USAGE OF MEDICAL JOURNAL ARTICLES IN BIOSTATISTICAL TRAINING FOR RESIDENTS ®

Reena Deutsch
University of California, San Diego
USA

Each year, a new crop of physicians enters residency training programs in medical teaching institutions worldwide. Second and subsequent year residents continue with the programs in which they have participated in a prior year. The educational curriculum may include a biostatistical component, where the instructor is presented with an opportunity to focus on biostatistical issues bearing on various aspects of medical practice and research. This paper describes such a presentation in a university medical school residency training program. The training session centered on research findings published recently in the medical literature. Issues regarding topic, journal, and article selection, teaching aids, approaches to illustrating aspects of study design, power analysis, statistical methodology, and interpretation of results, promoting contact with a biostatistical consultant, and feedback from lecture attendees are described.

INTRODUCTION

The technique of using medical journal articles as an aid for medical instruction is not new. Linzer (1987) described the historical background of journal clubs, defined as groups of individuals who meet to critically discuss articles in the current medical literature. The first reference to such a group goes back to 1875. Alguire (1998) summarized published reports of goals, organization, and teaching methods of journal clubs in five medical specialties over the previous decade. Woods and Winkel (1982) discussed various formats for journal clubs which strive to provide a forum for the education of residents in the techniques of critical reading.

Using medical literature to improve physician knowledge in clinical study design and biostatistics is also well documented. Letterie and Morgenstern (2000) successfully developed a program to link experimental design, statistical concepts, and clinical studies. Markert (1989) concluded that a residents' journal club for research methods and statistics can be instructive and successful. Others have described mixed results in effectiveness demonstrated from the approach to linking study design, statistical methodologies, and clinical research (See Langkamp, Pascoe, and Nelson, 1992; Diserens et al., 1994). In the only randomized, controlled study cited by Alguire (1998), Linzer, Brown, and Frazier et al. (1988) reported that residents who were journal club members, but not a control group, significantly improved their reading habits and knowledge scores on epidemiology and biostatistics tests given pre- and post-intervention.

Residency training programs provide postgraduate physicians with a broad range of opportunities. These programs expose this population to a variety of situations that may serve them well in clinical practice as well as in medical research careers. To judge the validity of published research findings and applicability to clinical practice, physicians-in-training must be familiar with biostatistical terminology and concepts. Further, it is vital that clinicians have the skills required to read, understand, and critically evaluate the medical literature. For those residents who will conduct their own independent clinical research, a background in study design and biostatistical techniques will be beneficial. Without this knowledge, physicians are at risk of basing clinical decisions on their evaluation of research findings without adequate tools to guide them properly.

Journal clubs traditionally consist of regular meetings held over a sustained period of time and are usually led or facilitated by either a faculty physician in the specialty area of the residents, or residents are assigned or volunteer to present an article (see Alguire, 1998; Letterie & Morgenstern 2000; Markert, 1989). Incorporating the approach of a journal club into a single lecture to postgraduate physicians on a biostatistical topic presented by a biostatistician presents considerable challenges. The available timeframe is substantially abbreviated, and the instructor lacks clinical expertise in the medical area of the audience.

Several invited single biostatistical lectures to participants of various residency training programs at the University of California, San Diego (UCSD), School of Medicine have been

previously presented. These seminars, designed to apply successful features of journal club instruction described in the literature, allowed experimentation with different formats and degrees of incorporating publications into the presentations as a tool for biostatistical instruction. The usage of medical journal articles from a wide range of applications was a noted strength of a prior seminar series in biostatistics applied to clinical research taught to medical faculty, postdoctoral fellows, and clinical research staff (see Deutsch, in press). The articles were offered as case studies and ran throughout entire seminar sessions. They promoted active participation by the attendees and stimulated classroom interest and discussion. Inclusion of current published research findings in the curriculum received overwhelmingly positive reactions from lecture attendees and maintained the attention of the participants. They contributed direct application to the understanding and interpretation of publications in the participants' own field.

An hour-long lecture to introduce and discuss a biostatistical topic as part of the Fall 2001 seminar schedule for residents in Family Medicine at UCSD was developed. It was built upon pedagogic lessons learned from earlier experiences. A medical journal article was an integral part of the presentation. The single article review format described by Woods and Winkel (1982) was utilized throughout the session. Preparation, presentation, challenges, and relevant issues regarding this seminar are described.

GOALS

The one-hour formal exposure to a biostatistical topic reported here included three goals: (1) to convey that study design and analysis techniques in clinical research studies can greatly influence the results; (2) to provide adequate background and examples so that residents will develop skills in evaluating published results in their field of practice within the biostatistical context and further develop their existing interpretative skills; and (3) to illustrate the process of forming a statistical analysis plan when conducting medical research.

It is difficult to meet these goals in a one hour session. Lessons learned from previous biostatistical training sessions for residents in other medical specialties at UCSD revealed that there are few prescribed restrictions but many challenges to overcome. Efforts were directed towards one specific topic encompassing enough depth to achieve the specified goals to the maximum extent possible. A recent article published in the medical literature that reports research findings illustrating aspects of the chosen topic was intended to focus discussion on linking the research objectives, study design, data analytical techniques, results, and conclusions.

TOPIC SELECTION

After conferring with the faculty organizer of the residency training session, diagnostic testing was selected as the lecture topic. Residents in Family Medicine are exposed to published reports about diagnostic and screening tests used in clinical practice, yet they typically have no formal training regarding assessment of test accuracy. The program organizer requested that the basic terminology and definitions related to diagnostic testing be addressed in as simple, yet meaningful, manner as possible. Tips and guidance on the interpretation of diagnostic test performance assessment were planned for the presentation.

JOURNAL AND ARTICLE SELECTION

As Letterie and Morgenstern (2000) report, the greatest obstacle to implementing teaching through the use of clinical literature is that of finding clinical studies that best and clearly demonstrate the targeted concepts. Additionally, the selected publication needs to be balanced with the clinical interest of the residents.

A Medline search was performed using keywords *diagnostic test* and *screening test.* A review of potentially relevant online journals was conducted, and current publications offered by the UCSD School of Medicine library were browsed. A clear statement and description of study hypotheses, design, and results were among the criteria used to identify an appropriate paper. Further, usage of common measures related to assessing accuracy of diagnostic tests and integration of basic conventional applications of diagnostic test assessments within a clinical research study were sought. A recent publication date of 2000 or 2001 was preferable, but a slightly older publication date would have been acceptable.

The publication selected as the example for the lecture, found among the current journals displayed at the university library, was in the Clinical Prediction Guide section of the *American College of Physicians Journal Club* (*ACPJC*) (2001). It summarized the clinical study and results reported in an article from *The New England Journal of Medicine* (Haydel et al., 2000). The clarity of the report and general application to diagnostic and screening tests provided a satisfactory example, and it met all pre-requisites for selection.

The process of selecting an appropriate publication proved to be tedious, time consuming, and frustrating. Many articles reported *sensitivity* and *specificity* but no other useful measures of diagnostic testing. Other articles provided insufficient detail to allow critical evaluation of the reported study design and methods. Some articles contained complicated aspects that might prove to be problematic within the setting of an introductory overview of the topic. Obtaining a journal article that adequately illustrated the topic in a straightforward manner was critical to the success of the lecture and to making it meaningful and interesting to the residents. Excessively complex designs were expected to confound the planned presentation and were avoided.

LECTURE CONTENT

Terms commonly associated with the general topic of diagnostic test assessment were listed for the lecture attendees. These included:

- Incidence
- Prevalence
- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value
- Likelihood ratio of a positive test
- Likelihood ratio of a negative test
- Misclassification rate.

The first two terms, incidence and prevalence, were introduced, defined, and contrasted to provide the global setting for disease diagnosis.

Immediately thereafter, the selected journal article was presented. Briefly, the article reports on a study where a set of criteria was established for non-comatose patients presenting to an emergency room with minor head injury. The aim of the study was to distinguish, based on the specified criteria, between those who need to have computed tomography (CT) and those who do not. For this study, all subjects in a training and a validation sample had CTs, and those subjects with abnormalities found with CT scanning were considered to have "needed" the CT scan. The process of developing the diagnostic model described in the article was discussed during the presentation, and the importance of validating the ensuing model was underscored.

The lecture emphasized the need to identify well-defined diagnostic criteria and a specific outcome. Issues relating to study design were introduced in the article and explained within the framework of the reported study. Applications of study design and methods to research findings in general were discussed. The impact of sample size was pointed out.

Results published in the article were closely examined. Corresponding to the diagnostic test assessments contained in the article summary, the measures listed above were introduced, predominantly in pairs (sensitivity vs. specificity, positive predictive value vs. negative predictive value, likelihood ratio of a positive test vs. likelihood ratio of a negative test). The terms were defined and contrasted, and calculations were presented. Normally, avoidance of mathematical computations is advisable in training sessions for residents (see Simpson, 1995), but the calculations, based only on the entries of a 2 x 2 table, were extremely simple and highly illustrative of the concepts behind the measures, thus they were casually presented. Moreover, since 95% confidence intervals were presented in the publication, this analysis method was also introduced regarding its application and general meaning. Lastly, findings in the paper were systematically critiqued.

TEACHING AIDS

A slide presentation, available electronically from the author, was created with Microsoft Powerpoint 2000 software (1987-99). Colorful graphics, tables, and lists were presented in a simple, yet visually stimulating, way. Avoidance of bland black-and-white presentation media, such as with overhead transparencies, was expected to be more likely to acquire and maintain the attention of the residents. Figure 1 is a copy of a slide (black and white version for publication) which shows an example of how the study data referred to during the lecture were used to illustrate measures common in diagnostic test assessment. Figure 2 reflects a slide that lists and compares pertinent information about positive and negative predictive values. Calculations for the data and diagnostic test measures reflected in Figures 1 and 2, respectively, are displayed as Figure 3. All other diagnostic test measures were presented in a similar fashion.

	normal CT	abnormal CT
0 findings	121	0
≥1 finding	363	36
	484	36

	No disease	Disease
Predict no disease	true negative (TN)	false negative (FN)
Predict disease	false positive (FP)	true positive (TP)
	(ND)	(D)

Figure 1. Slide Showing Sample Data and Terminology/Notation for Diagnostic Testing

positive vs. negative predictive values	
positive predictive value	**negative predictive value**
▪ "PPV"	▪ "NPV"
▪ probability [0-1] or % [0-100]	▪ probability [0-1] or % [0-100]
▪ the probability of the presence of disease when disease is predicted.	▪ the probability of the absence of disease when absence of disease is predicted.
▪ Prob(diseased if disease predicted)	▪ Prob(not diseased if disease absence predicted)
▪ large values are good	▪ large values are good

Figure 2. Slide Showing Characteristics and Pertinent Information about Predictive Values

	ND	D			
Predict ND	TN	FN	*derivation set*		
Predict D	FP	TP	normal CT	abnormal CT	
0 findings			121	0	121
≥1 finding			363	36	399
			484	36	520

Positive predictive value = #TP ÷ #pred(+) = 9%
    36/399 = 0.9 or 9%

Negative predictive value = #TN ÷ #pred(−) = 100%
    121/121 = 1.0 or 100%

Figure 3. Slide Showing Calculations for Positive and Negative Predictive Values

A copy of the *ACPJC* summary article (2001) was distributed as a handout to all residents in attendance. The article handout had selected text annotated with color-coded highlight marker to attract the attention of participants to specific parts of the report during the discussion. A hard

copy version of the electronic slide presentation was also distributed to each participant to reduce the need for note-taking, to assist in following the lecture, and to be available as a reference after the seminar. Included in the handout of slides was a listing of web sites containing electronic tutorials and automated diagnostic test measure calculators, as well as citations for related journal articles for future reference.

RESULTS AND EVALUATION

Approximately 13 residents attended the lecture. During the presentation, questions were solicited, and some attendees participated enthusiastically. Among questions asked by the audience was one requesting a simple way to conceptualize likelihood ratios. Alternative versions of the definition were furnished, and participants offered other suggestions. One attendee inquired about the process of determining the sample size needed to assess diagnostic performance adequately. This led to a discussion of the important factors involved in power and the appropriateness of consulting someone with expertise in study design and biostatistics.

A few attendees gave unsolicited verbal feedback after the lecture and offered positive appraisal for the journal article approach and overall presentation. Written evaluations were submitted by all attendees and revealed an overall multiple-choice rating of "good" (choices were excellent, good, fair, poor, no opinion). Five of the 13 respondents gave no additional comments. Written feedback that was received included statements that the presentation was clear, well organized, good, and helpful. Two evaluations suggested even more clinical examples. Among the responses aimed with a more narrow focus, three responders wrote that the topic was too dry, and three commented on the sleep-deprived state of the residents.

DISCUSSION

Despite the restrictive one-hour limit, the three aims stated in the Goals section above seem, at least in part, to have been achieved. First, the usage of the selected published research report was ideal to emphasize the impact of study design and analysis methodologies on reported findings and subsequent conclusions. The publication provided an excellent example to follow the traditional flow of clinical research, clearly illustrating how each step of the process is dependent upon the previous assumptions and methods. Secondly, awareness of technical aspects of assessing diagnostic test performance will likely assist residents in future assessments of measures that they are exposed to in clinical practice or research. Finally, the last goal was, to some extent, achieved through discussion of issues that arise when developing a research plan. Based on feedback, though, inclusion of additional clinical examples should be considered in future lectures.

Discussion of actual research cases has a tendency to offer suitable occasions to point out the benefits of consulting with a biostatistician. These can occur when describing the research design phase or when interpreting complicated or unfamiliar analysis results. This tendency was demonstrated in the recent experience, where ample opportunities arose during the presentation of the research article to encourage residents to consult with a biostatistician for specific circumstances. Similar opportunities to encourage consultation for study design and biostatistical issues arose in previous lectures where published articles from the medical literature were part of the curriculum. The need for this resource must be reinforced because of the tendency of some physicians to under-appreciate the technical aspects of study design, data collection, and statistical analysis and overestimate their expertise in these areas.

It is unknown whether efforts to promote critical reading and interpretation of findings published in medical journals, or to provide a deeper and broader understanding of statistical methods and study design, have been successful. The perception of some attendees that the presented topic was "dry" may negatively impact on understanding and retention. Nevertheless, usage of a relevant research publication from the medical literature is supported by reported findings of their positive influence and has consistently met with encouraging feedback from postgraduate physicians. The underlying benefits of the general approach of using the journal article as the foundation for lectures to residents remain promising.

REFERENCES

Alguire P.C. (1998). A review of journal clubs in postgraduate medical education. *Journal of General Internal Medicine*, *13,* 347-353.

Deutsch R. (in press) A seminar series in applied biostatistics for clinical research fellows, faculty and staff. *Statistics in Medicine*.

Diserens D., Giger U., Bull R., Cotter S., Meyers K. Smith J., & Wardrop J. (1994). The lecture vs. the journal article. *Academic Medicine*, *69*, 976-977.

Haydel M.J., Preston C.A., Mills T.J., Luber S., Blaudeau E., & DeBlieux P.M.C. (2000) Indications for computed tomography in patients with minor head injury. *The New England Journal of Medicine, 343,* 100-105.

Langkamp D.L., Pascoe J.M., & Nelson D.B. (1992). The effect of a medical journal club on residents' knowledge of clinical epidemiology and biostatistics. *Family Medicine, 24,* 528-530.

Letterie G.S., & Morgenstern L.S. (2000). The Journal Club, Teaching critical evaluation of clinical literature in an evidence-based environment. *Journal of Reproductive Medicine*, *45,* 299-304.

Linzer M. (1987). The journal club and medical education: over one hundred years of unrecorded history. *Postgraduate Medical Journal, 63*, 475-478.

Linzer M., Brown J.T., Frazier L.M., et al. (1988). Impact of a medical journal club on house-staff reading habits, knowledge, and critical appraisal skills. JAMA, *260,* 2537-2541.

Markert, R.J. (1989). A research methods and statistics journal club for residents. *Academic Medicine*, *64*, 223-224.

Microsoft®PowerPoint®2000 (1987-1999). Microsoft Corporation.

Simpson J.M. (1995) Teaching statistics to non-specialists. *Statistics in Medicine, 14*, 199-208.

Woods J.R., Jr., & Winkel C.E. (1982) Journal club format emphasizing techniques of critical reading. *Journal of Medical Education*, *57,* 799-801.

Clinical Prediction Guide (2001). A decision rule with 7 clinical criteria identified patients with minor head injury who did not need computed tomography. *ACP Journal Club, 134*, 74.