VALUE IN HEALTH 16 (2013) 655–669

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

METHODOLOGICAL ARTICLES

A Checklist for Ascertaining Study Cohorts in Oncology Health Services Research Using Secondary Data: Report of the ISPOR Oncology Good Outcomes Research Practices Working Group

Kathy L. Schulman, MA[1,], Karina Berenson, MPH[2], Ya-Chen (Tina) Shih, PhD[3], Kathleen A. Foley, PhD[4], Arijit Ganguli, MBA, PhD[5], Jonas de Souza, MD[3], Nicholas A. Yaghmour, MPP[3], Alex Shteynshlyuger, MD[6]*

[1]*Outcomes Research Solutions, Inc., Bolton, MA, USA;* [2]*Covance, Gaithersburg, MD, USA;* [3]*University of Chicago, Chicago, IL, USA;* [4]*Truven Health Analytics, Philadelphia, PA, USA;* [5]*Abbott Laboratories, Abbot Park, IL, USA;* [6]*Phoenixville Urologic Associates, Phoenixville, PA, USA*

ABSTRACT

Objectives: The ISPOR Oncology Special Interest Group formed a working group at the end of 2010 to develop standards for conducting oncology health services research using secondary data. The first mission of the group was to develop a checklist focused on issues specific to selection of a sample of oncology patients using a secondary data source. **Methods:** A systematic review of the published literature from 2006 to 2010 was conducted to characterize the use of secondary data sources in oncology and inform the leadership of the working group prior to the construction of the checklist. A draft checklist was subsequently presented to the ISPOR membership in 2011 with subsequent feedback from the larger Oncology Special Interest Group also incorporated into the final checklist. **Results:** The checklist includes six elements: identification of the cancer to be studied, selection of an appropriate data source, evaluation of the applicability of published algorithms, development of custom algorithms (if needed), validation of the custom algorithm, and reporting and discussions of the ascertainment criteria. The checklist was intended to be applicable to various types of secondary data sources, including cancer registries, claims databases, electronic medical records, and others. **Conclusions:** This checklist makes two important contributions to oncology health services research. First, it can assist decision makers and reviewers in evaluating the quality of studies using secondary data. Second, it highlights methodological issues to be considered when researchers are constructing a study cohort from a secondary data source.

Keywords: cohort, ISPOR checklist, oncology, sample selection, secondary data.

Purpose

The use of secondary data sources is an increasingly accepted approach in oncology health services research, as exemplified by studies using such data to profile care patterns, measure patient outcomes, and estimate cancer-related costs [1–6]. Ascertainment of study cohorts from these data sources, however, can be difficult. Selection of a sample of cancer patients from secondary data sources is challenging when coding systems (e.g., *International Classification of Diseases, Ninth Revision, Clinical Modification*) used to identify these patients lack adequate clinical precision (e.g., cancer stage) or when clinical interventions are masked by payment system protocols. In these circumstances, the cohort selected may not be reflective of the larger population with the disease (poor sensitivity) or may contain large numbers of patients who do not actually have the disease at all (poor specificity). In either case, this may lead to inconsistent, spurious, or erroneous results and compromise the clinical relevancy of study findings. While this is a common problem in observational studies, the complexity of oncologic disease and its inexact representation in clinical data systems may exacerbate both selection and misclassification bias.

The purpose of this article is to both provide researchers who conduct oncology health services research using secondary data a list of methodological issues to consider when selecting study cohorts and assist researchers as well as journal reviewers/readers, payers, and policymakers in evaluating the quality of published studies involving secondary data analyses.

A working group to develop standards for oncology health services research using secondary data was established through the ISPOR Oncology Special Interest Group (SIG) to review and address issues related to this line of research. The first task of our working group was to develop a checklist for use in selecting samples of patients with a specific type of cancer from secondary data sets. We completed this task in two phases: a systematic literature review followed by development of the draft and final checklist. The formal review of the literature was conducted to

* *Address correspondence to:* Kathy L. Schulman, Outcomes Research Solutions, Inc., 44 Whitcomb Road Bolton, MA 01740, USA.
E-mail: kathy.schulman@orsolutionsinc.com.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2013.02.006

inform SIG leadership as to the current state of science in the literature prior to the development of the checklist. It was intended to complement the knowledge base and experience of the larger SIG as well as the existing validation literature. It was not conducted, as are many systematic reviews, to support meta-analysis.

SIG consensus was shaped first by discussions regarding the extent to which the literature review should be structured and subsequently the development of the elements of the structured review. SIG members were involved in the construction of search parameters, in the development of the abstraction form, in the actual abstraction of selected articles, and of course in the summary of findings. This process, which took place over the course of a year, included multiple meetings in which the literature search, its findings, and the potential importance to the checklist were discussed. It should be noted that disagreement between abstractors was resolved by third-party review. Consensus was also formed by discussions with the larger ISPOR membership during the presentation of these results at the annual meeting and also during dissemination of the draft deliverable to the larger SIG membership.

The literature review resulting from the above efforts reflects a concerted decision on the part of the SIG leadership to either confirm or refute a priori assumptions about the current state of oncology research using secondary data prior to drafting a checklist. For example, SIG members underestimated the frequency with which claims data were being used in isolation from other data sources and overestimated researcher reliance on previously published algorithms. Moreover, the checklist itself was highly influenced by the degree of difficulty in abstracting the necessary detail about sample selection in general because key elements either were missing altogether or were vague or nonspecific in their presentation. Finally, consensus regarding key elements of the checklist was inferred on the basis of results of the review in conjunction with the experience of SIG members, most notably the importance of identifying the key clinical elements of the cancer of interest.

The literature search was conducted in PubMed on English language articles published between January 1, 2006, and December 31, 2010. We chose the most recent 5-year time period at the time the literature review was performed because it captures the majority of work done with secondary data sources while also reflecting the most recent methodological thinking in this topic area. Search terms included "claims," "hospital discharge," "electronic medical records (EMR)," "registry," and "administrative" in conjunction with the Mesh term "Neoplasms." No other restrictions were placed on the query. Two reviewers (K.S. and K.B.) assessed abstracts for potential eligibility with subsequent full-text article review by the larger working group leadership team to determine appropriateness of inclusion. Findings from the literature review were used to characterize the use of secondary data sources in oncology, to identify potential methodological and reporting issues for oncology health services researchers, and to inform the second phase, in which a draft checklist was developed by consensus among the leadership working group. The draft checklist, presented at the ISPOR international meeting in 2011, and subsequently reviewed by the larger Oncology SIG, was then modified by incorporating comments and feedback to produce the final checklist.

Review of Current Literature

The literature search yielded 863 abstracts, of which 321 were eligible for full article review and 294 [6–298] were eligible for abstraction (Fig. 1).

The literature search demonstrated that secondary data sources are widely used for oncology research and that the ascertainment of study cohorts largely relies on registry or claims-based data. As evidenced in Table 1, 72.8% of the studies identified in our review relied to some degree on claims data. Of the 294 articles selected for review, 21.8% (n = 64) were claims-only studies [7–70], 5.1% (n = 15) were registry-only [71–85], 51% (n = 150) used claims and registry data together without any other supplemental data set [6,24,151–298], and the remaining 22.1% (n = 65) used other data sources, primarily chart-based systems or hospital discharge data sets, either alone or in combination with claims and/or registry data [86–150]. Chart-based systems included paper or electronic health records (EHRs) as well as electronic medical record (EMR) databases, programmatically generated subsets of EHRs. The Surveillance Epidemiology and End Results (SEER)/Medicare database was the single most commonly used (N = 93) data set, representing 62% of combination claims and registry studies and 31.6% of all eligible studies.

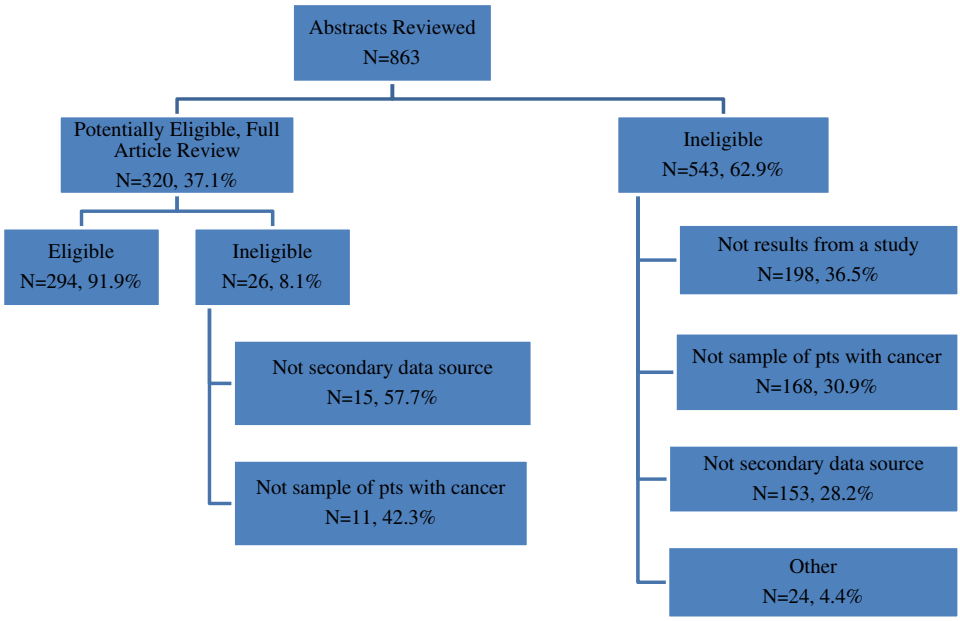

Fig. 1 – Abstract review and article selection.