

Scientific Writing: Discriminating Good from Bad

PART I: GENERAL CONSIDERATIONS

Arthur Croft, DC, MS, MPH, FACO

Despite the increasing use of statistics in medicine and chiropractic, a significant number of published reports contain serious errors. In an analysis of the papers published during 1982 in Arthritis and Rheumatism, it was found that 66 percent suffered from statistical errors.¹ A similar report has pointed to a surprising number of errors in clinical trials reported in such august journals as the Annals of Internal Medicine, British Medical Journal, Lancet and the New England Journal of Medicine during 1987.² And since specialty journals and association journals are generally subjected to even less scrutiny, it is not surprising that statistical and other errors are far more common in them.

In this five part series, we will explore some of the common errors found in clinical and scientific writing and we will develop a method of reading the literature with a critical eye. Before we look at the framework of this analysis let's start with an example in which you can search for flaws.

Fictitious Study

During their first year of chiropractic college 400 women were given the opportunity to participate in a yearly health check-up. Half of the students agreed to participate in the health check-up (becoming the study group) while the other 200 nonparticipants were assigned to the control group. The study group were given yearly breast exams, Pap smears, and blood tests, in addition to their general chiropractic care, for each of the subsequent three years of chiropractic college.

Upon graduation all members of both the study group and the control group were given an extensive history, physical examination and laboratory tests (including a Pap). The results from both groups were compared to determine if the yearly exams were of any real benefit to the women in the study group.

The results were as follows: 1) the study group had about half the rate of overweight problems compared to the control group; 2) the study group had graduated with a mean GPA nearly one point higher than the control group; 3) there were no statistically significant differences between groups for breast cancer; 4) the study group women had two times as many diagnoses given to them during their training; and 5) there were no differences between the groups with regard to cervical cancer.

The conclusions drawn by the authors were the following:

1. A yearly screening can reduce the rate of weight problems in female chiropractic students by half.
2. Since the mean GPA was higher in the study group, the yearly check-ups probably contributed to both physical and mental health.

3. Since the rate of developing breast cancer was the same for both groups, this screening can probably be omitted from yearly exams on women.
4. Since nearly twice the number of diagnoses were made for the women in the study group, it was assumed that diseases were being discovered at an earlier stage where therapy was more likely to be effective.
5. Since there was no difference in the occurrence of cervical cancer in the two groups, the researchers postulated that routine screening for this disease was unnecessary.
6. Regular chiropractic care can reduce the rate of both cervical and breast cancer.

Now test your powers of discernment and perspicuity by answering these questions:

Question 1) Was this study designed to allow such conclusions to be drawn?

Answer: No. The authors did not state a hypothesis which they intended to test. The goal of this project was quite vague. The study population may not be reflective of a general cross section of American woman since chiropractic students might tend to be younger, healthier and representative of a narrower socioeconomic group. Education level may be another factor the authors haven't considered. Since breast and cervical cancer are less common in younger women, the small sample size and narrow age stratum will limit any conclusions that can be drawn. Finally, the authors have not attempted to account for specific risk factors of these cancers.

Question 2) Was a proper method used to assign these women to the study and control groups?

Answer: No. Special care must be taken in the assignment to study or control groups. Ideally a randomization process should be used. In this case, students volunteered themselves into the study group, and therefore we must suspect that their health habits and risk factors may differ from the control (nonparticipant) group. And, since no baseline examination was performed on the control group, we have no way of knowing if the two groups were actually comparable.

Question 3) Did the authors appropriately assess the results in both groups?

Answer: Not entirely. For one thing the final figures were reported only for those women who graduated. Some (although it is unlikely) may have died of breast or cervical cancer but would not have been included in the final tally. Loss of follow up is a very common problem in clinical trials and to ignore the problems is to invite severe criticism of a study. As to the increased number of diagnoses given to the study group students, this may very well be an artifact of a more intense monitoring along with the active imaginations of clinic interns. It is quite possible that the control group, under similar scrutiny, would have been equally "afflicted."

Question 4) Did they make an adequate comparison between groups regarding outcome?

Answer: No again. Since baseline measures of overweight, obesity and other indices were not recorded for the control group, comparison is neither valid nor reasonable. Therefore it is quite possible that the weight profiles remained static throughout the study period and, in that case, the authors' conclusions are erroneous. It's also possible that the mean weight of the control group lowered during the study period while that of the study group increased.

Question 5) Based on these comparisons, were the interpretations drawn by the authors valid?

Answer: Let's take each of their conclusions:

1) Yearly screening reduces overweight problems.

While possible, the authors have not proven this.

2) Yearly check-ups contribute to improved mental health (hence the higher mean GPA in the study group).

The authors did not make any effort to assess the mental health of these students. Their conclusion is based on the assumption that the GPA is reflective of mental health when, in fact, there is little evidence to support this. Also, it would be important to compare IQ levels and previous GPA levels in order to ensure that any measured effect was actually due to this study.

3) Yearly breast exams can probably be safely omitted.

The authors did not consider women of all age strata, nor did they account for known risk factors carefully, and therefore, this extrapolation is not valid. The small study size, coupled with the fact that this suggestion seems counter to conventional thought should be a red flag.

4) Diseases were discovered at an earlier stage in the study group.

The fact that more diagnoses were made does not necessarily imply that they were correct (i.e., confirmed). Since no data was given regarding the types of conditions found, their stage, and the resulting treatment rendered, we cannot interpret the results this way. Besides, many of the conditions may have been self-limiting and may not have significantly affected the overall health picture.

5) Cervical cancer screening can be omitted.

The same comments as mentioned for #3 apply here. Authorities may not agree on the timing of these tests but few would consider eliminating them on all regular examinations, especially in women at higher risk.

6) Regular chiropractic care can reduce both breast and cervical cancer.

While there may be some truth in this, the authors have not provided any compelling evidence here. It is not unlikely that none of these 400 young and healthy women would have developed cancer in this four year period without chiropractic care. Also, there is no control group to evaluate this possible effect. Both groups received regular chiropractic care. The authors also did not quantify the frequency or type of care given, which again, limits the conclusions that can be drawn from this study.

References

1. Felson DT, Cupples LA, Meenan RF: Misuse of statistical methods in Arthritis and Rheumatism. 1982 vs. 1967-68. *Arthritis and Rheumatism* 27: 1018-1022, 1984.
2. Altman DG, Dore CJ: Randomization and baseline comparisons in clinical trials. *Lancet* 335: 149-153, 1990.

SEPTEMBER 1993