Metrology in Medicine

M.J. Turner[*], P.C. Kam[#], A.B. Baker[*]

* Department of Anaesthetics, University of Sydney,
# Department of Anaesthetics, Emergency Medicine & Intensive Care, University of New South Wales.

Abstract

Modern medicine employs a wide variety of physiological measurements made by instruments whose performance is difficult to verify by inspection. In Australia, medical devices are assessed by the TGA before they are licensed, but ongoing maintenance and calibration are not controlled.

Common characteristics of most medical measurements include wide inter- and intra-individual variability and large uncertainties (up to 20%). Different measurement techniques can yield results that differ widely. However, diagnoses rarely depend on a single measurement and experienced clinicians assign less weight to suspect measurements. All medical diagnostic and treatment procedures have associated risks. Clinicians are accustomed to the combination of variability and risk and many do not perceive lack of traceable calibration as a serious problem. Very few medical measurement systems are traceably calibrated.

One diagnosis that does depend on a single measurement is hypertension. Twenty-nine percent of adult Australians are hypertensive. Only one in 54 UK GP practices regularly calibrates their sphygmomanometers, and one in ten sphygmomanometers is in error by more than 5 mm Hg. We found no evidence that the situation is different in Australia. Overestimation of blood pressure (BP) by three mm Hg can almost double the number of patients diagnosed as hypertensive. AS3551 (1996) (Technical management programs for medical devices) suggests that test equipment should be *regularly calibrated … to achieve traceability*. We suggest that a system that ensures that medical measurement systems are traceably calibrated would improve the quality of health care and reduce overall costs to society in the long term.

Keywords: physiological measurements, blood pressure, spirometry, temperature, medical devices.

1. Introduction

Most medical diagnoses and treatments are based on a description of symptoms by the patient, clinical examination and measurements of one or more of an ever-widening variety of physiological variables. Physiological measurements are sometimes indirect and often exhibit high inter- and intra-individual variability. In many cases high measurement uncertainties (up to 20%) are acceptable for routine clinical work. Good clinicians usually disregard unusual or apparently erroneous measurements in the presence of conflicting evidence from other sources.

Historically, medical devices were simple and it was easy for experienced practitioners to detect malfunction or inadequate performance. Modern physiological measurement systems are becoming more sophisticated and the spectrum of measurements is broadening. Medical practitioners increasingly rely on quantitative measurements for the early detection of disease, and for diagnosis and treatment. It is becoming more difficult to detect malfunction or measurement error in modern electronic instruments.

A degree of risk is associated with most medical investigations and treatment. Medical practitioners are accustomed to the combination of variability and risk and many do not see a need for traceable calibration of medical instruments. We review the status of medical metrology in Australia and, as a case study, investigate the sensitivity of the diagnosis of hypertension to systematic errors in blood pressure measurement.

2. Laws, regulations, standards and accreditation

No medical device may be used in routine clinical practice in Australia unless it is registered by the Therapeutic Goods Authority (TGA). TGA regulations [1] require that the traceability of the calibration of instruments used in the manufacture and evaluation of medical devices be documented. The same regulations require that measurements given by medical devices with a 'measuring function':

- 'must be compared to at least one point of reference indicated in Australian legal units of measurement',
- must 'be accurate to enable the device to achieve its intended purpose', and
- 'must be designed and produced in a way that ensures that the device provides accurate, precise and stable measurements within the

limits indicated by the manufacturer and having regard to the intended purpose of the device'.

TGA regulations provide no guidance concerning the methods by which these objective should achieved.

Australian standard AS/NZS 3551 [2] specifies 'procedures required to develop equipment management programs for medical devices'. This standard requires that 'test equipment used shall be regularly calibrated … to achieve traceability of measurement'. This standard is not compulsory and the requirement for traceable calibration of test equipment is ignored in some hospitals. Some day-surgery hospitals are ISO 9001 certified and consequently are required to have all the instruments used to assess and control the quality of their work traceably calibrated [3]. Most medical pathology and biochemistry laboratories in Australia are classed as Medical Testing Laboratories and are accredited to ISO 17025 by NATA.

Many professional bodies publish guidelines relevant to their field [4]. Some professional medical societies accredit laboratories in their field and certify technicians. For example, the Thoracic Society of Australia and New Zealand provides an accreditation service for lung function laboratories, and the Australian & New Zealand Society Of Respiratory Science certifies respiratory function scientists.

Once a medical device is registered with the TGA, there appears to be no law or regulation requiring the device to be adequately maintained and traceably calibrated during its service lifetime. Maintenance and calibration programmes appear to be at the discretion of each institution or medical practitioner.

3. Cost-benefit and evidence

The additional costs of traceable calibration would be highly visible to individual medical practitioners and healthcare administrators, while potential benefits such as improved quality of life and long-term patient outcomes are very difficult to quantify in dollar terms. At present traceable calibration is not viewed as a substantial or potential cause of litigation [5]. In our experience, many medical measurement systems in Australia are not traceably calibrated.

Evidence-based medicine (EBM) is the 'conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients' [6]. Medical practitioners who practice EBM base their decisions on statistically proven evidence that a proposed treatment is better than an alternative or not treating at all. At present there is very little quantitative evidence in the medical literature that the use of uncalibrated or poorly calibrated measurement systems leads to adverse outcomes. A recent study of preventable adverse events in Australian hospitals did not consider preventable measurement error as a potential source of adverse outcomes [7]. The attention of healthcare administrators and practitioners would be drawn to traceable calibration if strong evidence proves that it leads to significantly improved outcomes and/or reduced overall costs or if the threat of litigation increased [8]. As a first step in the process of creating awareness of calibration issues, we analysed the sensitivity of the diagnosis of hypertension to errors in blood pressure measurements.

4. Case study: non-invasive blood pressure measurement.

Blood pressure (BP) is possibly the most common physiological measurement made by medical practitioners. Hypertension (high blood pressure) is a well-known independent risk factor for coronary heart disease and cardiovascular disease. Cardiovascular disease is the leading cause of death among Australians (42% of all deaths in 1996). Hypertension is one of the few diseases that is diagnosed by measurement of a single variable. A recent Australian study reported that 29% of Australians aged 25 years and older are hypertensive (≥140/90 mm Hg) or on anti-hypertensive medication [9]. A 1996 UK survey found 23% of adults to be hypertensive (SBP≥160 mmHg or DBP≥95 mmHg) or on antihypertensive drugs [10].

BP measurement is known to be prone to error and guidelines are available for the assessment of sphygmomanometers [11] and for the measurement process itself [12]. The Australian standard for sphygmomanometers requires that the static error due to the pressure sensor alone be less than ±3 mm Hg [13]. Current European guidelines for automatic sphygmomanometers allow the use of instruments that indicate pressures that are in error by 10 mm Hg in 20 % of measurements [11]. None of the guidelines recommend traceable calibration of sphygmomanometers.

In establishing guidelines for the calibration of sphygmomanometers it would be helpful if the sensitivity of diagnosis of hypertension to measurement error were known. Campbell and MacKay [14] suggest, based on a Canadian study [15], that systematic overestimation of diastolic pressure by five mm Hg more than doubles the number of patients classified as hypertensive, while a similar underestimation reduces the number by 62%. We analysed the results of a recent survey to estimate the sensitivity of hypertension diagnosis to systematic error in BP measurement.

Methods

Distributions of diastolic and systolic blood pressures of groups of male and female adults aged 16 to 75 years were obtained from a recently

published survey [15]. Cumulative distributions indicating the percentage of subjects who had BPs greater than the abscissa value were calculated by summing the published distribution data and subtracting the sums from 100. Second order polynomials were fitted to the logarithms of the cumulative distributions by linear regression, yielding the following equation:

$$p(x) = 10^{ax^2+bx+c}$$

where: *p(x)* is the percentage of subjects with diastolic or systolic BP greater than *x*, and *a*, *b* and *c* are the coefficients of the polynomial.

We assume that a subject is classified as hypertensive if his/her diastolic or systolic pressure exceeds x_o mm Hg. The likelihood that a subject selected at random from the survey group has a pressure greater than x_o is

$$p(x_o) = 10^{ax_o^2+bx_o+c} \quad (1)$$

If BP is over-estimated by Δx the likelihood of a randomly selected subject having an *estimated* BP greater than x_o is

$$p(x_o - \Delta x) = 10^{a(x_o-\Delta x)^2+b(x_o-\Delta x)+c} \quad (2)$$

Use of a sphygmomanometer that over-reads by Δx results in the proportion of subjects classified as hypertensive increasing by a factor *f* given by:

$$f = \frac{p(x_o - \Delta x)}{p(x_o)} \quad (3)$$

Substituting equations (1) and (2) into (3) and simplifying yields:

$$f = 10^{a(\Delta x)^2 - 2ax_o\Delta x - b\Delta x}$$

The percentage (*P*) by which the number of subjects diagnosed as hypertensive increases when BP is overestimated by Δx is given by:

$$P = 100(f - 1)$$

If the polynomial that describes the logarithm of the cumulative distribution is 2nd order or higher, *P* is a function of both the threshold (x_o) and the measurement error (Δx). If, however, a straight line fits the logarithm of the cumulative distribution well then the parameter *a* is zero and *P* is independent of the threshold, depending only on the measurement error.

The variance of *P* $\left(u_p^2\right)$ due to lack of fit of the polynomial to the logarithm of the cumulative distribution was estimated by using the propagation of error equation [16]

$$u_P^2 = u_a^2\left(\frac{\partial P}{\partial a}\right)^2 + u_b^2\left(\frac{\partial P}{\partial b}\right)^2 + 2u_{ab}^2\left(\frac{\partial P}{\partial a}\right)\left(\frac{\partial P}{\partial b}\right)$$

where: u_a^2 and u_b^2 are the variances of parameters *a* and *b* respectively

u_{ab}^2 is the covariance between parameters *a* and *b*.

$\left(\frac{\partial P}{\partial a}\right)$ is the sensitivity of *P* to uncertainty in *a*

$\left(\frac{\partial P}{\partial b}\right)$ is the sensitivity of *P* to uncertainty in *b*.

u_a^2 u_b^2 and u_{ab}^2 were estimated from the covariance matrix [17] of the regression.

The residuals of the regression were assumed to be normally distributed. We estimated the confidence limits for *P* as $P \pm t(\nu, 1-\alpha/2)u_P$ where *t* is the cumulative Student's *t* distribution, α = 0.05 and ν is the degrees of freedom of u_P. Data analysis was performed using software written in Matlab (Mathworks Nattick, USA).

Results

The cumulative distributions of diastolic and systolic data are shown in Figure 1 on semi-log axes. The diastolic distribution is described by a 2nd order polynomial, while a straight line fits the systolic distribution well. Figures 2 and 3 show the estimated change in hypertension diagnoses that would result from systematic errors in BP measurements. The width of each line represents the 95% confidence intervals associated with lack of fit of the polynomials to the logarithm of the cumulative distributions. Table 1 indicates the percentage change in the number of patients whose BP would exceed index values in a clinical practice using a sphygmomanometer that exhibits systematic measurement errors of 1, 3 and 5 mm Hg.

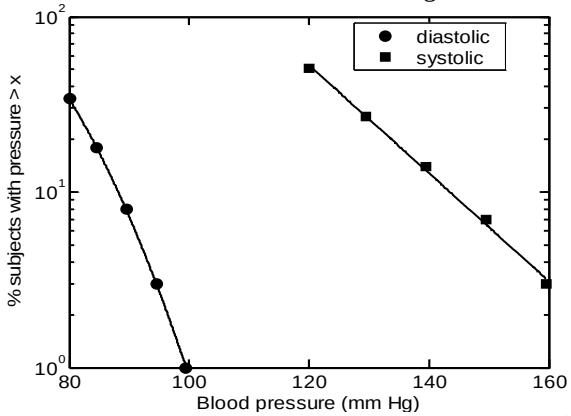

Figure 1. Cumulative distributions showing the percentage of subjects with BP greater than the abscissa. The regression lines are shown.

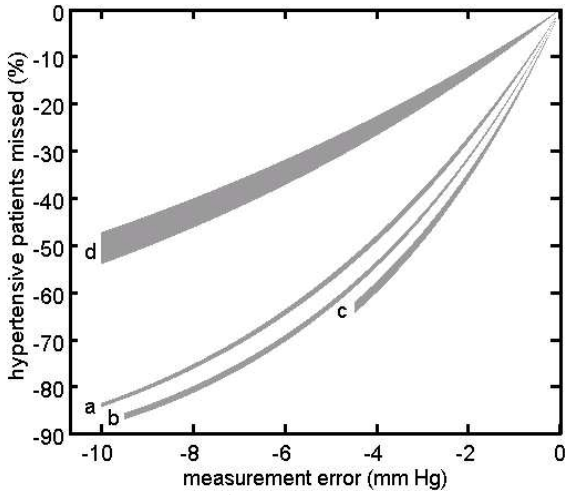

Figure 2. The number of hypertensive subjects who would be missed if BP were under-estimated, expressed as a percentage of the number correctly classified. The widths of each line represent the 95% CI associated with lack of fit of the regressions. a,b,c: diastolic with thresholds of 85, 90, 95 mm Hg respectively. d: systolic (independent of threshold).

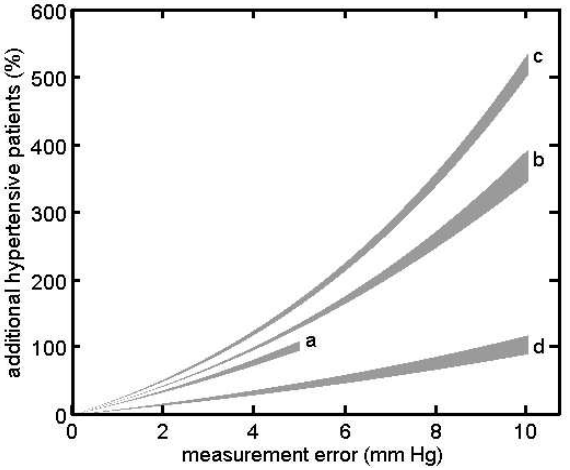

Figure 3. Additional number of patients who would be classified as hypertensive if BP were over-estimated, expressed as a percentage of those correctly classified. The widths of each line represent the 95% CI associated with lack of fit of the regressions. See Figure 2 for label definitions.

Table 1. Change in the number of patients classified as hypertensive due to systematic measurement error. Changes are expressed as a percentage of the number that would be classified in the absence of measurement error.

Systematic error (mm Hg)	Diastolic			Systolic*
	85†	**90**	**95**	**–**
–5	–57(1)‡	–62(1)	–67(1)	–30(2)
–3	–39(1)	–44(1)	–48(1)	–19(2)
–1	–15(0·4)	–17(0·2)	–19(0·5)	–7(1)
+1	16(1)	20(0.3)	23(1)	7(1)
+3	55(3)	68(1)	83(2)	24(3)
+5	102(7)	132(4)	166(5)	43(5)

* Systolic results are independent of threshold.
† Bold numbers in the second row are thresholds (mm Hg).
‡ The approximate 95% confidence range is in parentheses.

5. Other physiological measurements

Many other physiological measurements are made every day in hospitals and medical practitioners' rooms in Australia. We briefly review temperature measurement and spirometry as further examples of a much wider problem.

Infra-red (IR) aural canal thermometers have replaced mercury-in-glass thermometers for measuring body temperature. Imamura et al. [18] compared four commonly used tympanic IR thermometers with simultaneous calibrated thermocouple measurements in ten adults undergoing hyperthermia (32ºC) during cardiac surgery. Mean biases and precision (mean and standard deviation of the paired differences between IR and reference thermocouple) were 0.0(0.8)ºC, -0.1(0.8)ºC, 1.1(0.9)ºC and 2.3(0.8)ºC. A core temperature 1–2ºC above normal is often interpreted as indicating an infection, which may be benign, but may be life-threatening, e.g. septicaemia.

Over 2.2 million people in Australia have asthma. In 2001, 422 people died from asthma (National Asthma Council figures). Asthma is usually diagnosed from a medical history, clinical examination and spirometry before and after challenge tests. Most spirometers available in Australia are type-tested in accordance with American Thoracic Society guidelines [4] using a set of forced expiratory flow waveforms generated by a computer-controlled servo-driven syringe. Gas compression within the syringe can severely distort the output waveforms [19,20]. Most routine spirometer calibration is performed using hand-driven syringes. Many syringes are not regularly calibrated, and there is evidence to suggest that while *'the results from volume calibration may meet ATS criteria, … this is no guarantee that data from forced manoeuvres are accurate'* [21].

6. Discussion

Blood pressure measurement case study

This case study shows that the diagnosis of hypertension in a large number of patients is extremely sensitive to systematic errors in the measurement of blood pressure. Lack of calibration or untraceable calibration of sphygmomanometers is likely to result in systematic errors in BP readings. A medical practitioner using a sphygmomanometer that consistently overestimates diastolic BP by three mm Hg will experience an 83% increase in the number of patients with diastolic pressure exceeding 95 mm Hg (Table 1, Fig 3). For every five patients correctly diagnosed with hypertension, another four would be incorrectly diagnosed as hypertensive. If BP is consistently under-estimated by three mm Hg, almost half the patients with true diastolic BP greater than 95 mm Hg would be missed (Table 1, Fig 2).

Similar errors in systolic pressure measurements result in smaller changes in diagnoses.

Rouse and Marshall [5] found that 3·4% of 1462 sphygmomanometers (949 mercury, 513 aneroid) in English general practices over-read and 5·9% under-read by five mm Hg or more. A recent UK study [22] found that 28% of mercury and 42% of aneroid sphygmomanometers were in error by more than four mm Hg. Only one in 54 GP practices in the UK has an arrangement for the maintenance and calibration of sphygmomanometers [5]. We have no evidence to suggest that the situation is any better in Australia. Consistently over-estimating BP by five mm Hg may result in up to five patients being incorrectly assigned diastolic BP over 95 mm Hg for every three patients correctly diagnosed. Consistently under-estimating BP by five mm Hg may result in two-thirds of patients with true diastolic BP over 95 mm Hg being missed. If systematic errors were limited to ±1 mm Hg, one patient in five would be classified incorrectly as hypertensive or one hypertensive patient in five missed, at a 95 mm Hg threshold. Lower thresholds result in lower sensitivity to measurement error.

The sensitivity of the diagnosis of hypertension to errors in the measurement of BP depends on the slope of the cumulative distribution of BP. Random error in the survey BP readings widens the BP distribution and reduces the magnitude of the slope of the cumulative distribution, hence the sensitivity to error may be under-estimated in this study.

In the survey some subjects were undergoing treatment for hypertension. Successful treatment would shift subjects to the left in the BP distributions, possibly causing the magnitude of the slopes of the cumulative distributions to increase and thereby increase the apparent sensitivity to measurement errors. Nevertheless the survey data represent estimates of the BP of the population that medical practitioners see, and decisions regarding the initiation, continuation, discontinuation, reduction or increase in treatment are based on similar measurements.

This case study suggests that a tolerance of three mm Hg static pressure is too wide for reliable diagnosis of diastolic hypertension. Diastolic BP should be measured with an uncertainty of one mm Hg to avoid misclassifying too many subjects as normotensive or hypertensive. An uncertainty of three mm Hg may be adequate for detecting systolic hypertension.

If we assume that the systematic errors in BP measurements in Australia are similar to those reported by Rouse and Marshall, then we can deduce that a significant number of Australians takes anti-hypertensive drugs unnecessarily, and a similar number has untreated hypertension. Treating patients unnecessarily is costly and may be associated with adverse effects caused by unnecessary administration of medication. Not treating hypertensive patients results in increased incidences of cardiovascular disease (including strokes), coronary heart disease, and damage to the kidneys and eyes. Cardiovascular diseases cost Australia $3.9 billion in 1993/4 [23]. A study of the performance of sphygmomanometers in Australia would allow Monte-Carlo simulation to be used to estimate the cost in lives and dollars of not calibrating sphygmomanometers.

Other physiological measurements

The evidence suggests that avoidable and significant errors may occur in many other physiological measurements that are commonly made in hospitals and medical practices. The sensitivity to errors in spirometry and in the measurement of temperature has not been studied.

7. Conclusions

Uncalibrated and inadequately calibrated medical instruments are an unrecognised cause of preventable medical errors. We suggest that a system for ensuring that medical measurement systems are traceably calibrated would improve the quality of health care and reduce overall costs to society in the long term.

8. References

1. Therapeutic Goods (Medical Devices) Regulations 2002 Statutory Rules 2002 No. 236 as amended.
2. AS/NZS 3551:1996:Technical management programs for medical devices. Standards Australia.
3. AS/NZS ISO 9001:2000. Quality management systems– Requirements. Standards Australia.
4. American Thoracic Society guidelines. Standardization of Spirometry 1994 Update. Am J Respir Crit Care Med 1995 152:1107-36.
5. Rouse A and Marshall T. The extent and implications of sphygmomanometer calibration error in primary care. J Hum Hyperten 2001 15:587-591
6. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 312(7023):71-2.
7. Wilson RMcL, Harrison BT, Gibberd RW, Hamilton JD. An analysis of the causes of adverse events from the Quality in Australian Health Care Study. Med J Aust 1999 170:411-415.
8. Marshall T, Rouse A. Blood pressure measurement: Doctors who cannot calibrate sphygmomanometers should stop taking blood pressures BMJ 2001 323(7316):806.

9. Australian Institute of Health and Welfare (AIHW) 2002. Epidemic of coronary heart disease and its treatment in Australia. AIHW cat. No. CVD 21 Canberra: AIHW (Cardiovascular Disease Series No. 20).
10. Department of Health. Health Survey for England '96. Volume 1: Findings. Stationery Office: London. 1998.
11. O'Brien E et al. Working Group on Blood Pressure Monitoring of the European Society of Hypertension. Working Group on Blood Pressure Monitoring of the European Society of Hypertension International Protocol for validation of blood pressure measuring devices in adults. Blood Press Monit 2002 7:3-17.
12. O'Brien E et al., on behalf of the European Society of Hypertension Working Group on Blood Pressure Monitoring. European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement. J Hypertension 2003 21:821–848.
13. AS EN 1060.1. 2002: Non-invasive sphygmomanometers - General requirements. Standards Australia.
14. Campbell N.R.C. and McKay D.W. Accurate blood pressure measurement: Why does it matter? Can Med Assoc J 1999 161:277–278.
15. Joffres MR, Hamet P, Rabkin SW, Gelskey D, Hogan K, Fodor G. Prevalence, control and awareness of high blood pressure among Canadian adults. Can Med Assoc J 1992 146:1997–2005.
16. Bevington PR and Robinson DK. Data reduction and analysis for the physical sciences. McGraw-Hill Singapore 1994 p43.
17. Draper N and Smith H. Applied regression analysis 2nd Ed. New York: Wiley, 1981 p89.
18. Imamura M, Matsukawa T, Ozaki M, Sessler DI, Nishiyama T, Kumazawa T. The accuracy and precision of four infrared aural canal thermometers during cardiac surgery. Acta Anaesthesiol Scand 1998 42:1222-6.
19. Reynolds et al. Nonlinear model for offline correction of pulmonary waveform generators. IEEE Trans BME 2002 49(12 Pt 2):1567-73.
20. Miller et al. Peak expiratory flow profiles delivered by pump systems - limitations due to wave action. Am J Respir & Crit Care Med 2002 162:1887-1896.
21. van den Boom G. van der Star LM. Folgering H. van Schayck CP. van Weel C. Volume calibration alone may be misleading. Respiratory Medicine. 1999 93(9):643-7.
22. Waugh JJ, Gupta M, Rushbrook J, Halligan A, Shennan AH. Hidden errors of aneroid sphygmomanometers. Blood Press Monit 2002 7:309-12 2002.
23. Mathers C & Penm R. Health system costs of cardiovascular diseases and diabetes in Australia 1993-94. AIHW cat. no. HWE 11. 1999. Canberra: Australian Institute of Health and Welfare (Health and Welfare Expenditure Series no. 5).