A Rapid Outcomes Ascertainment System Improves the Quality of Prognostic and Pharmacogenetic Outcomes from Observational Studies

Penelope A. Bradbury,1 Rebecca Suk Heist,3,4,5 Matthew H. Kulke,3,6 Wei Zhou,4 Ariela L. Marshall,3,4 David P. Miller,4 Li Su,4 Sohee Park,4 Jennifer Temel,3,4,5 Panos Fidias,3,4,5 Lecia Sequist,3,5 Thomas J. Lynch,3,5 John C. Wain,3,5 Frances A. Shepherd,1 David C. Christiani,3,4,5 and Geoffrey Liu1,2,3,4,5

1Princess Margaret Hospital; 2Ontario Cancer Institute, Toronto, Ontario, Canada; 3Harvard Medical School; 4Harvard School of Public Health; 5Massachusetts General Hospital; and 6Dana Farber Cancer Institute, Boston, Massachusetts

Abstract

Purpose: Case-control and observational studies are popular choices for evaluating molecular prognostic/pharmacogenetic outcomes, but data quality is rarely tested. Using clinical trial and epidemiologic methods, we assessed the quality of prognostic and outcomes data obtainable from a large case-control study of lung cancer.

Methods: We developed an explicit algorithm (set of standard operating procedures forming a rapid outcomes ascertainment system) that encompassed multiple tests of quality assurance, and quality of data for a range of prognostic and outcomes variables, in several cancers, across several centers and two countries were assessed. Based on these assessments, the algorithm was revised and physicians‘ clinical practice changed. We reevaluated the quality of outcomes after these revisions.

Results: Development of an algorithm with internal quality controls showed specific patterns of data collection errors, which were fixable. Although the major discrepancy rate in retrospective data collection was low (0.6%) when compared with external validated sources, complete data were found in <50% of patients for treatment response rate, toxicity, and documentation of patient palliative symptoms. Prospective data collection and changes to clinical practice led to significantly improved data quality. Complete data on response rate increased from 45% to 76% (P = 0.01, Fisher’s exact test), for toxicity data, from 26% to 56% (P = 0.02), and for palliative symptoms, from 25% to 70% (P < 0.05), in one large lung cancer case-control study.

Conclusions: Observational studies can be a useful source for studying molecular prognostic and pharmacogenetic outcomes. A rapid outcomes ascertainment system with strict ongoing quality control measures is an excellent means of monitoring key variables.

(Cancer Epidemiol Biomarkers Prev 2008;17(1):204–11)

Introduction

The identification of molecular prognostic and predictive factors is an area of intense oncolgic research, with the potential for patient-tailored screening and treatment (1–6). Patients may be able to avoid the toxicities of receiving therapies inappropriate for their specific tumor or genetic background (7), improving cost-effectiveness. To undertake this research, biological specimens, patient outcomes, and clinical prognostic information are required. The creation of new case-series can be expensive and time-consuming, particularly for rare malignancies. Case-series derived from existing case-control risk studies offer a number of advantages. Many such studies already collect biological samples (6, 8–11). Case-control studies are frequently of large scale, with adequate intrinsic power for survival analysis. As part of the case-control study, cases are identified and recruited, detailed epidemiologic information is routinely collected, and databases are created and populated. Quality control mechanisms are in place for the questionnaire and biological data. The only missing component seems to be the collection of outcomes data. Thus, superficially, case-control studies seem to be perfect as a backdrop for studying molecular prognostic or pharmacogenetic factors. Indeed, a recent editorial supports a paradigm shift whereby molecular outcomes studies are integrated into the fabric of a molecular epidemiologic case-control study (12).

Despite these advantages, carrying out outcomes analyses on the foundation of a case-control study has potential pitfalls. Prognostic and risk factors for the same disease do not overlap completely (13), so that there may be missing data on some clinically important prognostic variables. Unlike randomized control studies, case-control studies are observational studies; their outcomes are typically collected retrospectively. Patients are not treated uniformly. Whereas population-based samples are a standard in the case-control setting (14), the same
population-based samples may yield a heterogeneous group inappropriate for survival analyses of the entire data set.

Between 1999 and 2001, the Harvard Lung Cancer Susceptibility Case-Control Study, which began in 1992, began to address the issue of conducting a clinical outcomes study within the case-control framework. We developed a rapid outcomes ascertainment system (ROAS) to allow us to collect these outcomes. ROAS involves a set of specific algorithmic procedures that incorporate multiple quality assurance tests. Thus, this Harvard study also enabled us to compare differences between a typical case-outcomes retrospective approach to data collection and ROAS. We hypothesized that case-control studies can yield high-quality outcomes analyses if properly monitored and conducted using a system similar to ROAS. After application of ROAS to Harvard Lung Cancer Study, we also applied ROAS to separate esophageal and pancreatic cancer studies at two institutions, and recently we started implementation of ROAS in a third institution. We hypothesized that the lessons learned from the Harvard Lung Cancer Study were generalizable to new tumor sites and new institutions, even in new health care systems of another country.

Materials and Methods

Case Control Studies. The Harvard Lung Cancer Susceptibility Study (D. Christiani, principal investigator) is a hospital-based single-institution [Massachusetts General Hospital (MGH)] case-control study with more than 2,500 cases and 1,500 healthy spouse and friend controls. This case-control study was initiated in 1992 and continues to accrue today. Between 1992 and 1997, only surgically resected lung cancer cases of all histologic subtypes were recruited. From 1997 onward, the eligibility criteria expanded to include all patients with histologically confirmed primary lung cancer. Blood and tissue specimens were collected at the time of diagnosis for cases. Access to all clinical and follow-up medical information was allowed. Demographic details for this study, up to the year 2000-2001, are found in Liu et al. (15).

The Dana-Farber Harvard Cancer Center studies of esophageal and pancreatic cancer involved not only MGH but also recruitment at Dana Farber Cancer Institute (G. Liu, D. Christiani, and M. Kulke; co-principal investigators); these were prospective studies developed under ROAS. The esophageal cancer study at Princess Margaret Hospital in Toronto (G. Liu, principal investigator) is currently implementing ROAS procedures.

Outcomes Data Collection Timeline. In 2000 to 2001, the feasibility of evaluating outcomes in this study was assessed. Because >80% were non–small-cell lung cancers, we limited the assessment of outcomes to this subgroup. The timeline included literature search (step 1), January to June 2000; feasibility pilot (step 2), June to September 2000; the development of standard operating procedures (ROAS) for collecting clinical prognostic and outcomes variables (step 3), August 2000 through June 2001; and ROAS quality control assessments (steps 4-6), March 2001 through January 2002. Prognostic and outcomes data collection using ROAS began in earnest in June 2002 and continues to the present.

Identification of Important Clinical Prognostic Factors and Outcomes (Step 1). PubMed search used the search terms “lung cancer” and “prognosis.” To avoid inclusion of outdated prognostic variables, we limited the search to 1990 through 1999. Articles were restricted to “English language,” “core clinical journals,” and “with available abstracts.” Through a separate search, we compiled phase II and III studies in the same period to identify the important clinical outcomes variables.

Pilot Feasibility Study (Step 2). In 1999, the feasibility of collecting outcomes data from this case-control study was uncertain. To assess feasibility, a small number (n = 40) of non–small-cell lung cancer patients with early-stage disease and a small cohort (n = 40) with advanced-stage disease were randomly selected. A basic qualitative assessment of feasibility (e.g., were charts available?, etc.) was done.

Development of Algorithm for Data Collection (Step 3). We developed standard operating procedures for prognostic and outcomes data collection in the form of an algorithm incorporating two internal quality control measures (Fig. 1 and steps 4-7). In summary, research assistants involved in data abstraction underwent initial training. Data were abstracted from a range of sources: hospital computerized and paper patient records, Social Security Death Index, referring physicians notes, and death certificates. If required, patients/patients’ families were approached, but only when data could not be obtained from the sources listed above. Abstracted data were subsequently computerized. The algorithm incorporated two internal quality control measures (step 4) and validation with external “gold standard” sources (steps 5 and 6). A key element to this algorithm was that it was accompanied by a paper trail (a procedure manual), so that historical procedural details were always documented. This algorithm ensured a consistently high level of quality control even when multiple individuals collected outcomes data.

Precision of Collected Data (Step 4). Internal assessment of precision involved (a) having an oncologist (G.L.) reabstract data using the algorithm from a random sample of 10% of cases, and (b) double data entry in 10% of randomly chosen cases. Discrepancies led to a review, followed by retraining of study personnel or modification to the database.

Adequacy of Follow-up and Outcomes (Step 5). In an observational study, both the patient and treating physician control the frequency and completeness of follow-up; treating physicians control the documentation of key outcomes. We assessed the quality of these processes. A panel of local oncology experts first established definitions of what constituted complete, substandard, and missing prognostic and outcomes variables. We used the standard Response Evaluation Criteria in Solid Tumors for outcomes data (16) and the National Cancer Institute Common Toxicity Criteria version 2.0 for grading toxicity.

Cross-Validation with External Sources (Step 6). A random sample of the prognostic and outcomes data was cross-checked against secondary sources of information:7

(a) an independent oncologist blinded to the algorithm results, using all available information and not confined to the algorithm protocol (R.S.H. and G.L.); (b) for patients concurrently recruited to a clinical trial, shadow charts of the clinical trials coordinators; and (c) the MGH Cancer Registry. Our comparisons were categorized as follows: (a) identical results between algorithmic and secondary sources; (b) minor discrepancies in documentation between the algorithmic and secondary sources that would not affect most outcomes analyses; (c) major discrepancies or missing data that could materially alter analytic results between the algorithmic and secondary sources.

Changing Clinical Practice (Step 7). Results from steps 4 to 6 were presented to clinicians. Roundtable discussions were held to discuss methods of improving the data collection process. After changes were made to the algorithm or to clinical practice and a sufficient time had elapsed, steps 4 to 7 were repeated.

Statistical Analysis. The majority of analysis consists of descriptive analyses and tabulations. Where appropriate, Fisher’s exact tests were done, comparing categories of data accuracy and completeness at different time points of data collection.

Results

Steps 1 and 2. In January 2000, a literature search of articles from January 1990 through December 1999 yielded 264 articles (34 reviews). A parallel literature search yielded 94 non–small-cell lung cancer phase II/III clinical trial publications. The key clinical prognostic factors were disease stage, performance status, weight loss, and complete resection. Key outcomes were overall survival, disease-free survival or progression-free survival, response rate, postoperative complication rates, and treatment toxicities. In the pilot feasibility study (early stage, n = 40; late stage, n = 40), for early stage, overall survival/disease-free survival was available in 93% of cases; for late disease, overall survival/disease-free survival was available in 85%.

Steps 3 and 4. Figure 1 shows the algorithm developed for collecting prognostic and outcomes data. We chose 15 cases per year from each of the years 1993 to 1999 for review (n = 105). There were significantly more missing data in 1993 and 1994 because the older charts were not systematically stored off-site. Complete data on overall survival and disease-free survival/progression-free survival were found in MGH computerized records in 67 (64%) cases; an additional 13 (12%) cases required data from MGH physicians offices; 3 (3%) cases also required contact of referring or primary care physicians and 2 (2%) cases required contact of families or patients. Twenty cases had missing data. Of these 20 cases, an additional 9 cases had overall survival data obtained from either the social security death index or death certificates. A data entry error rate of 1.4% was reduced to 0.4% after we modified the database. After assessment of the algorithm by an oncologist, specific patterns of errors were found. First, the date of first clinic visit was mistakenly used to calculate overall survival, rather than the actual date of diagnosis. Second, some of the apparently missing toxicity data were found in nursing notes. A written procedure manual improved the efficiency of the data abstraction process, reducing the mean time per case review by 15 min (60-45 min).

Step 5. The quality of data was assessed on key clinical prognostic variables. Data were graded as complete, substandard, or missing based on definitions developed by an oncology expert panel (Table 1; R.S.H., G.L., J.T., L.S., T.J.L., P.F., and M.H.K.). For example, overall survival data were deemed complete in stage I to II non–small-cell lung cancer if vital status was known to the most recent 6 months or we were able to obtain at least 4 years of follow-up data (or until first event).

Both general lung cancer cases and specific subsets were evaluated (Fig. 2). Of the variables assessed, data on symptoms, response rate, and treatment-related toxicity (chemotherapy or radiation) were of poorest quality. Failure to assign a corresponding Common Toxicity Criteria grade accounted for the majority of substandard/missing toxicity data. Symptoms were also poorly recorded for palliative cases; physician documentation of the presence or absence of at least three of four major non–small-cell lung cancer symptoms (fatigue, chest pain, dyspnea, and cough) was complete in only 25% of cases.

Figure 1. Outcomes data collection algorithm. Algorithm used for data abstraction by research assistant (RA) incorporating internal quality control measures. (a) in order of review; MGH computerized records; MGH paper records; Social Security Death Index; referring/primary care records*; contact of families/patients; death certificates* (*, only if required). (b) examples of data abstracted: demographics; performance status; stage; weight loss; pathologic variables; date of diagnosis, progression, or death; last date known without progression; last date known alive; toxicities/grade; response data. (c) assessment of precision of data collected; assessment of data quality and completeness; comparing data with secondary sources (steps 4-6); roundtable discussions with clinicians (step 7).
Step 6. The accuracy of data collected using the algorithm was cross-checked using secondary sources (Fig. 3). Overall, the accuracy of data abstracted using the algorithm was high. When missing data were excluded, major discrepancies between data observed using the algorithm as compared with data abstracted by an independent assessor were only 0.6%. However, comparison with external sources highlighted specific areas of inaccuracy. Major discrepancies occurred in 4% of data on the type of surgical resection. This resulted from the inaccurate recording of a wedge resection as a lobectomy. Furthermore, there was a 2% discrepancy in the recording of the number of metastatic sites at diagnosis, which arose from inclusion of new metastatic sites that occurred outside a 3-month window allowable for this variable. The accuracy of data on performance status was low with just 15% agreement with the independent assessor.

The majority of the discrepancies were minor (assessing patients’ Eastern Cooperative Oncology Group performance status as 0-1 instead of 1). All major discrepancies were due to differences in interpreting the performance status scale. Typically, medical students and junior house staff rated the performance status as being worse, compared with senior medical oncology staff. We took the assessment of the most senior staff being worse, compared with senior medical oncology and junior house staff rated the performance status as 0-1 instead of 1 (Fig. 3). Overall, the accuracy of data abstracted using the algorithm was high. When missing data were excluded, the percentages were 60% and 76% (P = 0.02). For pancreatic cancer, data quality was poorest but also improved the most for chemotherapy toxicity and follow-up variables.

An esophageal case control study was recently initiated at Princess Margaret Hospital, Canada (G. Liu, principal investigator). Results from the first 15 patients in ROAS showed good quality data for stage and performance status (87% and 93% complete, respectively)

### Table 1. Definition of complete, substandard, and missing data in non–small-cell lung cancer patients

<table>
<thead>
<tr>
<th>Stage</th>
<th>Prognostic/outcomes variable</th>
<th>Adequacy of prognostic/outcomes data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>OS/DFS</td>
<td>Data complete</td>
</tr>
<tr>
<td></td>
<td>Status known within last 6 mo or lost to follow-up after 4 y</td>
<td>Lost to follow-up after 1-4 y</td>
</tr>
<tr>
<td>Follow-up</td>
<td>≥1 radiographic evaluation of the chest in each of the first 5 y after diagnosis</td>
<td>At least 1 radiographic evaluation of the chest in the first 3 y</td>
</tr>
<tr>
<td>Documentation</td>
<td>Documentation in pathologic records present</td>
<td>Documentation in pathologic records absent</td>
</tr>
<tr>
<td>Documentation</td>
<td>Documentation in surgical records present</td>
<td>Documentation in surgical records absent</td>
</tr>
<tr>
<td>Early/late</td>
<td>Response rate</td>
<td>A response is documented in notes but there is inadequate information to assess RECIST criteria</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>Negative toxicity noted and positive toxicity graded using CTC/RTOG criteria</td>
<td>Negative toxicity not noted or positive toxicity not graded</td>
</tr>
<tr>
<td>Radiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>toxicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late</td>
<td>OS</td>
<td>Lost to follow-up after 0.5-2 y</td>
</tr>
<tr>
<td></td>
<td>Vital status known to within last 6 mo or lost to follow-up after 2 y</td>
<td>Lost to follow-up within 6 mo of diagnosis</td>
</tr>
<tr>
<td></td>
<td>Status known to within last 6 mo or lost to follow-up after 2 y</td>
<td>Lost to follow-up after 0.5-2 y or clinical determination of progression without radiological or pathologic verification</td>
</tr>
<tr>
<td>Palliative</td>
<td>Review of systems includes at least 3 major symptoms (pain, dyspnea, cough, fatigue)</td>
<td>The mention of at least 1 but less than 3 major symptoms (pain, dyspnea, cough, fatigue)</td>
</tr>
</tbody>
</table>

NOTE: Definitions were developed by an oncology expert panel.
Abbreviations: OS, overall survival; DFS, disease-free survival; PFS, progression-free survival; RECIST, Response Evaluation Criteria in Solid Tumors; CTC, Common Toxicity Criteria; RTOG, Radiation Therapy Oncology Group.
respectively; \( n = 15 \)). Data on chemotherapy toxicity again were of poor quality (complete toxicity data in 27% of cases).

**Discussion**

The number of molecular prognostic and pharmacogenetic outcomes studies derived from case-control studies is increasing, as many case-control studies now collect biological samples (6, 9, 17-19). The present study evaluated the quality of molecular prognostic and pharmacogenetic outcomes obtainable from an observational (case-control) study. We showed that certain outcomes were of sufficient quality even when ascertained using standard retrospective methods. Specifically, in early-stage lung cancers, overall survival, R0 resection rates, postoperative complications, and follow-up were of reasonable quality. However, retrospective ascertainment of important late-stage outcomes such as symptoms, toxicity, and response rates were poor, and there were some problems with obtaining progression-free survival/disease-free survival. We discovered that the problems encountered in our standard retrospective data collection helped us to greatly improve our ROAS protocols. ROAS consisted of several steps: a development of an algorithm for collecting outcomes data that is comprehensive yet reproducible; testing and modifying the algorithm using real outcomes data to ensure precise and accurate information; and finally, development of systematic improvements in clinical practice that enhances the effectiveness of ROAS. This process of reevaluation is ongoing. ROAS incorporated quality assurance mechanisms, allowing us to test the quality of our data on a continual basis and make changes in real time to the data collection procedures.
Case-control studies are established for risk analysis (14). To undertake pharmacogenetic or molecular prognostic research using a case-series design derived from a case-control study, data on prognostic factors must be captured. However, prognostic factors are not synonymous with disease risk factors and thus may not have been collected as part of the original case-control study; retrospective collection of missing prognostic variables becomes necessary. Unlike risk studies, where the goal is to capture every case, outcomes studies may only analyze specific subgroups based on a common disease stage or treatments. These fundamental differences in study design may affect the quality and range of outcomes research that can be undertaken from an initial case-control study.

This study also highlights the difficulty both of standardizing retrospective data collection procedures and of using retrospective data derived from standard clinical practice. Data were missing or substandard for key prognostic variables including performance status, toxicity, and disease symptoms. Results from this research have led to changes in the conduct of clinical practice, which we outlined in Table 2. In addition, we learned that quality must be evaluated on a continuous basis. In many case-control studies, data cleaning occurs toward the end of the data collection period, but for outcomes analysis, data cleaning should be ongoing. For example, we discovered that changes in clinic personnel had a considerable effect on the quality of data; thus, all new personnel (including physicians-in-training) receive instruction on how to use the clinical note templates. Outside the scope of this study, but just as important, is the need to carry out ongoing quality control measures for the biological samples collected.

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Comparisons of prognostic factors with secondary sources. Data abstracted using algorithm were compared with data obtained by an independent assessor (oncologist blinded to algorithm results); data collected as part of concurrent clinical trial; or data collected by the Massachusetts General Hospital Cancer Registry. PS, performance status. *, number of metastatic sites.
Changes made to our data collection procedures have increased the range of pharmacogenetic research that can be undertaken using the case-control study. We now publish on toxicity, response rate, and progression-free survival in addition to overall survival, which would not have been possible before the institution of our quality assurance mechanisms (20, 21). In addition, we have successfully adapted this approach for use in parallel esophageal and pancreatic case-series studies at MGH. Similar procedures are being used in case-control and case-series studies at Dana-Farber Cancer Institute and at Princess Margaret Hospital.

There are limitations to applying the results of this study to other case-control studies seeking to carry out outcomes analyses. Although we evaluated ROAS in multiple tumor sites across three institutions and two countries, all the centers were large institutions with computerized order entry, results of diagnostic tests, and treatment data, which are essential elements to improving data quality. The plan for outcomes collection occurred early in each study, and early consultation with clinicians helped to push many of the procedural changes into the patient clinics quickly. However, even when these enhanced practice guidelines are ingrained into clinical practice, we still struggle to ensure that toxicity and performance status templates attached to every clinical note are properly filled out and with every visit; we have devised random checks of the quality of these data and report back to clinicians on their individual performances. Whereas these specific issues may not be relevant to every center, the importance of critically evaluating data collection techniques and data quality is relevant to all molecular prognostic and pharmacogenetic studies.

One concern is whether the dedication of time and resources to establishing ROAS was worth the outcome, given that the majority of variables were of at least moderately good quality. We believe that it is worth-while. The resources required for implementation were actually much lower each time we implemented it at a new center, tumor site, or institution. We found that the quality issues were similar across sites and centers, as were the solutions (see Table 2 and Figs. 2 and 3). This helped us to anticipate problems and focus resources to the right places early. As expected, efficiency was much higher if a problem was anticipated and fixed early, rather than several years later.

Understanding the role of molecular cancer prognostic factors is of great importance. Promising results must be translated into the clinical setting to enable tailored treatment approaches. Large case-control studies provide a ready source of patients and information for such case-series analyses. However, attention must be paid to the fundamental differences between the two study types. Our experience presented here shows that it is possible to use a case-control study for outcomes analysis; however, the implementation of standardized procedures (such as ROAS) and ongoing monitoring of data quality is key to the success of this approach.

### Acknowledgments

We thank Peggy Suen, Andrea Shafer, Daisy Chiu, Salvatore Mucci, Lucy-Ann Principe, and Drs. Rihong Zhai, Zhaoxi Wang, and Darren Tse. We also thank Dr. Bruce Chabner and all the staff at Massachusetts General Hospital Cancer Center who support the daily functioning of the research studies.

### References


A Rapid Outcomes Ascertainment System Improves the Quality of Prognostic and Pharmacogenetic Outcomes from Observational Studies

Penelope A. Bradbury, Rebecca Suk Heist, Matthew H. Kulke, et al.

Cancer Epidemiol Biomarkers Prev 2008;17:204-211.