**Chapter 9: Matched Case Control and Comparative Effectiveness**

Farrokh Alemi, Ph.D.,

Austin Brown

Arthur Williams, Ph.D.,

Etienne Pracht, Ph.D.

**Learning Objectives**

After reading this chapter, students will be able to:

* Identify data in electronic health records that can be used to evaluate the comparative effectiveness of different interventions
* Define a case that has received an intervention and matched controls who have not received an intervention
* Contrast outcomes for cases that have received an intervention to matched controls
* Test if the difference of cases and matched controls is statistically significant
* Visually display data for outcomes of cases and matched controls over time
* Discuss the variables that can be used to match controls to intervention cases.

**Key Concepts**

* Cases
* Matched controls
* Observation period
* Enrollment event
* Follow-up period
* Exposure to treatment
* Time to event
* Days in range
* Probability of adverse events
* Odds ratio
* Verification of matching
* Confidence interval
* Normal distribution
* Standard error.

**Chapter in a Glance**

Managers often need to compare the effectiveness of different interventions. One method of doing so is using matched case controls. This retrospective approach can be used in a variety of settings, including evaluating the impact of a program, the productivity of employees, or the market penetration of varying initiatives. Cases are selected from those who have received an intervention; controls are selected from those who have not. Cases are matched to controls based on relevant characteristics, after which the outcome of specific interventions is assessed. Data on which the analysis is based often come from electronic health records.

This chapter describes a nested, matched, case-control design using retrospective data. It defines the enrollment, observation and follow-up time periods, as well as how cases and controls are matched. Finally, it describes statistical procedures for verifying that matching was done correctly, and evaluating the statistical significance of an intervention.

**Widespread Application**

Managers are often called upon to make judgments on the comparative effectiveness of various interventions. Here are some examples:

* **Marketing Decisions:** In marketing, cases and matched controls are needed to examine the effectiveness of competing marketing initiatives. For example, Hollon and colleagues used this technique to evaluate the effectiveness of direct-to-consumer marketing efforts.[[1]](#endnote-1)
* **Strategic Planning:** In strategic planning, this method can be used to assess the likelihood of success of different plans. For example, Mattes and colleagues used this method to evaluate the likelihood of commercial success of new inventions.[[2]](#endnote-2)
* **Quality Control and Process Improvement:** Case control studies are often used to assess quality of care. Consider the situation where there is a change in a clinical process: cases represent patient experience after a change is made; matched controls represent patients who were treated before the change was put in place. For example, Sundberg et al. contrasted the cost and effectiveness of integrated pain management. They identified pain patients who had received integrated management and contrasted them to pain patients who had received conventional care but had the same pain diagnosis, age, gender and socio-demographics.[[3]](#endnote-3) Others have used matched case control approaches to study quality factors that lead to unplanned readmissions.[[4]](#endnote-4) In another study, Danielsen and Rosenberg showed how patient education could reduce the cost of care using a matched case control study.[[5]](#endnote-5) In still another study, Grammatico-Guillon and colleagues used matched case control to monitor a hospital discharge database for hip and knee arthroplasty-related infections.[[6]](#endnote-6) Anantha and colleagues used matched case control to examine the cost and timing of care (day time or night time) for emergency general surgery.[[7]](#endnote-7) Dykes and colleagues used it to improve patient safety.[[8]](#endnote-8)
* **Health Information System Evaluation:** Case control approaches can be used to evaluate the effectiveness of electronic health record systems. For example, the relationship between computerized provider order entry and pediatric adverse drug events can be assessed,[[9]](#endnote-9) as can the effectiveness of care received remotely via telemedicine compared with that received in a typical hospital or physician office setting.[[10]](#endnote-10) Matched case controls have also been used to evaluate public health and occupational health programs.[[11]](#endnote-11)
* **Finance and Cost Effectiveness:** The use of case control studies to conduct cost effectiveness analysis is common.[[12]](#endnote-12),[[13]](#endnote-13),[[14]](#endnote-14) A retrospective matched case-control study was conducted to assess the financial impact of treating ventilator-associated pneumonia. The analysis provided the first demonstration of significant, sustained reductions in pediatric ventilator-associated pneumonia rates following the implementation of a costly prevention bundle.[[15]](#endnote-15) Others have used this method to examine the profitability of business operations,[[16]](#endnote-16) to examine hospital closure,[[17]](#endnote-17) to examine the cost effectiveness of robotic surgery, [[18]](#endnote-18) and to examine expenditures before and after surgical interventions.[[19]](#endnote-19)
* **Predictive Medicine:** A novel application was made in predictive medicine using the matched case control approach. Data from an electronic health record were used to select cases from Geisinger Clinical primary care patients with a diagnosis of heart failure. Controls were randomly selected matched based on sex, age, and clinic. The study demonstrated that it was possible to predict heart failure six months before a clinical diagnosis was made.
* **Human Resource Decisions**: The U.S. Army used the matched case control method to assess risk factors for disability retirement among its personnel.[[20]](#endnote-20) Matched case controls have also been used to evaluate the effectiveness of pre-employment screening.[[21]](#endnote-21)

As this demonstrates, the matched case control approach has broad application. Learning this method can lay a strong foundation for effective managerial decisions.

**Comparative Effectiveness Studies**

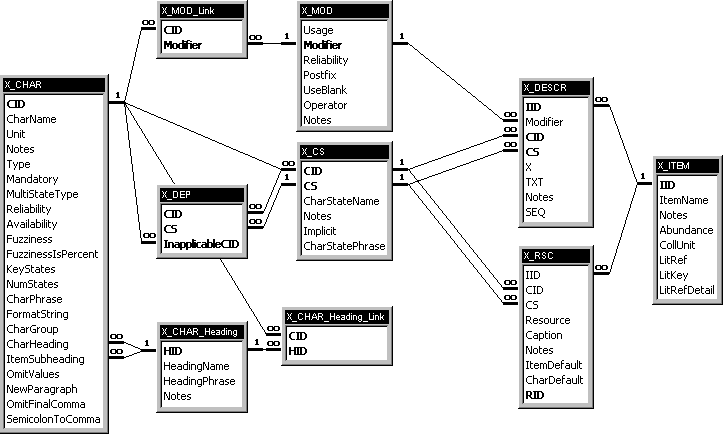
In recent years there has been growing interest in comparative effectiveness studies, partially due to the increased use of electronic health records which have made these techniques more accessible to a wider group of practitioners and researchers. The gold standard for medical research is the prospective randomized clinical trial (RCT), a rigorous approach that provides unbiased information about the impact of an intervention. The RCT does, however, have several drawbacks: (1) it involves costly data collection, (2) it restricts study to pre-defined eligible populations such as those without comorbidities, and (3) it denies access to some level of care for patients in the control group. By comparison, comparative effectiveness research (CER) allows use of data collected in the course of caring for patients. Although CER provides less rigorous conclusions, its retrospective approach enables all patients to be considered for inclusion in the study. Data from electronic medical records (EMR) and other electronic data sources are used to evaluate the impact of interventions with statistical methods. Although there are limitations, these techniques have yielded surprising and important insights into clinical care. Moreover, studies based on use of electronic and administrative data are generally less expensive and can be completed more quickly than studies based on randomized clinical trials.

Many different techniques have been developed to conduct comparative effectiveness studies, [[22]](#endnote-22) and none is without its critics.[[23]](#endnote-23) The chief complaint is often that different comparative effectiveness methods can lead to contradictory conclusions.[[24]](#endnote-24) Contradictions can occur because conclusions are based on nonrandom data and observations drawn from a wide variety of disparate sources including databases for insurance claims, prescription histories, national registries, and patient treatment records. This illustrates both the problem and its solution: lacking true random sampling, studies must be carefully designed to ensure that data are representative of the larger population for the characteristics being assessed; moreover, it must be possible to measure outcomes with variables available in the database.[[25]](#endnote-25) This chapter describes procedures for conducting a retrospective matched case control comparative effectiveness study.

**Source of Data**

Data for retrospective comparative effectiveness research (CER) is usually obtained from electronic health records. These data may include prescriptions, diagnoses, records from hospitalizations and outpatient care, clinician’s notes and dates of encounters. Data are usually obtained for a well-defined number of recent years that exceed both the planned observation prior to enrollment in the program, and the follow-up years after enrollment.

**Figure 1: Example of a Relational Database**[[26]](#endnote-26)



Statisticians are used to matrix data structures with cases in rows of a table and variables in columns. These types of data structures have sparse entries since many variables are not relevant in every case. In contrast, data in electronic health records are distributed in numerous smaller but dense tables. For example, all information about patient characteristics (e.g. date of birth, date of death) is available in one table (see left side of Table 1); information about encounters is available in other tables (see right side of Table 1), and yet another table provides information about laboratory findings. In modern electronic health record systems, millions of data elements can be distributed in thousands of tables. The analysis of data starts with becoming familiar with the data structure. The first challenge in performing a comparative effectiveness study is to aggregate data in a format that can be used for statistical analysis.[[27]](#endnote-27)

**Table 1: Patient Data & Visit Data Are in Two Different Tables**



In a relational database, each table is a set of information about a specific variable or primary key. As an example, for a table of diagnosis codes the primary key is DIAGNOSIS\_CODE, and the information in the table are possible values for the variable. In another example, a table on patients lists the patient’s medical record number as the primary key and patient’s name and birthday as other variables (see left side of Table 1). A table on visits (see right side of Table 1) has encounter ID and diagnoses ID but not the description of the diagnosis, a patient ID but not the patient characteristics, a provider ID but no other information about the provider.

Standard Query Language, (SQL) is used to prepare the data for analysis. Using SQL, the investigator uses the JOIN command to include data from multiple tables using each table’s primary key. This allows the investigator to connect the visit table to the patient table and thus be able to read the date of birth of the patient. It also enables joining the visit table to the diagnoses table, allowing the description of the patient’s diagnoses to be read. Knowledge of SQL is needed for preparing data electronic health records for statistical analysis.

Besides the JOIN function, SQL provides other functions to filter, count and average data. These commands can be learned quickly, and enable preparation of complex data in formats suitable for statistical analysis. Detailed instructions on use of SQL can be found at different locations including <http://openonlinecourses.com/databases>. Moreover, almost all common errors and methods of combining data can be found on-line, and there are many sites where experienced SQL programmers will help novices solve data transformation problems.

**Study Design and Methods**

In observational studies there is no random assignment of patients to groups. Consequently, observed outcomes may be due to a patient’s condition and not related to treatment. A matched case control study provides a comparison group for patients who have received the treatment, and thus reduces the possibility of erroneous attribution of findings.

The approach taken in case control studies has a long history. One of the earliest examples comes from the famous 1854 cholera epidemic in London in which it was demonstrated that most of those who died drew water from the same Broad Street pump.[[28]](#endnote-28) Case control studies also were used in many important studies in the 1920s, but truly came to prominence in the 1950s with studies that demonstrated the unexpectedly strong relationships between smoking and cancer.[[29]](#endnote-29) Use of matched case control studies in the analysis of data from electronic health records is common, and considerable advances in theory, methods, and practice of case control designs have been and are being made in epidemiology and biostatistics.[[30]](#endnote-30)

**Definition of Cases and Controls**

Patients who receive an intervention are referred to as “cases”. Patients who do not receive an intervention are referred to as “controls.” For example, patients who were admitted to the Veteran Administration’s Medical Foster Home (MFH) program (an alternative to nursing home care) may be considered cases and patients in the traditional nursing home program may be considered controls. MFH allows patients to rent their own room in a community home while receiving medical and social services from the VA in this community setting.

The identification of cases in the medical record can be difficult as these databases typically record utilization of services and not necessarily participation in a program or a need for care. There are at least two methods of identifying a case. First, a case can be identified by examining the medical record for a unique clinical event of interest. A clinical event can be a physician office visit, inpatient admission, or emergency room visit. For a study of heart failure, for example, a clinical event could be an initial diagnosis of congestive heart failure. Typically these events are defined using codified nomenclatures such as the International Classification of Diseases (ICD-9/10). The Healthcare Cost & Utilization Project of the Agency for Healthcare Research and Quality has defined how various diagnoses codes correspond to common disease categories.[[31]](#endnote-31). For example, heart failure can have one of the following ICD-9 codes: 402.01, 402.11, 402.91, 425.1, 425.4, 425.5, 425.7, 425.8, 425.9, 428.0, 428.1, 428.2, 428.21, 428.22, 428.23, 428.3, 428.31, 428.32, 428.33, 428.4, 428.41, 428.42, 428.43, or 428.9. Other examples include falls,[[32]](#endnote-32) injuries,[[33]](#endnote-33) medication errors,[[34]](#endnote-34) mood and anxiety problems,[[35]](#endnote-35) and hospitalization encounters.

Second, a case can be identified by examining admission to a program. For example, in the MFH project, the providers listed patients for whom they provided care. Patient’s scrambled social security numbers were used to identify them within the electronic health record. These patients were compared to patients in nursing homes, as MFH is an alternative to nursing home care. Nursing home patients were identified through admission and discharge dates for the nursing home information available in the medical record of the patients.

**Measurement of Exposure to Treatment**

In defining cases and controls, consideration should be given to the extent of “exposure” to an intervention. A sufficient exposure should be allowed so that a change in the outcome being evaluated can be expected. For example, the day after enrollment in MFH care, no change in patient outcome is expected since a person enrolled for one day is not considered to have received the full benefit of enrollment. Sometimes, patients enroll and dis-enroll shortly afterwards. In an on-going VA study, it was assumed that three months of enrollment is needed before a patient can be considered an MFH patient. A similar timeframe is used for controls in nursing homes. This excludes short stays – those that reside in nursing homes for less than three months.

Some patients receive both the intervention and the control programs. For example, a patient may enroll for MFH at first but after months of enrollment leave it for care in a nursing home. A patient’s enrollment in a case or control group is for a specific period. Since the same patient has spent time in both groups, they may appear to be an ideal match for themselves. The case and control match many features, with one exception - the case and control are examined in different timeframes. Unfortunately, transition from one intervention to another is almost always accompanied with a major crisis that affects patient’s health. In these situations, the same patient before and after has a different health status. For example, in Figure 2, we see information on the blood pressure of one patient. For seven years, this patient was in a nursing home. At end of the seventh year there was a hospitalization, shown as a circle. Following this hospitalization, the patient was discharged to the Medical Foster Home. The blood pressure values during year eight show the patient’s condition in the Medical Foster Home program. The conditions immediately prior year eight show blood pressure when the patient was in a nursing home. The patient’s condition worsened right before the transfer from nursing home to medical foster home program. Blood pressure can be compared before and after transfer if these data exist; without them, the patient’s experience in the two timeframes cannot be compared.

When patients are classified into cases or controls over different time periods the analytical methods for assessing treatment effects become statistically and conceptually complex. Analysis must explicitly consider “person-time”; that is, the amount of time that each patient spends as a case or control. Additionally, matching patients over time is often difficult, since time can be an important covariate with treatment results, and matching eliminates opportunities for statistical analyses of this variable.

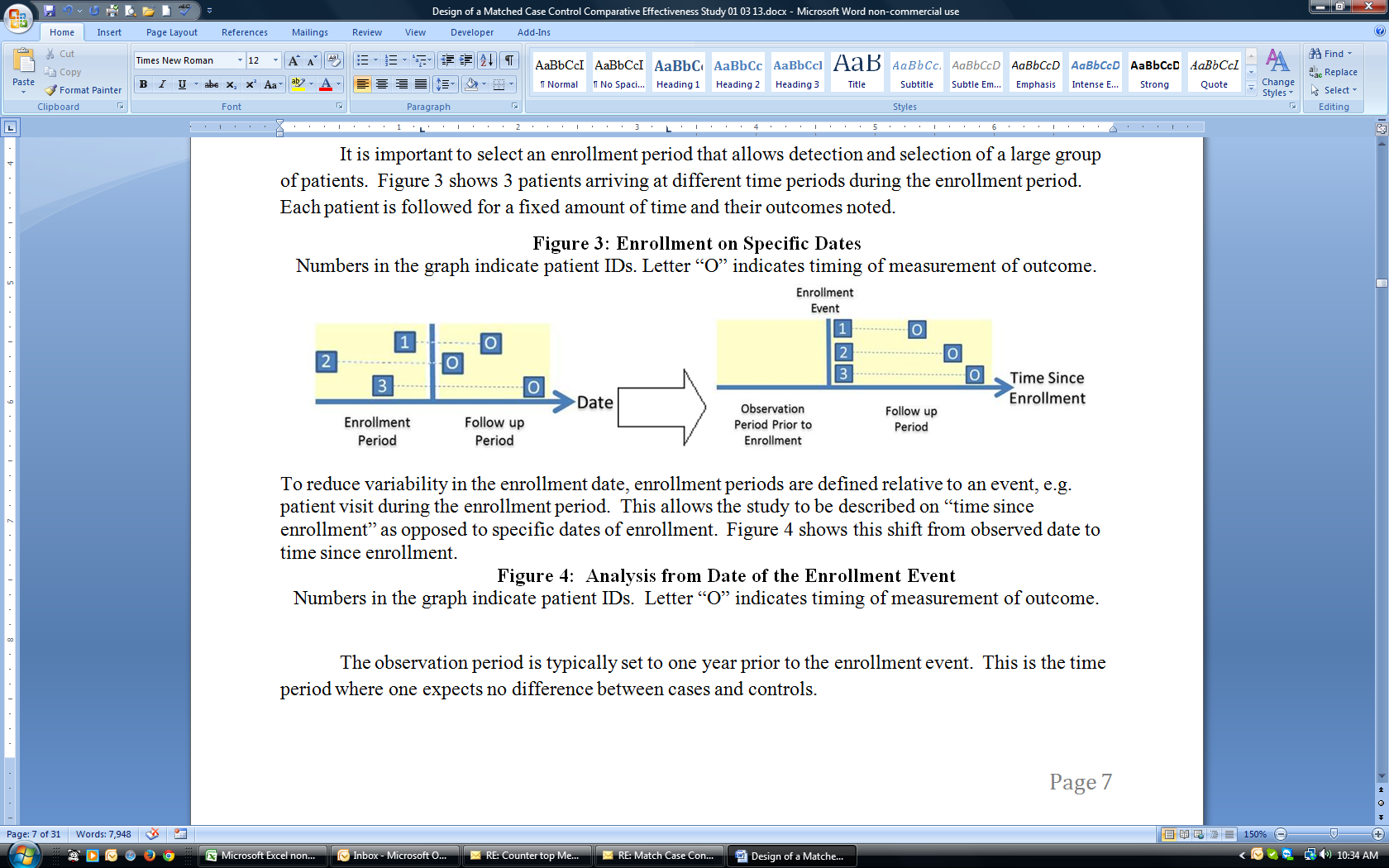
**Figure 2: Patient Transitions among Care Venues**



**Enrollment and Observation Period**

It is important to choose an enrollment period that allows selection of a large group of patients. On the left side of Figure 3, patients arrive at different times during the enrollment period. Each patient is followed for an amount of time and their outcomes noted. The enrollment period is defined relative to the enrollment event. The left side of Figure 3 shows a graph of data based on date of visits. The right side of Figure 3 shows the same data based on time since first visit, demonstrating that patients are followed for different intervals until the outcome of interest occurs.

**Figure 3: Enrollment on Specific Dates**Numbers in the graph show patient IDs. Letter “O” indicates timing of measurement of outcome.



The observation period is typically set to one year prior to the enrollment event. During this period, no difference between cases and controls is expected. In fact, by design, controls are selected so that there is no significant difference between cases and controls prior to enrollment.

As an example, consider the data in Table 2. Dates of various events are given as well as the calculated time since enrollment in the program. Suppose enrollment is any diabetic visit after January 1; the follow up period is any relevant event within one year of enrollment. The patient depicted in Table 2 visits the physician on January 9. This is the first diabetic visit during the enrollment period, so this becomes the enrollment event. Over the next thirteen months, the patient has several encounters with their physician as well as two ED visits. On both occasions, the ED visits were the results of falls. The time since enrollment in the study is the number of days between these fall events and the enrollment date. As Table 2 shows, one of these fall events was within one year of enrollment and therefore during the follow-up period. In analyzing data from electronic health records, it is important specify the enrollment event and follow up and observation periods as these time intervals change the data.

**Table 2: A Sample Patient and His Encounters**



**Matching Criteria**

Controls differ from cases in distinct ways. For example, controls may be older or have different comorbidities - differences that could change outcomes. Consequently, it is important to match controls to cases to ensure that such differences are not attributed to the intervention. The variables used to match controls to cases differ from study to study, but often involve one or more of the following:

1. Same observation period. Controls are often selected from the same enrollment period. When this is not done, the differences in case and controls could be due to changes that have occurred over time in the environment, or provider’s learning.
2. Same age on admission. Controls are often randomly selected so that patient ages are the same as that for cases.
3. Same gender on admission. The gender of the control and the case is often set to match.
4. Same comorbidities. Controls are randomly selected so that comorbidities match those of cases.

Another way of controlling for differences prior to enrollment is to choose controls so that they match cases in outcomes observed in the period prior to enrollment. In this fashion, case and controls have the same history. The variables typically used in this fashion include history of hospitalization, medical errors, falls, and use of mental health services. Matching by history, however, can be difficult when all relevant medical conditions data are not in the EMR and cannot be easily found in other sources.

For continuous variables, a control can be said to match the case if controls are within one standard deviation of the case. This general rule may not apply when the literature or previous research indicates absolute differences or other criteria that should be used for matching. For discrete variables, the two match if the control and the case have the same value. For each case, three matching controls may be selected, because literature and research show that the statistical power is not improved beyond a ratio of three controls to each case.[[36]](#endnote-36) If more than three controls are available, the decision on which three should be selected can be done randomly. To illustrate an example of how matching occurs, examine the data in Table 3. This table shows data for two cases who received an intervention and seven who did not. These patients are to be matched based on age, identifying as potential controls patients within one standard deviation. Control patients whose age is within one standard deviation for each case patient are eligible; selection of specific control patients is accomplished by choosing those with the lowest random number.

**Table 3: A Procedure for Matching Controls to Cases**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Patient ID** | **Received Intervention** | **Age** | **Months to Fall** | | **Random Number** | **Case or control?** |
| **Observation Period** | **Follow-up Period** |
| 1 | Yes | 65 |  | 3 | 0.24 | Case |
| 2 | No | 60 |  | 2 | 0.85 |  |
| 3 | Yes | 84 | 2 |  | 0.64 | Case |
| 4 | No | 82 | 4 |  | 0.7 | Control |
| 5 | No | 78 |  |  | 0.87 |  |
| 6 | No | 80 | 3 |  | 0.72 | Control |
| 7 | No | 79 |  |  | 0.86 |  |
| 8 | No | 64 |  |  | 0.16 | Control |
| 9 | No | 70 |  | 2 | 0.17 | Control |
|  | | | | | | |
| **Standard Deviation** | | 8.89 |  | | | |

Suppose we need to randomly select two controls for each case. For case 1, we choose patients 8 and 9. For case 2, we also choose patients 4 and 6. Let’s look at the calculations for this example step-by-step.

**Step 1**: Calculate the standard deviation for age across all nine patients. The average age in Table 2 is 73.6 years. The standard deviation, S, can be calculated using the formula:

Where is the ith observation from n observations is the average of the observations.

**Step 2**: For each patient that received the intervention, calculate a one standard deviation interval based on age (see Table 4).

**Table 4: Sample Calculations of Matching Controls to Age of Cases**

|  |  |  |  |
| --- | --- | --- | --- |
| Case Patients | Control Case Age | Age – One Std. Deviation | Age + One Std. Deviation |
| Patient 1 | 65 | 65 – 8.89 = 56.1 | 65 + 8.89 = 73.9 |
| Patient 3 | 84 | 84 – 8.89 = 75.1 | 84 + 8.89 = 92.9 |

**Step 3**: Identify eligible controls that fall within one standard deviation of the age of the case. Select two from among the eligible controls for each case with the lowest random numbers.

* Patient 1 is the first Case Patient with an age of 65. Patients 2, 8, and 9 fall within one standard deviation of this age. Patients 8 and 9 have the lowest random numbers, and are selected as controls.

* Patient 3 is the second Case patient with an age of 84. Patients 4, 5, 6 and 7 fall within one standard deviation of this age. Patients 4 and 6 have the lowest random numbers, and are selected as controls.

These selections provide us with assignments of controls to cases matched on age. Other variables could also be used for matching, and the decision of which matching variables should be used is a crucial decision. One common method to reduce uncertainty about the appropriate variable to match is to rely on the outcomes observed during the observation period. In this fashion, cases are matched to the control on history of the outcomes that are being monitored.

**Measurement of Outcomes**

Outcomes are typically defined over a range of values over time. Patients may be in and out of the therapeutic range on several days; consequently, it is important to calculate the percent of time patients are within the therapeutic range; this procedure was first proposed by Rosendaal.[[37]](#endnote-37) Figure 5 shows an example of a patient going in and out of range for blood pressure. In Figure 5, times when systolic blood pressures are within the range of 120 mmHG to 140 mmHG are indicated.

**Figure 5: Examples of Blood Pressures in and out of Therapeutic Range**

For two consecutive values that are in the range, the calculation is straightforward. For two consecutive values where one is in the range and another is out of range a linear extrapolation is made to determine the day the patient moved out of range. A preset maximum (e.g. 60 days) is used to reduce the influence of the linear extrapolation for two values that are very far apart. The percent of days the patient is in the therapeutic range is calculated as the sum of all estimated days in range divided by the number of days from first to last measure. For two values at two consecutive measurements, the extrapolation is based on the following formula:

“Unknown days” refer to the difference of “days in range” and the maximum number of days that can be estimated from two measures. These are days that are neither in nor out of range. The percent of days in range is then calculated using the following approximation:

For example, in Table 5 we see six blood pressures, some in and some out of range. In the first two measurements, the patient goes from blood pressure of 125 to out of range pressure of 150. These two measures are 20 days apart; this patient is estimated to be in the range for 12 of these 20 days. Similar estimates are made for every two consecutive blood pressure. For thesecond and third measurement the estimate of days in the range is 20 days. If we assume that days more than 2 weeks away from the measurement are “unknown days”, then 6 out of these 20 days are unknown days. In the last two measurements, both measures are within range; there are 30 days between these measures, 16 of which are unknown days. The total number of days the patient is in the range is 77 days, the total number of unknown days is 22 and the percent of days in the therapeutic range are 56%.

**Table 5: Calculation of Percent Days in Therapeutic Range**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Days Since Enrollment** | **BP** | **Therapeutic Range** | | **In range** | **Days in Range** | **Unknown Days > Max of 14 Days** |
|
| **Lower** | **Upper** |
| 0 | 125 | 120 | 140 | Yes | 12 | 0 |
| 20 | 150 | 120 | 140 | No | 20 | 6 |
| 60 | 130 | 120 | 140 | Yes | 5 | 0 |
| 70 | 110 | 120 | 140 | No | 10 | 0 |
| 90 | 130 | 120 | 140 | Yes | 30 | 16 |
| 120 | 135 | 120 | 140 | Yes |  |  |
|  |  | **Total Days in Range** | | | 77 | 22 |
|  | **Percent Days in Therapeutic Range** | | | | 56% |  |

At least two observations over time are needed before the “Percent Days in Therapeutic Range” can be calculated. A detailed example of these calculations is provided in Appendix A.

**Verification of Matching**

Cases and controls are compared before cases are enrolled to verify that matching was successful. The null hypothesis is that the cases and the matched controls have the same rate for outcomes during the pre-enrollment period. The alternative hypothesis is that there is a statistically significant difference between these two groups. First, the difference between each case and the mean of the controls that match the case are calculated; then the average and the standard deviation of the differences are calculated. The t-statistic is used to test if the mean of the differences is significantly different from zero. If so, the null hypothesis is rejected, and the matching process has not led to equivalent groups. If equivalent groups are not available, the procedure is repeated until two statistically equivalent groups are identified.

In many studies, the calculation of the statistics is made in error.[[38]](#endnote-38) A paired t-test can be done when cases have 1:1 match to controls and have a symmetrical distribution, the Wilcoxon, signed ranks test is used for measured non-parametric results with 1:1 matching, linear or logistic regression is used to handle matched sets other than 1:1 matching.[[39]](#endnote-39) The section below shows how to test the statistical significant of results when many controls have been matched to one case. Applying the same procedures to pre-intervention period is one way of testing that the method for matching controls to cases has been effective.

**Analysis of Outcomes**

Outcomes of case and control patients can be analyzed using the odds ratio of observed outcomes.  First, the number of outcomes in cases and controls during the follow-up period are counted.  Table 6 shows the presence of these outcomes with positive and absence with negative symbols.   So and is the number of cases and controls that had positive outcome.  Likewise, and   are the number of cases and controls that had negative outcomes. 

**Table 6:  Contingency Table for Adverse Outcomes**

|  |  |  |
| --- | --- | --- |
| **Outcome =Fall** | **Cases** | **Controls** |
| **Yes** | = 1 | = 1 |
| **No** | = 1 | = 3 |
| **Total** | = 2 | = 4 |

The probability of the outcomes among the cases is calculated as the ratio of the number of observed outcomes divided by the number of possible outcomes:

The odds ratio of positive outcomes is calculated as the ratio of these two probabilities and is written as:

An odds ratio greater than 1 suggests more positive outcomes among the cases than controls; an odds ratio less than one suggests the reverse. An odds ratio of 1 suggests that the probability of the positive outcomes in Cases and Controls is the same.

The odds ratio is a point estimate.  A more useful concept is to a have a confidence interval around this ratio.  Confidence intervals are created so that the probability of the true odds ratio falling outside of the interval is very small, say less than 5%.  If there are a large number of cases in the study, the distribution of the natural log odds ratio, shown below as L, is approximately Normal.

In a Normal distribution, 95% of the data would fall within 1.96 standard deviation of the mean.  The standard deviation of the log of the odds ratio can be calculated as:

In a Normal distribution, values that are more than 1.96 standard deviations away from the mean are relatively rare and occur less than 5% of time. Therefore, the approximate 95% confidence interval for the population log odds ratio was estimated as:[[40]](#endnote-40)

The above confidence limits are for the natural log of the odds ratio.  The confidence limit for the odds ratio, itself, is then given by:

If this confidence interval does not include 1, then the hypothesis that cases and controls have the same outcomes has less than 5% chance of occurrence.  Then, one could reject the hypothesis with relative confidence.  The above procedure shows how the statistical significance of the difference between cases and controls are examined.

Table 7 applies these concepts to our data. Clearly, we have very few cases and controls and one would not expect the findings to be significant nor appropriate for test of confidence interval. We apply the proposed method to these data only to explain the use of the formulas. We can examine the outcomes during the follow-up period among cases (i.e. 1 and 3) and controls (i.e. 2, 4, 8, and 9). From Table 5, the test statistic L is calculated as:

= (-2.5, 4.7)

This confidence interval must now be translated back into the original units:

Confidence interface = (, = (0.1, 107.4)

The confidence interval for the ratio excludes 1. Therefore, the hypothesis of significant difference between cases and controls can be rejected.

**Table 7: Calculation of Odds Ratio for Example Cases and Controls**

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Cases** | **Controls** |
| **Yes** | 1 | 1 |
| **No** | 1 | 3 |
| **Total** | 2 | 4 |
|  |  |  |
| **L=** | 1.1 |  |
| **Variance=** | 2.08 |  |
| **Std Dev.** | 1.44 |  |
| **1.96 Std. Dev. P** | 3.58 |  |
| **Upper Limit=** | 107.45 |  |
| **Lower Limit=** | 0.08 |  |

Many policymakers prefer visual displays. Typically one plots the data using the probability of the outcome before and after enrollment. Before enrollment in the program, the rate is 50% positive for cases and 50% positive for controls. This is the overlapping point in Figure 6 and shows that before enrollment in the program cases and controls had the same rate of falls. In the follow-up period, the percent of patients with fall changes. Both groups have a decline in the rate of falls but the decline is more in the cases than controls.

**Figure 6: Impact of Enrollment in Program on Patient Falls**

The problem with Figure 6 is that it does not show the period of time that elapses before a fall occurs. Since adverse outcomes often happen eventually it is important to trace the time to the event.

**Analysis of Time to Event**

Sometimes the outcome of interest is not a count of events (e.g. falls), but days, or in our case, months to the event (person-time in epidemiology). The analysis of outcomes over time is made using a Kaplan-Meier estimator.[[41]](#endnote-41) It is used to report the time until adverse outcomes occur. The advantage of using this estimator is that it can accommodate changes in patients available to the study. For example, if patients die before the outcome of interest is measured, they can be left out of the calculation for periods that they were not available. For another example, if patients change their place of care, then they can be excluded from the study from the time that they changed providers. During each period, one can recalculate the number of patients at risk and therefore patients who drop out of the study do not affect the calculations.

During each period, the number of patients at risk of falling (number of patients in the study minus the patients that are censored) and the number of patients that do not fall (number at risk minus number that have fallen) is calculated. This is the conditional probability of falling given that the patient has not fallen in previous time periods. The probability of falling in the period is calculated as the probability of falling in the past time period times the conditional probability of not falling. Table 8 shows the calculation of these probabilities for our data. Among the 2 cases, there was one fall in the 3rd month. Among the 4 controls, there was 1 fall in the 2nd month. There was one patient in the control group that died in the 1st month and therefore was not available to the study for the following months.

**Table 8: Probability of Falling at Different Time Periods**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Period** | **Number at risk** | **Number censored** | **Number Fell** | **Number No Fall** | **No Fall Probability** |
| **Cases** | 1 | 2 | 0 | 0 | 2 | 1 |
| 2 | 2 | 0 | 0 | 2 | 1 |
| 3 | 2 | 0 | 1 | 1 | 0.5 |
| **Controls** | 1 | 4 | 1 | 0 | 4 | 1 |
| 2 | 3 | 0 | 1 | 2 | 0.67 |
| 3 | 2 | 0 | 0 | 2 | 0.67 |

Figure 7 shows the resulting percent of patients having no falls at different time intervals. Cases had a fall sooner than the controls. The statistical significance of differences in rates of falls over time can be established by using the variance for Kaplan Meier statistic.[[42]](#endnote-42)

**Figure 7: Impact of Enrollment on Months to Falls**

**Measurement of Severity**

In comparative effectiveness research it is important to statistically control for differences among severity of illness of cases and controls. Without adequate control for severity of illness, one may mistakenly attribute differences in cases and controls during the follow-up period to the intervention, as opposed to initial differences in severity of illness. The matched case control tries to control for possible differences in the two groups prior to the intervention. It is reasonable to include severity as one of the variables on which cases and controls are matched. Using the history of patient outcomes is another way of also measuring severity as these two variables are often related. Unfortunately, measures of severity can differ in terms of reliability, and reliability is often ignored in such analyses when it should be specifically modeled or adjusted when known measures of reliability exist. It has already been suggested above that uses of patient history can have its own problems. No matter what is used for matching, it is important to verify as best as possible that the cases and controls did not differ at start in their severity of illness.

There are many approaches to measuring severity of illness from patient information available in medical records. One method uses patient laboratory findings, and another method uses patient diagnoses as coded in the International Classification of Diseases.[[43]](#endnote-43),[[44]](#endnote-44) Since data warehouses have a large number of cases, it is often possible to statistically derive relationships between the patient characteristics and outcomes from the data. In these circumstances, the features and the statistical model can be used as a measure of severity of illness. Also, if difficulties arise in the analyses, simulations, bootstrapping, or other techniques might be used to assess potential effects of the reliability of severity measures.

**Propensity Scoring**

We have shown how matched case control studies are organized. An alternative to matching is to use all data but weight the control patients to match the proportion of the cases on specific variables. Propensity scoring allows one to use all of the data while matching discards patients that do not match to cases, but the appropriateness of propensity scoring and its uses is still being debated in the literature.

**Summary**

This chapter has described how matched case controls can be organized and applied to data within electronic health records. These types of studies control for a number of possible ways in which observational studies may lead to erroneous conclusions. Like a quasi-experimental study, the matched case control protects against variables that affect both cases and controls, and the effect of time and enrollment period can be appropriately handled. There are a number of ways in which these methods could lead to erroneous conclusions. These include:

1. Poor selection of variables on which cases and controls are matched,
2. Inadequate measurement of severity of illness of patients,
3. Incomplete medical records, when patients visits outside the healthcare system is not reflected in the record,
4. Inadequate measurement of outcomes over time.

**References**

1. Hollon MF, Larson EB, Koepsell TD, Downer AE. Direct-to-consumer marketing of osteoporosis drugs and bone densitometry. Ann Pharmacother. 2003 Jul-Aug;37(7-8):976-81. [↑](#endnote-ref-1)
2. Mattes E, Stacey MC, Marinova D. Predicting commercial success for Australian medical inventions patented in the United States: a cross sectional survey of Australian inventors. Med J Aust. 2006 Jan 2;184(1):33-8. [↑](#endnote-ref-2)
3. Sundberg T, Petzold M, Kohls N, Falkenberg T. Opposite Drug Prescription and Cost Trajectories following Integrative and Conventional Care for Pain - A Case-Control Study. [↑](#endnote-ref-3)
4. Scott IA, Shohag H, Ahmed M. Quality of care factors associated with unplanned readmissions of older medical patients: a case-control study. Intern Med J. 2014 Feb;44(2):161-70. [↑](#endnote-ref-4)
5. Danielsen AK, Rosenberg J. Patient education after stoma creation may reduce health-care costs. Dan Med J. 2014 Apr;61(4):A4659. [↑](#endnote-ref-5)
6. Grammatico-Guillon L, Baron S, Gaborit C, Rusch E, Astagneau P. Quality assessment of hospital discharge database for routine surveillance of hip and knee arthroplasty-related infections. Infect Control Hosp Epidemiol. 2014 Jun;35(6):646-51. [↑](#endnote-ref-6)
7. Anantha RV, Parry N, Vogt K, Jain V, Crawford S, Leslie K. Implementation of an acute care emergency surgical service: a cost analysis from the surgeon's perspective. Anantha RV, Parry N, Vogt K, Jain V, Crawford S, Leslie K. Can J Surg. 2014 Apr;57(2):E9-14. [↑](#endnote-ref-7)
8. Dykes PC, I-Ching EH, Soukup JR, Chang F, Lipsitz S. A case control study to improve accuracy of an electronic fall prevention toolkit. AMIA Annu Symp Proc. 2012;2012:170-9 [↑](#endnote-ref-8)
9. Yu F, Salas M, Kim YI, Menachemi N. The relationship between computerized physician order entry and pediatric adverse drug events: a nested matched case-control study. Pharmacoepidemiol Drug Saf. 2009 Aug;18(8):751-5. [↑](#endnote-ref-9)
10. Palen TE1, Price D, Shetterly S, Wallace KB. Comparing virtual consults to traditional consults using an electronic health record: an observational case-control study. BMC Med Inform Decis Mak. 2012 Jul 8;12:65. [↑](#endnote-ref-10)
11. Cullen MR, Checkoway H, Alexander BH. Investigation of a cluster of pituitary adenomas in workers in the aluminum industry. Occup Environ Med. 1996 Nov;53(11):782-6. [↑](#endnote-ref-11)
12. Knijff-Dutmer EA1, Postma MJ, van der Palen J, Brouwers JR, van de Laar MA. Incremental cost-effectiveness of cyclooxygenase 2-selective versus nonselective nonsteroidal anti-inflammatory drugs in a cohort of coumarin users: a pharmacoeconomic analysis linked to a case-control study. Clin Ther. 2004 Jul;26(7):1160-7. [↑](#endnote-ref-12)
13. Preston NJ1, Fazio S. Establishing the efficacy and cost effectiveness of community intensive case management of long-term mentally ill: a matched control group study. Aust N Z J Psychiatry. 2000 Feb;34(1):114-21. [↑](#endnote-ref-13)
14. Vonkeman HE1, Braakman-Jansen LM, Klok RM, Postma MJ, Brouwers JR, van de Laar MA. Incremental cost effectiveness of proton pump inhibitors for the prevention of non-steroidal anti-inflammatory drug ulcers: a pharmacoeconomic analysis linked to a case-control study. Arthritis Res Ther. 2008;10(6):R144. [↑](#endnote-ref-14)
15. Brilli RJ1, Sparling KW, Lake MR, Butcher J, Myers SS, Clark MD, Helpling A, Stutler ME. The business case for preventing ventilator-associated pneumonia in pediatric intensive care unit patients. Jt Comm J Qual Patient Saf. 2008 Nov;34(11):629-38. [↑](#endnote-ref-15)
16. Bijl R, Kooistra SR, Hogeveen H. The profitability of automatic milking on Dutch dairy farms. J Dairy Sci. 2007 Jan;90(1):239-48. [↑](#endnote-ref-16)
17. Longo DR, Chase GA. Structural determinants of hospital closure. Med Care. 1984 May;22(5):388-402. [↑](#endnote-ref-17)
18. Sarlos D, Kots L, Stevanovic N, Schaer G. Robotic hysterectomy versus conventional laparoscopic hysterectomy: outcome and cost analyses of a matched case-control study. Eur J Obstet Gynecol Reprod Biol. 2010 May;150(1):92-6. [↑](#endnote-ref-18)
19. Relph S, Bell A, Sivashanmugarajan V, Munro K, Chigwidden K, Lloyd S, Fakokunde A, Yoong W. Cost effectiveness of enhanced recovery after surgery programme for vaginal hysterectomy: a comparison of pre and post-implementation expenditures. Int J Health Plann Manage. 2013 May 10. [↑](#endnote-ref-19)
20. Niebuhr DW, Krampf RL, Mayo JA, Blandford CD, Levin LI, Cowan DN. Risk factors for disability retirement among healthy adults joining the U.S. Army. Mil Med. 2011 Feb;176(2):170-5. [↑](#endnote-ref-20)
21. Sorgdrager B, Hulshof CT, van Dijk FJ. Evaluation of the effectiveness of pre-employment screening. Int Arch Occup Environ Health. 2004 May;77(4):271-6. [↑](#endnote-ref-21)
22. IOM. Digital data improvement priorities for continuous learning in health and health care. The National Academic Press, Washington DC, 2013. [↑](#endnote-ref-22)
23. Cuttino LW, Khan A, Wazer DE, Arthur DW, Vicini FA. When retrospective comparative effectiveness research hinders science and patient-centered care. J Clin Oncol. 2013 Jun 10;31(17):2226-7. [↑](#endnote-ref-23)
24. Lohr KN. Comparative effectiveness research methods: symposium overview and summary. Med Care. 2010 Jun; 48(6 Suppl): S3-6. [↑](#endnote-ref-24)
25. Berger M., Mamdani M, Atkins D., Johnson M. 2009. “Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I,” Value in Health, Volume 12, Number 8. [↑](#endnote-ref-25)
26. Hagedorn, G. (2002). *Preview of deltaaccess* . Retrieved from http://diversityworkbench.net/OldModels/Descriptions/Docu160/PREVIEW\_DELTAACCESS.HTML [↑](#endnote-ref-26)
27. Schwarz , J. (2010, October 12). [Web log message]. Retrieved from http://healthcaresecurity.wordpress.com/2010/10/12/why-are-hierarchical-databases-like-mumps-still-popular-in-healthcare/ [↑](#endnote-ref-27)
28. Newsom, S. W. B. (2006). Pioneers in infection control: John Snow, Henry [???]. *Journal of Hospital Infection*, *64*, 210-216. doi: 10.1016/j.jhin.2006.05.020 [↑](#endnote-ref-28)
29. Paneth, N., Susser, E., & Susser, M. (2002). Origins and early development of the case-control study:. *Journal of Epidemiology and Community Health*, *47*, 359–365. Retrieved from http://www.epidemiology.ch/history/papers/SPM 47(6) 359-65 Paneth et al. \_ Part 2.pdf [↑](#endnote-ref-29)
30. Moyses Szklo and F. Javier Nieto, *Epidemiology: Beyond the basics*, 3rd edition: Burlington, MA: Jones & Bartlett, 2014 [↑](#endnote-ref-30)
31. Elixhauser A., Steiner C., Harris R., Coffey R.M. 1998. "Comorbidity Measures for Use with Administrative Data," Medical Care 36:8-27. [↑](#endnote-ref-31)
32. Patient falls were classified using the following E codes in the International Classification of Disease: E8800 E8801 E8809 E8810 E8811 E882 E8830 E8831 E8832 E8839 E8840 E8841 E8842 E8843 E8844 E8845 E8846 E8849 E885 E8850 E8851 E8852 E8853 E8854 E8859 E8860 E8869 E888 E8880 E8881 E8888 E8889 E9681 E9870 E9871 E9872 E9879. [↑](#endnote-ref-32)
33. Patient injuries were measured using the following codes: 7960 7990 79901 79902 9071 9073 9074 9075 9079 9085 9086 9089 9092 9094 9099 9300 9301 9302 9308 9309 931 932 9330 9331 9340 9341 9348 9349 9350 9351 9352 936 937 938 9390 9391 9392 9393 9399 9500 9501 9502 9503 9509 9510 9511 9512 9513 9514 9515 9516 9517 9518 9519 9530 9531 9532 9533 9534 9535 9538 9539 9540 9541 9548 9549 9550 9551 9552 9553 9554 9555 9556 9557 9558 9559 9560 9561 9562 9563 9564 9565 9568 9569 9570 9571 9578 9579 9580 9581 9582 9583 9584 9585 9586 9587 9588 95890 95891 95892 95893 95899 9590 95901 95909 9591 95911 95912 95913 95914 95919 9592 9593 9594 9595 9596 9597 9598 9599 990 9910 9911 9912 9913 9914 9915 9916 9918 9919 9920 9921 9922 9923 9924 9925 9926 9927 9928 9929 9930 9931 9932 9933 9934 9938 9939 9940 9941 9942 9943 9944 9945 9946 9947 9948 9949 9951 9955 99550 99551 99552 99553 99554 99555 99559 99580 99581 99582 99583 99584 99585 99589 99590 99591 99592 99593 99594 V155 V1551 V1559 V156 V1588 V713 V714 V715 V716 V9010 V9011 V9012 V902 V9031 V9032 V9033 V9039 V9081 V9083 V9089 V909. [↑](#endnote-ref-33)
34. The following codes were used to determine if the patient has had an adverse medication effect: E9300 E9301 E9302 E9303 E9304 E9305 E9306 E9307 E9308 E9309 E9310 E9311 E9312 E9313 E9314 E9315 E9316 E9317 E9318 E9319 E9320 E9321 E9322 E9323 E9324 E9325 E9326 E9327 E9328 E9329 E9330 E9331 E9332 E9333 E9334 E9335 E9336 E9337 E9338 E9339 E9340 E9341 E9342 E9343 E9344 E9345 E9346 E9347 E9348 E9349 E9350 E9351 E9352 E9353 E9354 E9355 E9356 E9357 E9358 E9359 E9360 E9361 E9362 E9363 E9364 E9370 E9371 E9372 E9373 E9374 E9375 E9376 E9378 E9379 E9380 E9381 E9382 E9383 E9384 E9385 E9386 E9387 E9389 E9390 E9391 E9392 E9393 E9394 E9395 E9396 E9397 E9398 E9399 E9400 E9401 E9408 E9409 E9410 E9411 E9412 E9413 E9419 E9420 E9421 E9422 E9423 E9424 E9425 E9426 E9427 E9428 E9429 E9430 E9431 E9432 E9433 E9434 E9435 E9436 E9438 E9439 E9440 E9441 E9442 E9443 E9444 E9445 E9446 E9447 E9450 E9451 E9452 E9453 E9454 E9455 E9456 E9457 E9458 E9460 E9461 E9462 E9463 E9464 E9465 E9466 E9467 E9468 E9469 E9470 E9471 E9472 E9473 E9474 E9478 E9479 E9480 E9481 E9482 E9483 E9484 E9485 E9486 E9488 E9489 E9490 E9491 E9492 E9493 E9494 E9495 E9496 E9497 E9499. [↑](#endnote-ref-34)
35. The following codes were used to determine mood disorders: 29383 29600 29601 29602 29603 29604 29605 29606 29610 29611 29612 29613 29614 29615 29616 29620 29621 29622 29623 29624 29625 29626 29630 29631 29632 29633 29634 29635 29636 29640 29641 29642 29643 29644 29645 29646 29650 29651 29652 29653 29654 29655 29656 29660 29661 29662 29663 29664 29665 29666 2967 29680 29681 29682 29689 29690 29699 3004 311. The following codes were used to identify anxiety disorders: 29384 30000 30001 30002 30009 30010 30020 30021 30022 30023 30029 3003 3005 30089 3009 3080 3081 3082 3083 3084 3089 30981 3130 3131 31321 31322 3133 31382 31383. [↑](#endnote-ref-35)
36. James J. Schlesselman, Case-control studies: design, conduct, analysis. New York: Oxford University Press, 1982 [↑](#endnote-ref-36)
37. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost. 1993 Mar 1;69(3):236-9. [↑](#endnote-ref-37)
38. Niven DJ, Berthiaume LR, Fick GH, Laupland KB. Matched case-control studies: a review of reported statistical methodology. Clin Epidemiol. 2012;4:99-110. [↑](#endnote-ref-38)
39. Breslow NE, Day NE. Statistical Methods in Cancer Research; Volume 1 – The Analysis of Case-Control Studies. Lyon, France: International Agency for Research on Cancer; 1980. [↑](#endnote-ref-39)
40. Morris and Gardner; Gardner, MJ (1988). "Calculating confidence intervals for relative risks (odds ratios) and standardized ratios and rates". British Medical Journal 296 (6632): 1313–1316 [↑](#endnote-ref-40)
41. Kaplan, E. L.; Meier, P.: Nonparametric estimation from incomplete observations. J. Amer. Statist. Assn. 53:457–481, 1958. [↑](#endnote-ref-41)
42. Greenwood M. The natural duration of cancer. Reports on Public Health and Medical Subjects. London: Her Majesty's Stationery Office 1926;33:1–26. [↑](#endnote-ref-42)
43. Alemi F, Walters SR. A mathematical theory for identifying and measuring severity of episodes of care. Qual Manag Health Care. 2006 Apr-Jun;15(2):72-82. [↑](#endnote-ref-43)
44. Alemi F, Uriyo M. Accuracy of Claims-Based Measures of Severity of Childhood Illnesses. Health Outcomes Research in Medicine 2012, 2 (2): e71-e78. [↑](#endnote-ref-44)