Measurement error in the Hawksley random zero sphygmomanometer: what damage has been done and what can we learn?

Rónán M Conroy, Eoin O'Brien, Kevin O'Malley, Neil Atkins

The Hawksley random zero sphygmomanometer is used in all aspects of blood pressure research, from clinical trials to evaluation of new blood pressure recorders. It is designed to reduce observer bias in blood pressure measurement. The problem is that it also underestimates blood pressure. Furthermore, this was first reported more than two decades ago. In this paper Rónán Conroy and colleagues explore the consequences of using an inaccurate instrument for important research and why prestigious organisations like the World Health Organisation continue to use it.

But if the salt have lost his savour, wherewith shall it be salted?

MATTHEW V, 13.

We recently reported that the Hawksley random zero sphygmomanometer underestimated both systolic and diastolic blood pressure when compared with a mercury sphygmomanometer. Following the protocol of the British Hypertension Society, using large numbers of measurements with interchanged observers, we found that when measuring systolic pressure the Hawksley instrument read on average 3-5 mm Hg lower than the mercury instrument (95% confidence interval 3-4 to 4-2 mm Hg) and that when measuring diastolic pressure it read 7-5 mm Hg lower (7-0 to 8-0 mm Hg). These differences have serious implications for research results based on data gathered with the Hawksley instrument. What is more serious, however, is that doubts about the Hawksley random zero sphygmomanometer had been voiced as long ago as 1970, so that a large body of published research has subsequently accumulated whose interpretation may be in doubt.

Brief history of Hawksley random zero instrument

In 1963 Garrow described a “zero-muddler for unprejudiced sphygmanometry,” which was subsequently developed and marketed as the Hawksley random zero sphygmomanometer (Hawksley and Sons, Lancing, Sussex). The essential principle of the instrument is that it blinds the observer to the actual blood pressure until after the reading has been made. It does this by adding a random amount of mercury to the manometer as the reading is being taken. When the instrument falls back to zero the amount of extra mercury can be read and subtracted. The design is simple and ingenious. It really ought to work.

In 1970, almost as soon as the random zero sphygmomanometer became available, Evans and Prior reported that it read slightly but significantly lower than a mercury sphygmomanometer. This problem was also noted in 1973 in a paper reviewing the comparative performance of several blood pressure measuring devices. No concern seems to have been raised by these findings, however, and in the course of time the Hawksley instrument became popular in epidemiological research, favoured by such prestigious institutions as the Medical Research Council and the World Health Organisation.

In 1985 Gaudemaris et al reported a significant correlation between the zero level used by the Hawksley sphygmomanometer and the difference between the resulting blood pressure and the blood pressure obtained with a standard mercury instrument. In the same year Silman showed that although the Hawksley instrument reduced terminal digit preference, it did not eliminate it as the originators had claimed. These results were quickly corroborated by Hosie and Hosie. Three years later Parker et al published data suggesting that there was a serious problem with accuracy, and in 1990 a group of prominent American researchers concluded that the “random zero instrument does not function the way it is understood to” and called for a formal study of the Hawksley device.

That formal evaluation, carried out with both currently accepted protocols for evaluating blood pressure measuring devices, was published by us in 1990. It showed that the concerns expressed by other workers were amply justified: the Hawksley random zero sphygmomanometer systematically underestimates blood pressure. The study protocol was the most stringent evaluation to which the instrument had been subjected. The study was run with two machines—one British model, one American—selected from three of each type supplied by the manufacturers, who knew that they were required for formal evaluation. The machines had previously been compared to verify that they did not differ among themselves. As it was most unlikely that the manufacturers unwittingly supplied six similarly defective machines our findings must be taken as evidence of a generic fault.

Despite the accumulation of such comprehensive evidence against the Hawksley instrument, there is little sign that its popularity is waning. Two important questions arise: what are the consequences of the measurement error of the instrument for interpreting the findings of the many studies which have used it? and why has the instrument continued in use, unmodified, in the face of a growing and now definitive body of evidence that it is inaccurate?

Implications for published research

The most common uses of the Hawksley sphygmomanometer are in research directed at improving blood pressure recording protocols, in community surveys and other epidemiological studies, in defining patient populations for treatment trials in hypertension, and in assessing the performance of automated blood pressure measuring devices.

Improving blood pressure recording protocols

In studies aimed at improving protocols for blood pressure measurement use of the Hawksley sphygmomanometer is unlikely to have adverse effects. Typical studies in which the instrument was used include one by Whincup et al, which examined the effect of cuff size on blood pressure measurement in children; one by Canner et al, which examined the sources of variability in clinical blood pressure readings; and
would have erred on the low side, the study data were not have been affected by measurement bias. The conclusions reached would have been unbiased and thereby produced descriptive statistics which are biased. The situation is compounded by the fact that, although the MONICA protocol specifies a Hawksley sphygmomanometer, some countries that have not used mercury instruments. The use of a conventional mercury instrument rather than a Hawksley in measuring the blood pressure of an individual would therefore result in only a very small change in absolute risk, so it is unlikely that people are being seriously misinformed as a result of inaccuracies in logistic functions derived from blood pressure measured by the Hawksley.

DEFINING PATIENT POPULATIONS FOR TREATMENT TRIALS

The use of the Hawksley instrument for selecting patients to be included in treatment trials of antihypertensive agents has more serious consequences. Such trials generally recruit patients based on a minimum blood pressure level, such as the trial reported by Verdecchia et al,\(^1\) who compared duration of effect of two antihypertensive drugs. They used the Hawksley to recruit patients whose diastolic pressure was greater than 100 mm Hg. Other trials using the Hawksley, such as that reported by Kane et al have used a cut off point of 90 mm Hg. While the end points studied appear uninfluenced by the bias inherent in the use of the Hawksley, as they involved assessing and comparing changes in blood pressure, the use of the Hawksley introduces a selection bias of unknown effect. These studies were not, as reported, on patients with diastolic pressures of between 90 and 100 mm Hg, but other groups have reported an increase of 0.24% to 0.33% in relative risk.\(^14\) The use of a conventional mercury instrument would lead to a 0.4% increase in risk in blood pressure measured by Keys et al from data on middle aged men collected in the twin cities study,\(^19\) but other groups have both pressures over 100 and 90 mm Hg, respectively, but on patients with pressures over 107.5 and 97.5 mm Hg. As the treatment of mild hypertension, characterised by diastolic pressures of between 90 and 100 mm Hg continues to be the subject of debate, it is
unfortunate that several of the large scale studies designed to clarify the value of treatment in this group have been conducted on patients recruited on the basis of Hawksley sphygmomanometer blood pressure readings. We must face the prospect that the knowledge we thought we had acquired about the treatment of patients with diastolic pressures between 90 and 100 mm Hg actually relates to those with pressures between 97-7 and 107-5 mm Hg.

The cost of this loss of data is incalculable and some results may need extensive reanalysis. For example, the Australian therapeutic trial in mild hypertension used the Hawksley random zero sphygmomanometer and the London School of Hygiene sphygmomanometer interchangeably.46 Not only is there the likelihood of bad treatment in patients in the 90-100 mm Hg diastolic range wasted time and money by including the wrong patients but what we know of the benefits of treatment in this group may also be wrong. Guidelines for the treatment of mild hypertension may have been formulated on the basis of data gathered from people with moderate hypertension.

The primary prevention with metoprolol in patients with hypertension (MAPHY) study,67 for example, included 1209 patients over 20 years of follow-up. Patients were selected on the basis of diastolic pressure readings taken with a Hawksley random zero sphygmomanometer. Those included in the trial had diastolic pressures of 100-129 mm Hg but, had they been assessed with a standard sphygmomanometer, would have been in the range 107.5-136.5 mm Hg—clearly a more hypertensive group than intended.

The task of reviewing the methodology of large scale treatment trials in hypertension for selection bias associated with the use of the Hawksley instrument is one which bodies responsible for setting treatment guidelines should undertake without delay.

ASSESSING PERFORMANCE OF AUTOMATED DEVICES

The final and a commonly reported use of the Hawksley is in assessing the accuracy of automated blood pressure measuring devices. The Hawksley is used as the standard against which other techniques are assessed. Not surprisingly, given what we know now about the performance of the Hawksley, many such studies reported favourably on the automated instrument but criticised it for overestimating blood pressure. Rogers et al, for example, found that the Copal UA-231 read 2.3 mm Hg higher than the Hawksley, and the 1-9 mm Hg higher for diastolic pressure.68 Over-estimation of systolic pressure was reported by Maheswaran et al, who studied the Copal UA-251 and Dinamap 1848 machines.70 Fouqueray et al,71 who also studied agreement between the UA-251 and the Hawksley, found both systolic and diastolic pressures were higher when measured by the electronic instrument. Whincup et al, comparing a Dinamap 1846SX instrument with the Hawksley, reported that the Dinamap overestimated systolic pressure by a mean of 8.1 mm Hg. They also reported that differences between the two machines were related to blood pressure level, an effect which may be due to the relation of the inaccuracy of the Hawksley instrument to blood pressure level, and may not reflect a genuine inaccuracy in the Dinamap 1846SX.

A study reported by Ornstein et al is especially significant. They found that the Dinamap automated blood pressure measuring device seriously over-estimated blood pressure readings when compared with the Hawksley and concluded, “The results of this study would suggest that routine use of the Dinamap 8100 would lead to serious misclassification errors in screening for hypertension and in the follow-up of known hypertensive patients.” We do not know, on the basis of the data reported in the study, whether the Dinamap 8100 does, indeed, have a high rate of misclassification, but it should have been a result of the use of the Hawksley. To assess the consequences of using a Hawksley as a standard against which to assess automated instruments we reanalysed data gathered during the validation of the SpaceLabs 90202.72 We concluded that, validated against a Hawksley, the SpaceLabs 90202 would have achieved a British Hypertension Association grade of D for both systolic and diastolic pressure measurements—the lowest grade possible. In fact, assessed against a mercury instrument it achieved grades of C for systolic and B for diastolic pressure measurements (unpublished data). The potential damage done to commercial interests by using an inaccurate device as a standard should not be underestimated.

Automated blood pressure measurement is a growing and competitive science, and the use of the Hawksley as a standard is paradoxically giving an advantage to instruments whose bias coincides with that of the Hawksley. What we do not know is the extent to which manufacturers are using the Hawksley in the development and calibration of automated blood pressure 200 mmHg devices, leading to machines which will perform well in comparison with the Hawksley but which perpetuate the errors of the Hawksley. Not only is it not possible to estimate the cost of the studies whose results have been thrown into doubt by the finding that their standard was inappropriate but, in addition, the costs of these studies to manufacturers of what may well be excellent blood pressure measuring devices may only be guessed at.

Why is the Hawksley still in use?

When will we learn to extend the reasonable doubt that we have of the objectivity of the human observer to the accuracy of the instruments we use? It is remarkable that an instrument so widely used in the measurement of blood pressure was not itself subjected to assessment. If the loss of valuable data which we have described had happened as a result of a problem that no one knew about it would be tragic. But reports questioning the accuracy of the Hawksley data from 1970 and the accumulated evidence is now such that the use of the Hawksley as a blood pressure measuring instrument cannot be justified. Perhaps the vast body of data collected with use of the Hawksley is a deterrent to reanalysis—is it a psychological, emotional block at the WHO is good enough for me.”

Nevertheless, it verges on scandalous that the accumulating evidence has been ignored for two decades and vast sums in scarce medical research resources have been spent gathering data with an inaccurate instrument. And there is no sign that the popularity of the Hawksley is waning. Research protocols pass through the hands of many people, ethics committees, funding bodies, journal referees, and editors—and it is sadly apparent that nowhere in this chain of review is there anyone with the initiative to say stop. Where is John Snow now that we need him?
Seven years' experience in teaching communication skills to first year clinical students at St Mary's Hospital School of Medicine is described. The first component consists of a one-day introductory clinical course; this is divided into a lecture and small seminar groups and involves behavioural scientists and clinicians from many departments. The second component uses simulated patients and video feedback and takes place in small groups later in the year. Participation of the students through active critical discussion, role play, and interactive video feedback are important aspects in the success of the course. The methods have been refined through evaluation by students and tutors. This article aims to allow others, already running or planning similar courses, to develop effective courses within the practical constraints of their own institutions.

Communication skills are now widely acknowledged as having a central role in clinical practice. The General Medical Council has stated that communication skills are fundamental to patient care and include history taking, involving patients in decision making, and giving treatment, advice, support, and counselling. Most essential diagnostic information arises from the interview, and good communication increases patient's knowledge, satisfaction, and compliance and positively influences health. Poor communication has been implicated in medical accidents and in subsequent litigation. Complaints about doctors by the public usually do not deal with medical competence but with problems of communication.

Teaching of communication seems to be both feasible and effective. Although some early research findings have been criticized, other studies have clearly shown that training, medical students' communication skills seem to decline during medical training. Without specific training, medical students' communication skills seem to decline during medical training. Teaching of communication skills has been slow to develop. In 1983 Wakeford found that one third of British medical schools offered no communication skills teaching, and training in the others typically amounted to only one or two hours of video recording and replay. By the time of the 1992 GMC survey there was some improvement; only three out of 28 medical schools did not teach any communication skills, although it was usually taught as part of general practice or psychiatry, with only four schools including teaching within general medicine and surgery. In 1991 Whitehouse reported that teaching of communication skills accounted for less than 2% of curriculum time. Methods varied from formal lectures to analysis of student interviews. Video feedback was a part of only some courses, and students seldom had the opportunity to experience more than one or two sessions of feedback. Role play was commonly used in group work but the use of simulated patients was rare. Assessment was usually subjective and given only by the doctor or tutor. The GMC survey found that only seven schools included communication skills as a formal part of final examinations. In 1992 the University of London stated that newly qualified doctors should be competent in communication skills and have received formal training in the subject.

St Mary's Hospital Medical School has provided general communication skills training to first year students.

How to Do It

Teaching communication skills to clinical students

I C McManus, C A Vincent, S Thom, J Kidd

Seven years' experience in teaching communication skills to first year clinical students at St Mary's Hospital School of Medicine is described. The first component consists of a one-day introductory clinical course; this is divided into a lecture and small seminar groups and involves behavioural scientists and clinicians from many departments. The second component uses simulated patients and video feedback and takes place in small groups later in the year. Participation of the students through active critical discussion, role play, and interactive video feedback are important aspects in the success of the course. The methods have been refined through evaluation by students and tutors. This article aims to allow others, already running or considering such a course, to develop effective courses within the practical constraints of their own institutions.

Communication skills are now widely acknowledged as having a central role in clinical practice. The General Medical Council has stated that communication skills are fundamental to patient care and include history taking, involving patients in decision making, and giving treatment, advice, support, and counselling. Most essential diagnostic information arises from the interview, and good communication increases patient's knowledge, satisfaction, and compliance and positively influences health. Poor communication has been implicated in medical accidents and in subsequent litigation. Complaints about doctors by the public usually do not deal with medical competence but with problems of communication.

Teaching of communication seems to be both feasible and effective. Although some early research findings have been criticized, other studies have clearly shown that training, medical students' communication skills seem to decline during medical training. Without specific training, medical students' communication skills seem to decline during medical training. Teaching of communication skills has been slow to develop. In 1983 Wakeford found that one third of British medical schools offered no communication skills teaching, and training in the others typically amounted to only one or two hours of video recording and replay. By the time of the 1992 GMC survey there was some improvement; only three out of 28 medical schools did not teach any communication skills, although it was usually taught as part of general practice or psychiatry, with only four schools including teaching within general medicine and surgery. In 1991 Whitehouse reported that teaching of communication skills accounted for less than 2% of curriculum time. Methods varied from formal lectures to analysis of student interviews. Video feedback was a part of only some courses, and students seldom had the opportunity to experience more than one or two sessions of feedback. Role play was commonly used in group work but the use of simulated patients was rare. Assessment was usually subjective and given only by the doctor or tutor. The GMC survey found that only seven schools included communication skills as a formal part of final examinations. In 1992 the University of London stated that newly qualified doctors should be competent in communication skills and have received formal training in the subject.

St Mary's Hospital Medical School has provided general communication skills training to first year students.