MEDICINA NEI SECOLI ARTE E SCIENZA, 14/2 (2002) 515-527
Journal of History of Medicine

Articoli/*Articles*

OBSERVATION, SHERLOCK HOLMES, AND EVIDENCE BASED MEDICINE

JOHN OSBORN
Dipartimento di Scienze di Sanità Pubblica
Università degli Studi di Roma, "La Sapienza", I

SUMMARY

Sir Arthur Conan Doyle, the creator of the fictional detective Sherlock Holmes, studied medicine at the University of Edinburgh between 1876 and 1881 under Doctor Joseph Bell who emphasised in his teaching the importance of observation, deduction and evidence. Sherlock Holmes was modelled on Joseph Bell. The modern notions of Evidence Based Medicine (EBM) are not new. A very brief indication of some of the history of EBM is presented including a discussion of the important and usually overlooked contribution of statisticians to the Popperian philosophy of EBM.

Introduction

If one were to go by the explosion of interest in evidence based clinical practice in the last decade of the second millenium, one could be forgiven for thinking that the idea were new. It is claimed[1] that the origins of EBM date back to mid 19th century Paris or earlier, although the name EBM was coined in 1992. The inventor of the randomised controlled clinical trial and founder of the modern ideas of medical statistics, Sir Austin Bradford Hill, in the 1950s set out the statistical foundations of Evidence Based Medicine (EBM).

In an important editorial[2] entitled "Evidence Based Medicine: What it is and what it isn't", David Sackett, one of the pioneers

Key words: Holmes – EBM – Statistics – Popper

of the new movement for the practice of EBM, and his colleagues emphasise that EBM has two components

> *The practice of EBM means integrating individual clinical expertise with the best available external clinical evidence from systematic research.*

Individual clinical expertise is acquired as a result of clinical practice and means that a clinician is not expected to slavishly follow rules dictated by others when it comes to the treatment a particular patient. The clinician is likely to know much more about the medical history and needs of an individual patient, about the history of the condition, about the social context of the patient including his way of life, his family background, employment situation etc, than can be found by reading and learning from research reports, whose main objective is to reach generalised conclusions about "patients of this type". As Sherlock Holmes said, *"There is nothing like first hand evidence"*[3]. On the other hand, the results of excellent relevant clinical research provide a scientifically valid framework for patient care. According to Sackett et al[4],

> *External clinical evidence both invalidates previously accepted diagnostic tests and treatments and replaces them with new ones that are more powerful, more accurate, more efficacious and safer.*

Or in the words of Holmes,

> *The mystery gradually clears away as each new discovery furnishes a step which leads to the complete truth*[5].

Clearly both components are necessary; clinical expertise without the application of the results of new research is likely to stagnate and cannot be expected to progress without the continuing education provided by good clinical publications.

> *Education never ends, Watson. It is a series of lessons with the greatest for the last*[6].

The nature of external evidence

Evidence is the ultimate product of the analysis of a series of observations. Such a statement may appear banal, but in fact,

precise observations are a necessary ingredient for the improvement of clinical expertise and the production of good research. There is a great tendency for all of us to observe what we expect to see rather than what actually occurs. Sometimes this problem can be ameliorated in clinical research by blinding the patient and the clinical observer (perhaps using a double-dummy), and even blinding the statistician. Sir Arthur Conan Doyle was a medical practitioner and it is said that he modelled his fictional detective, Sherlock Holmes, on one of his professors at Edinburgh Medical School, Dr Joseph Bell (1837-1911). Bell, who was famed for being able to diagnose patients before they even told him their symptoms, was thought by his students to be a magician. In Doyle's words,

> *Dr Bell would sit in a receiving room, with a face like a Red Indian, and diagnose people as they came in, before they even opened their mouths. He would tell them their symptoms and even give them details of their past life and hardly ever would make a mistake.*

According to one of Doyle's classmates, Dr Harold Emery Jones[7], Bell was as full of dry humour and satire, and he was as jealous of his reputation as the detective Sherlock Holmes ever thought of being. One day, in the lecture theatre, Bell gave the students a long talk on the necessity for members of the medical profession to cultivate their senses, sight, smell, taste and hearing. Before him on the table stood a large tumbler filled with a dark, amber coloured liquid.

> *This, gentlemen - announced the professor - contains a very potent drug. To the taste it is very bitter. It is most offensive to the sense of smell. Yet as far as the sense of sight is concerned, that is its colour, it is no different to dozens of other liquids. Now I want to see how many of you gentlemen have educated your powers of perception. Of course we might easily analyse this chemically and find out what it is. But I want you to test it by smell and taste; and, as I don't ask anything of my students that I wouldn't be willing to do myself, I will taste it before passing it round.*

He dipped his finger in the liquid, and placed it in his mouth. The tumbler was passed round. With wry and sour faces the students followed the professor's lead. One after another tasted the

vile concoction; varied and amusing were the grimaces made. The tumbler, having gone the round, was returned to the professor.

> *Gentlemen - said he, with a laugh - I am deeply grieved to find that not one of you has developed this power of perception, which I so often speak about; for if you had watched me closely, you would have found that, while I placed my forefinger in the medicine, it was the middle finger which found its way to my mouth.*

Variations on this story substitute urine for the drug, which is tasted for its sweetness in order to diagnose diabetes.

If Joseph Bell, by his teaching of what was essentially evidence based diagnosis, was the model for Doyle's character Holmes, it was another of his teachers at Edinburgh, Sir Henry Littlejohn who gave Doyle the idea to cast his new character as a meticulous detective. Littlejohn was the Police Surgeon and the Medical Officer of Health to the City of Edinburgh and also Lecturer in Forensic Medicine and Public Health at the Royal College of Surgeons. As Police Surgeon, he had unequalled facilities for the study of crime and criminals and while Bell was lecturing on deduction and perception, Littlejohn was giving Doyle material for his detective stories.

Evidence Based Medicine, however, did not originate in Edinburgh. One of the earlier debates on the merits of evidence for the practice of clinical medicine arose as a result of the analysis by Jean Civiale (born 1792, died 1867) of a large body of data relating to two methods of treating patients suffering from calculi of the bladder. The results of his analyses were reviewed by a Commission whose report was presented to the Paris Academy of Science. This report has recently been translated and republished[8] and all the modern day arguments for and against the practice of EBM can be seen. On the etiology of calculi the Commission comments

> *The detailed study of the causes likely to produce calculi disprove a certain number of statements issued in relation to different foods and some beverages that were too hastily declared to be likely to cause the disease. Whatever the research taken into consideration, everything remains obscure, there is nothing but uncertainty on this point.*

On the criticism of the application of statistics in clinical practice, one finds comments not much different to those one hears now-a-days from some clinicians who harbour reservations about EBM:

> *In the field of statistics.....the first task is to lose sight of the individual seen in isolation, to consider him only as a fraction of the species. He must be stripped of his individuality so as to eliminate anything accidental that this individuality might introduce into the issue at hand. In applied medicine, on the contrary, the problem is always individual, facts to which a solution must be found only present themselves one by one; it is always the patient's individual personality that is in question, and in the end it is always a single man with all his idiosyncrasies that the physician must treat. For us, the masses are quite irrelevant to the issue.*

However, on the merits of the new method of extracting calculi via the urethra without any incision, instead of the old method of extraction after surgical incision, the Commission has little doubt:

> *Today, we must say that his new work, as it stands, will have provided new evidence for the advantages of the substitution.....of a simple operation presenting few dangers for another serious alarming and painful one which until now constituted the only resource of medical art.*

Although the philosophical origins of EBM date back to mid-19th century, or before, its legal status in Britain was implied by the Apothecaries Act of 1815, which licensed apothecaries in order to protect the public from the growing number of unqualified druggists and herbalists. The Medical Act of 1858 lead to the creation of the medical register which contained the names of all doctors with recognised medical qualifications. The 1858 Act restricted the practice of medicine to those doctors included in the register. There was also the implication that these doctors should practice "real" medicine, that is, the medicine taught and learned in medical schools, and the public would be protected against charlatans. The Act was not successful in eliminating complementary or alternative medicine, and indeed, apart from a short period in the middle of the 20th century, the number of

people who seek medical help outside the official medical profession, particularly from herbalists, has continued to increase. (Paradoxically, alternative medicine is still promoted and supplied by chemist's shops, the very place where a patient, having consulted a regular doctor, is required to go to collect his prescription! Boots, the chemists, even publish and distribute free a booklet[9] in which complementary medicine is stated to be safe, and it is implied that orthodox medical help need be sought only where symptoms are severe and persistent). The 1858 Act implied that the medicine practised by registered doctors was based on evidence while the alternative was based on hearsay, old-wives-tales, grannies' remedies etc. If this distinction was one of the objectives of the 1858 Act, it most certainly was not very successful; there are many examples, in all Medical specialities, of practice, which either for lack of evidence or ignorance is not based on evidence. Bandolier[10] is provocative by asking,

> *What do you do when there is no evidence? Carry on with what you are doing because you have no evidence to stop, or stop what you are doing because there is no evidence to carry on?*

The weight of the evidence derived from a clinical study will depend on its design and how well it has been conducted. A simple case series reporting a new treatment may not provide very strong evidence of the effect of the treatment unless the observed effect is exceedingly different from the natural progress of the disease or condition. On the other hand a case series may be sufficient to generate a hypothesis, which might be investigated by more rigorous studies. A control group will always increase the validity of a study based on a case series.

Randomisation of the patients to the treatment groups has the great advantage that it will tend to remove the effect of confounding factors especially if the trial is not too small. Thus in terms of a single study, the randomised controlled trial (RCT) provides the best evidence that a treatment has an effect in comparison with the control group. This evidence is usually presented in the form of a statistical significance test, and a confidence interval, which is the associated estimate of the treatment effect. When there are several studies of the effect of a treatment,

the results may be aggregated using the techniques of meta-analysis, another new name for an old idea. *"There is nothing new under the sun, it has all been done before."*[11]. To learn of the pitfalls of combining evidence in a meta analysis, there is no better starting point than the article by Daniels and Bradford Hill[12] published as long ago as 1952. A good meta analysis should take account of the study designs, involve a well defined strategy for literature searches, assessment of quality, inclusion and exclusion criteria, tests of homogeneity etc., although in 1991, Thompson and Pocock[13] felt that it was necessary to pose the question, can meta analysis be trusted? In the true spirit of meta analysis, Holmes pleads, *"Any truth is better than infinite doubt"*[14]. The inclusion and exclusion criteria are vital to the validity and the interpretation of the summary result.

> *Some facts should be suppressed, or at least a sense of proportion should be observed in treating them. The only point in the case which deserved mention was the curious analytic reasoning from effects to causes...*[15]

A review should bring together all the evidence for and against the effectiveness of a treatment, and there may be no simple clear-cut result. Further, there may be more than one review, and what to do if the reviews differ in their conclusions? For a given patient, the clinician must make a decision and may not have the luxury Holmes enjoyed when he said in honesty to Watson, *"No, no; I never guess. It is a shocking habit - destructive to the logical faculty"* [16].

If the reviews do agree, a review of the reviews may evolve into a clinical guideline. One might be forgiven for thinking that at this point there would no longer be controversy, but not so. Whether created locally or nationally or internationally, guidelines are generally an aggregation of research evidence, expert opinion and clinical experience. The existence of a clinical guideline may intentionally have the effect of limiting the freedom of action of a clinician in the treatment of his patient, and this could have legal consequences and ethical implications. Holmes is mistaken when he says of Dr Grimesby Roylott *"When a doctor does go wrong, he is the first of criminals. He has nerve and the knowledge."*[17]. Holmes is speaking of going wrong

in a legal sense rather than making a mistaken clinical judgement, but unfortunately a clinician rarely has all the knowledge, and errors will occur. Hurwitz[18] expounds a comprehensive and highly readable account of the possible legal implications of following or not following clinical guidelines in his book appropriately titled *Clinical Guidelines and the Law; Negligence, Discretion and Judgment*. These implications are important because the existence of guidelines neither implies that they will be followed in practice nor that their effectiveness will be formally evaluated. A survey[19] on the use of clinical guidelines of 270 senior hospital staff in the UK produced 202 replies. Among these, although 99% thought that clinical guidelines were a good idea, only 19% had a clinical guidelines strategy, although another 45% said that they had plans to develop one. In another study[20], in Australia, 92% thought that guidelines were good educational tools, but 85% qualified that praise by agreeing that guidelines "were developed by experts who don't understand general practice"!

Evidence and the philosophy of scientific progress

The title of this paragraph is nothing but presumptuous when one thinks of the miles of shelves of books and other publications on this subject produced over the past 100 years, but Statistics has played an important, under-rated and often overlooked role in the theories propounded by professional philosophers. Healy[21] has recently published an entertaining but serious discussion of the role of Statistics in the philosophy of science, and the philosophy of science in the practice of Statistics. In essence, the modern subject, Statistics has its origins at University College, London around the start of the twentieth century when Karl Pearson began studying the theory of distributions and applying statistical methods to study biological problems, and for example, discovered the chi-squared distribution and began thinking in terms of the significance test. Pearson's ideas were expanded and developed by Fisher and Gossett (the ever famous "Student" who first described the t-test) in the 1920s and 1930s. The latter two, Fisher and Gossett, perfected the idea of the statistical significance test, which has remained

with us, virtually without change until today. Basically the logical procedure followed in a statistical significance test is:

1. A stimulus provokes the need to perform an experiment to compare the effects of say, two treatments A and B on an outcome. The origin and form of the stimulus is not important and may come from a clinical observation, a hunch, hearsay, complimentary medicine etc. If the stimulus is based on evidence, this evidence cannot be used further in the experiment, and the experiment to compare the two treatments will be interpreted with a completely open mind, ignoring all that is known before (unless a Bayesian approach is used).
2. A null hypothesis is formulated, which states that there is no difference on average between the two treatments. This hypothesis represents the state of knowledge at the start of the experiment.
3. The results of the experiment are analysed to discover if they provide sufficient evidence to reject the hypothesis and thus change the state of knowledge by concluding that one treatment is better than the other.

> *It is a capital mistake to theorise before one has the data. Insensibly one begins to twist the facts to suit the theories, instead of the theories to suit the facts.*[22]

The decision to reject the null hypothesis however is based on probabilistic reasoning. (Actually it is the frequency or repeated experiment approach to probability as opposed to subjective probability or *a priori* probability reasoning); a single patient cannot of him/herself disprove the null hypothesis.

> *We balance probabilities and choose the most likely. It is the scientific use of the imagination*[23].

4. A confidence interval for the effect of interest, the average difference between the treatments, is constructed. This should enable the researcher to determine whether or not there is sufficient evidence to conclude that the difference between the treatments is of clinical importance.

In fact, it was some decades later that the most influential philosopher of science of the 20th century, Sir Karl Popper, re-proposed that science advances, a step at a time, by the refutation of hypotheses[24,25]. When one hypothesis is rejected, the alternative becomes the new state of knowledge. *"One new truth invariably suggests others"* [26]. However, Popper's reasoning was based on the idea that even a single observation could lead to the rejection of the hypothesis. Holmes concurs: *"I never make exceptions. An exception proves the rule"*[27]. However, it seems that the statisticians were expert Popperians long before Popper's theories became popular! In 1937 Fisher[28], anticipating Popper by almost twenty years, stated

> *Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis*

and this assertion was based on a complex form of probabilistic reasoning.

Later, Popper's theory was challenged by that of Thomas Kuhn[29], who argued that while science generally progresses slowly and steadily, there were events of dramatic importance, or revolutions: which totally change the state of knowledge. One can easily think of examples of such revolutions which have transformed scientific thinking: the introduction into Europe of the decimal number system by Leonardo di Pisa (Fibonacci) in 1202, enabling complex arithmetic to be performed and providing the trigger for the start of the renaissance, Galileo, Newton, Einstein's theory of relativity etc. However, revolutions also occur within specialities and in medicine one can think of examples, such as the discovery of effective anaesthetics, the discovery and development of antibiotics. In Medical Statistics the introduction of randomisation in clinical trials by Bradford Hill was a revolution, and it seems that maybe we are experiencing now a Kuhnian revolution in the form of the Bayesian approach to evidence from clinical studies. If indeed the subject, Statistics, is transformed totally by the adoption of Bayesian techniques, it will be necessary to re-think what is meant by evidence based on Popperian inference in relation to medical practice. However, for the

near or medium future, the validity and strength of evidence will continue to be based on Fisher-Popper statistical significance tests and their associated confidence intervals. Perhaps sadly, we are likely to witness for many more years the spectacle of our normally calm, serious, reserved research worker, suddenly triumphant and exuberant as his computer prints out the long awaited and much desired $P<0.05$ value.

Conclusion

It is well known that the number of research journals and research papers increases at an alarming rate every year. It would be hoped that the growth in the number of good research reports is equally rapid. If this is in fact the case, in future it will be ever more difficult to identify good research and maintain a register of valid evidence. The Cochrane Foundation, the National Health Service Centre for Reviews and Dissemination at the University of York and others have taken an enormous step forward by trying to filter out the valid evidence from the bulk of less worthy research. Certainly individual clinicians cannot be expected to read all the latest research reports in their field, let alone evaluate them and classify the results as good evidence or not.

> *I consider that a man's brain is like a little empty attic, and you have to stock it with such furniture as you choose. A fool takes in all the lumber of every sort that he comes across, so that the knowledge which might be useful to him gets crowded out, or at best is jumbled up with a lot of other things, so that he has a difficulty in laying his hands upon it. Now the skilful workman is very careful indeed as to what he takes into his brain-attic. He will have nothing but the tools which may help him in doing his work, but of these he has a large assortment, and all in the most perfect order. It is a mistake to think that that little room has elastic walls and can distend to any extent. Depend upon it - there comes a time when for every addition of knowledge you forget something you knew before. It is of the highest importance, therefore, not to have useless facts elbowing out the useful ones*[30].

A clinician therefore, either must become ever more specialised and remember only the very important aspects of his narrow field, or he can remain a general practitioner but he has to

accept that in many situations he will have to consult his "library". Nowadays, it is almost essential to have a computer to keep the lumber-room in an accessible order and enable easy contact to be made with such organisations as the Cochrane Foundation and the NHS Centre for Reviews and Dissemination.

Sir Arthur Conan Doyle and his mentors, Dr. Joseph Bell and Sir Henry Littlejohn were acutely aware of the value of good evidence in medical practice and for the detective work of Sherlock Holmes. The 60 stories involving Holmes and his assistant Dr. Watson were published between 1887 and 1927 in the Strand Magazine, Colliers Weekly and other periodicals. They have given pleasure to generations of avid readers eager to discover something of the extraordinary ability of Sherlock Holmes to deduce the truth from whatever evidence was available. Pearson, Fisher, "Student", Popper and others have formalised the idea of the use of evidence to test hypotheses and enable science to progress. Bell, Sackett, his colleagues and others have sought to identify from the mass of available research evidence what is valid and can be realistically applied in the every day practice of clinical medicine.

BIBLIOGRAPHY AND NOTES

1. SACKETT D.L., ROSENBERG W.M.C., GRAY JAM, HAYNES R.B. and RICHARDSON W.S., *Evidence based medicine: what it is and what it isn't*. Brit. Med. J. 1996; 312: 71-2.
2. Idem.
3. CONAN DOYLE A., *A Study in Scarlet*. Beeton Christmas Annual 1887.
4. See n. 1.
5. CONAN DOYLE A ., *The Engineer's Thumb*. Strand Magazine 1892.
6. CONAN DOYLE A., *The Adventure of the Red Circle*, 1911.
7. JONES H.E., Date and origin unknown, but cited on the world wide web at http://snopes.simplenet.com/college/medical/urine.htm
8. POISSON S.D., DOUBLE FJ., et al., *Statistical Research on conditions caused by calculi, by Doctor Civiale*. SWAINE VERDIER A., (Translated by), Int. J Epid. 2001; 30;1246-1249.
9. *Anon. Complementary Medicine. Your guide to choosing holistic medicines*. 2000. The Boots Company PLC, Nottingham, England.
10. Bandolier. Prophylactic removal of impacted third molars. http://www.jr2.ox.ac.uk/Bandolier/band42/b42-2.html
11. CONAN DOYLE A., *A Study in Scarlet*. Beeton Christmas Annual 1888.

12. DANIELS M., HILL A.B., *Chemotherapy of pulmonary tuberculosis in young adults; an analysis of the combined results of three Medical Research Council Trials*. Brit. Med. J. 1952; i: 1162-1168.
13. THOMPSON S.G., POCOCK S., *Can meta analysis be trusted?* Lancet 1991; 338:1127-30.
14. CONAN DOYLE A ., *The Yellow Face*, 1893.
15. CONAN DOYLE A., *The Sign of the Four*. Lippincott's Magazine 1890.
16. CONAN DOYLE A., *The Sign of the Four*. Lippincott's Magazine 1890.
17. CONAN DOYLE A., *The Speckled Band*, 1892.
18. HURWITZ B., *Clinical Guidelines and the Law, Negligence, discretion and Judgment*. Radcliffe Medical Press, Abingdon, 1998.
19. RENVOISE E.B., HAMPSHAW S.M., PINDER J.M., AYRES P., *What are hospitals doing about clinical guidelines?* Qual Health Care 1997; 6(4): 187-91.
20. GUPTA L., WARD J.E., HAYWARD R.S., *Clinical Practice guidelines in general practice: a national survey of recall, attitudes and impact*. Med. J. Aust. 1997; 166(2): 69-72.
21. HEALY MJR., *Paradigms and Pragmatism: Approaches to Medical Statistics*. Ann Ig 2000; 12(4): 257-264.
22. CONAN DOYLE A., *A Scandal in Bohemia*, 1892.
23. CONAN DOYLE A., *The hound of the Baskerville*, 1901-2.
24. POPPER K., *The Logic of Scientific Discovery*. Hutchinson, London, 1959.
25. POPPER K., *Conjectures and Refutations*. 4th ed., London, Routeledge and Kegan Paul, 1972.
26. CONAN DOYLE A., *Silver Blaze*, 1892.
27. CONAN DOYLE A., *The Sign of the Four*, 1890.
28. FISHER R.A., *Prof Karl Pearson and the method of moments*. Ann Eugenics 1937; 7: 303-318.
29. KUHN T., *The structure of Scientific revolutions (3rd ed.)*. Chicago, The University of Chicago Press, 1966.
30. CONAN DOYLE A., *A study in Scarlet*, 1887.

Correspondence should be addressed to:
John Osborn, Dipartimento di Scienze di Sanità Pubblica, Università degli Studi di Roma, "La Sapienza", Piazzale Aldo Moro, 5 - 00161 Roma - I.