HHS Public Access

Author manuscript

Am J Ophthalmol. Author manuscript; available in PMC 2022 March 21.

Published in final edited form as:

Am J Ophthalmol. 2005 May ; 139(5): 900–905. doi:10.1016/j.ajo.2004.12.009.

Improving the Reporting of Clinical Case Series

DOUGLAS A. JABS, MD, MBA

Departments of Ophthalmology and Medicine, The Johns Hopkins University School of Medicine, and Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland.

Abstract

PURPOSE: To describe common errors in the analyses and data presentation of a clinical case series and to suggest simple solutions.

DESIGN: Instructional examples.

METHODS: Problems with commonly used data analysis and reporting techniques in clinical case series are described using both theoretical examples and those from the literature.

RESULTS: An analysis reporting the proportion of a series of patients with variable follow-up does not adequately account for the differential follow-up among patients and is a potentially misleading way to present data. Instead, the proportion of patients at presentation (or study entry) and the rate during follow-up should be reported. Similarly, an analysis in which the final visual acuity of a series of patients with variable follow-up is reported does not adequately account for the effect of time and also may be misleading. Reporting of the rates of visual acuity events during follow-up (e.g., falling below a specified threshold, such as 20/50 or worse) is preferred. Alternatively, when there is nearly complete follow-up, reporting the distribution of visual acuity at specified time points (e.g., 1 year after study presentation) is appropriate. Small case series should not be overinterpreted because of the effects of chance, and appropriate statistical analyses should be employed.

CONCLUSIONS: Clinical case series often suffer from several potential reporting flaws. Correction of these flaws would permit the proper interpretation of the data and allow for the ability to combine data from several case series to assemble more meaningful and reliable conclusions.

Although nearly every author of articles in the medical literature recognizes that randomized controlled clinical trials are the gold standard for evaluating treatments and that epidemiologic studies provide important information on risk factors and long-term outcomes, clinical case series still occupy a substantial proportion of the ophthalmic literature. Case series are attractive to clinicians because they often replicate their experience and are perceived to provide insights as to how to handle issues facing them in the management of their patients. Their data analysis appears easy to perform and to understand and provides numbers that are easy to remember. They often provide information on

Inquiries to Douglas Jabs, MD, The Wilmer Eye Institute, 550 North Broadway, Suite #700, Baltimore, MD 21205; Fax: (410) 955-0629; djabs@jhmi.edu.

new treatments and important preliminary data necessary for randomized clinical trials. Furthermore, for uncommon diseases, an individual center's experience may be limited, and thus data from a case series may be all that can be assembled for publication. Nevertheless, typical methodologic flaws in the data analysis of case series render these reports difficult to interpret at best and misleading at worst. The term "uncontrolled case series" has almost acquired a pejorative connotation, which has led to euphemisms (such as "retrospective open-label interventional trial") being used in the literature to disguise the nature of the study. However, many of the flaws in the analysis of clinical case series are often easy to rectify and do not require highly sophisticated biostatistical approaches. Correction of these flaws would allow not only the proper interpretation of the data but also the ability to combine data from several case series to assemble a more meaningful and reliable conclusions about the effect of the disease and its treatment.

This article attempts to demonstrate some of the more common flaws in data analysis of case series and propose simple solutions. If all authors writing case series adhered to these principles, the quality of the literature would improve, and the ability to compare or combine data from disparate sources would be improved, thereby enhancing the information available to physicians.

PROPORTIONS OF A POPULATION WITH VARIABLE FOLLOW-UP

One of the more common statements made in a clinical case series is one such as "In 28% of patients with some disease (e.g., uveitis) followed for a mean of some time (e.g., 9 months, range 3 months-5 years), some complication (e.g., elevated intraocular pressure) occurred." Although at face value, the statement seems reasonable, it actually conveys little or no information and may convey misleading information. As shown in Figure 1, two very different event curves both are consistent with this statement. In curve B, 28% of patients were affected at study entry, and no new cases of elevated intraocular pressure developed. In curve A, the majority of patients have been followed for a short time (mean follow-up is 9 months, but the range is large) and lack sufficient exposure to experience the side effect, whereas of those few patients followed for a long time nearly all experienced the side effect. Because most of the patients in the series have been followed for a short time and have not developed the complication, the overall percent of patients with the complication is low. The problem is that the proportion of a series of patients followed for varying lengths of time ("variable follow-up") does not account for the effect of time and therefore is a flawed concept. A more appropriate way to analyze these data is to report the proportion at presentation (or study entry) and the rate during follow-up (incidence). Rate is the number of events in those at risk for the event divided by total follow-up person time.[1] It is calculated by dividing the number of events (e.g., patients with elevated intraocular pressure) by the sum of the patients' follow-up time (typically expressed as person-years). A person-year is one patient followed for 1 year, two patients followed for 6 months, or three patients followed for 4 months, and so on. The rate typically is expressed per unit time (e.g., 0.50/person-year) or as a percent per unit time (e.g., 50%/person-year). With the proper reporting, the curve in Figure 1A has a proportion of 0% at presentation and a rate of 50%/person-year during follow-up (or alternatively 0.50/person-year). Conversely, the curve in Figure B has a proportion of 28% at presentation and a rate of 0%/person-year during

follow-up. Correct presentation of the data could lead to markedly different interpretations. Figure 1A suggests that the complication being analyzed might be a complication of the treatment rather than of the disease or that the treatment does not prevent this complication. Conversely, Figure 1B suggests that the complication may be a complication of the disease and the treatment may prevent further complications. As such, the correct analysis may lead to different interpretations of the data.

This issue is not simply theoretical. Early in the acquired immune deficiency syndrome epidemic, it was reported that 50% of patients not treated with systemic anticyto-megalovirus therapy developed second eye disease by 6 months of follow-up,[2] a result subsequently confirmed.[3] Rates of second eye disease therefore could be calculated as 1.0/person-year (or as 100%/person-year). However, case series of patients treated with repetitive intravitreous injections without concomitant systemic therapy typically reported that only 20% of patients in the series developed second eye disease.[4–6] Because their proportions were “lower,” the conclusion of many of those authors reporting a relatively low proportion of second eye disease with repetitive intravitreous injections was that the previously reported rates might be too high. However, the patients in these case series treated with repetitive intravitreous injections had variable follow-up times, which typically averaged about 2 months. Had these case series reported rates, the authors would have discovered that their rate was similar to the previously reported rate and that the different proportion reported was due to a shorter follow-up. Hence, the proportion of a series of patients with variable follow-up should not be reported. Instead, the proportion at presentation and the rate during follow-up should be reported. Although proportions on follow-up can be used in randomized clinical trials, *when there is complete and comparable follow-up*, rates remain superior because they can be compared with the results of other trials with different follow-up periods.

There are issues with rates. If an event may occur more than once in a given patient, such as a reversible side effect, an event rate (a rate in which all events are counted) may overestimate the actual number of patients affected. If 20 patients in a series of 100 patients with a mean follow-up of 1 year developed 60 events (e.g., leukopenia while on systemic ganciclovir), then the event rate would be 0.6/person-year, whereas the rate of patients being affected would be only 0.2/person-year. Ways to approach this issue include using life table analyses (Kaplan-Meier curves) calculating the proportion without an event during follow-up or life table analyses of those developing the first event or a two-state model that takes account of events and recoveries.[1,7] If the authors simultaneously report both the event rate and the rate at which patients are affected, they convey to the reader more information on the history of the disease and its treatment.

One seemingly attractive way to evaluate the introduction of a new treatment among patients in an uncontrolled case series is to look at event rates before and after the introduction of the treatment.[8] As such, one might read statements such as, “With immunosuppressive treatment with some drug under study for Behçet’s disease, the mean rate of attacks of uveitis dropped from 2.3/yr to one/yr.” This statement appears to suggest that the treatment being evaluated is highly, although not perfectly, effective. However, this approach requires an assumption, which must be known to properly evaluate the data. The assumption is that the rate is

constant over time. If the natural history of a disease was that the attack rate was decreasing over time, then a case series with an ineffective treatment would report a reduction in the rate of disease progression or the rate of attacks after the study drug's introduction, simply because of the natural history of the disease. Conversely, if the natural history of the disease was that the rate of progression or of attacks increased over time, then the introduction of the useful treatment might result in the rate of attacks being unchanged from that before the intervention but a decrease in the number of attacks compared with what was expected. However, the conclusion of an uncontrolled case series without adequate natural history data would be that this drug was not effective as the attack rate did not change. There are several examples of diseases in which the rate of progression or attacks changes over time. Larger cytomegalovirus retinitis lesions are less likely to progress than are smaller ones.[9,10] The rate of attacks of HLA-B27 associated acute anterior uveitis appears to be lower in older patients than in young adults.[11] Conversely, among patients with ocular mucous membrane pemphigoid, more advanced lesions appear more likely to progress.[12] As such, case series using this approach should be able to cite natural history studies supporting the constancy of the event rate over time or, at a minimum, acknowledge the limitations of the study.

A more subtle problem with analyzing attack rates in individuals before and after a new treatment is introduced without appropriate controls is regression to the mean. The number of attacks during a specified time period usually is variable and can be described as a probability distribution centered around the mean number of attacks expected during that time period. Studies evaluating the attack rate before and after a new intervention often require frequent attacks before study entry (and investigators are more likely to enter those with frequent attacks) to see an effect with a smaller population studied, a number that often is greater than the mean attack rate during the given time period. During the follow-up period, the most likely attack rate is the mean, which would be less than the entry attack rate and, hence, suggest efficacy when there might be none. If the effect of treatment is large, then the issue is less problematic, but if the effect is modest, then the reader must be concerned.

FINAL VISUAL ACUITY:

A variation of the problem of the proportion of a series with variable follow-up is the concept of "final visual acuity," an analysis used in more than 60% of papers reporting visual acuity outcomes.[13] The problem with final visual acuity is that visual acuity outcomes, as a whole, are influenced by the duration of follow-up. Final visual acuity uses an arbitrary and variable follow-up for each patient based on data set closure rather than taking follow-up time into account. As such, final visual acuity is typically misleading. The nature of the problem can be seen in Table 1 and Figure 2. Table 1 shows the visual acuity in the better eye of 280 patients with CMV retinitis reported in 1995.[14] The final visual acuity analysis suggests that 75% of patients with CMV retinitis maintain good visual acuity. Conversely, the Kaplan-Meier curves of visual acuity loss (Figure 2), which analyze the probability of visual loss over time, show that visual acuity loss is substantial. The "final" visual acuity analysis underestimates the rate of loss because of the relatively short follow-up.[13] Indeed, the evident conclusion of the proper analysis is that with sufficient follow-up most patients will lose good visual acuity.

A simple thought experiment will demonstrate the fallacy of final visual acuity. Suppose that in a given disease patients lose vision to worse than 20/40 at the rate of 100%/person-year. Also suppose that treatment A is use-less for preventing visual loss and kills 100% of patients after 1 month of follow-up. Treatment B decreases the rate of visual acuity loss by 50% (i.e., to 50%/person-year) and has no effect on survival. Patients treated with treatment B are followed for 6 months. Each of the two treatment groups consisted of 100 patients. A final visual acuity analysis would lead to the following conclusion: “At the end of follow-up 88% of patients treated with treatment A had a final visual acuity of 20/40 or better vs 75% of patients treated with treatment B ($P = .02$).” Hence, the conclusion based on final visual acuity would be that treatment A is better than treatment B, which is both obviously and absurdly wrong. Although this example is extreme, a more realistic problem is demonstrated in Figure 3.

Suppose that the standard treatment results in visual acuity loss to the level of 20/200 or worse at a rate of 25%/person-year, as shown in Figure 3A, and that patients typically are followed for 1 year. The “final visual acuity” in this group typically would be reported as being worse than 20/200 in 25%. Suppose now that a new treatment is evaluated in an uncontrolled case series in which patients had a mean follow-up of 6 months. The new treatment reports that only 15% had a final visual acuity 20/200 or worse, an outcome demonstrated in Figure 3B. The conclusion based on final visual acuity might be that the new treatment appears better than the old treatment as evidenced by the reduction in the proportion of patients with final visual acuity worse than 20/200. The problem is shown by Figure 3C, which demonstrates that the new treatment is actually worse and that the standard follow-up for the new treatment accounts for the misleading “final visual acuity” analysis. An analysis of the rates of visual acuity loss would have demonstrated that the rate of acuity loss with the new treatment is 30%/person-year and that the new treatment is not as good as the standard treatment. The reporting of rates would have avoided the incorrect conclusion made from analysis of final visual acuity. Indeed, if final visual acuity rather than rates of loss are reported, new treatments may look better than older standard treatments because of the shorter follow-up, even when the new treatments are inferior.

Although these discussions have focused on visual loss, the issues are similar for situations in which there is visual improvement. If rates of visual loss or gain are not analyzed, then the distribution of acuities at specified time points after presentation or study entry could be evaluated (“interval acuities”), but complete follow-up is needed. If there is substantial loss to follow-up, or variability of follow-up, then “interval acuities” are biased in unknown ways (consider the different effects of the two extremes of this problem: all of the patients with blindness no longer returned for follow-up, which overestimates the visual outcomes, vs all of the healthy patients with good acuities no longer returned for follow-up, which underestimates the visual outcomes).

Visual acuity results should be reported as rates, and if there is variable follow-up, life table (Kaplan-Meier) or two-state model analyses should be performed. The most appropriate ways to report visual acuity outcomes are rates falling below given thresholds, such as 20/50 (6/15) or worse or 20/200 (6/60) or worse, or as a doubling of the visual angle. It is the doubling of the visual angle that corresponds to a loss of “3 lines” on an Early

Treatment Diabetic Retinopathy Study or ETDRS chart, also known as a logarithmic visual acuity chart.[15] Reporting 1 or 2 lines of loss of visual acuity on a standard Snellen chart is problematic because of the inconstant relationship of "lines" with changes in the visual angle. For example, a decrease of 1 line of visual acuity from 20/40 to 20/50 represents a small decrement (25%) in the angle of resolution, whereas loss of 1 line from 20/100 to 20/200 represents a doubling of the visual angle. When logarithmic visual acuity charts are not available, reporting of visual acuity outcomes can be converted to LogMAR (logarithm of the minimal angle of resolution) and a doubling of the visual angle analyzed.

Even with these approaches, there remain issues that relate to the presence of bilateral disease. The problem is that the two eyes of a single patient often are linked (i.e., behave more similarly than chance alone), particularly when a systemic therapy is given and that reporting all eyes is statistically problematic because it violates the basic principle of statistical analysis that events are independent. The problem can be addressed by appropriate analyses that correct for linked events such as generalized estimating equations.[16] Simpler solutions are available, however.[17] One is to report both the results seen in all eyes affected by the disease ("involved eyes") and the results seen in the better eye (e.g., the eye with better acuity). This approach gives an estimate of what happens to each eye with disease. Because, there also are statistical problems with reporting "involved eyes," one also should report what happens to the "better" eye, which is a patient characteristic and also gives information as to how the patient will function over time.

SMALL STUDIES

A third problem present in clinical case series is the overinterpretation of small studies. This problem is not restricted to uncontrolled case series but also may affect underpowered clinical trials. One example is a small clinical trial that suggested a new treatment was ineffective for preventing relapse of uveitis because there was no significant difference in the proportion relapsing between the treatment group (30%) and the placebo group (50%, $P = .66$).[18] The problem with this study was that the sample size was 10 patients/group. If the sample size had been 100 patients/group and there was a similar event rate, then the conclusion would have been that the treatment was effective in preventing relapse with a P value of 0.0003. Indeed, with a 10 patient/group sample size, the only detectable "significant" difference would have occurred if there were no relapses in the new treatment group. Therefore, it is difficult to conclude equivalence from a small underpowered study. Indeed, one of the guidelines for reporting clinical trials promulgated by medical editors is that the sample size and a priori assumptions leading to that sample size be reported.[19] A more appropriate conclusion of the result from the study discussed here would have been something similar to the following: "Although the small sample size limits the ability to make definitive interpretations of these data, the relatively modest decrease in relapse rate observed suggests that this new treatment is unlikely to be highly effective for preventing relapse of uveitis."

A similar problem occurs in uncontrolled case series, where the results are compared with other small, uncontrolled case series. For example, one might say that the surgical outcomes of one group appear to be better than those of a another group because only 30% of

10 patients in the new study developed complications, whereas 70% of patients in the older study developed complications. Although this conclusion may appear reasonable, a statistical analysis of the data suggests that the two event rates may be indistinguishable. If the real event frequency is 50%, then the 95% confidence interval for a 10-patient series ranges from 19% to 81%, which would include both the 30% and 70% frequencies reported. Indeed, looking at the 95% confidence interval often will prevent the author from overinterpreting or misinterpreting the data. As shown in Table 2, as the size of a case series increases, the width of the 95% confidence intervals for an observed proportion of 50% progressively narrows.

CONCLUSION

Uncommon diseases may not be easily amenable to investigation in adequately powered randomized controlled clinical trials. The disease may not be judged to be important enough for the large amount of funding required for a large multicenter clinical trial, and the sample size may not be easily achievable. No one center or small group of collaborating centers may have enough patients to conduct a large epidemiologic study. A new treatment may be being used and pilot data needed. Thus, case series may provide the best information available. Reporting case series data in a standardized and statistically appropriate manner would allow for proper interpretation of the data available and for future “meta-analyses” combining these case series to produce better estimates of the long-term outcomes. The simple changes proposed include not reporting the proportion of a series of patients with variable follow-up but reporting both the proportion at presentation or study entry and follow-up event rates (incidence). Final visual acuity should not be reported. Annual rates or life table analyses of visual acuity events, such as falling below standard thresholds or doubling of the visual angle should be reported. Small case series should not be over interpreted. Finally, most of these problems can be solved by biostatistical assistance. The author’s approach to retrospective case series has been to involve biostatiticians at the time of designing the study and the data collection forms to make sure that biostatistical techniques (even simplistic ones) can be applied properly to enhance the accuracy of these reports and that findings are reported appropriately.

Acknowledgments

Supported in part by grant EY 00405 from the National Eye Institute, National Institutes of Health, Bethesda, MD.

REFERENCES

1. Armitage P, Berry G. Statistical methods in medical research, 2nd ed., Oxford: Blackwell Scientific.
2. Jabs DA, Enger C, Bartlett JG. Cytomegalovirus retinitis and acquired immunodeficiency syndrome. Arch Ophthalmol 1989;107:75–80. [PubMed: 2535932]
3. Martin DF, Parks DJ, Mellow SD, et al. Treatment of cytomegalovirus retinitis with an intraocular sustained-release ganciclovir implant. A randomized controlled clinical trial. Arch Ophthalmol 1994;112:1531–1539. [PubMed: 7993207]
4. Cantrill HL, Henry K, Melroe NH. Treatment of cytomegalovirus retinitis with intravitreal ganciclovir. Long-term results. Ophthalmology 1989;96:367–374. [PubMed: 2540470]
5. Heinemann MH. Long-term intravitreal ganciclovir therapy for cytomegalovirus retinopathy. Arch Ophthalmol 1989; 107:1767–1772. [PubMed: 2556990]

6. Cochereau-Massin I, Lehoang P, Lautier-Frau M, et al. Efficacy and tolerance of intravitreal ganciclovir in cytomegalovirus retinitis in acquired immune deficiency syndrome. Ophthalmology 1991;98:1348–1353. [PubMed: 1658703]
7. Hillis A, Maguire M, Hawkins BS, Newhouse MM. The Markov process as a general method for nonparametric analysis of right-censored medical data. J Chron Dis 1986; 39:595–604. [PubMed: 3525597]
8. Mudun BA, Ergen A, Ipcioglu SU, et al. Short-term chlorambucil for refractory uveitis in Behçet's disease. Ocul Immunol Inflamm 2001;9:219–229. [PubMed: 11935432]
9. Holbrook JT, Davis MD, Hubbard ??, et al. , for the Studies of Ocular Complications of AIDS Research Group. Risk factors for advancement of cytomegalovirus retinitis in patients with acquired immunodeficiency syndrome. Arch Ophthalmol 2000;118:1196–1204. [PubMed: 10980764]
10. Jabs DA, Van Natta ML, Thorne JE, et al. , for the Studies of Ocular Complications of AIDS Research Group. Course of cytomegalovirus retinitis in the era of highly active antiretroviral therapy: 1. Retinitis progression. Ophthalmology. 2004;111: 2224–2231. [PubMed: 15582078]
11. Monnet D, Breban M, Hurdy C, Dougados M, Brezin AP. Ophthalmic findings and frequency of extraocular manifestations in patients with HLA-B27 uveitis: a study of 175 cases. Ophthalmology 2004;11:802–809.
12. Mondino BJ, Brown SI. Ocular cicatricial pemphigoid. Ophthalmology 1981;88:95–100. [PubMed: 7015218]
13. DiLoreto DA Jr, Bressler NM, Bressler SB, Schachat AP. Use of best and final visual acuity outcomes in ophthalmological research. Arch Ophthalmol 2003;121:1586–1590. [PubMed: 14609916]
14. Jabs DA. Ocular manifestations of HIV infection. Trans Am Ophthalmol Soc 1995;93:623–683. [PubMed: 8719695]
15. Ferris FL III, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. Am J Ophthalmol 1982; 94:91–96. [PubMed: 7091289]
16. Liang KV, Zuger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
17. Studies of Ocular Complications of AIDS Research Group in collaboration with the AIDS Clinical Trials Group. Foscar-net-Ganciclovir Cytomegalovirus Retinitis Trial: 4. Visual outcomes. Ophthalmology 1994;101:1250–1261. [PubMed: 8035989]
18. Foster CS, Tufail F, Waheed NK, et al. Efficacy of etanercept in preventing relapse of uveitis controlled by methotrexate. Arch Ophthalmol 2003;121:437–440. [PubMed: 12695239]
19. Moher D, Schultz KF, Altman D, for the CONSORT Group. The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 1987;285:1987–1991.

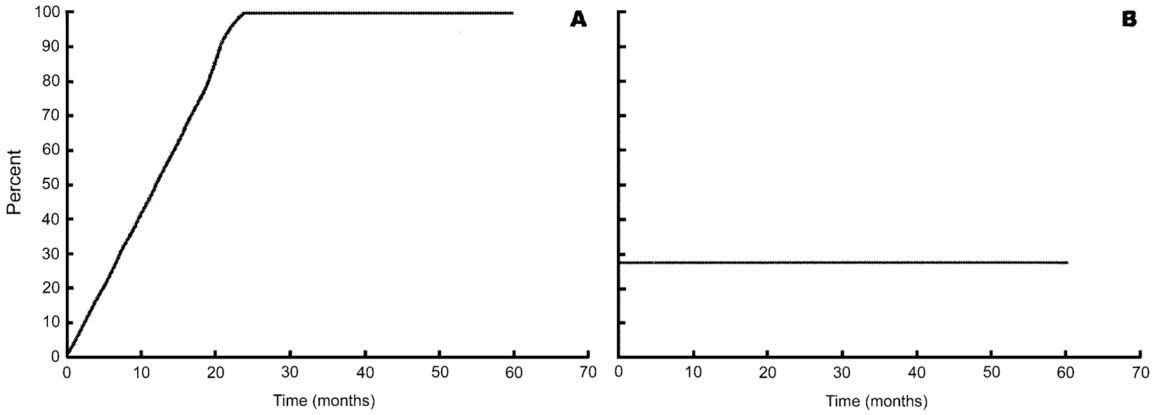

FIGURE 1.

Hypothetical event curves consistent with the statement “In 28% of patients followed for a mean of 9 months, a specific complication occurred.”

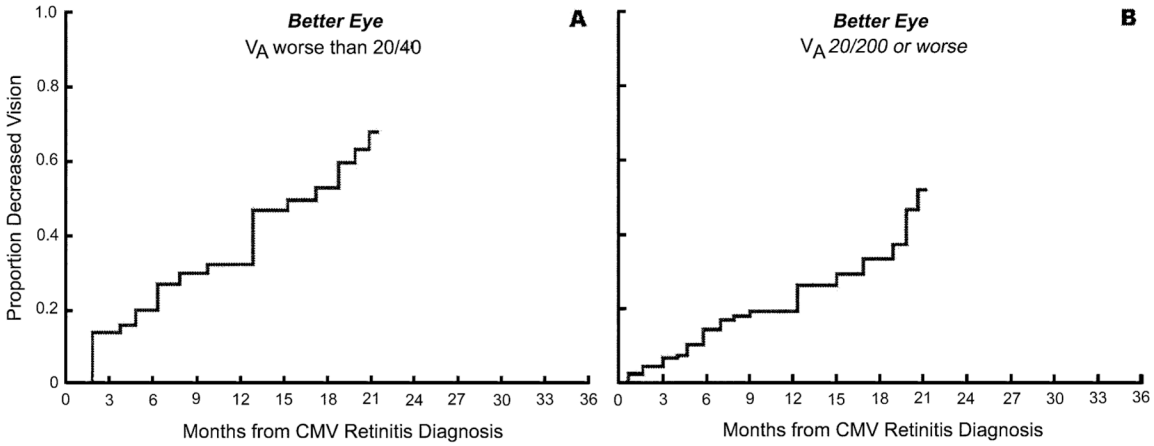

FIGURE 2.

Kaplan-Meier curves for the visual acuity in the better eye of patients with cytomegalovirus retinitis. A. Visual acuity worse than 20/40, and B. Visual acuity 20/200 or worse. From *Trans Am Ophthalmol Soc*1995;93:623. Used with permission.

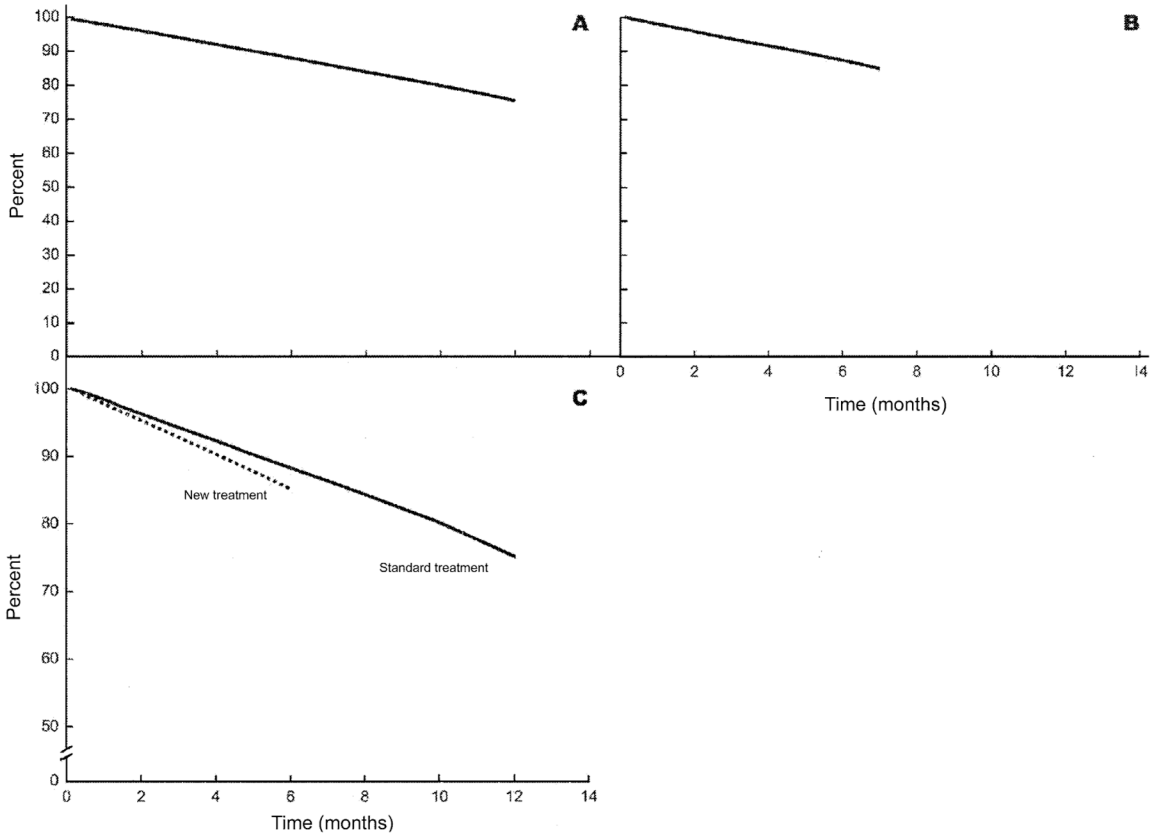

FIGURE 3.

Hypothetical events curves for loss of acuity with: A. standard therapy; B. investigational therapy; and C. comparing the two treatments.

TABLE 1.

Visual Acuity in the Better Eye of 280 Patients with CMV Retinitis

Visual Acuity Level	Initial (%)	Final (%)
≥20/40	90	75
20/50 to 20/160	8	11
≤20/200	2	14

From *Trans Am Ophthal Soc* 1995;93:623. Used with permission.

CMV = cytomegalovirus.

TABLE 2.

95% Confidence Intervals (CI) for a 50% Estimate

N	95% CI
10	19–81
100	40–60
1,000	47–53