Matched case-control studies: a review of reported statistical methodology

Niven, Daniel J.; Berthiaume, Luc R.; Fick, Gordon H.; Laupland, Kevin B.

Dove Medical Press

http://hdl.handle.net/1880/49178
journal article

Downloaded from PRISM: https://prism.ucalgary.ca
Matched case-control studies: a review of reported statistical methodology

Daniel J Niven 1
Luc R Berthiaume 1
Gordon H Fick 2
Kevin B Laupland 1

1 Department of Critical Care Medicine, Peter Lougheed Centre, Calgary, 2 Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

Background: Case-control studies are a common and efficient means of studying rare diseases or illnesses with long latency periods. Matching of cases and controls is frequently employed to control the effects of known potential confounding variables. The analysis of matched data requires specific statistical methods.

Methods: The objective of this study was to determine the proportion of published, peer-reviewed matched case-control studies that used statistical methods appropriate for matched data. Using a comprehensive set of search criteria we identified 37 matched case-control studies for detailed analysis.

Results: Among these 37 articles, only 16 studies were analyzed with proper statistical techniques (43%). Studies that were properly analyzed were more likely to have included case patients with cancer and cardiovascular disease compared to those that did not use proper statistics (10/16 or 63%, versus 5/21 or 24%, \( P = 0.02 \)). They were also more likely to have matched multiple controls for each case (14/16 or 88%, versus 13/21 or 62%, \( P = 0.08 \)). In addition, studies with properly analyzed data were more likely to have been published in a journal with an impact factor listed in the top 100 according to the Journal Citation Reports index (12/16 or 69%, versus 1/21 or 5%, \( P \leq 0.0001 \)).

Conclusion: The findings of this study raise concern that the majority of matched case-control studies report results that are derived from improper statistical analyses. This may lead to errors in estimating the relationship between a disease and exposure, as well as the incorrect adaptation of emerging medical literature.

Keywords: case-control, matched, dependent data, statistics

Introduction

Case-control studies provide a quick and efficient means of studying diseases with long latency periods or with low incidence in the population.1 Given their utility, it is not surprising that a Medline search of the English language literature with the words “case-control” in the title revealed more than 1000 articles published in 2010. Although they are convenient and common, there are several important considerations in the design of case-control studies.1,2 One consideration is the decision to match cases to controls and the subsequent selection of statistical techniques that are appropriate for the matched data. The analysis of matched (dependent) data is different from unmatched (independent) data and is described in detail by Breslow and Day.3 In their text on the analysis of case-control studies, they describe the use of paired \( t \)-tests for measured outcomes with 1:1 matching and with a symmetrical distribution, the Wilcoxon signed ranks test for measured outcomes with 1:1 matching as a non-parametric alternative,
model-based options such as the appropriate linear regression to handle modification and additional potential confounding, and matched sets other than 1:1 matching. For dichotomous outcomes, they describe McNemar’s test, a Mantel–Haenszel matched-pairs analysis, and additional methods to handle matched sets other than 1:1 matching. They also describe conditional logistic regression to handle all forms of matching as well as the consideration of modification and other potential confounding.

We recently published a study wherein matching was employed to control for known potential confounding variables. During the analysis, we noticed that many of the peer-reviewed publications that describe the use of matched data, employ inconsistent statistical methods of analysis. Data analyses that employ incorrect statistical methods will commonly result in inappropriate conclusions. Therefore, the current study was designed to evaluate the statistical methodology in a collection of matched case-control studies.

Methods
A literature review was conducted using PubMed from January 1, 2010 to December 1, 2010 with the goal of identifying articles that employed a matched case-control design. The search-term was “case-control” (found in the article title only) and the search was limited to human studies published in the English language literature in PubMed’s defined subset of “core clinical journals” (see Appendix 1 for a list of these core clinical journals). This search strategy was chosen to yield a representative sample of case-control studies in a variety of subject areas, published in peer-reviewed, mainstream journals. One of the authors screened all the abstracts for relevance and appropriate full-length articles were subsequently retrieved for appraisal. To maintain relative homogeneity among the final collection of articles, it was decided, a priori, to exclude articles that focused on subjects in the pediatric age group (under 18 years of age). We also excluded studies that used matching methods other than simple, individual criteria-based matching, ie, frequency-matching, and propensity-matching, as this type of data is analyzed with different statistical methods.

Each full length article was independently reviewed in detail by two of the authors. The goal of this review was to evaluate the appropriateness of the statistical methodology. Disagreements between the reviewers were resolved by the independent evaluation of a senior biostatistician. Inter-observer agreement was quantified using the kappa statistic, wherein a kappa value of 0.61 to 1.0 indicates substantial agreement. Statistical methods were appropriate for matched data if they were consistent with those described by Breslow and Day. Each article was appraised using the following analysis scoring system: (1) continuous outcomes analyzed using the paired t-test, or Wilcoxon signed ranks test, or others, as described above and dichotomous outcomes analyzed using McNemar’s test, a Mantel–Haenszel matched-pairs analysis, conditional logistic regression, or other, as described above; (2) investigators failed to analyze continuous outcomes with a paired t-test or the Wilcoxon signed ranks test, or did not use McNemar’s test, a Mantel–Haenszel matched-pairs analysis, or conditional logistic regression for dichotomous outcomes; and (3) the authors did not use any of the aforementioned statistical methods for continuous and dichotomous outcomes.

Following the review of their statistical methodology, the collection of matched case-control articles were reviewed a second time for factors that may be associated with the use of statistical methods appropriate for matched data. These factors included items that form common issues in the design of matched case-control studies, namely the case population definition, the number of matching variables, and the control-to-case ratio. In addition, we used the 2010 Journal Citation Reports (JCR) index to determine whether the appropriateness of the statistical methodology was associated with the impact factor of the publishing journal. Using the JCR index, we determined that an impact factor of at least 12.245 was required for a journal to be listed in the top 100 major journals. Data was analyzed using Stata (v11.0; Stata Corp, College Station, TX) and statistical significance was set at P ≤ 0.05.

Results
The initial search strategy yielded 74 articles (Figure 1). Upon review of these abstracts, 36 articles were excluded for reasons outlined in Figure 1. The remaining 38 articles were reviewed in detail. After reviewing the articles, two of the authors identified one study that employed frequency-matching rather than individual patient-matching. Exclusion of this study from further statistical evaluation left 37 studies for the overall analysis. Table 1 provides a summary of these 37 studies. The two authors reviewing the studies agreed on the appropriateness of the statistical methods in 36 of the 37 studies (97%) and the inter-observer agreement, as measured by the kappa statistic, was 0.94. Sixteen of the selected studies were analyzed with correct statistical methods (analysis score of 1, 43%), and 21 were analyzed with at least one incorrect statistical method (sum of analysis score 2 and 3, 57%).
Given the low number of studies with correctly analyzed data, each of the 37 articles was reviewed for factors that may be associated with the use of correct statistical tests (Table 2). Studies that were properly analyzed were more likely to have included case patients with cancer and cardiovascular disease compared to those that did not use proper statistics (10/16 or 63%, versus 5/21 or 24%, \(P = 0.02\)). Furthermore, properly analyzed studies were more likely to have matched multiple controls for each case (14/16 or 88%, versus 13/21 or 62%, \(P = 0.1\)).

Table 3 presents the data on each publishing journal’s impact factor, in addition to the association between the previously described study design characteristics and correct statistical methodology. From this table, it is clear that matched case-control articles published in the *British Medical Journal* (BMJ) were consistently analyzed with correct statistical techniques. Furthermore, the BMJ was responsible for publishing the greatest number of articles in this series of matched case-control studies (7/37 or 19%). This is in contrast to articles published in *Archives of Otolaryngology-Head and Neck Surgery*, *Journal of Clinical Endocrinology and Metabolism*, and *Neurology* wherein the data was frequently analyzed incorrectly. Moreover, matched case-control studies published in *Lancet* were notably inconsistent in their statistical methodology. From this table, it is also evident that more studies in the correctly analyzed collection were published in a journal with an impact factor within the top 100 listing on JCR (11/16 or 69%, versus 1/21 or 5%, \(P = 0.0001\)). The median (interquartile range [IQR]) impact factor among studies that were correctly analyzed was 13.471 (8.516–13.950), compared to 6.495 (4.231–8.017) for those with incorrect statistical methodology (Wilcoxon rank sum test, \(P = 0.0009\)).

**Discussion**

To our knowledge, this is the first study to assess the appropriateness of the statistical methodology used in a published series of matched case-control studies. From this structured review of studies published in a number of diverse mainstream, peer-reviewed journals it is clear that matched case-control studies are not consistently analyzed using appropriate statistical methods. More than 50% of the articles reviewed in this study present data that was analyzed with improper statistics. For many of these studies, this may simply change the strength of the association between the disease and exposure of interest; however, for studies with small numbers of discordant sets, the use of appropriate statistical methods may alter the significance of the findings. Unfortunately, none of the articles reviewed in this study with incorrect statistical methodology provided the data in a format whereby the magnitude of the difference between a proper and improper analysis could be assessed. This is important, as it is clear from recent reviews of the literature that the use of statistics in medical literature is increasing over time.\(^{45,46}\) However, if the analyses are being performed incorrectly, this increased use of statistics does not equate to an improvement in study quality and may lead to the adoption of incorrect medical literature.
<table>
<thead>
<tr>
<th>Study</th>
<th>Case population</th>
<th>Cases (n)</th>
<th>Controls (n)</th>
<th>Matching variables</th>
<th>Analysis score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ances et al⁷</td>
<td>Patients with human immunodeficiency virus (HIV)</td>
<td>10</td>
<td>20</td>
<td>Gender, education</td>
<td>3</td>
</tr>
<tr>
<td>Antonelli et al⁸</td>
<td>Patients with systemic lupus erythematosus</td>
<td>153</td>
<td>918</td>
<td>Gender, age</td>
<td>3</td>
</tr>
<tr>
<td>Carnaby-Mann et al⁹</td>
<td>Patients with dysphagia who entered the McNeill Dysphagia Therapy Program</td>
<td>8</td>
<td>16</td>
<td>Age, gender, primary medical diagnosis</td>
<td>2</td>
</tr>
<tr>
<td>Dubois et al¹⁰</td>
<td>Patients with previous thoracic radiation therapy undergoing coronary stenting</td>
<td>41</td>
<td>82</td>
<td>Gender, stented vessel, drug-eluting stent use, available duration of follow-up,</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>unstable coronary disease, renal insufficiency, diabetes mellitus, bifurcation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>disease, stent length and size, ejection fraction</td>
<td></td>
</tr>
<tr>
<td>Etminan et al¹¹</td>
<td>Patients who had a coronary revascularization procedure between 1995–2004 with</td>
<td>217</td>
<td>2170</td>
<td>Age, cohort entry day</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>a discharge diagnosis of epilepsy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedland et al¹²</td>
<td>Patients with cochlear implants</td>
<td>28</td>
<td>28</td>
<td>Pre-cochlear implant score on Hearing In Noise Test–Quiet, duration of deafness</td>
<td>3</td>
</tr>
<tr>
<td>Garg et al¹³</td>
<td>Patients 18–65 years old with migraine headaches onset before age 50</td>
<td>144</td>
<td>144</td>
<td>Age, gender</td>
<td>1</td>
</tr>
<tr>
<td>Green et al¹⁴</td>
<td>Patients ≥ 40 years old with incident esophageal, gastric, or colorectal cancer</td>
<td>15 613</td>
<td>77 750</td>
<td>Age, gender, general practice, observation time</td>
<td>1</td>
</tr>
<tr>
<td>Jenab et al¹⁵</td>
<td>Patients with incident colorectal cancer</td>
<td>1248</td>
<td>1248</td>
<td>Age, gender, study center, time of day at blood collection, duration of fasting at</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>blood collection; women were further matched by menopausal status, phase of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>menstrual cycle at time of blood collection, use of hormone replacement therapy</td>
<td></td>
</tr>
<tr>
<td>Kermani et al¹⁶</td>
<td>Incident cases with giant cell arteritis</td>
<td>204</td>
<td>407</td>
<td>Age, gender, length of medical history</td>
<td>3</td>
</tr>
<tr>
<td>Koscieiny et al¹⁷</td>
<td>Patients with claudication or critical ischemia undergoing surgery for occlusion</td>
<td>28</td>
<td>28</td>
<td>Occluded SFA, patent popliteal artery not amenable to endovascular intervention,</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>of the superficial femoral artery (SFA)</td>
<td></td>
<td></td>
<td>patent profunda femoris artery with less than 50% stenosis before the first branch,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>occluded SFA, patent crural outflow artery, American Society of Anesthesiologists</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Class I, II, or III, patient mobile before surgery, or expected to have</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>postoperative mobility, on aspirin and a statin</td>
<td></td>
</tr>
<tr>
<td>Lang et al¹⁸</td>
<td>Patients with HIV found to have an incident myocardial infarction</td>
<td>289</td>
<td>884</td>
<td>Age, gender, clinical center</td>
<td>3</td>
</tr>
<tr>
<td>Marshall et al¹⁹</td>
<td>Patients with HIV and Kaposi’s sarcoma</td>
<td>21</td>
<td>14</td>
<td>Unclear</td>
<td>3</td>
</tr>
<tr>
<td>Martinez et al²⁰</td>
<td>Patients with sudden cardiac death, or near death on antidepressants</td>
<td>568</td>
<td>14 812</td>
<td>Date of incident antidepressant prescription, year of birth, gender, indication</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>for antidepressant prescription</td>
<td></td>
</tr>
<tr>
<td>Nickel et al²¹</td>
<td>Females with interstitial cystitis/painful bladder syndrome</td>
<td>207</td>
<td>117</td>
<td>Age, partner status, education</td>
<td>3</td>
</tr>
<tr>
<td>O’ donnell et al²³</td>
<td>Patients with first acute stroke</td>
<td>3000</td>
<td>3000</td>
<td>Age, gender, ± ethnic origin</td>
<td>3</td>
</tr>
<tr>
<td>Parker et al²⁴</td>
<td>Patients with incident venous thromboembolism</td>
<td>25, 532</td>
<td>89, 491</td>
<td>Age, calendar time, gender, family practice</td>
<td>1</td>
</tr>
<tr>
<td>Persoon et al²⁵</td>
<td>Patients with limb shaking associated with internal carotid artery occlusion</td>
<td>34</td>
<td>68</td>
<td>Age, gender</td>
<td>3</td>
</tr>
<tr>
<td>Pouwels et al²⁶</td>
<td>Patients with incident admission for a hip fracture</td>
<td>6,763</td>
<td>26,341</td>
<td>Year of birth, gender, geographic location, index admission date</td>
<td>1</td>
</tr>
<tr>
<td>Renoux et al²⁷</td>
<td>Patients with incident diagnosis of stroke and on hormone replacement therapy</td>
<td>15,710</td>
<td>59,958</td>
<td>Age, general practice attended, year of joining the practice</td>
<td>1</td>
</tr>
<tr>
<td>Ripatti et al²⁸</td>
<td>Patients with coronary artery disease</td>
<td>2,101</td>
<td>3,914</td>
<td>Gender, birth year</td>
<td>1</td>
</tr>
<tr>
<td>Authors</td>
<td>Description</td>
<td>N</td>
<td>Cases</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Roder et al²⁹</td>
<td>Patients with total hip arthroplasty and signs of cup failure</td>
<td>809</td>
<td>3,986</td>
<td>Hospital site of total hip arthroplasty, date of surgery; a follow-up exam within 6 months of the follow-up interval of the case; cup design, size, and material</td>
<td></td>
</tr>
<tr>
<td>Schaer et al³⁰</td>
<td>Patients with atrial fibrillation</td>
<td>4661</td>
<td>18,642</td>
<td>Age, gender, general practice, calendar time</td>
<td></td>
</tr>
<tr>
<td>Schillaci et al³¹</td>
<td>Patients aged 18–50 years with migraine headaches</td>
<td>60</td>
<td>60</td>
<td>Age, gender, blood pressure</td>
<td></td>
</tr>
<tr>
<td>Sinwardena et al³²</td>
<td>Patients with incident myocardial infarction</td>
<td>16,012</td>
<td>62,694</td>
<td>Age, gender, general practice, calendar time</td>
<td></td>
</tr>
<tr>
<td>Talving et al³³</td>
<td>Patients with severe traumatic brain injury receiving erythropoiesis stimulating agent</td>
<td>89</td>
<td>178</td>
<td>Age, gender, mechanism, Glasgow Coma Scale, hypotension, injury severity score, abbreviated injury scale, presence of anemia</td>
<td></td>
</tr>
<tr>
<td>Tammemagi et al³⁴</td>
<td>Never smokers or former smokers with adult-onset chronic rhinosinusitis</td>
<td>306</td>
<td>306</td>
<td>Age, gender, race/ethnicity</td>
<td></td>
</tr>
<tr>
<td>Taran et al³⁵</td>
<td>Women with a diagnosis of cellular meiomyomas</td>
<td>99</td>
<td>198</td>
<td>Surgeon, surgical procedure, year of surgery</td>
<td></td>
</tr>
<tr>
<td>Telem et al³⁶</td>
<td>Patients with anastomotic leak following colorectal surgery</td>
<td>90</td>
<td>180</td>
<td>Procedure, surgeon</td>
<td></td>
</tr>
<tr>
<td>Trifiro et al³⁷</td>
<td>Elderly patients receiving an anti-psychotic drug prescription with an incident episode of community-acquired pneumonia</td>
<td>258</td>
<td>1,689</td>
<td>Year of birth, gender, index date</td>
<td></td>
</tr>
<tr>
<td>Vickers et al³⁸</td>
<td>Patients with prostate cancer</td>
<td>126</td>
<td>373</td>
<td>Date of birth, date of baseline blood tests</td>
<td></td>
</tr>
<tr>
<td>Vlaar et al³⁹</td>
<td>Patients with transfusion-associated lung injury admitted to an intensive care unit</td>
<td>109</td>
<td>327</td>
<td>Age, gender, admission diagnosis</td>
<td></td>
</tr>
<tr>
<td>Warenßjö et al⁴⁰</td>
<td>Consecutive patients with incident myocardial infarction</td>
<td>444</td>
<td>556</td>
<td>Gender, age, date of health survey, and geographic region</td>
<td></td>
</tr>
<tr>
<td>Wassenaar et al⁴¹</td>
<td>Patients with acromegaly undergoing screening colonoscopy</td>
<td>107</td>
<td>214</td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>White et al⁴²</td>
<td>Patients who discontinued zonisamide due to psychiatric or cognitive adverse events</td>
<td>59</td>
<td>118</td>
<td>Zonisamide initiation date</td>
<td></td>
</tr>
<tr>
<td>Wohl et al⁴³</td>
<td>Patients older than 20 years of age with a diagnosis of pemphigus</td>
<td>255</td>
<td>509</td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>Yates and James⁴⁴</td>
<td>Physicians with a finding of serious professional misconduct</td>
<td>59</td>
<td>236</td>
<td>Year of graduation from medical school</td>
<td></td>
</tr>
</tbody>
</table>

Notes: See Methods section for an outline of the statistical scoring system; Categorical variables assessed using the Cochran–Mantel–Haenszel test with the matched pair as a stratification variable; Used univariate conditional logistic regression to perform hypothesis test on baseline continuous variables; Authors mentioned deliberately analyzing the data as if it were not matched; Based on analysis of the nested COROGENE matched case-control sample; three separate control groups of 109 patients each.
perform complex statistical analyses in a matter of seconds, statistical software has allowed the novice researcher to perform proper analysis of matched data as shown in Table 2, generated solely by statisticians as they required careful thought and reflected a thoughtful, systematic approach to the study design. This same thought was likely extended to the analysis phase and resulted in the correct application of statistical methodology.

Second, a greater number of studies in the incorrectly analyzed collection focused on a surgical topic, compared to studies in the correctly analyzed collection. Although the reasons for this finding are unclear, one potential explanation might be that the use of inferential statistics in surgical studies is a more recent development when compared to studies focusing on other medical topics. Older surgical research involved smaller sample sizes and very few of these studies employed inferential statistics.46 The observation that studies with multiple controls matched to each case were more likely to use statistics that were appropriate for matched data was not surprising. The decision to match more than one control per case may increase the power of case-control studies,1 which, in turn, increases the strength of the study and reflects a thoughtful, systematic approach to the study design. This same thought was likely extended to the analysis phase and resulted in the correct application of statistical methodology.

Furthermore, the observations made with regard to the publishing journal’s impact factor were also not surprising. Recent studies have shown that there is a correlation with the strength of study design (including the use of recommended statistical reporting) and journal publication characteristics including the journal’s impact factor.50,51 Lee and colleagues reviewed 243 randomly selected articles from the general internal medicine literature published between January 1, 1999 and December 31, 1999 to determine if there was a link between methodological quality and journal characteristics.50 The authors found significant associations between quality scores and higher citation rates, higher impact factors, higher circulation, and lower manuscript acceptance rates. Similarly, Kuroki et al investigated the potential link between research methodology and statistical reporting in medical journals with a high impact factor compared to moderate-impact-factor obstetrics and gynecology journals.51 The high-impact-factor medical journals included: Journal of the American Medical Association, The Lancet, and the New England Journal of Medicine; whereas the moderate-impact-factor group included American Journal of Obstetrics and Gynecology, British Journal of Obstetrics and Gynecology, and Obstetrics and Gynecology. The authors found that the majority of studies included in the high-impact-factor group were randomized controlled trials (35%) and had high compliance with recommended statistical reporting (84% compared to 65%, P = 0.002). Therefore, our finding of an increased impact factor in the correct statistical analysis group aligns with trends observed in similar studies.

### Table 2 Study design and publication characteristics

<table>
<thead>
<tr>
<th>Factor</th>
<th>Correct analysis</th>
<th>Incorrect analysis</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 16)</td>
<td>(n = 21)</td>
<td></td>
</tr>
<tr>
<td>Case population, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer or cardiovascular disease</td>
<td>10 (63)</td>
<td>5 (24)</td>
<td>0.02</td>
</tr>
<tr>
<td>HIV</td>
<td>0 (0)</td>
<td>3 (14)</td>
<td>0.1</td>
</tr>
<tr>
<td>Other</td>
<td>6 (38)</td>
<td>13 (62)</td>
<td>0.1</td>
</tr>
<tr>
<td>Basic science topic, n (%)</td>
<td>1 (6)</td>
<td>1 (5)</td>
<td>0.8</td>
</tr>
<tr>
<td>Medical topic, n (%)</td>
<td>13 (81)</td>
<td>15 (71)</td>
<td>0.5</td>
</tr>
<tr>
<td>Surgical topic, n (%)</td>
<td>2 (13)</td>
<td>5 (24)</td>
<td>0.4</td>
</tr>
<tr>
<td>&gt;2 matching variables, n (%)</td>
<td>11 (63)</td>
<td>12 (57)</td>
<td>0.5</td>
</tr>
<tr>
<td>&gt;1:1 control-to-case ratio, n (%)</td>
<td>14 (88)</td>
<td>13 (62)</td>
<td>0.1</td>
</tr>
<tr>
<td>Publishing journal in top 100</td>
<td>11 (69)</td>
<td>1 (5)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**Notes:**
- *Correct statistical analysis: received a statistical analysis score of “1”;
- *Incorrect statistical analysis: received a statistical analysis score of “2” or “3”;
- *Comparisons were made via Fisher’s exact test; cardiovascular disease was defined as any disease of the coronary vascular, peripheral vascular, cerebrovascular, and cardiac electrical systems.

**Abbreviations:** HIV, human immunodeficiency virus; JCR, journal citation reports.

Although this study appears to be the first to evaluate the quality of the statistical methods employed in a series of matched case-control studies, it is not the first study to review the quality of statistical methods in medical journals. A number of these studies were published in the 1980’s and 1990’s at a time that coincided with a rapid rise in the use of statistics in medical research. They consistently found that a minority of studies reported unacceptable methods of data analysis and concluded that this is likely due to the fact that individuals leading these medical publications may not have a solid grasp on basic statistical concepts.45-47 This may not be any different in today’s medical literature and is likely further compounded by the fact that statisticians are not consistently involved in the peer review process,48 a step that often improves the quality of the statistical methodology of accepted articles.49 Furthermore, the advent of sophisticated statistical software has allowed the novice researcher to perform complex statistical analyses in a matter of seconds, whereas these complicated analyses were previously performed solely by statisticians as they required careful thought and complex mathematical formulae.

The study design characteristics and their relation to the proper analysis of matched data as shown in Table 2, generate a few interesting hypotheses. First, studies involving case populations with cancer or cardiovascular disease were more likely to employ statistical techniques that account for dependent data than studies involving other case definitions. The reason for this is not clear but it may be a reflection of the rigor with which these studies were designed.
Table 3 A list of the publishing journals and impact factors among the collection of articles reviewed in this study

<table>
<thead>
<tr>
<th>Journal</th>
<th>Study(s)</th>
<th>Impact factor*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analyzed correctly</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal of Bone and Joint Surgery, American Volume</td>
<td>Roder et al29</td>
<td>2.967</td>
</tr>
<tr>
<td>Heart</td>
<td>Dubois et al30</td>
<td>4.706</td>
</tr>
<tr>
<td>Journal of Clinical Endocrinology and Metabolism</td>
<td>Pouwels et al31</td>
<td>6.495</td>
</tr>
<tr>
<td>Neurology</td>
<td>Etmian et al31</td>
<td>8.017</td>
</tr>
<tr>
<td>Canadian Medical Association Journal</td>
<td>Siriwardena et al32</td>
<td>9.015</td>
</tr>
<tr>
<td>Circulation</td>
<td>Garg et al13</td>
<td>14.429</td>
</tr>
<tr>
<td>Annals of Internal Medicine</td>
<td>Schaer et al,30 Trifiro et al37</td>
<td>16.729</td>
</tr>
<tr>
<td>The Lancet</td>
<td>Ripatti et al38</td>
<td>33.633</td>
</tr>
<tr>
<td><strong>Analyzed incorrectly</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archives of Otolaryngology, Head and Neck Surgery</td>
<td>Friedland et al,12 Tammemagi et al34</td>
<td>1.571</td>
</tr>
<tr>
<td>Archives of Physical Medicine and Rehabilitation</td>
<td>Carnaby-Mann et al8</td>
<td>2.254</td>
</tr>
<tr>
<td>American Journal of Obstetrics and Gynecology</td>
<td>Taran et al35</td>
<td>3.313</td>
</tr>
<tr>
<td>Journal of Urology</td>
<td>Nickel et al32</td>
<td>3.862</td>
</tr>
<tr>
<td>Archives of Dermatology</td>
<td>Wohl et al41</td>
<td>4.231</td>
</tr>
<tr>
<td>British Journal of Surgery</td>
<td>Koscienski et al17</td>
<td>4.444</td>
</tr>
<tr>
<td>Archives of Surgery</td>
<td>Telem et al36</td>
<td>4.500</td>
</tr>
<tr>
<td>Critical Care Medicine</td>
<td>Vlaar et al37</td>
<td>6.254</td>
</tr>
<tr>
<td>Journal of Clinical Endocrinology and Metabolism</td>
<td>Antonelli et al,8 Wassenaar et al41</td>
<td>6.495</td>
</tr>
<tr>
<td>American Journal of Clinical Nutrition</td>
<td>Warensojo et al40</td>
<td>6.606</td>
</tr>
<tr>
<td>Annals of Surgery</td>
<td>Talving et al33</td>
<td>7.474</td>
</tr>
<tr>
<td>Neurology</td>
<td>Ances et al,7 Schillaci et al,31 White et al52</td>
<td>8.017</td>
</tr>
<tr>
<td>Arthritis and Rheumatism</td>
<td>Kermani et al36</td>
<td>8.435</td>
</tr>
<tr>
<td>Brain</td>
<td>Persson et al18</td>
<td>9.230</td>
</tr>
<tr>
<td>Archives of Internal Medicine</td>
<td>Lang et al38</td>
<td>10.639</td>
</tr>
<tr>
<td>The Lancet</td>
<td>O’donnell et al23</td>
<td>33.633</td>
</tr>
</tbody>
</table>

Notes: *Impact factor as listed on Journal Citation Reports for 2010; †Excellent example of the analysis of matched data.

This study has a number of pertinent strengths and limitations that warrant discussion. First, this study is novel; as to our knowledge, this is the first review of the appropriateness of the statistical methodology employed in a collection of matched case-control studies. Second, the studies selected for review focused on a broad range of topics in a number of different peer-reviewed journals, so, although we reviewed only 37 studies, this sample is representative of the much larger population of available articles. Furthermore, many of these articles were published in mainstream journals read by individuals from a variety of backgrounds such as the British Medical Journal, the Canadian Medical Association Journal, and the Lancet. The major limitation of this study is that the methodology of the articles was not assessed with a validated scoring system. To our knowledge, such a scoring system does not exist for matched case-control studies, and other validated scoring systems were not applicable to the current study’s objectives.52 In spite of this limitation, the inter-observer agreement between the authors reviewing the studies was high. This suggests that the scoring system used in this study was applied in a consistent manner, and our conclusion regarding inconsistent use of proper statistical methods in matched case-control studies is valid. Another important limitation is the inability to determine whether the use of proper statistics would change the conclusions presented by studies that used improper statistical methods. This is due to the fact that very few case-control studies presented tables outlining the number of discordant sets. This limitation notwithstanding, it is possible that the use of proper statistical methods will at least decrease the strength of the association between the outcome and exposure variables when compared to that obtained from improper statistical methodology.

Conclusion

The majority of matched case-control studies reviewed in this investigation used improper statistical methods. Although matching cases to controls provides a means of controlling for known potential confounding variables, it is a complicated process that requires a great deal of thought in order to be effective. Improper application of this methodology can distort a study’s power and possibly lead to the reporting of
incorrect disease-exposure associations. The acceptance of invalid conclusions and subsequent adaptation into medical practice may lead to the inappropriate use of resources and even worse, harm to individuals. This is why guidelines such as CONSORT (Consolidated Standards of Reporting Trials) and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) were created.\(^5^,\(^6^\)\) These documents contain templates that are designed to create homogeneity in the reporting of randomized clinical trials and observational studies, respectively. The results of the current study suggest that although the STROBE checklist includes recommendations for outlining the matching methodology when reporting a matched case-control study, these comprehensive epidemiologic guidelines may require an additional section that outlines the proper statistical techniques to be employed when conducting a matched case-control study.

**Contributors**

Kevin B Laupland conceived the research question and Daniel J Niven and Kevin B Laupland designed the study. Daniel J Niven carried out the literature search, screened all relevant abstracts, and independently evaluated each study selected for the detailed review of the reported statistical methodology. Daniel J Niven also analyzed the data, and drafted the manuscript. Luc R Berthiaume independently evaluated the statistical methodology of the included studies, and contributed to manuscript revision. Gordon H Fick is a senior biostatistician and settled discrepancies between Daniel J Niven and Luc R Berthiaume during the review of the selected studies. Kevin B Laupland contributed to manuscript revision. All authors approved the final manuscript for publication.

**Ethics**

As this study did not involve collecting data from patients, formal approval from the regional ethics board was not required.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


Appendix 1 List of "core clinical journal" subset of PubMed journals

Academic Medicine: journal of the Association of American Medical Colleges
AJR: American Journal of Roentgenology
American Family Physician
American Heart Journal
The American Journal of Cardiology
The American Journal of Clinical Nutrition
American Journal of Clinical Pathology
The American Journal of Medicine
The American Journal of Nursing
American journal of Obstetrics and Gynecology
American Journal of Ophthalmology
American Journal of Pathology
American Journal of Physical Medicine and Rehabilitation
American Journal of Psychiatry
American Journal of Public Health
American Journal of Respiratory and Critical Care Medicine
American Journal of Surgery
The American Journal of the Medical Sciences
The American Journal of Tropical Medicine and Hygiene
Anaesthesia
Anaesthesia and Analgesia
Anesthesiology
Annals of Emergency Medicine
Annals of Internal Medicine
The Annals of Otology, Rhinology, and Laryngology
Annals of surgery
The Annals of Thoracic Surgery
Archives of Dermatology
Archives of Disease in Childhood
Archives of Disease in Childhood, fetal and neonatal edition
Archives of Environmental and Occupational Health [continues Archives of environmental health]
Archives of General Psychiatry
Archives of Internal Medicine
Archives of Neurology
Archives of Ophthalmology
Archives of Otolaryngology, head and neck surgery
Archives of Pathology and Laboratory Medicine
Archives of Pediatrics and Adolescent Medicine
Archives of Physical Medicine and Rehabilitation
Archives of Surgery (Chicago, Ill : 1960)
Arthritis and Rheumatism
BJOG: an international journal of obstetrics and gynaecology [continues British journal of obstetrics and gynaecology]
Blood
BMJ (Clinical research ed)
Brain: a journal of neurology
The British Journal of Radiology
The British Journal of Surgery
CA: A cancer Journal for Clinicians
Cancer
Chest
Circulation
Clinical Orthopedics and Related Research
Clinical Pediatrics
Clinical Toxicology: the official journal of the American Academy of Clinical Toxicology and European Association of Poisons Centers and Clinical Toxicologists [continues Journal of toxicology. Clinical toxicology]
Clinical Pharmacology and Therapeutics
CMAJ: Canadian Medical Association Journal/Journal de l’Association Medicale Canadienne
Critical Care Medicine

(Continued)
Appendix 1 (Continued)

Current Problems in Surgery
Diabetes
Digestive Diseases and Sciences
DM: Disease-a-month
Endocrinology
Gastroenterology
Gut
Heart and Lung: the journal of critical care
Heart (British Cardiac Society)
Hospital Practice (1995) [Formed by the union of: Hospital practice (Family practice ed.; Hospital practice (Hospital ed.; and Hospital practice (Office ed). No issues published between 2001 Sep 15;36(9) and 2009 Dec;37(1)]
Hospitals and Health Networks/AHA
JAMA: Journal of the American Medical Association
The Journal of Allergy and Clinical Immunology
The Journal of Bone and Joint Surgery, American volume
The Journal of Bone and Joint Surgery, British volume
The Journal of Clinical Endocrinology and Metabolism
The Journal of Clinical Investigation
Journal of Clinical Pathology
The Journal of Family Practice
Journal of Immunology (Baltimore, Md : 1950)
The Journal of Infectious Diseases
The Journal of Laryngology and Otology
The Journal of Nervous and Mental Disease
The Journal of Neurosurgery
The Journal of Nursing Administration
The Journal of Pediatrics
Journal of the American College of Cardiology
Journal of the American College of Surgeons
Journal of the American Dietetic Association
The Journal of Thoracic and Cardiovascular Surgery
The Journal of Trauma
The Journal of Urology
The Journals of Gerontology. Series A, Biological sciences and medical sciences
The Journals of Gerontology. Series B, Psychological sciences and social sciences
Lancet
Mayo Clinic Proceedings
The Medical Clinics of North America
The Medical Letter on Drugs and Therapeutics
Medicine
Neurology
The New England Journal of Medicine
The Nursing Clinics of North America
Nursing Outlook
Nursing Research
Obstetrics and Gynecology
The Orthopedic Clinics of North America
Pediatric Clinics of North America
Pediatrics
Physical Therapy
Plastic and Reconstructive Surgery
Postgraduate Medicine
Progress in Cardiovascular Diseases
Public Health Reports (Washington, DC : 1974)
Radiologic Clinics of North America
Radiology
Rheumatology (Oxford, England) [continues British Journal of Rheumatology]

(Continued)
Appendix I (Continued)

Southern Medical Journal
Surgery
The Surgical Clinics of North America
Translational Research: the journal of laboratory and clinical medicine [continues The Journal of Laboratory and Clinical Medicine]
The Urologic Clinics of North America