REASONS TO BELIEVE—BIOSTATISTICS & METHODOLOGY FOR THE NEUROSURGEON

What Isn’t a Case-Control Study?

**BACKGROUND:** Confusion exists among neurosurgeons when choosing and implementing an appropriate study design and statistical methods when conducting research. We noticed particular difficulty with mislabeled and inappropriate case-control studies in the neurosurgical literature.

**OBJECTIVE:** To quantify and to rigorously review this issue for appropriateness in publication and to establish quality of the manuscripts using a rigorous technique.

**METHODS:** Following a literature search, pairs drawn from 5 independent reviewers evaluated a complete sample of 125 manuscripts claiming to be case-control studies with respect to basic case-control criteria. Seventy-five papers were then subjected to a more rigorous appraisal for quality using the SIGN Methodology Checklist for case-control studies.

**RESULTS:** Fifty publications were rejected based on basic criteria used to identify case-control design. Of the 75 subjected to quality analysis, 46 were felt to be acceptable for publication. Only 11 papers (9%) achieved the designation of high quality. Of the original 125 papers evaluated, 79 (63%) were inappropriately labeled case-control studies.

**CONCLUSION:** Mislabeling and use of inappropriate study design are common in the neurosurgical literature. Manuscripts should be evaluated rigorously by reviewers and readers, and neurosurgical training programs should include instruction on choice of appropriate study design and critical appraisal of the literature.

**KEY WORDS:** Case control, Observational study, Study design, Quality review

---

As neurosurgeons and neurological trainees become more experienced in critically reading their pertinent practice literature, knowledge about comparative study design becomes increasingly important. Understanding differences in the design, conduct, and strengths and weaknesses of various study types is essential to using the findings of these studies. Unlike cohort studies, including randomized controlled trials (RCTs), case-control studies utilize a specific methodological study design in which groups are divided based on outcome of interest (cases—specifically referring to subjects with the outcome under investigation, and controls—subjects proven to not have the outcome in question), and exposure (or treatment or risk factor) is assessed to determine a correlation between the exposure rates in each outcome group. Many authors mislabel any study that invokes cases and controls as a case-control study, though often these do not adhere to the case-control methodology as a study design. The most commonly confused study design is a cohort study, where groups are divided on exposure (eg, treatment) and the rates of a specific outcome are assessed in each exposure group (Figure 1).

**METHODS**

A complete sample of articles was used for this detailed qualitative review. An online search was
undertaken using the phrase “impact factors for neurosurgery journals” to provide a specialty list of journals that would be used by neurosurgeons. This search produced a site entitled “neurosurgic.com,” which provided a cross-sectional list of neurosurgical journals (Supplemental Digital Content 1, Appendix A). Thirty-three journals were featured, and these were chosen as index journals, with 2010 impact factors, where known, from 0.103 (Neurosurgery Quarterly) to 4.764 (Journal of Neurology, Neurosurgery and Psychiatry). A National Library of Medicine online search (PubMed) was carried out for each of the journals from January 1988 to December 2012 using the search terms “case-control” and “case control” in the title or abstract. This produced a total of 132 references.

The assessment team included 5 members of the author group which included an attending neurosurgeon with a master’s degree in clinical epidemiology (B.C.W.), 3 neurosurgical residents with training in critical appraisal (K.P.K., E.C.D., A.S.G.), and a doctorate-level epidemiologist with a special expertise in case-control design for a final decision. Papers excluded at the first pass qualitative review were not subjected to further qualitative review. The final papers that met the methodological criteria were applied to both cases and controls, controls were proven not to be cases, and confidence intervals were included in the results. In particular, instructions were given in the checklist with respect to evaluating the papers as follows based on the SIGN checklist instructions: “High quality (+): Majority of criteria met. Little or no risk of bias. Results unlikely to be changed by further research. Acceptable (+): Most criteria met. Some flaws in the study with an associated risk of bias. Conclusions may change in the light of further studies. Rejected (0): Either most criteria not met, or significant flaws relating to key aspects of study design. Conclusions likely to change in the light of further studies.” Again, each paper was assigned to a pair of raters who independently used the checklist to assign a final rating of high quality (HQ), acceptable, or rejected. To understand our interpretation and application of the SIGN instructions, in a HQ paper, both raters were required to declare the study HQ for sound study design, data analysis, and conclusions, with no more than one-failed category on the 11-point assessment. The failed category, if any, was minor related to presentation of the results (eg, inclusion of confidence intervals). The failure of a major methodological category was not allowed in HQ papers. HQ papers can be looked to as practice-defining publications due to their methodological rigor. Papers rated acceptable fit the definition of case control but often drew conclusions that weren’t wholly proven by their results and study design, most often choosing an inappropriate control population which was either not appropriately proven to be a noncase, or did not differ appropriately from the case population in order to investigate the clinical question. Acceptable papers are able to guide clinical decision making as long as an understanding of their shortcomings are appreciated. Papers that were rejected in this round featured a major construct error in methodology, data analysis, or conclusions, and should not be used to influence clinical practice. Papers that did not include a clear definition of the study population, failed to state outcome measures, or omitted a discussion of the possibility of confounding were included in this category. No manuscript that met the methodological criteria of a case-control study received a rating of reject based on the rating scheme previously discussed.

In the event of a disagreement between the 2 raters between accept and reject, the publication was assessed by the remaining raters (4 at this time) and a consensus reached. In the event of a tie or strong debate, papers were sent to an independent epidemiologist (N.E.M.) with special expertise in case-control design for a final decision. Papers with 1 acceptable and 1 HQ rating were assigned the label of acceptable. Only papers earning unanimous HQ ratings were granted the label of HQ. There were no discrepancies between reject and HQ.

RESULTS

Of the 125 studies examined in detail, 50 did not qualify as case-control studies based on first pass (abstract) evaluation. The most common study design in this category was cohort study (30), followed by cross-sectional studies (10), reliability studies (4), case series (1), and meta-analysis (5; Table 1).

Of the 75 studies assessed for quality, 11 were determined to be HQ. Thirty-five papers earned an acceptable rating, while 29 submitted to a detailed critical appraisal checklist to establish quality using the SIGN Methodology Checklist for case-control studies (second pass qualitative review; see Figure 3). The 11-item checklist focuses on objective qualities of the paper such as assessing that the same exclusion criteria were applied to both cases and controls, controls were proven not to be cases, and confidence intervals were included in the results. In particular, instructions were given in the checklist with respect to evaluating the papers as follows based on the SIGN checklist instructions: “High quality (+): Majority of criteria met. Little or no risk of bias. Results unlikely to be changed by further research. Acceptable (+): Most criteria met. Some flaws in the study with an associated risk of bias. Conclusions may change in the light of further studies. Rejected (0): Either most criteria not met, or significant flaws relating to key aspects of study design. Conclusions likely to change in the light of further studies.” Again, each paper was assigned to a pair of raters who independently used the checklist to assign a final rating of high quality (HQ), acceptable, or rejected. To understand our interpretation and application of the SIGN instructions, in a HQ paper, both raters were required to declare the study HQ for sound study design, data analysis, and conclusions, with no more than one-failed category on the 11-point assessment. The failed category, if any, was minor related to presentation of the results (eg, inclusion of confidence intervals). The failure of a major methodological category was not allowed in HQ papers. HQ papers can be looked to as practice-defining publications due to their methodological rigor. Papers rated acceptable fit the definition of case control but often drew conclusions that weren’t wholly proven by their results and study design, most often choosing an inappropriate control population which was either not appropriately proven to be a noncase, or did not differ appropriately from the case population in order to investigate the clinical question. Acceptable papers are able to guide clinical decision making as long as an understanding of their shortcomings are appreciated. Papers that were rejected in this round featured a major construct error in methodology, data analysis, or conclusions, and should not be used to influence clinical practice. Papers that did not include a clear definition of the study population, failed to state outcome measures, or omitted a discussion of the possibility of confounding were included in this category. No manuscript that met the methodological criteria of a case-control study received a rating of reject based on the rating scheme previously discussed.

In the event of a disagreement between the 2 raters between accept and reject, the publication was assessed by the remaining raters (4 at this time) and a consensus reached. In the event of a tie or strong debate, papers were sent to an independent epidemiologist (N.E.M.) with special expertise in case-control design for a final decision. Papers with 1 acceptable and 1 HQ rating were assigned the label of acceptable. Only papers earning unanimous HQ ratings were granted the label of HQ. There were no discrepancies between reject and HQ.

RESULTS

Of the 125 studies examined in detail, 50 did not qualify as case-control studies based on first pass (abstract) evaluation. The most common study design in this category was cohort study (30), followed by cross-sectional studies (10), reliability studies (4), case series (1), and meta-analysis (5; Table 1).

Of the 75 studies assessed for quality, 11 were determined to be HQ. Thirty-five papers earned an acceptable rating, while 29

---

**FIGURE 1.** Case control vs cohort and cross-sectional design.
were rejected. Of the original 125 papers claiming to be case-control studies by PubMed search, 79 were found not to be case-control studies using standard criteria. Fifty papers were rejected strictly based on abstract and another 29 were rejected at the point of assessment for quality, giving the neurosurgical literature surveyed a 63% rate of mislabeled case-control studies. The rejected manuscripts featured study designs that were primarily cross-sectional studies (23), as well as cohort (5) and case-cross over (1). A flowchart of the study results can be found in Figure 4 and results in Table 2.

The most common difference between acceptable and HQ papers was in their treatment of the controls. Acceptable papers often failed to use the same exclusion criteria for cases and controls. Additionally, acceptable papers often did not prove that controls were not cases, which would have required some sort of testing rather than a convenience sample of controls. Though manuscripts may have addressed these issues in their design, many did not specifically state so in the manuscript, and thus failed any category for which supportive information was not clearly defined.

High-Quality Papers

The 1 HQ papers were further investigated to find any defining publication characteristics. These papers were not of a specific type or genre or clustered in familiar journals. HQ papers were scattered throughout journals of varied impact factors (Table 3).

Discrepancy

On the first round of assessments (abstract review for case control methodology), 8 (6%) of the 125 papers assessed purely on methodology incurred discrepancies between the pair of raters and required review by the expanded panel for consensus rating. On the second round of appraisal (manuscript review for quality), 12 (16%) of the 75 papers reviewed for quality earned disparate ratings by the assigned assessment pair and required review by all team members for consensus rating. Six (8%) of those 75 papers could not be satisfactorily categorized by the initial assessment team and required appraisal by an independent epidemiologist (N.E.M.). In all, 20 (15.9%) of the initial 126 papers required some additional or expanded review due to disagreement among the raters.
DISCUSSION

The results of our study show that there is an abundance of mislabeled case-control studies in the neurosurgical literature. We found 63% of studies to be mislabeled as case-control studies. Previously published studies have found a mislabeled case-control study rate of 97% in the physical medicine and rehabilitation literature and 30% in the obstetrics and gynecology literature. A recent study of the neurosurgical literature found a similar case control labeling inaccuracy of 52%. Esene and colleagues recently reviewed 224 neurosurgical publications purporting to be case-control studies using the STROBE criteria and found 59.38% to be misclassified. Similar to our experience, cohort studies were the most common study design among misclassified case-control studies. Our work did find a higher proportion of cross-sectional studies among the misclassified studies (41.8% vs 2.7%). One reason for this may be the exclusion of genetic studies from the case-control definition in our study, feeling they are better classified as cross-sectional studies. Since genetic mutation is possible, there is no confirmation that exposure preceded outcome when genetic samples are obtained at or after the time of outcome. Since routine cataloging of genetic information prior to outcome onset is not standard, it is nearly impossible to carry out a true case-control study using genetic exposures unless a DNA source prior to the outcome was identified and confirmed to be free of degradation or mutation. While some may feel that genetic studies can be effectively classified as case controls since a priori exposure collection is not routinely attainable, likening them to case-control studies with retrospective exposure collection, we felt adhering to this strict definition was most appropriate in light of applying the same criteria to other, nongenetic...
WHAT ISN'T A CASE-CONTROL STUDY?

Of the 33 cross-sectional rejected studies, 12 were genetic papers.

Our study is the first to assign an established quality rating to case-control studies in the neurosurgical literature. More startling than the low number of actual case-control studies, fewer than 10% of the studies we assessed met the criteria for a HQ case-control study. This points not only to a fundamental misunderstanding of case-control studies and what appropriate methodology is among the neurosurgical research community as it pertains to study design and analysis, but also manuscript review and approval for publication by journal reviewers. Of the true case-control studies identified, no journal had more than 60% rate of HQ, practice-defining, case-control studies, suggesting nearly half of all case-control studies need to be applied with caution and appreciation for limitations. \(\text{The European Journal of Pediatric Surgery}\) appears to have a 100% HQ rating, but only 1 paper found by the authors was included for review and an exhaustive search of this journal was not undertaken.

With a constantly growing body of literature, rife with conflicting studies, these incorrectly designed or inappropriately analyzed studies add to the noise that detracts from the information relied upon by neurosurgical clinicians in search of high fidelity data with reliable and sound conclusions on which to base their practices. Busy professionals lack the time to perform exhaustive searches and sift through papers to assess their quality, leading to the use of faulty information for clinical decision-making. With the majority of clinicians lacking formalized training in critical appraisal of the literature, many assume anything passing peer review can be trusted. Many journals are employing methodological reviewers to clear studies prior to acceptance for publications to stem the epidemic of publishing clinically relevant but inaccurately designed and faulty studies. These inaccurately designed studies can yield conclusions that are incorrect and misleading, having an adverse effect on patient care and future research.

In addition, there is a bias toward assuming that the contents of a published paper can be trusted de facto. Readers rely on the journal and reviewers to have cleared a peer-reviewed and published manuscript of any egregious methodological errors. There is increasing evidence and outside media attention on the failure of this process, including the lack of reproducibility and major holes in previously considered landmark studies.\(^{18}\) This mounting evidence will serve to erode faith in the publication process and literature unless more rigorous reviews are conducted prior to publication. Our study also supports the existing literature that the quality of papers does not correlate with the prestige or impact factor of a journal.\(^{19}\)

### Table 1. Reasons for Rejection

<table>
<thead>
<tr>
<th>Correct-study design</th>
<th>Rejected 1st pass ((n = 50))</th>
<th>Rejected 2nd pass ((n = 29))</th>
<th>Total ((n = 75))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>30</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>Cross sectional</td>
<td>10</td>
<td>23</td>
<td>33</td>
</tr>
<tr>
<td>Reliability/Diagnostic test</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Case series</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Case-cross over</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Figure 4. Study design and results flow chart.
Correctly assessing a study and its utility to influence practice decisions can be difficult. The discrepancy rate of 15.9% among our raters, trained and attuned to this practice, highlights the somewhat subjective and challenging nature of this process. Also staggering, 29 papers were rejected at the point of quality assessments. Of the 79 mislabeled papers, 36% progressed into the second round of appraisal by our team of specially trained raters, and only at the point of a full review of the manuscript was it clear that the paper was not appropriate to be designated as a case-control study. This highlights the misleading nature of some abstracts and the need for caution when reviewing the literature. Existing studies suggest that inclusion of a trained epidemiologist or biostatistician improve the likelihood of correctly carrying out a case-control study.17

The majority of papers that were misclassified as case control were in reality cohort studies or cross-sectional studies. Cohort studies differ from case-control studies in that they begin their investigation by dividing study groups based on differing exposures or treatments, and then follow the patients to compare the rates of outcomes among the groups. Cohort studies were almost always identified as such in the first pass/abstract review. Case-control studies start by identifying and separating patient groups based on outcome, and then comparing past exposure rates or treatments among cases and controls. The relevant test statistic is an odds ratio, which is calculated differently than the relative risk that applies to cohort studies. Mislabeling a study can lead to applying inappropriate statistics and drawing unfounded conclusions. While many case-control studies collect information on exposure after the outcome, the subjects are asked to recall and report on their exposures prior to the outcome so the data reflect pre-outcome exposures. To qualify as a case-control study, the exposure must predate the outcome, but data on the exposure may be collected prospectively or retrospectively. The opportunity for recall bias that may be differentially expressed by cases and controls because of their experience of a given outcome is a known limitation of case-control studies using a retrospective collection method. When exposure data are collected in this way, the possibility of recall bias needs to be addressed in the limitations of the study, and objective measures of exposure prior to the outcome are preferred when possible.

Cross-sectional studies are also observational studies, but examine patients at a single point in time, providing a snapshot of patient demographics and disease states, used to better understand the patient population. The majority of studies rejected on the second pass were cross-sectional studies, suggesting that special and rigorous attention is applied to differentiating this study design, as the nuances can be challenging for even trained reviewers to appreciate, especially from an abstract.

A common situation encountered in the second pass was genetic studies that looked at genes among patients with and without a specific pathology. The patients were genetically sampled at the time of the outcome/pathology, so it is impossible to say that the genetic variations found predate the outcome, as is necessary for an exposure in a case-control study. While the genetic code is felt to be stable, genetic mutations as a result of study design, as the nuances can be challenging for even trained special and rigorous attention is applied to differentiating this study design, as the nuances can be challenging for even trained reviewers to appreciate, especially from an abstract.

CONCLUSION

A large portion of case-control studies in the neurosurgical literature are mislabeled and, therefore, unusable. Readers must be cautious of these incorrectly executed studies when reviewing the literature. Publishers and reviewers should also be alert to this when assessing papers prior to publication. Researchers designing studies should also solidify their clinical question and study design, often with consultation of a trained epidemiologist

<table>
<thead>
<tr>
<th>Journal</th>
<th>Impact factor 2010</th>
<th>Total number of papers assessed</th>
<th>Number of papers found to be case control studies</th>
<th>Number of high-quality papers</th>
<th>% High quality per total papers assessed by journal</th>
<th>% High quality per true case control studies per journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Neurology, Neurosurgery, and Psychiatry</td>
<td>4.924</td>
<td>10</td>
<td>6</td>
<td>2</td>
<td>20%</td>
<td>33%</td>
</tr>
<tr>
<td>Journal of Neurosurgery</td>
<td>3.148</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td>12.5%</td>
<td>50%</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>2.532</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>14%</td>
<td>50%</td>
</tr>
<tr>
<td>Journal of Clinical Neuroscience</td>
<td>1.253</td>
<td>11</td>
<td>6</td>
<td>3</td>
<td>27%</td>
<td>50%</td>
</tr>
<tr>
<td>Child's Nervous System</td>
<td>1.241</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>33%</td>
<td>60%</td>
</tr>
<tr>
<td>European Journal of Pediatric Surgery</td>
<td>0.839</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
or biostatistician, prior to initiation of data collection. Additional training on case-control methodology and interpretation is warranted in the neurosurgical community.

Disclosure

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES


2. SIGN Notes on Methodology Checklist 4: Case Control Studies. Available at: https://www.sign.ac.uk/assets/notes_on_case_control_studies_checklist.docx.


Supplemental digital content is available for this article at www.neurosurgery-online.com.


Supplemental Digital Content 2. Appendix B. Bibliography of Relevant Journal Articles Assessed

COMMENT

The authors expand on their prior publication “What Is a Case Control Study?”1 with a systematic review of published studies called “case control” by their authors. Their most important finding is that a substantial proportion of such self-identified “case control” studies used other designs (cohort and cross sectional being the most common). As they point out, appropriate statistical analysis depends on study design and therefore the statistical support for the results of these misclassified studies may be called into question.

These findings, supported by the report of Esene cited by the authors, emphasize the importance of having at least a basic understanding of the fundamental requirements for various study designs and of independently evaluating authors’ claims about their own work rather than relying entirely on the peer review and editorial process for assessment of the quality of published studies. Caveat emptor.

The authors take a firm position on genetic studies using the case-control methodology, raising the concern that it is often not possible to be sure of the timing of a genetic mutation and therefore considering such studies cross-sectional rather than case-control. This is a controversial position in the field and further work to better understand the most appropriate analysis of such studies is required.

Stephen J. Haines

Minneapolis, Minnesota

---