Chapter 12

Methods for Correlational Studies

Francis Lau

12.1 Introduction

Correlational studies aim to find out if there are differences in the characteristics of a population depending on whether or not its subjects have been exposed to an event of interest in the naturalistic setting. In eHealth, correlational studies are often used to determine whether the use of an eHealth system is associated with a particular set of user characteristics and/or quality of care patterns (Friedman & Wyatt, 2006). An example is a computerized provider order entry (CPOE) study to differentiate the background, usage and performance between clinical users and non-users of the CPOE system after its implementation in a hospital.

Correlational studies are different from comparative studies in that the evaluator does not control the allocation of subjects into comparison groups or assignment of the intervention to specific groups. Instead, the evaluator defines a set of variables including an outcome of interest then tests for hypothesized relations among these variables. The outcome is known as the dependent variable and the variables being tested for association are the independent variables. Correlational studies are similar to comparative studies in that they take on an objectivist view where the variables can be defined, measured and analyzed for the presence of hypothesized relations. As such, correlational studies face the same challenges as comparative studies in terms of their internal and external validity. Of particular importance are the issues of design choices, selection bias, confounders, and reporting consistency.

In this chapter we describe the basic types of correlational studies seen in the eHealth literature and their methodological considerations. Also included are three case examples to show how these studies are done.

12.2 Types of Correlational Studies

Correlational studies, better known as observational studies in epidemiology, are used to examine event exposure, disease prevalence and risk factors in a population (Elwood, 2007). In eHealth, the exposure typically refers to the use of an eHealth system by a population of subjects in a given setting. These subjects may be patients, providers or organizations identified through a set of variables that are thought to differ in their measured values depending on whether or not the subjects were "exposed" to the eHealth system.

There are three basic types of correlational studies that are used in eHealth evaluation: cohort, cross-sectional, and case-control studies (Vandenbroucke et al., 2014). These are described below.

- *Cohort studies* – A sample of subjects is observed over time where those exposed and not exposed to the eHealth system are compared for differences in one or more predefined outcomes, such as adverse event rates. Cohort studies may be prospective in nature where subjects are followed for a time period into the future or retrospective for a period into the past. The comparisons are typically made at the beginning of the study as baseline measures, then repeated over time at predefined intervals for differences and trends. Some cohort studies involve only a single group of subjects. Their focus is to describe the characteristics of subjects based on a set of variables, such as the pattern of EHR use by providers and their quality of care in an organization over a given time period.

- *Cross-sectional studies* – These are considered a type of cohort study where only one comparison is made between exposed and unexposed subjects. They provide a snapshot of the outcome and the associated characteristics of the cohort at a specific point in time.

- *Case-control studies* – Subjects in a sample that are exposed to the eHealth system are matched with those not exposed but otherwise similar in composition, then compared for differences in some predefined outcomes. Case-control studies are retrospective in nature where subjects already exposed to the event are selected then matched with unexposed subjects, using historical cases to ensure they have similar characteristics.

A cross-sectional survey is a type of cross-sectional study where the data source is drawn from postal questionnaires and interviews. This topic will be covered in the chapter on methods for survey studies.

12.3 Methodological Considerations

While correlational studies are considered less rigorous than RCTs, they are the preferred designs when it is neither feasible nor ethical to conduct experimental trials. Key methodological issues arise in terms of: (a) design options, (b) biases and confounders, (c) controlling for confounding effects, (d) adherence to good practices, and (e) reporting consistency. These issues are discussed below.

12.3.1 Design Options

There are growing populations with multiple chronic conditions and healthcare interventions. They have made it difficult to design RCTs with sufficient sample size and long-term follow-up to account for all the variability this phenomenon entails. Also RCTs are intended to test the efficacy of an intervention in a restricted sample of subjects under ideal settings. They have limited generalizability to the population at large in routine settings (Fleurence, Naci, & Jansen, 2010). As such, correlational studies, especially those involving the use of routinely collected EHR data from the general population, have become viable alternatives to RCTs. There are advantages and disadvantages to each of the three design options presented above. They are listed below.

- *Cohort studies* – These studies typically follow the cohorts over time, which allow one to examine causal relationships between exposure and one or more outcomes. They also allow one to measure change in exposure and outcomes over time. However, these studies can be costly and time-consuming to conduct if the outcomes are rare or occur in the future. With prospective cohorts they can be prone to dropout. With retrospective cohorts accurate historical records are required which may not be available or complete (Levin, 2003a).

- *Case-control studies* – These studies are suited to examine infrequent or rare outcomes since they are selected at the outset to ensure sufficient cases. Yet the selection of exposed and matching cases can be problematic, as not all relevant characteristics are known. Moreover, the cases may not be representative of the population of interest. The focus on exposed cases that occur infrequently may overestimate their risks (Levin, 2003b).

- *Cross-sectional studies* – These studies are easier and quicker to conduct than others as they involve a one-time effort over a short period using a sample from the population of interest. They can be used to generate hypotheses and examine multiple outcomes and characteristics at the same time with no loss to follow-up. On the other hand, these studies only give a snapshot of the situation at one time point, making it difficult for causal inference of the ex-

posure and outcomes. The results might be different had another time period been chosen (Levin, 2006).

12.3.2 Biases and Confounders

Shamliyan, Kane, and Dickinson (2010) conducted a systematic review on tools used to assess the quality of observational studies. Despite the large number of quality scales and checklists found in the literature, they concluded that the universal concerns are in the areas of selection bias, confounding, and misclassification. These concerns, also mentioned by Vandenbroucke and colleagues (2014) in their reporting guidelines for observational studies, are summarized below.

- *Selection bias* – When subjects are selected through their exposure to the event rather than by random or concealed allocation, there is a risk that the subjects are not comparable due to the presence of systematic differences in their baseline characteristics. For example, a correlational study that examines the association between EHR use and quality of care may have younger providers with more computer savvy in the exposed group because they use EHR more and with more facility than those in the unexposed group. It is also possible to have sicker patients in the exposed group since they require more frequent EHR use than unexposed patients who may be healthier and have less need for the EHR. This is sometimes referred to as response bias, where the characteristics of subjects agreed to be in the study are different from those who declined to take part.

- *Confounding* – Extraneous factors that influence the outcome but are also associated with the exposure are said to have a confounding effect. One such type is confounding by indication where sicker patients are both more likely to receive treatments and also more likely to have adverse outcomes. For example, a study of CDS alerts and adverse drug events may find a positive but spurious association due to the inclusion of sicker patients with multiple conditions and medications, which increases their chance of adverse events regardless of CDS alerts.

- *Misclassification* – When there are systematic differences in the completeness or accuracy of the data recorded on the subjects, there is a risk of misclassification in their exposures or outcomes. This is also known as information or detection bias. An example is where sicker patients may have more complete EHR data because they received more tests, treatments and outcome tracking than those who are healthier and require less attention. As such, the exposure and outcomes of sicker patients may be overestimated.

It is important to note that bias and confounding are not synonymous. Bias is caused by finding the wrong association from flawed information or subject selection. Confounding is factually correct with respect to the relationship found, but is incorrect in its interpretation due to an extraneous factor that is associated with both the exposure and outcome.

12.3.3 Controlling for Confounding Effects

There are three common methods to control for confounding effects. These are by matching, stratification, and modelling. They are described below (Higgins & Green, 2011).

- *Matching* – The selection of subjects with similar characteristics so that they are comparable; the matching can be done at the individual subject level where each exposed subject is matched with one or more unexposed subjects as controls. It can also be done at the group level with equal numbers of exposed and unexposed subjects. Another way to match subjects is by propensity score, that is, a measure derived from a set of characteristics in the subjects. An example is the retrospective cohort study by Zhou, Leith, Li, and Tom (2015) to examine the association between caregiver PHR use and healthcare utilization by pediatric patients. In that study, a propensity score-matching algorithm was used to match PHR-registered children to non-registered children. The matching model used registration as the outcome variable and all child and caregiver characteristics as the independent variables.

- *Stratification* – Subjects are categorized into subgroups based on a set of characteristics such as age and sex then analyzed for the effect within each subgroup. An example is the retrospective cohort study by Staes et al. (2008), examining the impact of computerized alerts on the quality of outpatient lab monitoring for transplant patients. In that study, the before/after comparison of the timeliness of reporting and clinician responses was stratified by the type of test (creatinine, cyclosporine A, and tacrolimus) and report source (hospital laboratory or other labs).

- *Modelling* – The use of statistical models to compute adjusted effects while accounting for relevant characteristics such as age and sex differences among subjects. An example is the retrospective cohort study by Beck and colleagues (2012) to compare documentation consistency and care plan improvement before and after the implementation of an electronic asthma-specific history and phys-

ical template. In that study, before/after group characteristics were compared for differences using *t*-tests for continuous variables and χ^2 statistics for categorical variables. Logistic regression was used to adjust for group differences in age, gender, insurance, albuterol use at admission, and previous hospitalization.

12.3.4 Adherence to Good Practices in Prospective Observational Studies

The ISPOR[1] Good Research Practices Task Force published a set of recommendations in designing, conducting and reporting prospective observational studies for comparative effectiveness research (Berger et al., 2012) that are relevant to eHealth evaluation. Their key recommendations are listed below.

- Key policy questions should be defined to allow inferences to be drawn.
- Hypothesis testing protocol design to include the hypothesis/questions, treatment groups and outcomes, measured and unmeasured confounders, primary analyses, and required sample size.
- Rationale for prospective observational study design over others (e.g., RCT) is based on question, feasibility, intervention characteristics and ability to answer the question versus cost and timeliness.
- Study design choice is able to address potential biases and confounders through the use of inception cohorts, multiple comparator groups, matching designs and unaffected outcomes.
- Explanation of study design and analytic choices is transparent.
- Study execution is carried out in ways that ensure relevance and reasonable follow-up is not different from the usual practice.
- Study registration takes place on publicly available sites prior to its initiation.

12.3.5 The Need for Reporting Consistency

Vandenbroucke et al. (2014) published an expanded version of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement to improve the reporting of observational studies that can be applied in eHealth evaluation. It is made up of 22 items, of which 18 are com-

1 ISPOR – International Society for Pharmacoeconomics and Outcomes Research

mon to cohort, case-control and cross-sectional studies, with four being specific to each of the three designs. The 22 reporting items are listed below (for details refer to the cited reference).

- *Title and abstract* – one item that covers the type of design used, and a summary of what was done and found.
- *Introduction* – two items on study background/rationale, objectives and/or hypotheses.
- *Methods* – nine items on design, setting, participants, variables, data sources/measurement, bias, study size, quantitative variables and statistical methods used.
- *Results* – five items on participants, descriptive, outcome data, main results and other analyses.
- *Discussion* – four items on key results, limitations, interpretation and generalizability.
- *Other information* – one item on funding source.

The four items specific to study design relate to the reporting of participants, statistical methods, descriptive results and outcome data. They are briefly described below for the three types of designs.

- *Cohort studies* – Participant eligibility criteria and sources, methods of selection, follow-up and handling dropouts, description of follow-up time and duration, and number of outcome events or summary measures over time. For matched studies include matching criteria and number of exposed and unexposed subjects.
- *Cross-sectional studies* – Participant eligibility criteria, sources and methods of selection, analytical methods accounting for sampling strategy as needed, and number of outcome events or summary measures.
- *Case-control studies* – Participant eligibility criteria, sources and methods of case/control selection with rationale for choices, methods of matching cases/controls, and number of exposures by category or summary measures of exposures. For matched studies include matching criteria and number of controls per case.

12.4 Case Examples

12.4.1 Cohort Study of Automated Immunosuppressive Care

Park and colleagues (2010) conducted a retrospective cohort study to examine the association between the use of a CDS (clinical decision support) system in post-liver transplant immunosuppressive care and the rates of rejection episode and drug toxicity. The study is summarized below.

- *Setting* – A liver transplant program in the United States that had implemented an automated CDS system to manage immunosuppressive therapy for its post-liver transplant recipients after discharge. The system consolidated all clinical information to expedite immunosuppressive review, ordering, and follow-up with recipients. Prior to automation, a paper charting system was used that involved manually tracking lab tests, transcribing results into a paper spreadsheet, finding physicians to review results and orders, and contacting recipients to notify them of changes.

- *Subjects* – The study population included recipients of liver transplants between 2004 and 2008 who received outpatient immunosuppressive therapy that included tacrolimus medications.

- *Design* – A retrospective cohort study with a before/after design to compare recipients managed by the paper charting system against those managed by the CDS system for up to one year after discharge.

- *Measures* – The outcome variables were the percentages of recipients with at least one rejection and/or tacrolimus toxicity episode during the one-year follow-up period. The independent variables included recipient, intraoperative, donor and postoperative characteristics, and use of paper charting or CDS. Examples of recipient variables were age, gender, body mass index, presence of diabetes and hypertension, and pre-transplant lab results. Examples of intraoperative data were blood type match, type of transplant and volume of blood transfused. Examples of donor data included percentage of fat in the liver. Examples of post-transplantation data included the type of immunosuppressive induction therapy and the management method.

- *Analysis* – Mean, standard deviation and *t*-tests were computed for continuous variables after checking for normal distribution. Percentages and Fisher's exact test were computed for categorical variables. Autoregressive integrated moving average analysis was

done to determine change in outcomes over time. Logistic regression with variables thought to be clinically relevant was used to identify significant univariable and multivariable factors associated with the outcomes. *P* values of less than 0.05 were considered significant.

- *Findings* – Overall, the CDS system was associated with significantly fewer episodes of rejection and tacrolimus toxicity. The integrated moving average analysis showed a significant decrease in outcome rates after the CDS system was implemented compared with paper charting. Multivariable analysis showed the CDS system had lower odds of a rejection episode than paper charting (OR 0.20; $p < 0.01$) and lower odds of tacrolimus toxicity (OR 0.5; $p < 0.01$). Other significant non-system related factors included the use of specific drugs, the percentage of fat in the donor liver and the volume of packed red cells transfused.

12.4.2 Cross-sectional Analysis of EHR Documentation and Care Quality

Linder, Schnipper, and Middleton (2012) conducted a cross-sectional study to examine the association between the type of EHR documentation used by physicians and the quality of care provided. The study is summarized below.

- *Setting* – An integrated primary care practice-based research network affiliated with an academic centre in the United States. The network uses an in-house EHR system with decision support for preventive services, chronic care management, and medication monitoring and alerts. The EHR data include problem and medication lists, coded allergies and lab tests.

- *Subjects* – Physicians and patients from 10 primary care practices that were part of an RCT to examine the use of a decision support tool to manage patients with coronary artery disease and diabetes (CAD/DM). Eligible patients were those with CAD/DM in their EHR problem list prior to the RCT start date.

- *Design* – A nine-month retrospective cross-sectional analysis of EHR data collected from the RCT. Three physician documentation styles were defined based on 188,554 visit notes in the EHR: (a) dictation, (b) structured documentation, and (c) free text note. Physicians were divided into three groups based on their predominant style defined as more than 25% of their notes composed by a given method.

- *Measures* – The outcome variables were 15 EHR-based CAD/DM quality measures assessed 30 days after primary care visits. They covered quality of documentation, medication use, lab testing, physiologic measures, and vaccinations. Measures collected prior to the day of visit were eligible and considered fulfilled with the presence of coded EHR data on vital signs, medications, allergies, problem lists, lab tests, and vaccinations. Independent variables on physicians and patients were included as covariates. For physicians, they included age, gender, training level, proportion of CAD/DM patients in their panel, total patient visits, and self-reported experience with the EHR. For patients, they included sociodemographic factors, the number of clinic visits and hospitalizations, the number of problems and medications in the EHR, and whether their physician was in the intervention group.

- *Analysis* – Baseline characteristics of physicians and patients were compared using descriptive statistics. Continuous variables were compared using ANOVA. For categorical variables, Fisher's exact test was used for physician variables and χ^2 test for patient variables. Multivariate logistic regression models were used for each quality measure to adjust for patient and physician clustering and potential confounders. Bonferroni procedure was used to account for multiple comparisons for the 15 quality measures.

- *Findings* – During the study period, 234 physicians documented 18,569 visits from 7,000 CAD/DM patients. Of these physicians, 146 (62%) typed free-text notes, 68 (25%) used structured documentation, and 20 (9%) dictated notes. After adjusting for cluster effect, physicians who dictated their notes had the worst quality of care in all 15 measures. In particular, physicians who dictated notes were significantly worse in three of 15 measures (antiplatelet medication, tobacco use, diabetic eye exam); physicians who used structured documentation were better in three measures (blood pressure, body mass, diabetic foot exam); and those who used free-text were better in one measure (influenza vaccination). In summary, physicians who dictated notes had worse quality of care than those with structured documentation.

12.4.3 Case-control Comparison of Internet Portal Use

Nielsen, Halamka, and Kinkel (2012) conducted a case-control study to evaluate whether there was an association between active Internet patient portal use by Multiple Sclerosis (MS) patients and medical resource utilization. Patient predictors and barriers to portal use were also identified. The study is summarized below.

- *Setting* – An academic MS centre in the United States with an in-house Internet patient portal site that was accessed by MS patients to schedule clinic appointments, request prescription refills and referrals, view test results, upload personal health information, and communicate with providers via secure e-mails.

- *Subjects* – 240 adult MS patients actively followed during 2008 and 2009 were randomly selected from the EHR; 120 of these patients had submitted at least one message during that period and were defined as portal users. Another 120 patients who did not enrol in the portal or send any message were selected as non-users for comparison.

- *Design* – A retrospective case-control study facilitated through a chart review comparing portal users against non-users from the same period. Patient demographic and clinical information was extracted from the EHR, while portal usage, including feature access type and frequency and e-mail message content, were provided by IT staff.

- *Measures* – Patient variables included age, gender, race, insurance type, employment status, number of medical problems, disease duration, psychiatric history, number of medications, and physical disability scores. Provider variables included prescription type and frequency. Portal usage variables included feature access type and frequency for test results, appointments, prescription requests and logins, and categorized messaging contents.

- *Analysis* – Comparison of patient demographic, clinical and medical resource utilization data from users and non-users were made using descriptive statistics, Wilcoxon rank sum test, Fisher's exact test and χ^2 test. Multivariate logistic regression was used to identify patient predictors and barriers to portal use. Provider prescribing habits against patient's psychiatric history and portal use were examined by two-way analysis of variance. All statistical tests used p value of 0.05 with no adjustment made for multiple comparisons. A logistic multivariate regression model was created to predict portal use based on patient demographics, clinical condition, socio-economic status, and physical disability metrics.

- *Findings* – Portal users were mostly young professionals with little physical disability. The most frequently used feature was secure patient-provider messaging, often for medication requests or refills, and self-reported side effects. Predictors and barriers of portal

use were the number of medications prescribed (OR 1.69, $p < 0.0001$), Caucasian ethnicity (OR 5.04, $p = 0.007$), arm and hand disability (OR 0.23, $p = 0.01$), and impaired vision (OR 0.31, $p = 0.01$). For medical resource utilization, portal users had more frequent clinic visits, medication use and prescriptions from centre staff providers. Patients with a history of psychiatric disease were prescribed more MS medications than those without any history ($p < 0.0001$). In summary, MS patients used the Internet more than the general population, but physical disability limited their access and need to be addressed.

12.4.4 Limitations

A general limitation of a correlational study is that it can determine association between exposure and outcomes but cannot predict causation. The more specific limitations of the three case examples cited by the authors are listed below.

- *Automated immunosuppressive care* – Baseline differences existed between groups with unknown effects; possible other unmeasured confounders; possible Hawthorne effects from focus on immunosuppressive care.

- EHR *documentation and care quality* – Small sample size; only three documentation styles were considered (e.g., scribe and voice recognition software were excluded) and unsure if they were stable during study period; quality measures specific to CAD/DM conditions only; complex methods of adjusting for clustering and confounding that did not account for unmeasured confounders; the level of physician training (e.g., attending versus residents) not adjusted.

- *Internet portal use* – Small sample size not representative of the study population; referral centre site could over-represent complex patients requiring advanced care; all patients had health insurance.

12.5 Summary

In this chapter we described cohort, case-control and cross-sectional studies as three types of correlational studies used in eHealth evaluation. The methodological issues addressed include bias and confounding, controlling for confounders, adherence to good practices and consistency in reporting. Three case examples were included to show how eHealth correlational studies are done.

References

Beck, A. F., Sauers, H. S., Kahn, R. S., Yau, C., Weiser, J., & Simmons, J. M. (2012). Improved documentation and care planning with an asthma-specific history and physical. *Hospital Pediatrics, 2*(4), 194–201.

Berger, M. L., Dreyer, N., Anderson, F., Towse, A., Sedrakyan, A., & Normand, S. L. (2012). Prospective observational studies to address comparative effectiveness: The ISPOR good research practices task force report. *Value in Health, 15*(2), 217–230. Retrieved from http://www.sciencedirect.com/science/article/pii/S1098301512000071

Elwood, M. (2007). *Critical appraisal of epidemiological studies and clinical studies* (3rd ed.). Oxford: Oxford University Press.

Fleurence, R. L., Naci, H., & Jansen, J. P. (2010). The critical role of observational evidence in comparative effectiveness research. *Health Affairs, 29*(10), 1826–1833.

Friedman, C. P., & Wyatt, J. C. (2006). *Evaluation methods in biomedical informatics* (2nd ed.). New York: Springer Science + Business Media, Inc.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0, updated March 2011). London: The Cochrane Collaboration. Retrieved from http://handbook.cochrane.org/

Levin, K. A. (2003a). Study design IV: Cohort studies. *Evidence-based Dentistry, 7*, 51–52.

Levin, K. A. (2003b). Study design V: Case-control studies. *Evidence-based Dentistry, 7*, 83–84.

Levin, K. A. (2006). Study design *III*: Cross-sectional studies. *Evidence-based Dentistry, 7*, 24–25.

Linder, J. A., Schnipper, J. L., & Middleton, B. (2012). Method of electronic health record documentation and quality of primary care. *Journal of the American Medical Informatics Association, 19*(6), 1019–1024.

Nielsen, A. S., Halamka, J. D., & Kinkel, R. P. (2012). Internet portal use in an academic multiple sclerosis center. *Journal of the American Medical Informatics Association, 19*(1), 128–133.

Park, E. S., Peccoud, M. R., Wicks, K. A., Halldorson, J. B., Carithers Jr., R. L., Reyes, J. D., & Perkins, J. D. (2010). Use of an automated clinical management system improves outpatient immunosuppressive care following liver transplantation. *Journal of the American Medical Informatics Association, 17*(4), 396–402.

Shamliyan, T., Kane, R. L., & Dickinson, S. (2010). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology, 63*(10), 1061–1070.

Staes, C. J., Evans, R. S., Rocha, B. H. S. C., Sorensen, J. B., Huff, S. M., Arata, J., & Narus, S. P. (2008). Computerized alerts improve outpatient laboratory monitoring of transplant patients. *Journal of the American Medical Informatics Association, 15*(3), 324–332.

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gotzsche, P. C., Mulrow, C. D., Pocock, S. J., … Egger, M., for the STROBE Initiative. (2014). Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *International Journal of Surgery, 12*(12), 1500–1524. Retrieved from http://www.sciencedirect.com/science/article/pii/s174391911400212x

Zhou, Y. Y., Leith, W. M., Li, H., & Tom, J. O. (2015). Personal health record use for children and health care utilization: propensity score-matched cohort analysis. *Journal of the American Medical Informatics Association, 22*(4), 748–754.