Chapter 2

METHODS OF MEDICAL RESEARCH

If you have had the good fortune to see the film "The madness of King George", the eighteenth century British monarch who presided over the loss of the American colonies, you will have noted the activities of his physicians. They applied scalding hot cups to his back and were fascinated with the form and texture of his faeces and the colour of his urine. No doubt they let his blood as well. In those days, it generally boded ill for you if you fell into the hands of the medical profession. Nowadays, there is a very good chance that your doctor may relieve you from some dreadful affliction or even save your life.

What is the reason for this change? It is neatly summed up in the names of the final examinations that I sat for nearly 40 years ago as a medical student. One subject was called "The principles and practice of medicine". The other was entitled "The science and art of surgery". Medicine is now founded on scientific principles. It has repeatedly shown itself able to predict, to develop and to explain the many valuable treatments on which its reputation depends.[1] If alternative medicine is to make the same contribution as orthodox medicine, it has to be tested and refined at the bar of the scientific method. That is what this book is about. It is concerned with an assessment of various forms of alternative medicine using scientific principles.

What is the scientific method? We need to have an understanding of the philosophy of science whereby an hypothesis is formulated and tested. Secondly, we need to know how to decide whether such tests have answered the question; this often involves at least an elementary understanding of statistics. Finally, it is useful to be able to access the scientific literature so that you are able to pursue the details of a subject which is of particular interest.

The basic principles of science

Science is the systematic gathering of knowledge leading to an understanding of the world around us. In fact, the word itself is derived from the Latin word *scientia* meaning "knowledge". Before we discuss the scientific method, we need to know something about the concepts of deduction and induction.

---

[1]Charlton BG. Philosophy of medicine: alternative or scientific. *Journal of the Royal Society of Medicine* 85: 436-438, 1992

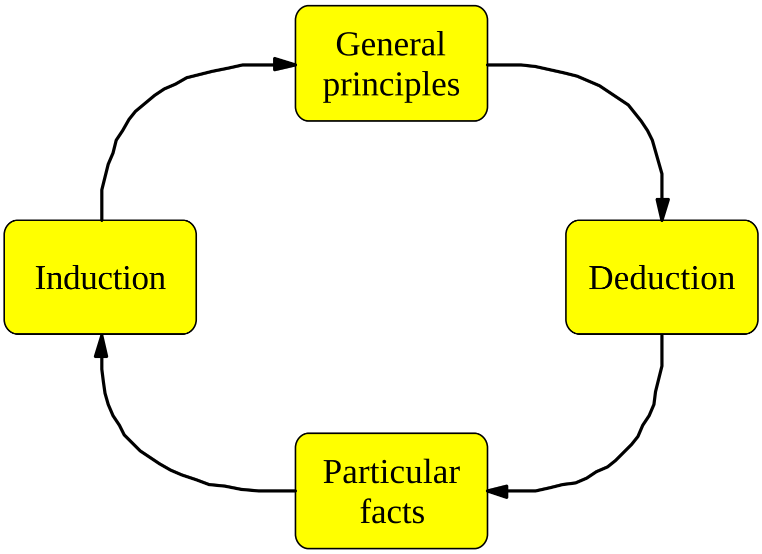

Figure 1. The places of deduction and induction in scientific reasoning.

Deductive reasoning

In deduction, we reason from the general to the particular (Figure 1). Thus, conclusions are reached by reasoning from certain general principles that are assumed to be true. This approach of logic by argumentation was expounded by the famous Greek philosopher, Aristotle, 350 years or so before Christ. An illustration of such a process is as follows; if we know that humans have two legs and if we know that John Brown is a human, then it is reasonable to assume or deduce that he has two legs.

Deduction is the cornerstone of mathematics. Principles such as these were used by Pythagoras to prove his well-known theorem that in the case of a right-angled triangle, the area described by the square of the hypotenuse is equal to the sum of the squares of the other two sides. Thus, deduction clearly gives rise to a better understanding of the reality with which we are familiar but usually does not give rise to new knowledge.

Inductive reasoning

With induction, on the other hand, we reason from the particular to the general (Figure 1). This technique was described by Francis Bacon in the early seventeenth century. Charles Darwin provided a well-known example of this process. He made

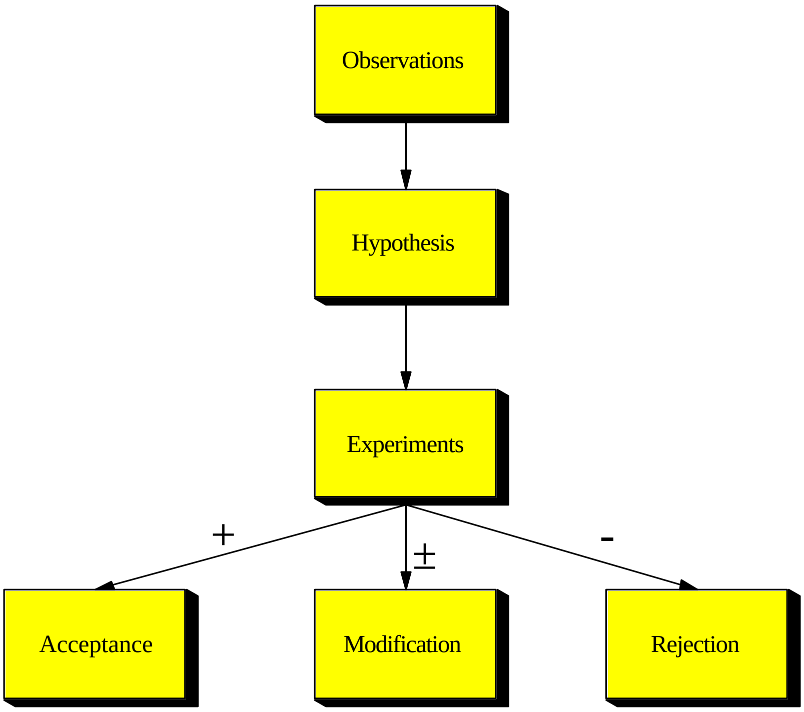

Figure 2. Processes involved in the scientific method. + = successful experiment; ± = suggestive or partially successful experiment; - = negative experiment.

many observations on variations between different members of many species of plants and animals. He eventually conceived an explanation for his observations and formulated his theory of evolution by natural selection.

Induction is therefore the foundation of the observational sciences. But this produces a problem. Any given situation cannot be described as a certainty until all the possible options have been observed and this is frequently impossible. The inductive sciences will always be tentative and remain open to revision. We will come back to this problem when we discuss statistics.

Observation and experimentation

The scientific method involves a systematic series of steps (Figure 2). First, information, also called data, is gathered by observation. Next, an hypothesis is generated by inductive

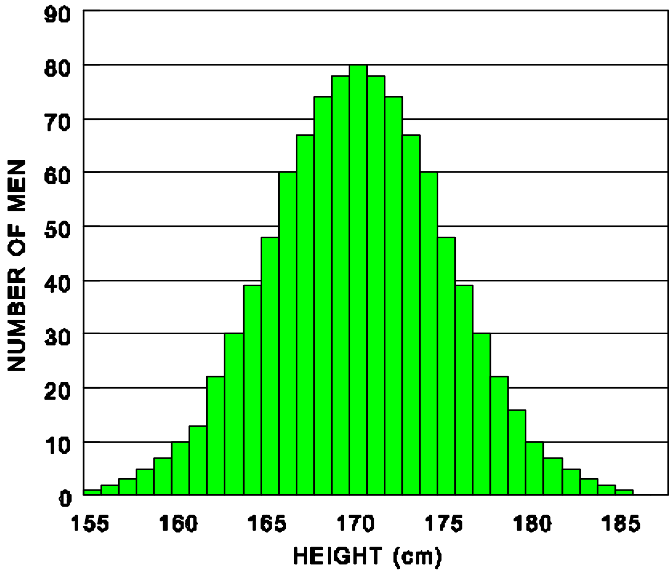

Figure 3. Histogram showing a normal distribution. It plots the heights of 1,000 men.

reasoning to explain these observations. Third, a number of implications that can be tested are then deduced logically from this hypothesis. Finally, experiments, especially those that provide measurements, are then devised to test these postulates. Depending upon the results of the experiments, the hypothesis is either supported or refuted. In the latter case, the hypothesis is either modified or discarded. In the former instance, if enough evidence is gathered, the hypothesis becomes accepted as a theory or law.[2]

Why do we need statistics?

When the dentist William Morton in 1846 persuaded some surgeons at the Massachusetts General Hospital in Boston, USA to administer ether to a young man who then had an operation performed painlessly on his neck, one of the assembled throng is said to have remarked “Gentlemen, this is no humbug!”. No statistics were necessary; the anaesthetic clearly worked.

Likewise, when Howard Florey in 1940 injected penicillin into half of a group of mice infected with streptococcal bacteria, no statistical analysis was performed. All of the mice given penicillin survived whereas the rest all died. The difference was obvious and the

---

[2]Lafferty K. The burden of proof. *Today’s Life Science* March, pp. 10-15, 1992

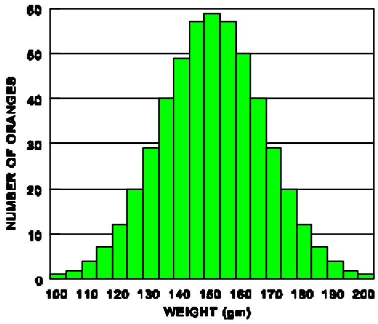

Figure 4. Oranges collected from tree A.

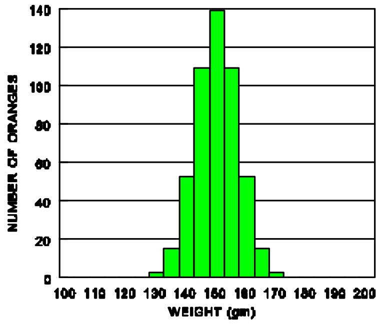

Figure 5. Oranges collected from tree B.

antibiotic era had arrived. However, there have been many advances in medicine which are incremental rather than representing a dramatic breakthrough. A method was needed to decide whether small differences represented a real improvement or could have occurred by chance. The answer was the development of statistics.

Variation

The basic problem which needs to be addressed by statistics is the phenomenon of variation. Populations are rarely exactly uniform. There are usually variations between members of the population, whatever characteristic is being measured. If you measure the heights of 1,000 men, they will not all be the same. If you categorise everybody to the nearest centimetre and count the numbers in each group, you will generate what is called a histogram or bar graph (Figure 3). A histogram such as this is often stylised into a curve. When such a curve is symmetrical; it is called a "bell-shaped" curve or a "normal" curve. In a normal curve, the "mode", the "median" and the "mean" values are all the same. The mode is the most commonly occurring value. The median is the middle value, that is, there are equal numbers of values less than and greater than it. The mean is also called the "average"; this is calculated by adding up the sum of all the values and dividing the total by the number of observations.

But look at Figures 4 and 5 which show the numbers of oranges of various weights when 500 oranges were collected from each of two trees. There are two plots which have the same median, mode and mean (150 grams), but they are clearly different. Tree A has a wide spread while the Tree B has a narrow spread. There are some very small oranges from Tree A but also some very large ones whereas the oranges from Tree B are much more consistent in size and weight. The spread from one side to the other is called the range. Although it is useful for you and me to "eyeball" the degree of spread, statisticians have devised an arithmetical calculation which defines the amount of spread. This is called

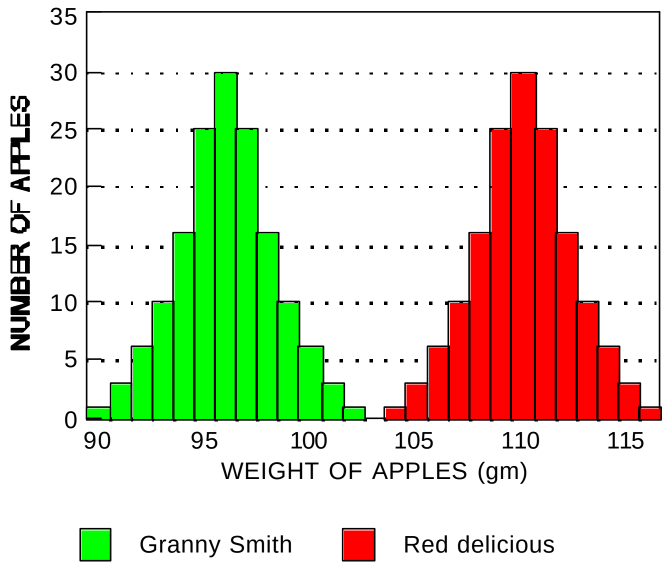

Figure 6. Numbers of granny smith apples and red delicious apples of

the standard deviation. We do not need to understand how statisticians do it; we simply need to know that they use this tool.

Every population then can be measured by three parameters - the number of individuals in the population, the mean value and the standard deviation. This provides the basis for statistical analysis.

Consider Figure 6. There are two groups (or populations) of 150 apples, one of granny smiths, and the other of red delicious. The smallest red delicious apple weighs more than the largest granny smith. Clearly, the red delicious apples are bigger. Statistical analysis will simply confirm that assumption. But what about Figure 7 which compares the weights of red delicious and yellow delicious apples? Are red delicious apples really heavier than yellow delicious apples or could the higher average weight of red delicious simply be a sampling error and just occurred by chance? By knowing the number of apples in each group, the average weight, and after calculating the standard deviation by measuring all the individual weights, a statistician can estimate whether or not the weights of red and yellow delicious apples are really likely to be different.

Notice that I said "likely". Statisticians are not definite; they talk in terms of probability. They generate what is called a "P" or "probability" value. When a statistician considers that two populations are probably the same, he says that any differences observed between the two groups are "not statistically significant". On the other hand, if

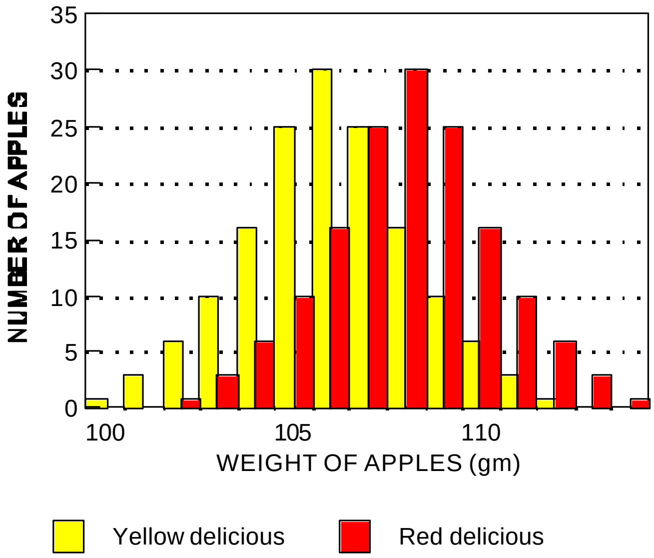

Figure 7. Numbers of granny smith apples and red delicious apples of various weights.

they are likely to be different, he quantifies the likelihood. Just to make life difficult, statisticians work backwards and use what is called the "Null Hypothesis". Thus, a P value of 0.05 means that there is one chance in 20 that the two populations studied are really the same. To put it in the way that you and I think, this means there are 19 chances out of 20 that an observed difference is real rather than occurring by chance. Even more convincing is a P value of 0.001 which means that there are 999 chances out 1,000 that a difference is real. By convention, a P value of 0.05 or less is considered to represent statistical significance.

Sometimes, however, we do not measure changes in shades of grey, as in weights of apples, but a simple "yes" or 'no". For example, suppose there were 200 people with a serious infection. One hundred patients were given antibiotic A and 100 received antibiotic B. You can then count how many people in each group were cured and how many died. Look at Table 1. This is called a "two-by-two" or "contingency" table. Clearly, in this case, antibiotic A was no better than antibiotic B.

But what about Table 2? Is antibiotic C better than antibiotic D or could this result have occurred by chance? Fortunately, statisticians are able to help again. They have special tests such as the $?^2$ (chi-squared) test or Fisher's exact test. Again, they are able to derive a "P" value which indicates the chances of a difference being real. In the

Table 1. Number of patients who lived or died after being treated with antibiotics A or B.

	Number cured	Number died
Antibiotic C	50	50
Antibiotic D	50	50

Table 2. Number of patients who lived or died after being treated with antibiotics C or D.

	Number cured	Number died
Antibiotic C	65	35
Antibiotic D	50	50

example cited above, P = 0.045; this means that it is probable that antibiotic C really is likely to be better than antibiotic D but there is still a 1 in 22 chance that it is no better than D.

Rather than get bogged down with statistics in the rest of this book, I will attempt to show you the data and simply indicate whether or not a significant difference was found. Sometimes results are shown in tabular form, but in most instances I have presented the result graphically.

How does one read graphs

Graphs are sometimes called charts. There are various forms of graphs such as bar graphs (histograms), line graphs and pie charts. Bar graphs may be illustrated in two- or three-dimensional formats. Most of the graphs used in this book are bar graphs. How does one read such a graph?. An example is shown in Figure 8 which illustrates the effect of yoga on asthma. A graph has two axes. Along the bottom line (the horizontal or "X" axis), the population being studied is divided into groups and these groups are labelled in the legend at the bottom. In this case, there are two groups, one given yoga, and the control group which received no intervention. Furthermore, each group is represented twice, one before starting yoga, and again after 54 months. The other important axis is the vertical axis (sometimes called the "Y" axis) along the left hand side. This usually quantifies something. In this case, it indicates the number of attacks of asthma per week. A graph such as this

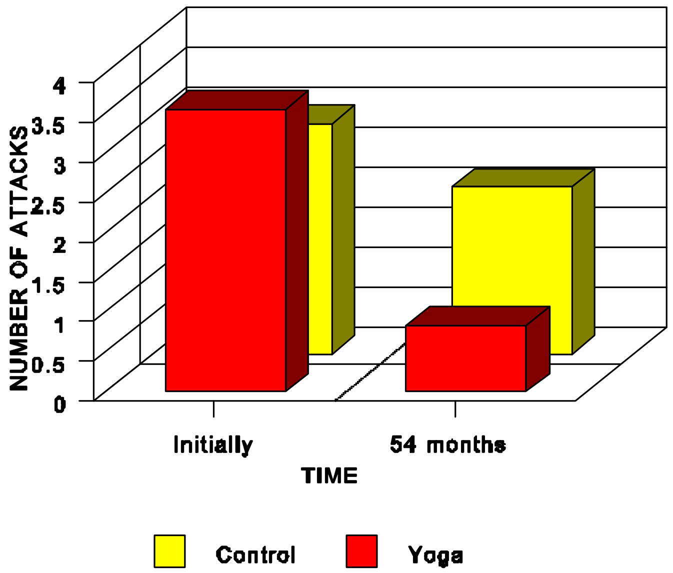

Figure 7. Number of attacks per week in patients who practised yoga or in a control group, initially and after 54 weeks of yoga.

can be drawn two-dimensionally or three-dimensionally. The latter perhaps looks a little more artistic, but the former is easier to read accurately.

The randomised, double-blind, placebo-controlled trial

With this background knowledge, we are now ready to examine the randomised double-blind placebo-controlled trial. Since World War II, such trials have become the cornerstone of the measurement of experiments in clinical medicine. What do these words mean?

What is a placebo?

A placebo has been defined as any treatment used for its non-specific effects.[3] This means that it does not work specifically against the disease for which it is used but generally tones up or stimulates the body's innate healing processes. Thus, the action and effectiveness of a placebo depends largely on subconscious interactions between the doctor, the patient and the treatment process. In most drug trials, the placebo looks just

[3]Oh VMS The placebo effect: can we use it better? *British Medical Journal* 309: 69-70, 1994

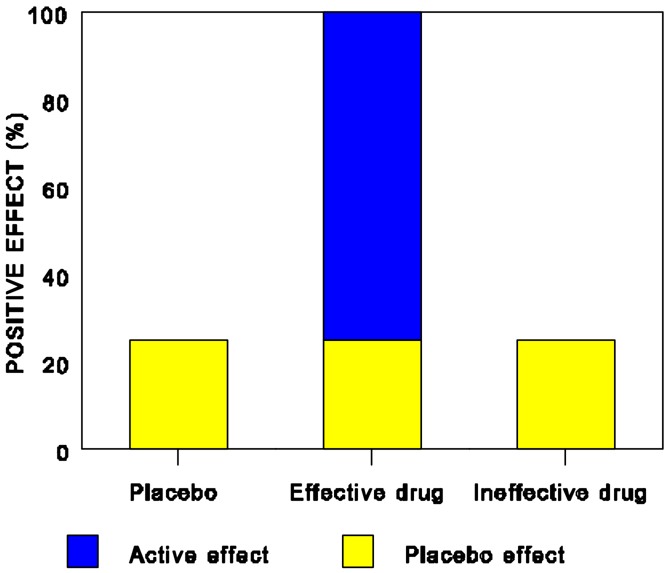

Figure 9. Interactions of placebo and specific actions in placebo, effective and ineffective drugs.

like the drug being tested but does not contain any active ingredient.

Placebos are very important in clinical trials as they allow the investigators to account for non-specific effects of treatment. Any drug or treatment has a placebo component and may or may not have an active component (Figure 9). All of this has two implications. Firstly, trials of a new treatment should include giving a placebo to a control group so that the placebo component of the drug being trialled is removed from the comparison and only its specific effects are left. This is called a placebo-controlled trial. Secondly, it has been questioned whether the effects of alternative medical treatments are only placebo effects.[4] The answer to this question, of course, is provided by setting up trials of alternative medicine which include, whenever possible, giving a placebo to the control group. Of course, the intrinsic placebo effects of both orthodox and alternative medicine at times be valuable simply by the reassurance they may give.[5]

---

[4]Lynöe N. Is the effect of alternative medical treatment only a placebo effect? *Scandinavian Journal of Social Medicine* 18: 149-153, 1990

[5]Kaptchuk TJ. The placebo effect in alternative medicine: can the performance of a healing ritual have clinical significance. *Annals of Internal Medicine* 136: 817-825, 2002

Why do trials need to be randomised?

When we have two groups of patients being investigated, we have to be very careful to ensure that we do not introduce bias. If we have 100 consecutive patients and give the test treatment to the first fifty and placebo to the second fifty (or *vice versa*), something might happen during the course of the trial that confounds the results. For example, the population being studied may change its characteristics or the investigators may subtly and unknowingly change the way they measure things.

It is far better to mix patients up so that some receive the treatment under investigation and others the placebo throughout the course of the trial. This could be done by allocating alternative patients to each group. Even this could introduce bias. Better is to randomly allot patients to each group by a pre-determined method such as a table of random numbers or numbers generated randomly by a computer.

What are single-blind and double-blind trials?

Even if we randomly allocate patients, bias can still creep in. The patients or the investigators might consciously or subconsciously alter their responses. This can be avoided by conducting a double-blind trial. This means that neither the patient nor the researcher knows which treatment is being given and a third party holds the code which is broken later. This is a technique commonly used in drug trials and can be employed in trials of homeopathy, for example.

Sometimes it is not possible to blind the investigator. Thus, a researcher would need to know whether he used true or sham acupuncture. This is called a single-blind trial in which only the patient does not know which treatment is being given. Sometimes this can be partially got around by dividing the researchers into two groups- one that administers the treatment and the other which assesses the patient without being aware of the nature of the treatment that had been given.

On other occasions, however, it is not possible to hide from either the investigators or the patients which group they are in.[6] In these instances, the trial is not blind at all.

Thus, the most powerful tool for analysing the effectiveness of any treatment, including those of alternative medicine, is the randomised double-blind placebo-controlled trial.[7] This sometimes raises ethical dilemmas when it is highly likely that a new drug is life-saving and it is with-held from one group of patients who are given either a placebo or the next best form of treatment.[8] This is most unlikely to be a problem when studying alternative

---

[6]Joyce CRB. Placebo and complementary medicine. *The Lancet* 344: 1279-1281, 1994

[7]Kleijnen J, de Craen JM, van Everdingen J, Krol L. Placebo effect in double-blind clinical trials: a review of interactions with medications. *The Lancet* 344: 1347-1349, 1994

[8]Hellman S, Hellman DS. Of mice but not men: problems of the randomized clinical trial. *New England Journal of Medicine* 324: 1585-1587, 1991

medicine as no alternative medical intervention is dramatically effective.

Alternative medicine in the light of modern science

In September 1841, the renowned British medical journal, *The Lancet*, observed in an editorial that there were two classes of individuals who bring discredit upon the medical profession:

> "The first may be described as the credulous ... who believe, upon the most shallow evidence, whatever is presented to them with the character of novelty or wonder... A little consideration will tend to the conclusion that credulity, that is incapacity to sift evidence in physic and religion, is the parent of absurdity and fanaticism. The second class includes those who have cleverness enough to detect the error or the preceding ... and despising the acquisition of knowledge, they start in life with a determination to prey upon the credulity of their fellow-men. These are your quacks of high and low degree."

Application of the scientific method has largely eliminated individuals such as these from within the ranks of the medical profession.

Should the same rules be applied to alternative medicine? Some have tried to argue that they should not. Two authors wrote that "the methods for obtaining knowledge in a healing art must be coherent with the art's understanding and theory of illness" and "insisted upon a more complete and coherent description and defense of the alternative epistemic methods and tools of these disciplines".[9] It is hard to know what these words mean. Such views have brought a torrent of criticism, with most commentators writing that the methods described in the chapter can and must be applied to alternative medicine.[10,11,12] In the view of the writer of this book, unless we wish to return to the mediaeval period, if not the Dark Ages, it is imperative that the same principles are applied to the practice and practitioners of alternative medicine as are required of medicine and the medical profession. To play upon the words in *The Doctors' Dilemma* of the famous playwright, George Bernard Shaw, we need to differentiate between:

Lies, damned lies and statistics

[9]Tonelli MR, Callahan TC. Why alternative medicine cannot be evidence-based. *Academic Medicine* 76: 1213-1220, 2001

[10]Bloom BS. What is this nonsense that complementary and alternative medicine is not amenable to controlled investigation with population effects? *Academic Medicine* 76: 1213-1220, 2001

[11]Hoffer LJ. Complementary or alternative medicine: the need for plausibility. *Canadian Medical Association Jourmal* 168: 180-182, 2003

[12]Miller FG, Emanuel EJ, Rosentein DL, Straus SE. Ethical issues concerning research in complementary and alternative medicine. *Journal of the American Medical Association* 291: 599-604, 2004

Last revision: 2004