codes: 279.50–279.53) or complications of transplant (ICD-9 codes: 996.85 and 996.88). As these diagnoses can only occur as a result of transplant, these patients were excluded.

**Both cohorts**—Some patients in both cohorts did not have primary or secondary diagnoses of AML on their chemotherapy claims. Some diagnoses were clearly problematic, as they indicated the patient was being treated for other blood cancers not necessarily related to AML—e.g., chronic myeloid leukemia (CML) or acute lymphoid leukemia (ALL). This led to the exclusion of patients if they had other blood disease on chemotherapy initiation date, an exclusion of 6 patients in the HCT group, and 46 in the chemotherapy group. In other instances, claims were for other diagnoses (e.g., nausea, chemotherapy toxicity or infections). To investigate this, histograms were created, illustrating the distribution of chemotherapy claims for patients with either AML or neither AML nor other blood diseases on their chemotherapy initiation date (the first chemotherapy claim after diagnosis of AML - presumably an indication of the intent-to-treat for AML). Examination of the percentages of the chemotherapy claims by primary diagnosis on chemotherapy initiation date resulted in the decision to include patient if the chemotherapy initiation claim was for AML, and more than half of their chemotherapy claims were for AML. If >75% of the chemotherapy claims were for AML and the chemotherapy initiation claim was for “neither AML nor other blood diseases”, the patient was included (resulting in 23 patients excluded from the HCT group, and 123 in the chemotherapy group).

### Enrollment Status and Cost Estimation

To accurately reflect costs from the payer’s perspective, the fifth criterion requires that costs reported in the claims dataset represent actual payments for all categories of health care utilization, net of discounts and the patient’s share of expenses. Our first challenge in meeting this criterion involved the presence of fully- and partially-capitated health insurance plans in Truven. In capitated payment arrangements, payment is made on a per-patient basis, and is not usually tied directly to individual services provided. This can limit the incentive to code all diagnoses properly, or the ability to determine the true cost of the care provided.<sup>33</sup> There was concern, based on previous studies<sup>38,39</sup> that this would induce bias if insurance



type was also related to overall cost. Investigation of the total costs and number of claims per patient suggested differences between capitated and non-capitated plans. This resulted in the exclusion of patients (n= 212) who had a plan type of health maintenance organization (HMO) with at least one claim with a capitated service, or a plan type of point of service with capitation (POS), and was consistent with other studies.<sup>40-43</sup>

### Addressing Treatment Selection Bias

Criterion 6 for valid claims data cost analyses requires accurate accounting of patient characteristics, including pre-existing comorbidities that may affect both the propensity to select one of the treatments being compared as well as subsequent costs. Not controlling for these characteristics could result in treatment selection bias substantially affecting the estimate of comparative costs. Under Andersen's Behavioral Model,<sup>44</sup> characteristics affecting the utilization of health services can be categorized as predisposing, enabling or need-related. Among the predisposing characteristics most likely to influence choice of treatment for AML are basic socio-demographic characteristics and physician and patient preferences. Unfortunately, the set of such characteristics available in claims data analysis is limited, though enrollment data, as part of administrative claims data generally has information on the patient's age, gender and other demographics. Enabling characteristics include the level of insurance coverage, available financial resources and social support. It is important to note that, by definition, patients in our dataset are those who have insurance coverage and therefore have cost estimates that are not generalizable to the uninsured population. Need-related factors include disease risk as well as the presence or absence of comorbidities that are likely to impact therapy decision making.

Additionally, lead-time bias can materially inflate any survival advantage associated with HCT. Patients undergoing HCT must live long enough to get to HCT. One limitation in using administrative claims data is that we cannot be sure that some of the patients in the chemotherapy-only group were in fact being considered for HCT but, due to chemotherapy-related complications or disease relapse, never made it to HCT. Sensitivity analyses may help provide assurances that the final results are insensitive to such a concern but cannot fully address the possibility that differences in cost result from inclusion of a disproportionate number of patients with early death in the chemotherapy only cohort. Ideally, this could be addressed by having survival data for all patients, allowing correction for time to transplant, but survival data are not available in the Truven data.

### Discussion

As with efficacy or effectiveness, the most internally valid design for estimating the difference in costs between treatment alternatives would be a randomized controlled trial (RCT). However, economic analysis has rarely been performed alongside RCTs for HCT (with some exceptions regarding upcoming clinical trials<sup>45</sup>), leaving an important gap in actionable information. Administrative claims data can fill that gap, but with important caveats. As in any research, analysis must be conducted and interpreted judiciously and assumptions well documented.

Our case study serves to highlight the use of administrative claims data for use in HCT research. In spite of the challenges identified, administrative claims data provide useful information on patterns of care utilization and costs, and has been used successfully in previous cancer-related studies.<sup>46–51</sup> Administrative claims analyses may offer important advantages over cost studies conducted alongside RCTs in terms of generalizability. In highly controlled efficacy RCTs, it is often desirable to standardize the provision of ancillary care or to restrict inclusion of individuals with comorbidities in order to more efficiently estimate an internally valid treatment effect. However, such controls may make cost data collected alongside RCTs less relevant to real world clinical practice. A substantial proportion of “protocol-driven” healthcare utilization may be used in order to obtain detailed and precise measurements of health outcomes.<sup>3</sup> Patient participation in RCTs is voluntary, and the so-called “healthy volunteer” bias (one may substitute “motivated” or “enabled” for healthy) may exist. Exclusion of patients with specific comorbidities may result in biased estimates of comparative health care utilization and cost in the target population if comorbidities interact with treatment for the condition of interest. Thus, the use of administrative claims versus collecting cost data alongside RCTs involves tradeoffs of internal and external validity.

The process of cohort definition in administrative claims data analyses also involves tradeoffs – between sensitivity and specificity of criteria for inclusion. In cases with relatively rare disease and/or multiple treatment options, highly specific selection criteria can result in a small sample size and leave the study underpowered. Increasing the sensitivity by including more cases with unclear coding can increase power, but risks bias. In general, we tended to err toward specificity. For example, the impact of the inclusion/exclusion criteria occasionally resulted in the exclusion of large numbers of patients. Other researchers may wish to investigate a formal analysis of the bias-variance tradeoff.

It is not possible to evaluate differences in costs or dosage between specific therapeutic regimens using administrative claims data. High cost drugs are covered under the medical benefit during an inpatient stay, but are generally covered under the “specialty pharmacy” benefit in the outpatient setting and would not be included in administrative claims data, thus the data is likely missing some portion of drug claims. In our study, we did not expect all patients to be treated with the same non-HCT regimens, and within HCT varying platforms of therapy and donor sources exist that can impact cost. However, this situation aligns with therapeutic decisions made throughout the course of a patient’s illness.

The transition to ICD-10, particularly the additional details this coding provides compared to ICD-9, may provide additional challenges. For example, in ICD-9, code 205.00: AML, without mention of having achieved remission, translates into 5 codes in ICD-10. Additionally, future research, especially longitudinal studies, will have to take into account changes in codes during the transition period from ICD-9 to ICD-10 in 2015.

Administrative data is currently a useful real-world source of data to identify cost and utilization from the payer’s perspective. Future opportunities to enrich administrative data by linking to medical records or registry data could further improve the usefulness of this data source, and could allow for validation of claims coding. Such data provide an opportunity to

obtain detailed information, such as health outcomes and patient characteristics that may drive heterogeneity in response to treatment. Previous studies have successfully linked registry and administrative claims data,<sup>52</sup> though studies verifying administrative claims data by comparing data to registry data show widely varying levels of agreement; registries showed signs of incomplete coding as well. The Center for International Blood and Marrow Registry (CIBMTR) collects data on all allogeneic HCTs performed in the U.S, which provides an opportunity to link registry/outcomes data with administrative claims data to provide information to help improve treatment and outcomes for patients, and for cost-effectiveness research.

## Conclusion

Scrutiny for decision making in health care is likely to continue. There will be a continued need for cost and utilization data reflecting real-world clinical experiences. Administrative claims data can fill this need, but it is important to recognize that with increased relevance comes the tradeoff of potentially reduced internal validity. Our study suggests that with care, it is feasible to construct cohorts for economic analysis of HCT and related treatments for hematological malignancies using administrative claims data. Perhaps our single most valuable piece of advice given the variety of challenges we faced is that it is absolutely necessary to have a collaborative multidisciplinary team of researchers possessing clinical, methodological and applied claims data analysis expertise. Other researchers using administrative data for assessing the costs of different alternatives for the treatment of blood diseases, or performing economic analysis of HCT, who are likely to face similar challenges, can potentially benefit from our experience, as can those assessing the validity of published cost analyses using such data.

## Acknowledgments

Our sincere thanks to Mary Horowitz, MD, MS, Chief Scientific Director of the Center for International Blood and Marrow Transplant Research, and Jackie Foster, MPH, RN, OCN, Patient Education Specialist, National Marrow Donor Program/Be The Match, for providing critical review of the draft manuscript.

**Disclosure:** CIBMTR<sup>®</sup> (Center for International Blood and Marrow Transplant Research<sup>®</sup>) is a research collaboration between the National Marrow Donor Program<sup>®</sup>/Be The Match<sup>®</sup> and Medical College of Wisconsin. The CIBMTR is supported by Public Health Service Grant/Cooperative Agreement U24-CA76518 from the National Cancer Institute (NCI), the National Heart, Lung and Blood Institute (NHLBI) and the National Institute of Allergy and Infectious Diseases (NIAID); a Grant/Cooperative Agreement 5U01HL069294 from NHLBI and NCI; a contract HSSH234200637015C with Health Resources and Services Administration (HRSA/DHHS); two Grants N00014-06-1-0704 and N00014-08-1-0058 from the Office of Naval Research; and grants from AABB; Allos, Inc.; Amgen, Inc.; Anonymous donation to the Medical College of Wisconsin; Astellas Pharma US, Inc.; Be the Match Foundation; Biogen IDEC; BioMarin Pharmaceutical, Inc.; Biovitrum AB; BloodCenter of Wisconsin; Blue Cross and Blue Shield Association; Bone Marrow Foundation; Buchanan Family Foundation; CaridianBCT; Celgene Corporation; CellGenix, GmbH; Children's Leukemia Research Association; ClinImmune Labs; CTI Clinical Trial and Consulting Services; Eisai, Inc.; Genentech, Inc.; Genzyme Corporation; Histogenetics, Inc.; HKS Medical Information Systems; Hospira, Inc.; Kirin Brewery Co., Ltd.; The Leukemia & Lymphoma Society; Merck & Company; The Medical College of Wisconsin; Millennium Pharmaceuticals, Inc.; Miller Pharmacal Group; Milliman USA, Inc.; Miltenyi Biotec, Inc.; National Marrow Donor Program; Nature Publishing Group; Novartis Oncology; Oncology Nursing Society; Osiris Therapeutics, Inc.; Otsuka America Pharmaceutical, Inc.; Pall Life Sciences; Pfizer Inc; Schering Corporation; Sigma-Tau Pharmaceuticals; Soligenix, Inc.; StemCyte, Inc.; StemSoft Software, Inc.; Sysmex America, Inc.; THERAKOS, Inc.; Vidacare Corporation; ViraCor Laboratories; ViroPharma, Inc.; and Wellpoint, Inc. The views expressed in this article do not reflect the official policy or position of the National Institutes of Health, the Department of the Navy, the Department of Defense, or any other agency of the U.S. Government.

The Health Services Research program is supported in part by Health Resources and Services Administration Contract No. HHS234200637018C. The views expressed in this article do not reflect the official policy or position of the Health Resources and Services Administration or the National Marrow Donor Program/Be The Match®.

## References

1. Schnipper, LE., Davidson, NE., Wollins, DS., et al. [Accessed November 25, 2015] American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. <http://jco.ascopubs.org>
2. Porter ME. What Is Value in Health Care? *N Engl J Med.* 2010; 363(26):2477–2481. DOI: 10.1056/NEJMp1011024 [PubMed: 21142528]
3. Drummond, MF., Sculpher, MJ., Torrance, GW., O'Brien, BJ., Stoddart, GL. *Methods for the Economic Evaluation of Health Care Programmes.* Third. Oxford University Press; 2005.
4. Burgard SA, Chen PV. Challenges of health measurement in studies of health disparities. *Soc Sci Med.* 2014; 106:143–150. DOI: 10.1016/j.socscimed.2014.01.045 [PubMed: 24561776]
5. Krumholz HM. Real-world Imperative of Outcomes Research. *J Am Med Assoc.* 2011; 306(7):754–755. DOI: 10.1001/jama.2011.1170
6. Krischer JP, Gopal-Srivastava R, Groft SC, Eckstein DJ. Network for the RDCR. The Rare Diseases Clinical Research Network's Organization and Approach to Observational Research and Health Outcomes Research. *J Gen Intern Med.* 2014; 29(3):739–744. DOI: 10.1007/s11606-014-2894-x
7. [Accessed May 14, 2015] ICD - ICD-9-CM - International Classification of Diseases, Ninth Revision, Clinical Modification. <http://www.cdc.gov/nchs/icd/icd9cm.htm>
8. [Accessed June 24, 2015] ICD - ICD-10 - International Classification of Diseases, Tenth Revision. <http://www.cdc.gov/nchs/icd/icd10.htm>
9. [Accessed May 14, 2015] CPT Coding, Medical Billing and Insurance. <http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance.page?>
10. [Accessed July 15, 2015] Medicare C for, Baltimore MS 7500 SB, Usa M. HCPCS\_Coding\_Questions. [http://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCS\\_Coding\\_Questions.html](http://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCS_Coding_Questions.html). Published July 22, 2013
11. [Accessed June 24, 2015] Research C for DE and. Drug Approvals and Databases - National Drug Code Directory. <http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm>
12. Glossary: J- Codes. [Accessed July 15, 2015] Centers for Medicare and Medicaid Services. <http://www.cms.gov/apps/glossary/default.asp?Letter=J&Language=English>. Published May 14, 2006
13. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med.* 1997; 127(8 Pt 2):666–674. [PubMed: 9382378]
14. Riley GF. Administrative and claims records as sources of health care cost data. *Med Care.* 2009; 47(7 Suppl 1):S51–S55. DOI: 10.1097/MLR.0b013e31819c95aa [PubMed: 19536019]
15. Etzioni R, Riley GF, Ramsey SD, Brown M. Measuring costs: administrative claims data, clinical trials, and beyond. *Med Care.* 2002; 40(6 Suppl):III63–III72. [PubMed: 12064760]
16. Stranges, E., Russo, A., Friedman, B. [Accessed January 9, 2012] Procedures with the Most Rapidly Increasing Hospital Costs, 2004–2007 - Healthcare Cost and Utilization Project (HCUP) Statistical Briefs - Statistical Brief #82. <http://www.ncbi.nlm.nih.gov/books/NBK53597/>
17. Mirkin D, Murphy-Barron C, Iwasaki K. Actuarial analysis of private payer administrative claims data for women with endometriosis. *J Manag Care Pharm JMCP.* 2007; 13(3):262–272. [PubMed: 17407392]
18. Hsia RY, Akosa Antwi Y, Weber E. Analysis of variation in charges and prices paid for vaginal and caesarean section births: a cross-sectional study. *BMJ Open.* 2014; 4(1):e004017.doi: 10.1136/bmjopen-2013-004017
19. Levit KR, Friedman B, Wong HS. Estimating inpatient hospital prices from state administrative data and hospital financial reports. *Health Serv Res.* 2013; 48(5):1779–1797. DOI: 10.1111/1475-6773.12065 [PubMed: 23662642]

20. LeMaistre CF, Farnia SH. Goals for Pay for Performance in Hematopoietic Cell Transplantation: A Primer. *Biol Blood Marrow Transplant J Am Soc Blood Marrow Transplant*. 2015; 21(8):1367–1372. DOI: 10.1016/j.bbmt.2015.04.014
21. Gordis, L. *Epidemiology*, 3e. 3. Philadelphia, Pa: Saunders; 2004.
22. Pasquini, M., Zhu, X. [Accessed April 13, 2015] Current use and outcome of hematopoietic stem cell transplantation: CIBMTR Summary Slides. 2014. Available at: <http://www.cibmtr.org>
23. [Accessed June 5, 2013] Acute Myeloid Leukemia - SEER Stat Fact Sheets. <http://seer.cancer.gov/statfacts/html/amy1.html>
24. Majhail NS, Rizzo JD, Hahn T, et al. Pilot study of patient and caregiver out-of-pocket costs of allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant*. Dec.2012 doi: 10.1038/bmt.2012.248
25. Khera N, Chang Y-H, Hashmi S, et al. Financial burden in recipients of allogeneic hematopoietic cell transplantation. *Biol Blood Marrow Transplant J Am Soc Blood Marrow Transplant*. 2014; 20(9):1375–1381. DOI: 10.1016/j.bbmt.2014.05.011
26. Majhail NS, Mothukuri JM, Brunstein CG, Weisdorf DJ. Costs of Hematopoietic Cell Transplantation: Comparison of Umbilical Cord Blood and Matched Related Donor Transplantation and the Impact of Posttransplant Complications. *Biol Blood Marrow Transplant*. 2009; 15(5):564–573. DOI: 10.1016/j.bbmt.2009.01.011 [PubMed: 19361748]
27. Saito AM, Zahrieh D, Cutler C, et al. Lower costs associated with hematopoietic cell transplantation using reduced intensity vs high-dose regimens for hematological malignancy. *Bone Marrow Transplant*. 2007; 40(3):209–217. DOI: 10.1038/sj.bmt.1705733 [PubMed: 17563734]
28. Saito AM, Cutler C, Zahrieh D, et al. Costs of Allogeneic Hematopoietic Cell Transplantation with High-Dose Regimens. *Biol Blood Marrow Transplant*. 2008; 14(2):197–207. DOI: 10.1016/j.bbmt.2007.10.010 [PubMed: 18215780]
29. Lee SJ, Klar N, Weeks JC, Antin JH. Predicting costs of stem-cell transplantation. *J Clin Oncol Off J Am Soc Clin Oncol*. 2000; 18(1):64–71.
30. Majhail, NS., Mothukuri, J., MacMillan, ML., et al. [Accessed September 14, 2011] Costs of pediatric allogeneic hematopoietic-cell transplantation - Majhail-- 2009 - *Pediatric Blood & Cancer* - Wiley Online Library. <http://onlinelibrary.wiley.com.ezp2.lib.umn.edu/doi/10.1002/pbc.22250/pdf>. <http://onlinelibrary.wiley.com.ezp2.lib.umn.edu/doi/10.1002/pbc.22250/pdf>. Published 2010
31. Khera N, Storer B, Sandmaier BM, Chapko MK, J Lee S. Costs of second allogeneic hematopoietic cell transplantation. *Transplantation*. 2013; 96(1):108–115. DOI: 10.1097/TP.0b013e318294caf1 [PubMed: 23694949]
32. Farag SS, Maharry K, Zhang M-J, et al. Comparison of reduced-intensity hematopoietic cell transplantation with chemotherapy in patients age 60–70 years with acute myelogenous leukemia in first remission. *Biol Blood Marrow Transplant J Am Soc Blood Marrow Transplant*. 2011; 17(12):1796–1803. DOI: 10.1016/j.bbmt.2011.06.005
33. Truven. Truven Health MarketScan<sup>®</sup> Research Databases.
34. Pickens, G., Moldwin, E., Marder, WD. Healthcare Spending Index for Employer-Sponsored Insurance: Methodology and Baseline Results. Truven Health Analytics; 2010. p. 24 [http://truvenhealth.com/Portals/0/Assets/HealthInsights/TRU\\_15667\\_0415\\_HSI\\_ESI\\_WP.pdf](http://truvenhealth.com/Portals/0/Assets/HealthInsights/TRU_15667_0415_HSI_ESI_WP.pdf)
35. Majhail NS, Mau LW, Denzen EM, Arneson TJ. Costs of autologous and allogeneic hematopoietic cell transplantation in the United States: a study using a large national private claims database. *Bone Marrow Transplant*. 2013; 48(2):294–300. DOI: 10.1038/bmt.2012.133 [PubMed: 22773126]
36. National Quality Forum (NQF). Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care. Washington, DC: 2009. [http://www.qualityforum.org/Publications/2010/01/Measurement\\_Framework\\_\\_Evaluating\\_Efficiency\\_Across\\_Patient-Focused\\_Episodes\\_of\\_Care.aspx](http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx)
37. Giordano SH, Duan Z, Kuo Y-F, Hortobagyi GN, Goodwin JS. Use and outcomes of adjuvant chemotherapy in older women with breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2006; 24(18):2750–2756. DOI: 10.1200/JCO.2005.02.3028

38. McKellar MR, Naimer S, Landrum MB, Gibson TB, Chandra A, Chernew M. Insurer Market Structure and Variation in Commercial Health Care Spending. *Health Serv Res.* 2014; 49(3):878–892. DOI: 10.1111/1475-6773.12131 [PubMed: 24303879]
39. Milliman. Comparing Episode of Cancer Care Costs in Different Settings: An Actuarial Analysis of Patients Receiving Chemotherapy. Aug. 2013 <http://www.milliman.com/uploadedFiles/insight/2013/comparing-episode-cancer-care.pdf>
40. Government Accountability Office. Geographic Variation in Spending for Certain High - Cost Procedures Driven by Inpatient Prices. 2014. <http://www.gao.gov/assets/670/667781.pdf>
41. Baser O. Modeling Transformed Health Care Cost with Unknown Heteroskedasticity. *Appl Econ Res Bull.* 2007; (01):1–6.
42. Truven Health Analytics MarketScan. The Cost of Having a Baby in the United States. 2013. <http://transform.childbirthconnection.org/wp-content/uploads/2013/01/Cost-of-Having-a-Baby1.pdf>
43. Ozminkowski RJ, Wang S, Walsh JK. The direct and indirect costs of untreated insomnia in adults in the United States. *Sleep.* 2007; 30(3):263–273. [PubMed: 17425222]
44. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav.* 1995; 36(1):1–10. [PubMed: 7738325]
45. Roth JA, Bensink ME, O'Donnell PV, Fuchs EJ, Eapen M, Ramsey SD. Design of a cost-effectiveness analysis alongside a randomized trial of transplantation using umbilical cord blood versus HLA-haploidentical related bone marrow in advanced hematologic cancer. *J Comp Eff Res.* 2014; 3(2):135–144. DOI: 10.2217/ce.13.95 [PubMed: 24645687]
46. Chang S, Long SR, Kutikova L, et al. Estimating the cost of cancer: results on the basis of claims data analyses for cancer patients diagnosed with seven types of cancer during 1999 to 2000. *J Clin Oncol Off J Am Soc Clin Oncol.* 2004; 22(17):3524–3530. DOI: 10.1200/JCO.2004.10.170
47. Dreyfus B, Kawabata HM, Gomez-Caminero A. Adverse events in patients with liver cancer. *Anticancer Drugs.* 2013; 24(6):630–635. DOI: 10.1097/CAD.0b013e3283607f4f [PubMed: 23525070]
48. Hagiwara M, Delea TE, Chung K. Healthcare costs associated with skeletal-related events in breast cancer patients with bone metastases. *J Med Econ.* 2014; 17(3):223–230. DOI: 10.3111/13696998.2014.890937 [PubMed: 24494707]
49. Pyenson B, Connor S, Fitch K, Kinzbrunner B. Medicare cost in matched hospice and non-hospice cohorts. *J Pain Symptom Manage.* 2004; 28(3):200–210. DOI: 10.1016/j.jpainsymman.2004.05.003 [PubMed: 15336332]
50. Shih Y-CT, Xu Y, Cormier JN, et al. Incidence, treatment costs, and complications of lymphedema after breast cancer among women of working age: a 2-year follow-up study. *J Clin Oncol Off J Am Soc Clin Oncol.* 2009; 27(12):2007–2014. DOI: 10.1200/JCO.2008.18.3517
51. Snyder CF, Frick KD, Blackford AL, et al. How does initial treatment choice affect short-term and long-term costs for clinically localized prostate cancer? *Cancer.* 2010; 116(23):5391–5399. DOI: 10.1002/cncr.25517 [PubMed: 20734396]
52. Gill AA, Zahm SH, Shriver CD, Stojadinovic A, McGlynn KA, Zhu K. Colon cancer lymph node evaluation among military health system beneficiaries: an analysis by race/ethnicity. *Ann Surg Oncol.* 2015; 22(1):195–202. DOI: 10.1245/s10434-014-3939-4 [PubMed: 25059789]

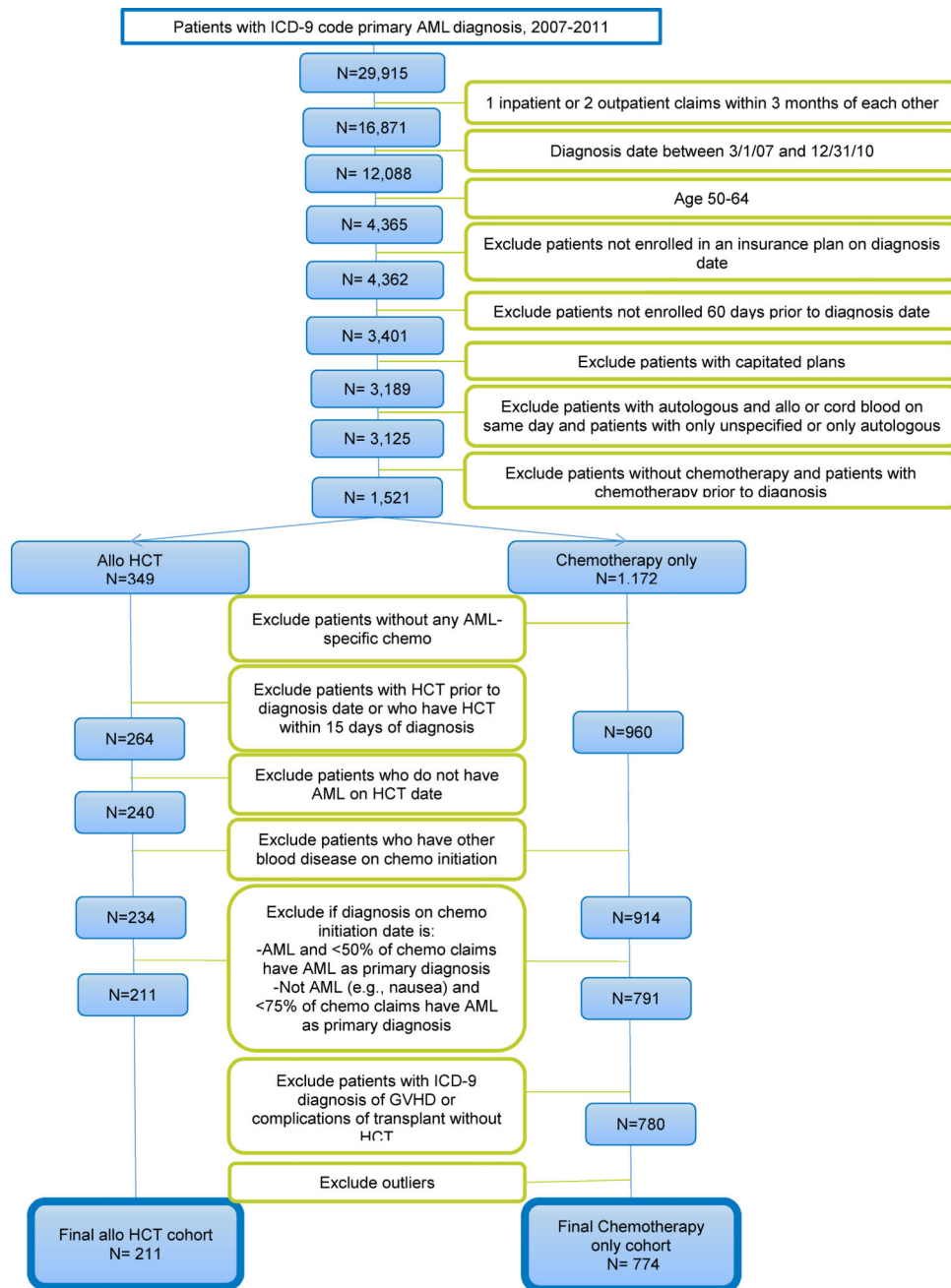
### Highlights

- Claims data are useful to study cost and utilization of HCT and chemotherapy
- An AML case study demonstrated opportunities and challenges of claims data
- Clinical knowledge and applied methods are critical to valid cohorts and measures

Criterion 1	The sample from which the cohort is derived should be representative of the population of interest
Criterion 2	Each service provided should have an accurate attribution of which diagnosis or diagnoses led to the service.
Criterion 3	Claims should capture all services actually provided to a patient throughout the episode of care related to the condition of interest.
Criterion 4	The identification of individuals with the condition of interest receiving the treatments being compared/examined should be both accurate and yield large enough treatment groups to provide adequate statistical power to detect meaningful differences in costs
Criterion 5	Estimates of cost are comprehensive with respect to type of service delivered and site of care and reflect actual payments made by payers
Criterion 6	Any and all concurrent or previous diagnoses or other individual factors that make the delivery of the interventions being compared/examined more complex or costly or alter the likelihood that specific treatment be chosen (relative to other available treatments) should be available.

**Figure 1.**  
Criteria for using administrative claims datasets for research





**Figure 2.**  
Patient inclusion and exclusion criteria

**Table 1**

Challenges and solutions using administrative claims data

Challenges	Solutions
Ascertain AML diagnosis	Require 1 inpatient or 2 outpatient claims (within 3 months) with AML primary diagnosis codes
Define date of diagnosis	Base on first inpatient or outpatient claim with AML diagnosis
Variation in pre-existing conditions and treatment, follow-up timeframe	Include claims 2 months before and 1-year post AML diagnosis Exclude patients with chemotherapy prior to AML diagnosis Exclude patients with HCT with no chemotherapy prior to HCT
Presence of capitated health plans	Exclude patients who had specific plan types with at least one claim with capitated service plans
AML not listed as primary diagnosis on therapy date	Determine primary diagnosis on chemo initiation claim, exclude those with other blood diseases Calculate percentage of AML chemo claims Include enrollees if diagnosis on therapy date: <ul style="list-style-type: none"> <li>• Is AML and &gt;50% of chemo claims have AML as primary diagnosis</li> <li>• Is not AML or other blood disease yet &gt;75% of chemo claims have AML as primary diagnosis</li> </ul>
Define the cohort of patients treated with chemo only Define chemo associated with AML Lack of chemo drug codes (J code or NDC) for inpatient claims	Determine how chemo identifiers and codes are used (inpatient v. outpatient) Specify chemo identifiers and codes for AML treatment Develop algorithm of code combinations for inclusion criteria (DRG, revenue, J code, NDC, administration, and diagnosis codes) Ensure comparable diagnoses in both study cohorts; specify timing of therapy in relation to date of diagnosis
Timing of HCT and HCT prior to diagnosis	Exclude patients with HCT prior to AML diagnosis and patients who had HCT within 15 days of AML diagnosis
GVHD and HCT complications existing within patients treated with chemotherapy alone	Excluded patients with these codes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript