

The Need for Collaboration Between Clinicians and Statisticians: Some Experience and Examples

Y B Cheung,*^{PhD}, S B Tan,*^{PhD}, K S Khoo,**^{FAMS, M Med, MRCP}

Abstract

Introduction: Very often we see poor communication and collaboration between clinicians and statisticians. Both sides may fail to realise the importance of a truly collaborative effort. In this paper we give examples to illustrate some problems clinicians and statisticians may encounter when they do not have the full support of each other. The aim is to convince both parties the importance of a truly collaborative effort. **Methods:** Real examples in various medical research areas are drawn from the authors' practical experience for illustration. The examples cover various research aspects such as the use of computer software, regression analysis and interpretation of findings. **Results:** Superficial collaboration between clinicians and statisticians may lead to serious problems and sub-optimal research practice that may not be obvious in the first sight. Some of the barriers to effective communication and collaboration are discussed. **Conclusions:** On the one hand, robust statistical practice is vital in many medical research projects. On the other hand, medical thinking is important in the formulation and application of statistical strategies. Statistical inputs should be integrated into medical research projects throughout the whole research process. Sporadic contacts between clinicians and statisticians are not enough. Both parties must learn to communicate more effectively and to be willing to collaborate with each other.

Ann Acad Med Singapore 2001; 30:552-5

Key words: Communication, Consultation, Medical statistics, Multidisciplinary, Research methods

Introduction

Many areas of medical research require the application of statistical techniques. Although most clinicians are taught some statistics as part of their basic medical training, the important role of statistics in medical research has led to many pharmaceutical companies, hospitals and medical research institutions employing full time statisticians to contribute to their research projects. Furthermore, the International Conference on Harmonisation (ICH) guideline E9 (see <http://www.ich.org/ich5e.html>),¹ which is the internationally accepted standard of statistical principles for clinical trials, assumes that "the actual responsibility for all statistical work associated with clinical trials will lie with an appropriately qualified and experienced statistician".

Clinicians and statisticians do not always agree on their roles within the context of a medical research project. In this paper, we share some of our thoughts and experience on this matter. Some clinicians consult a statistician at the beginning of a project to get advice on issues such as the recommended sample size for a study. Once this is done, they may proceed with the project without any further

consultation with the statistician until the data are ready for analysis. They then pass the data to the statistician and expect the statistician to carry out the necessary analysis and prepare a quantitative report on the data for the investigator. Both parties go about their tasks independently, with little communication between them.

Many statisticians on the other hand, particularly those who are active in statistical methodology research, often frown upon having to provide too much "statistical consultations" or forming "unnecessary" collaborations with clinicians. They would rather devote their time to working on statistical projects, in which they or other statisticians are the main investigators.

The above situations describe what we see in some medical research institutions. In this paper, we emphasise that both clinicians and statisticians will benefit greatly if they involve each other in their work. In the next section, we highlight the importance of good statistical techniques in medical research and give some examples on how statisticians can add value to the research work of clinicians. In the third section, we will give some examples to illustrate

* Biostatistician

** Director

Clinical Trials and Epidemiological Sciences
National Cancer Centre, Singapore

Address for Reprints: Dr Cheung Yin Bun, Clinical Trials and Epidemiological Sciences, National Cancer Centre, 11 Hospital Drive, Singapore 169610.
E-mail: ctecyb@nccs.com.sg

why medical statisticians need the input and help of clinicians in their research. We conclude with a plea to both parties to appreciate the important role played by each other and to put in more effort in communicating and genuinely collaborating with each other.

Clinicians Need Medical Statisticians

In his article on statistical reviewing for medical journals, Altman² stated that: “The main reason for the plethora of statistical errors is that the majority of statistical analyses are performed by people with an inadequate understanding of statistical methods. They are then peer reviewed by people who are generally no more knowledgeable.” He also gave a summary of the findings from selected reviews of the quality of statistics in medical journals, which highlighted the alarmingly poor quality of statistics included in articles published in medical journals. We now discuss some examples that highlight the value of a statistician’s input in clinician-led research projects.

Example: Use of Computer Software

The popularity of user-friendly statistical software has its advantages and disadvantages. It causes some clinicians to develop an over-simplified “do-it-yourself” mentality. One example of the danger of using popular software is as follows. There are two ways of using statistical software: interactive mode or batch mode. Popular software are usually designed for interactive mode, allowing the users to use a mouse to point and click, with the output “magically” appearing on the screen. In contrast, the batch mode requires the users to write a file (programme), consisting of command lines. The user then runs the file or part of it and produces the results. Using the interactive mode, it is very difficult to keep track of past analysis activities. Careless mistakes induced in one operation may become a permanent feature of the data. Imagine an inspector from a regulatory authority asking to see exactly what has been done to the data. The interactive mode users will have a difficult time, while the batch mode users can print and show the inspector the programme file and explain every command line to him. Errors can be identified. Using the batch mode is of course more difficult than using the interactive, point and click, mode.

Example: Regression Analysis of Categorical Variables

For the regression analysis of categorical outcome variables, a log-linear relation is often the default option in statistical software. In fact, some software packages can only implement a log-linear relation. Examples of software packages that estimate both linear and log-linear forms of relation are SAS³ and Stata.⁴ Suppose now we are studying the risk of skin cancer, and there are two binary risk factors (e.g., high level coded as 1; low level coded as 0), X1 and

X2, in the analysis. The log-linear form says that:

$$\log_e(\text{risk}) = \alpha + \beta_1 \times X_1 + \beta_2 \times X_2,$$

where α is the intercept and β_1 and β_2 are the regression coefficients to be estimated. Taking exponents to transform back to the original scale, the log-linear equation implies:

$$\text{risk} = \exp(\alpha) \times \exp(\beta_1 \times X_1) \times \exp(\beta_2 \times X_2),$$

$\exp(\beta_1)$ and $\exp(\beta_2)$ being the risk ratios associated with X1 and X2. The equation suggests that X1 and X2 interact with each other. This formulation may be unnoticed by many users of the software. If X1 and X2 are indicators of sun exposure and use of sunscreens, defining them as terms that interact may make behavioural/clinical sense as they are on the same behavioural pathway. On the other hand, if X1 and X2 are indicators of, say, sun exposure and smoking, the risks associated with them are probably on different pathways. The linear form:

$$\text{risk} = \alpha + \beta_1 \times X_1 + \beta_2 \times X_2$$

may then be preferred. This allows the impact of the two variables to work independently (It is possible to combine the above two formulations in a single equation.). The correct choice and implementation of the form of relationship requires not only familiarity with the functions available in software packages, but also with statistics itself. Sometimes new programmes have to be tailor-made in order to estimate a relationship that makes medical sense.

Example: Interaction

The choice of linear and log-linear form of relation also has great implication to the studies of interaction. For instance, Rothman⁵ discussed the problem in a study of thrombotic stroke in details. Table I imitates the risk profiles of sub-groups in that study. Taking normotensive women (H = 0; H = 1 if hypertensive) who do not use oral contraceptive (OC = 0; OC = 1 for users) as the reference group, the pattern of risk can be described either by a linear relation of:

$$\text{risk} = 0.05 + 0.10 \times \text{OC} + 0.30 \times \text{H} + 0.25 \times \text{OC} \times \text{H},$$

or by a log-linear relation of:

$$\log_e(\text{risk}) = -3.00 + 1.10 \times \text{OC} + 1.95 \times \text{H} - 0.40 \times \text{OC} \times \text{H}.$$

TABLE I: HYPOTHETICAL RISK PROFILES FOR THROMBOTIC STROKE ACCORDING TO ORAL CONTRACEPTIVE (OC) USE AND HYPERTENSION

	Use OC (OC = 1)	Do not use OC (OC = 0)
Hypertensive (H = 1)	0.70	0.35
Normotensive (H = 0)	0.15	0.05

In the linear model the coefficient for the interaction term is positive (0.25), suggesting a synergistic effect. However, in the log-linear model, the coefficient is negative (-0.40), suggesting an antagonistic effect. This self-contradiction clearly demonstrates that “statistical interaction” is model-dependent. Rothman emphasised that researchers should use their medical understanding to choose an appropriate model, on which an interaction term can be built. The prerequisites of utilising the medical understanding, however, are that the researchers must be aware of the statistical problem, statistical choices, and computing options.

Example: Interpretation of Findings

In medical research, a U-shape or J-shape relation sometimes arouses great interest (e.g., that between alcohol consumption and cardiovascular disease risk). However, such a U- or J-shape relation can be an artefact arising from many problems, for example, the use of a second-order polynomial i.e., $\beta_1 \times X + \beta_2 \times X^2$, in the analysis. This imposes a U- or J-shape upon the data. Perhaps the true relation is an initial decline in the outcome with respect to the independent variable, with the decline gradually levelling off. While clinicians are usually aware of the fact that using only a linear term imposes a linear relation on possibly non-linear data, they may be less aware of the imposition of a U- or J-shape relation. Statisticians would usually carry out model diagnostics and confirm a finding before reporting a potentially interesting relation.

Medical Statisticians Need Clinicians

Having emphasised the importance of statisticians to clinicians, we now discuss why medical statisticians require the input of clinicians. Clinicians are also reminded that they need to assist their statistician colleagues in various aspects of their collaborative research.

Example: Research Design

Suppose an investigator wants to evaluate the effect of a painkiller. A randomised controlled trial (RCT) is to be set up for comparison with a control group. One possible design is to use a placebo control group; another is to have non-placebo controls. A question that may be raised to the trial statistician is that of which is the correct design to use.

In a properly designed RCT, the only systematic difference between the treatment group and the placebo control group is the active substance in the painkiller. So the difference in the outcome, level of pain, can be attributed to the pharmaceutical effect of the painkiller. In the alternative, non-placebo design, the difference in the outcome includes two effects: a pharmaceutical effect and a psychological, or placebo effect.

If the investigator aims to study the pharmaceutical

effect, a placebo control group should be used. Statisticians called this an “explanatory attitude”. However, a clinician should not ignore the patient’s psychology. If the investigator wants to evaluate the overall impact of the drug, a non-placebo control group should be used. This is called a “pragmatic attitude”. The choice of design depends on which attitude is judged to be clinically relevant to the study. Statisticians cannot make this decision themselves.

Example: Data Analysis

Survival analysis is one of the most common statistical methods seen in the medical literature. To employ a parametric survival analysis method, one needs to specify the hazard function (or mortality rate in relation to time). For instance, Figure 1a shows the typical neonatal mortality hazard, which can be approximated by a “Weibull model”. Figure 1b shows the typical mortality hazard in the elderly (above 60 years old), which requires a “logistic model”.

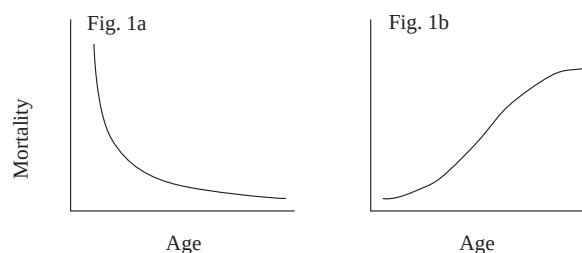


Fig. 1. Illustrations of typical patterns of mortality hazard in relation to age: a) neonatal mortality (0-12 months); b) old age mortality (60 years and above).

The dominant survival analysis method in the medical community is the Cox model, which is not parametric. There are pros and cons of using either parametric or Cox models. The justification most commonly given by users of the Cox model is that it prevents the risk of mis-specifying the hazard. However, investigators familiar with the clinical phenomenon being studied may find this justification irrelevant. How can an expert in his or her field mis-specify the hazard function? In the first author’s own research projects, he often prefers parametric survival analysis methods. However, when asked to give statistical advice to others, he is sometimes more inclined to recommend the Cox model, simply because, as a statistician, he is not always sufficiently familiar with the subject matter at hand to recommend a suitable parametric model. Thus, statisticians need to constantly remind themselves of the importance of eliciting clinical information from clinical collaborators.

In a previous section, we discussed the importance of model selection in the studies of interaction. We said

clinicians need statistical support to make informed choices of models. In fact, this also illustrates how difficult it is for statisticians themselves to set up good regression models.

Example: Interpretation of Findings

An important role of clinicians in the interpretation of quantitative findings is to assess the “clinical significance” of the results, in contrast to the “statistical significance” which is assessed by statisticians. For instance, a statistically significant difference has been shown in the gestational duration of males and females. But this difference is less than a day. In this case the clinical insignificance is obvious. In some cases, however, it would be better to have a clinical opinion. In a study, we found a statistically significant non-linear relationship between body mass index and blood pressure. The decision was whether the non-linearity was clinically significant enough to be kept in the regression model at the expense of simplicity in the model and in the presentation of findings. In the end, we considered it not important and we dropped the non-linear term.

Working Together

Quite often we see poor communication and collaboration between clinicians and statisticians. On the one hand, some clinical investigators may pass a dataset to their statistical colleagues and think that they have no further responsibility until they receive the statistical report. They may also be tempted to think that simply using a user-friendly statistical package that allows them to point and click will eventually give them the right results. On the other hand, some statisticians are reluctant to seek the input of clinicians.

Some of the barriers to effective communication and collaboration between clinicians and statisticians that we can see are as follows. Firstly, both parties can be over confident about the strengths of statisticians and fail to appreciate the relevancy of medical thinking in statistics. Furthermore, the emphasis of specialisation does not encourage discussion across disciplines. Some clinical investigators therefore very much leave the statistical aspects to the statisticians alone. Secondly, and paradoxically, some clinicians do not realise the importance of statisticians’ participation in various stages of research. They may not think their statistical colleagues are justified to become co-authors in publications arising from the research. As a result they may hesitate to discuss with the statisticians since it has implication on authorship. Likewise, statisticians can have a similar attitude towards clinicians.

Thirdly, we have heard clinicians criticising statisticians as being unable to communicate in non-technical ways. Sometimes the thinking and language of statisticians are so unearthly that they seem to come from another planet. This dissatisfaction has certainly been heard by the professional statistician community. Some statistics degree programmes now include consultation and communication as one of their study modules. Exposure to medical and other real-life settings have also been emphasised by some statistics educators.⁶ Fourthly, clinicians and statisticians do not often have easy access to the service or collaboration of each other. Not all clinical research departments employ a statistician. Budget certainly matters; but the under-valuation of the potential contribution of a statistician also has an influence.

Our objective in writing this article is to convince our fellow clinicians and statisticians of the importance of working together. It is not our intention to criticise either party in any way. Our personal experience in participating in clinician-statistician collaborations has been largely positive. We have found the depreciation of value of publication due to having multiple authors compensated by an increase in quality, and the little trouble of asking advice from across disciplines rewarded by knowledge. We strongly feel that only by recognising the strengths and weaknesses of each other, and by learning to work together, can we truly make progress in our common areas of research. If a medical team appreciates and needs the participation of statisticians but does not have access to them, they should proceed to employ one. As always, the ultimate concern is the benefit of patients. By drawing the two professions closer together, we believe this can lead to a better outlook for patients.

REFERENCES

1. Lewis J A. Statistical principles for clinical trials (ICHE9): an introductory note on an international guideline. *Stat Med* 1999; 18:1903-4.
2. Altman D G. Statistical reviewing for medical journals. *Stat Med* 1998; 17:2661-74.
3. SAS Institute. SAS/STAT Software: Changes and Enhancements through Release 6.12. Cary, NC: SAS Institute, 1997:247-348.
4. Stata Corporation. Stata Reference Manual, Release 6, Volume 1 A-G. College Station, Texas: Stata Press, 1999:525-36.
5. Rothman K J. Modern epidemiology. Boston: Little, Brown and Company, 1986:311-26.
6. Hand D J. Deconstructing statistical questions (with discussion). *J R Stat Soc Series A* 1994; 157:317-56.