BIOAUTOMATION, 2009, 12, 82-87

Evidentiary Standards in Medicine with Particular Reference to Meta-analysis

Yee Hung Choy[1], Anthony Shannon[2*]

[1]*Hong Kong Polytechnic University*
Hung Hom, Kowloon, Hong Kong
E-mail: mayhchoy@inet.polyu.edu.hk

[2]*Warrane College, University of New South Wales*
PO Box 123, Kensington, NSW 1465, Australia
E-mail: t.shannon@warrane.unsw.edu.au

[*]*Corresponding author*

Received: January 25, 2009 **Accepted: February 27, 2009**

Published: March 16, 2009

***Abstract**: The purpose of this paper is to raise issues about what constitutes convincing medical evidence, particularly in meta-analysis. While meta-analysis, like any statistical approach, is open to abuse, the point here is how compelling does the quality of evidence need to be in order to be convincing?*

***Keywords**: Meta-analysis, Epidemiology, Controlled trials, Cohort, Case-control, Quasi-Experimentation, Intuitionistic Fuzzy Logic.*

Introduction

Increasingly, government approval of medical decisions requires Level One evidence (as outlined below). For examples, the extent of a government subsidy for an urgent life-saving operation or medical insurance benefits for a pharmaceutical product can depend on this level of evidence. While this paper is essentially theoretical, the problems it raises apply to some syntheses published in this journal from time to time.

Yet there are many situations where double-blind randomised cross-over clinical trials are neither possible nor appropriate. For instance, in considering the question "how will we test the efficacy and safety of new life-prolonging technologies ?", Kent [12] observes that "if senescence begins in one's 30s but the outcome (that is, death) can only be measured in one's 70s or 80s, how will researchers be able to perform timely clinical trials in humans?" Nor is Level One is always sensible, especially if the result is obvious as Smith and Pell satirise [22].

Moreover, as Newman [18] compellingly argues, we tend to be convinced less by pure logic, than by a convergence of probabilities, for which he posits the "illative sense". Hayes too [11] grapples with the questions: "How do you persuade yourself that a statement is true or an answer is correct? How do you persuade someone else?" Thus, Fisher was troubled by Mendel's experimental data because they fitted the theory too well! [6]. The presentation of facts to the human mind calls for persuasion; "and if you would persuade, you must have some idea of how people's minds work, of the ideals which move them and the prejudices which enchain them" [26].

Conjecture

Newman [18] argued persuasively that "absolute certitude … was the results of an assemblage of concurring and converging probabilities", and that "probabilities which did not reach to logical certainty might suffice for a logical certitude; that "the certitude thus brought about might equal in strength the certitude which was created by the strictest scientific demonstration". For Newman [17], "certitude is a mental state ... a quality of propositions but not … a passive impression made upon the mind from without … but … an active recognition of propositions as true".

"The central question is one of partial belief, belief that may be quite pronounced, but which stops short of demonstrative or 'mathematical' certainty. Very simply, what kind of evidence and how much of it ought to be necessary to persuade a reasonable inquirer that it is more appropriate to accept a proposition than to reject it? How ought we to order degrees of belief that lie somewhere between absolute conviction and utter dismissal? What degree of belief is necessary to justify an action with grave consequences? This is not the sort of question that mathematicians are given to worrying about, at least not when going about their mathematical business. But it is the central concern of 'practical reason', and in various forms it confronts us in many societal roles" [13].

As Franklin [7] also observes: "superficially, it is easy to know whether an herb cures a certain disease, by testing it on many cases and seeing if a cure results. Just occasionally, it is as easy as that, but almost always there are so many variables, spontaneous cures, extenuating circumstances, and possible excuses that it is almost impossible to extract a truth from any reasonable amount of data. It was only from the late nineteenth century that the sophisticated statistical techniques of biometry (modern statistics) were developed in agriculture and genetics and were applied in psychology, sociology, and drug trials."

Gordon Smith of Cambridge and Jill Pell of Glasgow in a celebrated paper [22], which is relevant to the theme of this article, humorously, but seriously, set out to asses outcomes in control groups and parachute groups "to determine whether parachutes are effective in preventing major trauma related to gravitation challenge"! Not surprisingly, their "search strategy did not find any randomised controlled trials of the parachute", and so their conclusion was that common sense be applied when considering the potential risks and benefits of interventions.

Meta-analysis

Meta-analysis, in particular, goes beyond primary and secondary analysis in order to compare studies (usually previously published) as statistical units themselves and to combine studies when justified [8].

Glass *et al*. [10] argue, and illustrate their arguments, that meta-analysis should include poorly designed studies as well as well-controlled ones because their general trend is in the same direction and thus they reinforce and emphasise such trends when they exist. This is illustrated in Fig. 1 which shows the results of meta-analysis of hundreds of studies on the controversial issue of the relationship between class-size and academic achievement in schools [9]. The graph also shows how the meta-analysis in this case has harmonized results from some apparently conflicting individual studies.

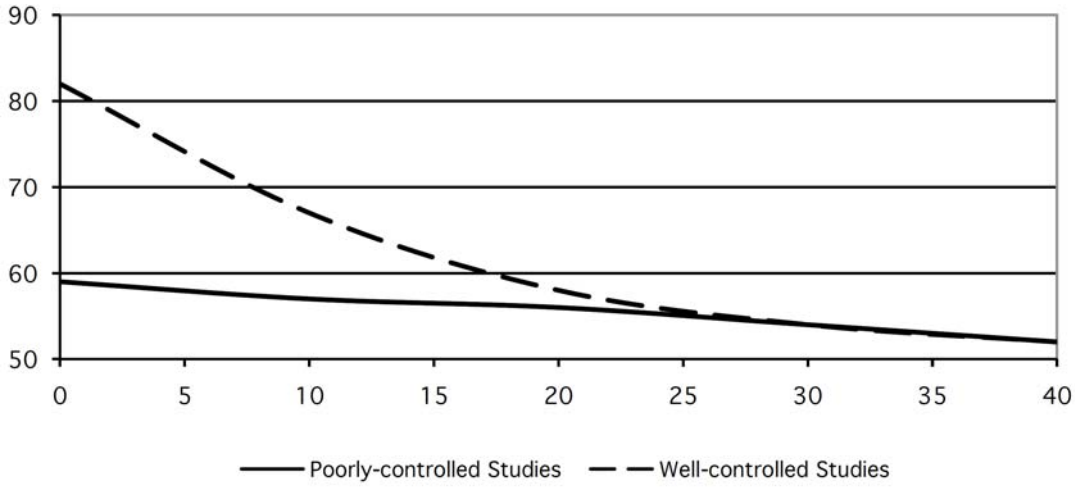


Fig. 1 Regression lines for achievement onto class size

Furthermore, it is becoming customary in meta-analyses of medical literature to qualify trends with the quality of evidence ratings as in Table 1.

Table 1. Quality of evidence for meta-analysis

Levels	Controlled Trials	Epidemiological Evidence
I	A systematic review of all relevant randomized controlled trials	Systematic review of all relevant population-based studies
II	At least one properly-designed randomized controlled trial	A well-designed population based study or representative cohort study
IIIA	Well-designed but not randomized, controlled trials	Well-designed case-control study, cohort study or less well-designed population based study
IIIB	Well-designed cohort or case-control analytic studies, preferably from more than one centre	
IIIC	Multiple time-series with or without intervention	
IV	Opinions of experts based on clinical experience or descriptive studies	Descriptive case series, clinical experiences, respected authorities

Pocock [18] has summarised the essential features of the meta-analytic approach as follows: Meta-analyses (systematic reviews, overviews) have become a dominant feature of the medical literature, and the best can get close to encapsulating all the essential features of a world wide research endeavour in a way that no other approach can achieve … There are three determining factors in any specific meta-analysis: problem definition … quality … comprehensiveness [2] … Another valuable development has been the increased use of patient data for meta-analysis [23] … The statistical basis of meta-analysis is now well established [4], the main controversy being over the choice between fixed and random effects models, the essentials of which have long been known [3]. At times, I have found such debate frustrating since (a) if there is no serious heterogeneity, then the difference is negligible and (b) if substantial statistical heterogeneity does exist, then both approaches require rather peculiar assumptions for valid interpretation of the estimated treatment effects. What is really needed, and tends to get neglected, is serious investigation into the sources of heterogeneity [24].

Hence, a meta-analysis can create a larger pool of subjects with greater statistical power, but it cannot improve the evidence levels of the original studies. When then does compelling evidence become convincing? [21].

Evidence

Thus, for example, it is becoming customary in meta-analyses of medical literature to qualify trends with the quality of evidence ratings as in Table 1. This is based on Liddle et al [14] and Mitchell and Wang [16]. This four point scale has been recommended by the National Health and Medical Research Council in Australia after adaptation from the United States Preventative Services Taskforce.

In order to assign a level of quality, the following are typical of criteria which can be applied to each study in a meta-analysis:

- What risk factors considered;
- What outcomes considered;
- Characteristics of study population;
- Study type: case-control, cohort, etc;
- Is study population well defined?;
- Are all prognostic factors included in analysis?;
- Existence of bias: direction of bias.

Smith and Pell's paper [22] also implicitly contains advice for potential meta-analysts that the use of powerful statistical tools might not always be justified. Considerations about the appropriate use of parametric tests or distribution-free assumptions cannot be ignored, though the cumulative effect of combining studies, even those which are poorly designed, can reveal a trend which individual studies might obscure or even contradict. The conclusions of a well-prepared meta-analysis can sensitise medical practitioners and alert medical administrators to conclusions about the management of disease which transcend current accepted wisdom.

Concluding comments

Glass, Pocock, Smith and Pell, and the Franklins are, in effect, asking us to look more closely at the conclusions we reach, and in questioning current norms to distinguish the relative from the absolute and the subjective from the objective.

There are also other more fundamental issues in the design and analysis of medical issues for field settings [4]. These include quasi-experimental designs and the modes of analyzing the data that result from them. They have been used effectively in education and psychology, but rarely referred to in medicine. "The designs serve to probe causal hypotheses about a wide variety of substantive issues in both basic and applied research." The study of the epistemology of causation shows the need to distinguish, for instance, the efficient causation of the mathematical physicist from the formal causation of the applied mathematician [15].

Finally we return to the review by Levitt with which we began this paper. "The word 'probable' as now used, even by scientists, rarely falls, within the rubric of quantitative probability calculus. It is easy to see this even within the discourse of mathematics itself. What do we mean when we aver that conjecture X is 'probably' true or that a given strategy is 'likely' to succeed in proving it? Pretty clearly, there is no way to give a quantitative significance to these assertions, nor, indeed, to translate them into any suitable formalism.

Worse, we really have no idea of a systematic epistemology that might justify them. Yet remarks like these are really the working discourse of research mathematicians; we work on conjectures that seem probable, using methods that seem likely to get somewhere, but all this 'seeming' is tied up in unaccountable subjective intuition, informed by analogy and experience. The philosophical status of all this is unclear. The same applies to science all down the line; we think of string theory as probable (or not) and likewise for anthropogenic warming or prions as the cause of Alzheimer's. These judgements are the stock-in-trade of everyday science. But there is no widely accepted justificatory *theory* of judgment that stands behind them" [13].

Perhaps it is time to embed more of the calculus of fuzzy logics in general [25] and intuitionistic fuzzy logic in particular [1] into the discourse of medical decision making.

References

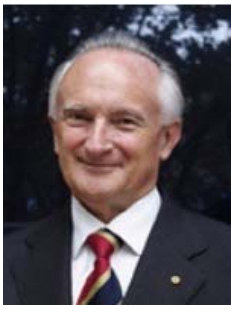
1. Atanassov K. (1999). Intuitionistic Fuzzy Sets, Heidelberg, Physica-Verlag.
2. Chalmers I. (1993). The Cochrane Collaboration: Preparing, Maintaining and Disseminating Systematic Reviews of the Effects of Health Care, Annals of the New York Academy of Sciences, 703, 156-165.
3. Cochrane W. (1954). The Combination of Estimates from Different Experiments, Biometrics, 10, 101-129.
4. Cook T. D., D. T. Campbell (1979). Quasi-experimentation: Design and Analysis Issues for Field Settings, Boston, Houghton-Mifflin, Ch. 1.
5. Fleiss J. L. (1993). The Statistical Basis of Meta-analysis, Statistical Methods in Medical Research, 2, 121-145.
6. Franklin A., A. W. F. Edwards, D. J. Fairbanks, D. L. Hartl, T. Seidenfeld (2008). Ending the Mendel-Fisher Controversy, University of Pittsburgh, Pittsburgh.
7. Franklin J. (2001). The Science of Conjecture: Evidence and Probability before Pascal, Baltimore, The Johns Hopkins University Press, 162.
8. Glass G. V. (1976). Primary, Secondary and Meta-analysis of Research, Educational Researcher, 5, 3-8.
9. Glass G. V., M. L. Smith (1979). Meta-analysis of Research on Class Size and Achievement, Education Evaluation and Policy Analysis, 1, 2-16.
10. Glass G. V., B. McGaw, M. L. Smith (1981). Meta-analysis in Social Research, Beverley Hills, Sage.
11. Hayes B. (2008). Monty Hall Redux, American Scientist, 96, 434-435.
12. Kent D. M. (2008). Listening to Resveratol, American Scientist, 96, 358-360.
13. Levitt N. (2004). Review of the Science of Conjecture, Mathematical Intelligencer, 26, 53-55.
14. Liddle, J., M. Williamson, L. Irwig (1995). Evidence Evaluation Test, Sydney, New South Wales Department of Health.
15. McCaughan J. B. T. (1987). Capillarity – A Lesson in the Epistemology of Physics, Physics Education, 22, 100-106.
16. Mitchell P., J. J. Wang (1996). Clinical Practice Guidelines for the Management of Diabetic Retinopathy. Sydney: Department of Ophthalmology, The University of Sydney, Westmead Hospital.
17. Newman J. H. (1985). An Essay in Aid of a Grammar of Assent, (Ed. I.T. Ker), Oxford, Clarendon Press, 337.
18. Newman, J. H. (1967). Apologia pro Vita Sua, (Ed. M.J. Svaglia), Oxford, Clarendon Press, 31.

19. Pocock S. J. (1996). Clinical Trials: A Statistician's Perspective, In P. Armitage, D. A. Herbert (Eds), Advances in Biometry, New York, Wiley, Ch. 20.
20. Sackett D. L., S. E. Straus, W. S. Richardson, W. Rosenberg, R. B Haynes (2000). Evidence-based Medicine, Second Edition, Edinburgh, Churchill Livingstone, 29-65.
21. Shannon A. G., D. K. Y. Chan, W. T. Hung, Y. H. Choy (2008). Meta-analysis of the Effect of Medication in Falls in the Elderly, Bioautomation, 10, 75-87.
22. Smith G. C. S., J. P. Pell (2003). Parachute Use to Prevent Death and Major Trauma Related to Gravitational Challenge: Systematic Review of Randomised Controlled Trials, British Medical Journal, 327, 1459-1461.
23. Stewart L. A., M. J. Clarke (1995). Practical Methodology of Meta-analyses (Overviews) using Updated Individual Patient Data, Statistics in Medicine, 14, 2057-2079.
24. Thompson S. G. (1994). Why Sources of Heterogeneity in Meta-analysis should be Investigated, British Medical Journal, 309, 1351-1355.
25. Turunen E. (1999). Mathematics Behind Fuzzy Logic, Heidelberg, Physica-Verlag.
26. Walsh M. (2007). Ronald Knox as Apologist: Wit, Laughter and the Popish Creed, San Francisco, Ignatius Press.

Yee Hung (Edward) Choy, Ph.D.
E-mail: mayhchoy@inet.polyu.edu.hk

Dr. Yee Hung (Edward) Choy is an Assistant Professor of the Department of Applied Mathematics, Hong Kong Polytechnic University. He has taught various subjects in Mathematics and Statistics to different levels of students, ranging from Associated degree up to Master levels. He holds the degrees of M.Sc. and Ph.D. and is also a Chartered Statistician of the Royal Statistical Society of United Kingdom. His research interests are in Medical Statistics, epidemiology, particularly through the application of Meta-analysis. He is the principal investigator of several projects funded by the research grants of the Department. He has also provided various kinds of consultancy commissioned by the government or public bodies. He enjoys reading, hiking, theatre, watching soccer matches, and history.

Prof. Anthony G. (Tony) Shannon, Ph.D., Ed.D., D.Sc.
E-mail: tony@warrane.unsw.edu.au, tonySHANNON@raffles.edu.au



Professor A. G. (Tony) Shannon AM is an Emeritus Professor of the University of Technology, Sydney, where he was Foundation Dean of the University Graduate School and Professor of Applied Mathematics, and where he is currently Chair of the Key University Research Centre for Health Technologies. He holds the degrees of Ph.D., Ed.D. and D.Sc. He is co-author of numerous books and articles in medicine, mathematics and education. His research interests are in the philosophy of education and epidemiology, particularly through the application of generalized nets and intuitionistic fuzzy logic. He has taught and mentored at all levels from primary school to post-doctoral. Prof. Shannon is a Fellow of several professional societies and a member of several course advisory committees at private higher education providers. He is on the Board of Trustees of Campion College, a liberal arts degree granting institution in Sydney. In June 1987 he was appointed a Member of the Order of Australia for services to education. He enjoys reading, walking, theatre, number theory, and thoroughbred racing.