Objective: To evaluate how often visual acuity outcomes are reported in the ophthalmological literature as best or final outcomes, despite potential bias with this type of analysis, as compared with interval outcomes, when a specific condition may continue to cause gain or loss of visual acuity beyond the time that the best or final outcome is determined.

Methods: Each article published in the 3 most frequently cited comprehensive clinical ophthalmological journals in the United States from January through December 2000 was reviewed. Clinical studies were identified in which visual acuity was used as an outcome measure. Visual acuity outcomes were examined throughout the articles and classified as follows: best visual acuity, defined as an outcome at any time during follow-up; final visual acuity, defined as an outcome at last follow-up; and interval visual acuity, defined as an outcome at specified follow-up times. A few factors that might be associated with the different types of outcome were evaluated. Reproducibility of the categorization between 2 ophthalmologists was determined by using the $\kappa$ statistic.

Results: A total of 527 clinical studies met the criteria. Authors of 195 reported visual acuity as an outcome measure. Authors of 1 article (0.5%) reported only best visual acuity, authors of 6 (3%) reported both best and final visual acuity, authors of 113 (58%) reported only final visual acuity, and authors of 73 (37%) reported interval visual acuity outcomes. Reproducibility of these categorizations between 2 ophthalmologists was considered excellent, as compared with chance alone ($\kappa=0.84$). Authors of only 2 of the 120 articles that used either best or final visual acuity outcomes discussed the limitations or potential bias of reporting outcomes in this way. Randomized trials and other prospective study designs more often were associated with interval outcomes than were nonrandomized and retrospective studies.

Conclusions: Despite potential bias with use of best or final visual acuity outcomes, these end points alone were used in most studies published during 2000 in the 3 most commonly cited journals. Authors of clinical studies should consider avoiding use of best or final visual acuity outcomes whenever possible to minimize possible data misinterpretation. If best or final outcomes are used, authors should consider discussing the limitations of these methods and their potential effect on the interpretation of results.

Arch Ophthalmol. 2003;121:1586-1590

CLINICAL (HUMAN) studies in ophthalmology often have visual acuity as the main or primary outcome measure. In randomized clinical trials in ophthalmology, visual acuity data points usually are collected at specific follow-up intervals that are selected prior to study initiation, usually within a time window around those intervals, such as 12 months after study entry plus or minus 1 month. Because outcomes are usually time dependent, the visual acuity outcome of each patient should be obtained at the same interval after study entry so that the patients in control and experimental arms of the study have similar prognoses. For example, in a study comparing laser photocoagulation with no laser photocoagulation for macular edema, if the group with no laser photocoagulation was followed up for only 3 months and visual acuity was maintained in 95% of the patients, while the group with laser photocoagulation was followed up for 12 months and visual acuity was maintained in only 75% of the patients, one might erroneously conclude that laser photocoagulation results in a worse outcome. To minimize these biases, randomized clinical trials usually compare outcomes of the experimental and control groups at the same interval after study entry.

If studies that are not randomized are to guide the feasibility, safety, and design of randomized clinical trials or influence...
clinicians in the management of ophthalmologic diseases, interval outcomes of visual acuity are needed to minimize the biases described. For example, when evaluating visual acuity outcomes after laser photocoagulation for choroidal neovascularization, if outcomes in 50 patients at 3 months after laser photocoagulation are combined with outcomes in 50 other patients at 24 months after such treatment, the results of these 100 patients with an average follow-up of 13.5 months may be biased toward better outcomes than what can be expected in reality because half of the patients are followed up before the natural history of the condition leads to some loss of visual acuity and before many instances of recurrent choroidal neovascularization lead to additional loss of visual acuity.

Authors of a large number of published ophthalmological articles used visual acuity end points that were at a time other than a specified interval after study entry as a main outcome measure. In these articles, the adjectives “best” or “final” often were used to describe these visual acuity end points. Final was used to group all visual acuities that were the last ones recorded in each patient, whether it was, for example, at 3 months or 24 months. Some articles listed best visual acuity, meaning any time at which the patient’s best visual acuity was recorded between study entry and a final visual acuity outcome.

A major drawback with best or final visual acuity outcomes is the potential bias of overestimating the number of people with good visual acuity outcomes or underestimating the number of people with bad visual acuity outcomes when a specific condition may degenerate, with loss of visual acuity, beyond the time that the best or final outcome was obtained. Alternatively, best or final visual acuity outcomes may lead to underestimation of the number of people with good visual acuity or overestimation of the number of people with bad visual acuity when a specific condition might improve, with gain in visual acuity, beyond the time that the best or final outcome was obtained. The purpose of this article is to evaluate the use of best or final visual acuity reporting, which potentially could bias conclusions regarding visual acuity outcomes in the ophthalmological literature.

METHODS

Articles from January through December 2000 in the 3 comprehensive clinical ophthalmological journals in the United States most cited on the basis of the average impact factor were searched for in the Wilmer Ophthalmological Institute (Baltimore, Md) electronic journal database. These were the American Journal of Ophthalmology, Archives of Ophthalmology, and Ophthalmology. Each article was examined after excluding nonhuman studies, case reports, and small case series with fewer than 5 cases. The number of published articles was counted as was the number of articles in which visual acuity was used as an outcome measure. All of the articles in which visual acuity was used as an outcome measure were then reviewed, and the visual acuity data points were examined.

Visual acuity data points were classified into 3 main categories: those reported as best, final, or interval visual acuity. Best visual acuity was defined as an outcome based on the patient’s best visual acuity any time during the posttreatment period. Final visual acuity was defined as an outcome based on the visual acuity recorded at the last follow-up. In these studies, the final time assessed varied among subjects, depending on the length of follow-up. Interval visual acuity was defined as an outcome based on a specified posttreatment time at which all patients available had their visual acuity evaluated and recorded. Use of last observation carried forward for missing data was considered an interval outcome if fewer than 10% of the observations were missing and imputed in this way. An article in which the authors used a time-to-event analysis, such as a Kaplan-Meier curve, was included as an interval visual acuity; however, a specific tally of such instances was recorded. Because a condition that is no longer expected to deteriorate or improve after a specified time is not likely to result in much bias when one compares final outcomes with interval outcomes, final outcomes that spanned a time when fewer than 10% of the cases would be expected to change also were tallied separately.

To assess the reproducibility of this categorization, a sample of the articles was reexamined as follows: 2 independent random numbers were generated from 1 to 12 by using a random number generator. These numbers were used to represent months of the year from January through December. The 2 months of the year chosen randomly then were reexamined by the author (D.A.D.) who performed the original search, this time using library copies of the journals to obtain the same information as originally gathered. A second author (N.M.B.), who was unaware of the data collected by the first author, then performed an independent search of these 2 months by using library copies of the journals. Reproducibility of answers was analyzed with the $k$ statistic.

Factors that may influence the type of visual acuity reporting then were analyzed and compared, including journal, study sample size (described later), whether the study was a randomized clinical trial, whether visual acuity data were collected in a prospective or retrospective fashion, and subspecialty (cornea or anterior segment, glaucoma, neuroophthalmology, oculoplastics, pediatrics, refractive surgery, retina, and uveitis). Some articles had components of more than 1 subspecialty. In these cases, each subspecialty involved in the article received credit for that type of visual acuity reporting. The percentage of articles in which either best or final visual acuity was used per total number of articles published in which visual acuity was used was compared from journal to journal and with all 3 journals combined. A visual acuity report ratio was then calculated for each subspecialty on the basis of the following formula: (number of articles with final visual acuity + number of articles with best visual acuity) / number of articles with interval visual acuity.

All participants in the study were placed into groups as follows: 5 to 10, 11 to 20, 21 to 40, 41 to 80, 81 to 160, and so on, with each category having an upper bound that was double the upper bound of the previous category. The statistical characteristics of the sample sizes were compared, including the mean, SD, range, median, and mode. Each article in which best or final visual acuity was used was reviewed to determine if the potential bias of best or final were mentioned in the discussion section of the article.

RESULTS

We identified a total of 527 articles, and authors of 195 (37%) used visual acuity as an outcome measure. These articles are listed alphabetically by first author at the following Web site: http://www.wilmer.jhu.edu/departments/RVC.HTM. Of the 195 articles in which visual acuity was used as an outcome measure, authors of 1 article (0.5%) reported only best visual acuity; authors of 6 (3%) reported both best and final visual acuity, authors of 113...
FACTORS ASSOCIATED WITH USE OF BEST OR FINAL VISUAL ACUITY OUTCOMES

For each journal searched, the type of visual acuity outcomes reported is shown in Figure 1. Results were similar across these journals, although the total number of articles in which visual acuity was used as an outcome varied among the journals. One of 7 articles in which best and final outcomes were used and 34 of 113 articles in which final outcomes were used had prospective study designs, as compared with 46 of 73 articles in which interval outcomes were used. Among these prospective studies, none of the articles in which best and final outcomes were used, 3 of 113 in which final outcomes were used, and 21 of 73 in which interval outcomes were used were randomized clinical trials. The subspecialty factor showed greater variation. The visual acuity report ratio (the number of articles with final visual acuity plus the number of articles with best visual acuity divided by the number of articles with interval visual acuity) ranged from 0.33 to 13.00, but the total number of articles across subspecialties varied to create these ratios (Figure 2). Statistical characteristics of the study sizes grouped according to type of visual acuity outcome reported are shown in the Table. The distribution of the range of sample sizes per type of visual acuity reporting is shown in Figure 3.

POTENTIAL BIAS OF BEST OR FINAL VISUAL ACUITY OUTCOMES

Authors of only 2 of 120 articles in which best or final visual acuity outcomes were used briefly mentioned the potential bias of reporting visual acuity this way.

The results of this study are limited by the fact that the study is retrospective in nature . . . best-corrected Snellen visual acuities were not obtained according to a standardized protocol. . . . Outcome of data consisted of visual acuity at last follow-up examination and were not obtained at scheduled follow-up intervals, therefore analysis of outcome data could not be performed.5

“However, the results of the current study are limited by the fact that the study is retrospective and . . . best-corrected Snellen visual acuities were not obtained according to a standardized protocol.”6

REPRODUCIBILITY OF CATEGORIZATIONS

For the 2 months sampled randomly to evaluate reproducibility of these results, 1 author (D.A.D.) identified 91 articles that excluded nonhuman studies, case reports, and small case series and identified 37 articles in which visual acuity was used as an outcome. Among these 91 articles, the second author (N.M.B.) identified 32 articles in which visual acuity was used as an outcome; all were the same articles identified by the first author. The agreement between the 2 authors for categorizing these 32 articles as final or not final (ie, best or interval) visual acuity outcomes had a κ value of 0.84, which is considered almost perfect agreement.4 At review of the 5 articles that were not categorized according to both reviews as having visual acuity as an outcome, none of the 5 had visual acuity mentioned as an outcome measure in the abstract. Authors of 4 of the articles mentioned visual acuity outcomes that were interpreted as a safety measure rather than an effectiveness outcome by the second reviewer. The fifth article showed limited visual acuity outcomes in a table that in retrospect the second reviewer would have categorized as an effectiveness outcome.
Large, multicenter, randomized clinical trials usually use standard intervals to record visual acuity outcomes because use of best or final visual acuity outcomes may mislead the reader regarding conclusions or bias the results unintentionally when the condition being studied may continue to improve or deteriorate beyond the time of the best or final outcome. Results of this study showed that among the most commonly cited comprehensive ophthalmological journals, best or final visual acuity outcomes were used in most clinical studies despite the potential biases of these methods. Interval outcomes were used more often in prospective studies, and among these, 24 were randomized clinical trials, 21 of which had interval outcome results. Use of best or final visual acuity outcome was more often associated with small studies that involved 5 to 20 participants, although such outcomes were also used in many articles with a larger number of participants. In this study, only a few factors (subspecialty, study size, and journal) that might influence the use of best or final visual acuity as an outcome measure were analyzed. Other factors may not have been analyzed or an insufficient number of articles may have been evaluated to detect an association when one actually exists.

Although this article quantitates the potential problem of using these methods, it does not explain why best or final visual acuity is reported in most ophthalmological research articles. Such reasons may include the following: authors may not be aware of such biases, reviewers or editors may not be aware of the potential limitations of these methods, costs of interval outcomes might be greater, or retrospective data at specific intervals may not be available. However, even if retrospective data are difficult to collate, it should be possible to evaluate patients at approximately the same posttreatment times. If there are outlying values for patients who have too short or long of a follow-up, they should be excluded from the analysis to strengthen the comparisons in patients with a similar length of follow-up.

Although interval outcomes are recommended to avoid potential biases in best and final visual acuities, interval outcomes may be limited if there are many missing data points at the intervals chosen. Time-to-event analyses do not necessarily overcome these potential biases. For example, in articles in which only time-to-event analysis was used, one might assume that the results are unbiased and can overcome some of the limitations of missing data for interval outcomes. However, such analyses possibly show overestimated event

---

**Table 1. Statistical Characteristics of Study Sizes as Grouped According to Type of Visual Acuity Outcome Reported**

<table>
<thead>
<tr>
<th>Statistical Characteristic</th>
<th>Best</th>
<th>Final</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>18.0</td>
<td>93.2</td>
<td>213.9</td>
</tr>
<tr>
<td>SD</td>
<td>18.0</td>
<td>322.9</td>
<td>474.7</td>
</tr>
<tr>
<td>Minimum</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Maximum</td>
<td>55</td>
<td>3442</td>
<td>3427</td>
</tr>
<tr>
<td>Median</td>
<td>11.0</td>
<td>29.5</td>
<td>71.0</td>
</tr>
<tr>
<td>Mode</td>
<td>5</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

*Best visual acuity was more apt to be used in smaller studies, while interval visual acuity was more often used in larger studies. Although it is difficult to compare the means because the data are nonuniform, which results in a large SD of the mean in each group, the mode and median reinforce the findings.*

---

**Figure 3.** Type of visual acuity outcome used according to study sample size. For best and final visual acuity outcomes, the greatest number of articles was in the smallest sample size category of 5 to 10. For best visual acuity outcome, the trend declined sharply, while for final visual acuity outcome the decline was more gradual. For interval visual acuity outcome, the greatest number of articles was in the 81 to 160 range. As opposed to the downward trends seen for articles in which best or final visual acuity outcomes were used, articles in which interval visual acuity outcome was used showed a bell-shaped curve that peaked in the smaller to medium study categories.
rates or visual acuity outcomes. Such analyses may benefit from supplementation with interval visual acuity outcomes that clearly define the number of participants available for follow-up at each interval.

If time-to-event analysis is used, then follow-up data have to be carefully censored to avoid overestimating an event rate. Time-to-event analyses also may lead to overestimating an event rate when the event rate is an end point from which one could recover, such as loss of at least 6 lines of visual acuity at baseline, in contrast to an event from which one could not recover, such as death, if one uses a survival analysis such as a Kaplan-Meier curve. Using a 2-state stochastic model for an event such as loss of at least 6 lines of visual acuity can take into account recovery from such loss.

Although authors of 62% of the articles evaluated reported best or final visual acuity outcomes, authors of only 2 articles discussed the potential bias of such reporting in the discussion section of the article. Without such a discussion, some readers may not recognize the potential of such reporting to invalidate conclusions of the study.

Several approaches could be considered to try to overcome these potential biases. First, authors, reviewers, and editors could recognize and highlight shortcomings of such articles to assist readers in interpreting the results. Second, once these potential biases are brought to their attention, researchers could strive to use interval reporting. In the methods section, authors could discuss the rationale for the windows chosen around these intervals (eg, 6-month, 12-month, 18-month, and 24-month follow-up with a window of 2 months around each interval to try to maximize use of even retrospective data that might be collected at a time near to but not at a specific interval). Because patients often have varying follow-ups (eg, some have only 6 months of follow-up, while others may have 24 months of follow-up), interval outcome data should indicate the actual number of study participants who returned at each specific interval. In that way, the reader can judge potential limitations or confidence in the results given the number of missing values at each interval. Third, if best or final visual acuity outcomes must be used, the rationale for using them rather than interval outcomes should be justified and the effect of their use on the validity of the results and conclusions should be made obvious to the reader in the abstract and discussion sections. With these suggestions, the quality of clinical articles and their effect on clinical care may improve.

Submitted for publication October 1, 2002; final revision received May 15, 2003; accepted June 9, 2003.

This study was supported by the HEED Foundation, Plainview, NY (Dr DiLoreto); Ronald G. Michels Fellowship Foundation, Scarsdale, NY (Dr DiLoreto); Herman Knapp Testimonial Fund, Cleveland, Ohio (Dr DiLoreto); Altsheler-Durrell Foundation; and Michael B. Panitch Stop Age-related Macular Degeneration Research Fund.

We thank Roy W. Beck, MD, PhD, Jaeb Center for Health Research, Tampa, Fla, who provided helpful review and editing of early drafts of the manuscript.

Corresponding author: Neil M. Bressler, MD, 550 N Broadway, Suite 115, Baltimore, MD 21205 (e-mail: nmboffice@jhmi.edu).

REFERENCES