Use of Health Plan Combined with Registry Data to Predict Clinical Trial Recruitment

Jeffrey R Curtis, MD MS MPH[1], **Nicole C Wright, PhD MPH**[2], **Fenglong Xie, MS**[1], **Lang Chen, PhD**[1], **Jie Zhang, PhD**[1], **Kenneth G Saag, MD MSc**[1], **Aseem Bharat, MPH**[1], **Joel Kremer, MD**[3], **Stacey Cofield, PhD**[2], **Kevin Winthrop, MD**[4], and **Elizabeth Delzell, ScD**[1]

[1]University of Alabama at Birmingham : 50 20th Street South FOT 802D Birmingham, AL 35294 United States

[2]University of Alabama School of Public Health:1720 2nd Ave N Birmingham, AL 35294 United States

[3]Albany Medical Center: Albany, NY United States

[4]Oregon Health & Science University: Portland, OR United States

Abstract

Background—Large pragmatic clinical trials (PCTs) are increasingly used to conduct comparative effectiveness research. In the context of planning a safety PCT of the live herpes zoster vaccine in rheumatoid arthritis (RA) patients age ≥ 50 receiving anti- tumor necrosis factor (TNF) therapy, we evaluated the use of health plan combined with registry data to assess the feasibility of recruiting the 4,000 patients needed for the trial and to facilitate site selection.

Methods—Using national United States data from Medicare, we identified older RA patients who received anti-TNF therapy in the last quarter of 2009. Extrapolations were made from the Medicare patient population to younger patients and those with other types of insurance using the Consortium of Rheumatology Researchers of North America (CORRONA) disease registry. Patients' treating rheumatologists were grouped into practices and sorted by size from the greatest to the least number of eligible patients.

Results—Approximately 50,000 RA patients receiving anti-TNF therapy were identified in the Medicare data, distributed across 1,980 physician practices. After augmenting Medicare data with information from CORRONA and extrapolating to younger patients and those with other types of insurance, more than 12,000 potentially eligible study subjects were identified from the 40-45 largest rheumatology practices.

Conclusion—Health plan and registry databases appear useful to assess feasibility of large pragmatic trials and to assist in selection of recruitment sites with the greatest number of

potentially eligible patients. This novel approach is applicable to trials with simple inclusion/ exclusion criteria that can be readily assessed in these data sources.

Background

The increasing importance of comparative effectiveness research (CER) continues to spotlight a need to develop and refine the methods that are used to generate new, real-world evidence [1]. Randomized controlled trials, and specifically, large pragmatic clinical trials (PCTs) are sometimes used to conduct CER [2]. Similar to traditional randomized clinical trials (RCT), PCTs are typically randomized; however they generally differ by having simple inclusion/exclusion criteria, yielding high generalizability to the target population, and PCTs examine hard outcomes that allow for simplified data collection for clinically relevant endpoints [3-6]. PCTs can be challenging to execute due to the large number of patients required. Failure to recruit patients as expected can have a major impact on the science and budget of CER studies.

The National Institutes of Health (NIH) and other sponsors typically require a multi-stage process in order to consider funding a RCT or PCT that begins with a planning grant. Among the activities required at this stage of trial design is specification of a detailed approach to patient recruitment and site selection in order to demonstrate feasibility. Accurate estimates of numbers of eligible patients and associated recruitment site selection are essential. However, efficient methods are needed to provide assurance that adequate numbers of patients can be recruited and to know which physician office practice settings have the greatest number of potentially eligible patients in order to select recruitment sites.

In the context of planning a PCT to evaluate the safety of the live herpes zoster vaccine in older rheumatoid arthritis (RA) patients receiving anti-tumor necrosis factor (TNF) therapy, we used both health plan and registry data to 1) assess the feasibility of recruiting the 4,000 patients expected to be needed for the trial based upon key inclusion criteria, and 2) to facilitate investigator site selection. This report describes our assessment methods as an example of a highly generalizable framework that might be used in planning future trials for a variety of rheumatic diseases and other chronic medical conditions.

Methods

Inclusion Criteria of the Example Trial

The proposed Varicella zostER VaccinE (VERVE) trial is a randomized, double-blind, placebo-controlled study to evaluate the safety, tolerability, immunogenicity, and long-term effectiveness of the live herpes zoster vaccine among patients age 50 years or older currently receiving anti-TNF therapy for RA or other inflammatory arthritis. Anti-TNF therapy improves clinical signs and symptoms of RA and reduces X-ray damage and associated disability. However, because of the potential risk that patients treated with these biologic

therapies are immunosuppressed, a live virus vaccine therefore poses a theoretical risk that it could trigger zoster reactivation [7].This restriction is not based upon clinical data and thus prompted the interest in a safety trial of the vaccine in this population. A total of 4,000 patients are required to attain 80% power to demonstrate non-inferiority of the clinical safety of the vaccine compared to placebo. As with most trials, there are expected costs related to initiating each trial site to cover IRB fees, site training, and data monitoring. For the VERVE trial, an additional site-related cost will be the deployment of tablet computers used to facilitate efficient patient identification, the consent process, and data collection. Therefore, minimizing the number of recruitment sites is preferred. Moreover, because RA is relatively rare and not all US rheumatologists conduct research, the feasibility of screening and enrolling the required number of patients needed to be demonstrated.

Health Plan Used to Assess Feasibility

As one example of a large health care database, we used the 2009 Medicare data of 100% of beneficiaries with inflammatory autoimmune diseases with traditional fee-for-service Medicare. Medicare data is made available to the public for qualified research purposes by the Center for Medicare and Medicaid Services (CMS) and is subject to privacy restrictions defined as part of a Data Use Agreement. Medicare administrative data includes monthly coverage indicators for all enrollees and encompasses inpatient and outpatient claims, prescriptions filled under Medicare part D, and information on all Medicare providers. We identified enrollees who were age 60 years with RA. A person was defined to have RA if they had at least 2 office visits claims with RA diagnosis codes (ICD9-CM 714.X) [8-10]. Because VERVE requires patients to be currently receiving anti-TNF therapy, we restricted patients to those who filled a prescription for, or received an infusion of, any anti-TNF agent (adalimumab, certolizumab, etanercept, golimumab, infliximab) in the last calendar quarter of 2009 identified from Part B (infusion/injection) or Part D (prescriptions filled). These patients were linked to treating rheumatologists through use of National Provider Identification (NPI) numbers and the physician specialty field available in the Medicare data for every submitted claim. Rheumatologists were grouped together into practices (offices) using the tax identification numbers used for billing purposes available in the CMS data. We calculated the ratio of the number of eligible patients with fee-for-service Medicare with Part D prescription coverage (Parts A, B and D) to the number of patients with any type of Medicare coverage (Part A, Part B, Part D, other type of drug plan, or Medicare Advantage).

Cross-referencing with Registry Data

Although RA patients younger than 65 years of age might qualify for Medicare on the basis of disability (which is an allowable reason based upon the U.S. Social Security Disability listing of Adult Impairments [11]), most younger patients were not expected to have Medicare coverage. Therefore, to obtain more information on younger patients (age >= 50 years) and patients older than 65 years who were not in Medicare, rheumatologists were cross-referenced with the list of physicians participating in the Consortium of Rheumatology Researchers of North America (CORRONA), the largest U.S. rheumatology registry with physician-derived data [12, 13]. CORRONA data are collected in a de-identified fashion. The participating physicians at each site are the only individuals who have access to the protected health information that would be needed to uniquely identify patients. The

CORRONA dataset is not publically available, and patient participants provide written informed consent as part of their enrollment in the registry. CORRONA includes 563 physicians covering 39 states, with high generalizability to the RA population in the US [Curtis JR et. al., American College of Rheumatology 2013, San Diego, CA].

The CORRONA data were searched for eligible patients in similar fashion as in Medicare. Within the CORRONA data across all participating rheumatology practices, we identified the proportion of eligible patients age >= 50 on anti-TNF therapy at the end of 2009 with Medicare insurance coverage (self-reported by patients at every CORRONA visit) compared to the total number of eligible patients with Medicare or non-Medicare insurance coverage. This reciprocal of this proportion was multiplied by the patient numbers obtained in the Medicare administrative data to estimate the total number of eligible patients in each rheumatologist's practice across all insurance types. For example, if 40% of eligible patients in CORRONA had Medicare insurance coverage, then the estimated number of patients obtained from the Medicare administrative data would be multiplied by $(1 / 0.4) = 2.5$.

Finally, we identified physicians currently participating or who had participated in clinical or RA- specific research in the past using publically available sources of information (e.g. investigator list of the TEAR trial [14]; acknowledgements sections of other published rheumatology RCTs listing participating sites; the Food and Drug Administration registry of physicians who conduct clinical research, available since 1992 [15]).

Statistical analysis

We multiplied the total number of RA patients who had fee-for-service Medicare with prescription drug coverage (Parts A, B and D) by the cumulative product of the various multipliers to calculate the total number of eligible patients for VERVE. The total number of eligible RA patients was plotted against the number of rheumatologist practices (which could consist of one or more rheumatologists) at which these patients received care. We sorted practices by descending number of eligible patients (e.g. Site 1 has the greatest number of patients). The resulting figure was examined in light of the needed sample size of 4,000 eligible patients, varying assumptions about the patient recruitment rate (e.g. 25%, 33%, 50% of all eligible patients). This range of recruitment assumptions were felt to have face validity given the simple nature of the trial that is not burdensome for participants and is minimally intrusive for physicians to conduct; for example, for most participants, only a single study visit is required.

We conducted a sensitivity analysis based on three key sources of variance in our estimates that might affect the generalizability of our estimates derived using 2009 data to the results we might expect in 2014 when the trial began recruiting. These three parameters included the proportion of Medicare patients who had Medicare Advantage (rather than traditional fee-for-service coverage) or a non-part D drug plan; changing prevalence of biologic use over time; and variability in the proportion of RA patients with Medicare coverage across CORRONA sites, accounting for practice-level variability.

Finally, we ran a simulation using bootstrap methods with 10,000 iterations to model these three sources of variance. We reported the inter-quartile range around our primary

recruitment estimates based upon the bootstrap result. All analyses were performed in SAS 9.2. Medicare data were governed by a Data Use Agreement with CMS, and the University of Alabama at Birmingham IRB approved the study protocol. CORRONA data were governed by a Data Use Agreement with CORRONA.

Results

A total of 47,623 RA patients identified in the Medicare fee-for-service administrative data met age criteria and received any anti-TNF therapy in the last quarter of 2009. These individuals were assigned to their 3,547 treating rheumatologists and the associated 1,980 rheumatology practices at which they received care.

Correction Factor Multipliers

As of January 1st 2009, we estimated that 35.8% of the RA patients in our Medicare data were enrolled in Medicare Advantage plans and/or were not enrolled in a Part D plan, indicating that the raw fee-for-service data available to us represented approximately 64% of the total Medicare patient population (with any type of coverage) who would be eligible for the trial (Table 1). This estimate was intentionally conservative given that only approximately forty percent of Medicare patients have traditional fee for service with part D coverage in any given month (data not shown; based upon the Medicare random 5% sample). To extrapolate to the entirety of the Medicare population, we then multiplied the number of eligible patients derived from the raw data for the traditional fee-for-service Medicare enrollees by its reciprocal, yielding a ratio of 1.56.

Overall, 47% of patients age 50 and older in CORRONA who met our inclusion criteria had any form of Medicare coverage. Our final multiplier (cumulative product of all multipliers, rounded to 2 digits after the decimal for display purposes), was 3.29 (Table 1). After applying the overall ratio to the raw numbers from Medicare fee-for-service data, we therefore estimated that 156,263 participants with any type of insurance coverage would potentially be eligible for VERVE (Figure 1).

The number of eligible patients identified was plotted for the 100 largest rheumatology practices sorted by the greatest to the least number of potential participants (Figure, dashed line). Almost all of the largest practices had evidence that they had participated in clinical research, as evidenced by this line being almost superimposed with the dotted line for the first 6,000 patients. Considering rheumatologists who conduct research (solid line) and with a hypothesized trial participation rate of 33% (which would require 12,000 eligible patients to be screened), 40-45 rheumatology practices would be required to successfully recruit the proposed PCT to yield the needed 4,000 patients.

Derivation of Estimates used for Sensitivity Analysis and Bootstrap Simulation Results

Based upon the literature [16], the proportion of Medicare beneficiaries enrolled in Medicare Advantage has increased over the past several years, but in the evaluation of Medicare Advantage plan enrollment from its inception, there has not been a consistent nor a linear increase over time [17]. Thus, for our sensitivity analysis, we estimated that the eligible

traditional fee for service Medicare population with part D drug coverage could increase or decrease by up to 3% compared to 2009 (Table 1).

Regarding trends in biologic use among RA patients, Zhang et al found no overall appreciable change in overall biologic use from 2007-2009 in the Medicare program [18]. Nevertheless, although we used no change in the prevalence over time for our base case estimates, we allowed that anti-TNF use among RA patients might increase or decrease by plus or minus 10% by the time VERVE begins recruitment.

Finally, we observed appreciable variability across CORRONA sites in the proportion of eligible patients who had Medicare coverage. Based upon the distribution of the proportion of eligible patients at each site who had Medicare coverage, and selecting a sampling distribution from the 10th and 90th percentiles, the inter-quartile range of the recruitment ratio estimated in our primary results spanned 2.73 to 3.73. Thus, based upon this result, the trial could potentially need up to ± 20% fewer or more patients to screen for VERVE compared to the base case estimate (Table 1).

Discussion

In the context of planning a large PCT, we used administrative data linked with a disease registry to assess the feasibility of recruiting the necessary number of eligible patients and to facilitate site selection for patient recruitment. We observed that a reasonable number of rheumatology practices (~ 50) would be required to participate in order to access a sufficiently large pool of potentially eligible patients. Using plausible assumptions about patient participation (e.g. 25-33% of all eligible patients), the trial was considered feasible with regard to attaining the recruitment goal and in light of the trial's resources.

Our novel approach leverages increasingly available data from health plans and insurance systems, as well as information from large registries. While these data are frequently used for CER and other research, acquiring these data from CMS or other sources (e.g. commercial insurance plans or health systems) and obtaining permission to use them are not without cost or administrative burden. Moreover, because our trial needed patients 50 years of age and older, Medicare data alone were suboptimal to provide the complete picture desired to support trial feasibility. Using data from CORRONA, a large US RA registry, to derive estimates of the number of patients with other types of health insurance was therefore valuable to improve the estimation of the number of eligible patients available nationwide. Conversely, having only the data from a registry would represent only the sites and patients enrolled and would miss both patients who declined registry participation as well as physicians not participating in the registry. Thus, the data sources are complementary.

The geographic and demographic generalizability of the 100% Medicare data used to identify Medicare enrollees is high. However, we recognize that information for patients enrolled in self-contained managed healthcare systems (e.g. Kaiser Permanente) will not be represented within CMS data. Another potential limitation of our approach to planning some trials by using health plan data relates to a lack of detailed clinical information in these information sources. Only a few administrative data sources contain lab test results (e.g. C-

reactive protein). Therefore, this approach of using administrative claims or health plan data is limited to assessing recruitment feasibility where precise phenotypic information is not required. However, since the VERVE trial, like most PCTs, has very simple inclusion criteria, the data sources used were well-suited for the needed feasibility assessment. A disease registry like CORRONA is of even greater value if detailed phenotypic information (e.g. disease activity) is required. Finally, although we used statistical techniques to model variance in our recruitment estimates, they could potentially be incorrect depending on the actual trends over time with respect to insurance coverage and biologic use, thus potentially under- or over-estimating the total number of eligible patients available for the VERVE trial. While this may be a limitation for VERVE, the main goal of this report was to showcase how administrative data along with registry data could be used in planning a PCT irrespective of the specific use case; researchers should permute the parameters they feel are most essential to obtain accurate recruitment targets for their particular study.

In conclusion, large health plan databases and registries appear useful to assess the feasibility of large pragmatic trials and to assist in selection of physician offices with the greatest number of eligible patients. This novel approach is applicable to trials with simple inclusion/exclusion criteria that can be readily assessed with these data sources.

Acknowledgments

References

1. Tinetti ME, Studenski SA. Comparative Effectiveness Research and Patients with Multiple Chronic Conditions. N Engl J Med. 2011

2. Saag KG, et al. Improving the efficiency and effectiveness of pragmatic clinical trials in older adults in the United States. Contemp Clin Trials. 2012; 33(6):1211–6. [PubMed: 22796098]

3. Thorpe KE, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. CMAJ. 2009; 180(10):E47–57. [PubMed: 19372436]

4. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Clin Epidemiol. 2009; 62(5):499–505. [PubMed: 19348976]

5. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003; 290(12):1624–32. [PubMed: 14506122]

6. Roland M, Torgerson DJ. What are pragmatic trials? BMJ. 1998; 316(7127):285. [PubMed: 9472515]

7. Singh JA, et al. 2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken). 2012; 64(5):625–39. [PubMed: 22473917]

8. Singh JA, Holmgren AR, Noorbaloochi S. Accuracy of Veterans Administration databases for a diagnosis of rheumatoid arthritis. Arthritis Rheum. 2004; 51(6):952–7. [PubMed: 15593102]

9. MacLean CH, et al. Positive predictive value (PPV) of an administrative data-based algorithm for the identi cation of patients with rheumatoid arthritis (RA). Arthritis & Rheumatism. 2001; 44:S106.

10. Kim SY, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. Arthritis Res Ther. 2011; 13(1):R32. [PubMed: 21345216]

11. Disability Evaluation Under Social Security: 14.00 Immune System Disorders - Adult. Available from:http://www.ssa.gov/disability/professionals/bluebook/14.00-Immune-Adult.htm#14_09

12. Curtis JR, et al. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. Semin Arthritis Rheum. 2010; 40(1):2–14 e1. [PubMed: 20674669]

13. Kremer JM. The CORRONA database. Autoimmun Rev. 2006; 5(1):46–54. [PubMed: 16338211]

14. Moreland LW, et al. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis: The treatment of early aggressive rheumatoid arthritis trial. Arthritis Rheum. 2012; 64(9):2824–35. [PubMed: 22508468]

15. Bioresearch Monitoring Information System (BMIS). 2012 Oct 31. Available from: http://www.fda.gov/Drugs/InformationOnDrugs/ucm135162.htm

16. Medicare Advantage 2012 Data Spotlight: Enrollment Market Update. Available from http://kff.org/health-costs/report/medicare-advantage-2012-enrollment-market-update/

17. McGuire TG, Newhouse JP, Sinaiko AD. An economic history of Medicare part C. Milbank Q. 2011; 89(2):289–332. [PubMed: 21676024]

18. Zhang J, et al. Trends in the Use of Biologic Therapies among Rheumatoid Arthritis Patients Enrolled in the U.S. Medicare Program. Arthritis Care Res (Hoboken). 2013
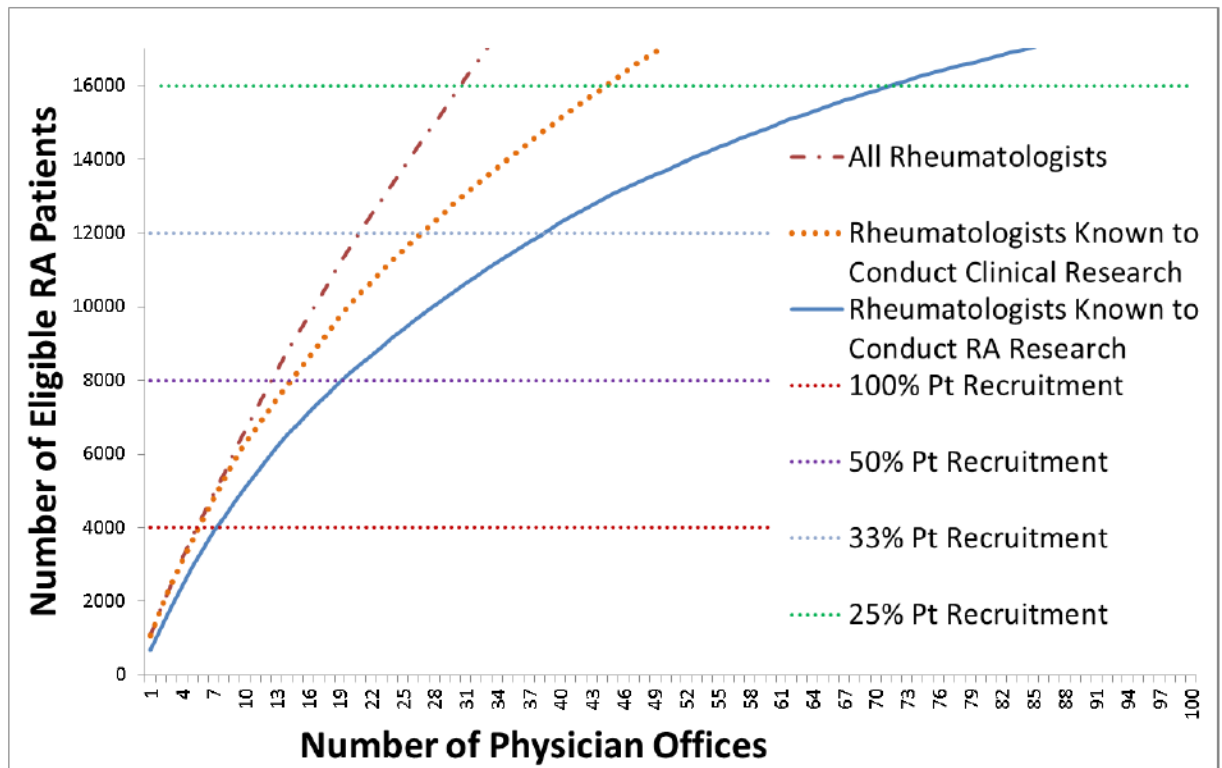
Figure 1.
Number of Potentially Eligible Patients Needed for Trial Recruitment According to the
Number of Physician Offices* Required to Participate
* physician offices in the figure were ranked from the largest to the smallest based upon the
number of patients eligible for the trial

Table 1

Sources of Variance in VERVE Multiplier Estimates Used for Bootstrap Simulation

	2009 Data	Multiplier applied to raw Medicare FFS+D data*	Sampling distribution to allow extrapolation to 2014 when recruitment might begin	Reference for sampling distribution
Source : Medicare Data				
% of patients with Fee-for-Service Medicare with part D, referent to patients with any type of Medicare coverage	64%	1.56	Mean =0.64; range from 0.61 – 0.67	Kaiser Family Foundation, 2012[16] McGuire et al, 2011[17]
Change in prevalence of biologic use	referent	1.00 (i.e. no change)	Mean = 1.0; 95% CI = 0.90, 1.10; normal distribution	Zhang et al, 2013[18]
Source : CORRONA				
Prevalence of Medicare coverage among CORRONA-enrolled RA patients meeting VERVE eligibility criteria in 2009	47%	2.11	Mean = 0.47; range from 0.31, 0.63	10th/90th percentile of Medicare coverage across CORRONA sites for RA patients in CORRONA meeting trial ligibility criteria
Final multiplier applied to 2009 raw Medicare data		**3.29**	**Mean (IQR) = 3.40 (2.73; 3.73)**	

FFS+D = fee-for-service (with part D); IQR = inter-quartile range

Note: numbers rounded to the 2nd digit after the decimal for simplicity of presentation

*
this multiplier is the reciprocal of the number in the column to the left. For example, (1 / 0.64) = 1.56