Editorials

The interpretation of uncontrolled before-and-after studies: as demonstrated by recent studies of the introduction of medical emergency teams

The recent publication of the MERIT study, which explored the treatment effect of the medical emergency team (MET), has presumably generated considerable interest within the critical care community. The MERIT investigators recorded the "notable limitations of previous studies...the use of historical controls and the absence of randomisation" and suggested that the MERIT study "...should have provided a reliable estimate of the treatment effect of a MET system". However, the published results of the MERIT study would not appear to have afforded substantial support for arguments proffered for the introduction of METs. As the MERIT investigators reported: "The MET system greatly increases emergency team calling, but does not substantially affect the incidence of cardiac arrest, unplanned ICU admissions, or unexpected death". A previous cohort comparison (1996) study by Bristow et al (MET in hospital 1 and conventional cardiac arrest teams in hospitals 2 and 3) had observed fewer unanticipated intensive care (ICU)/high dependency unit (HDU) admissions in the MET hospital, but no significant difference in the rates of cardiac arrests or total deaths between the three trial hospitals. Explanations offered by Bristow et al for the null findings were lack of sensitivity of calling criteria, the irreversibility of pathophysiological processes despite call-out, and MET underutilisation. These two studies stand in contrast to a number of non-controlled "before-after" studies which have shown "favourable" effects of the introduction of the MET, although this has not been consistent. How are we to understand this?

Before-after MET studies with no concurrent randomised control groups

Buist et al in a non-randomised population based study, before (1996) and after (1999) the introduction of the MET, reported a reduction (45.6%) in the incidence of "unexpected cardiac arrest" and subsequent mortality. Overall hospital mortality also declined (12.6%), although, discounted for post "cardiac arrest deaths", the death rate at the two time periods was similar (1996: 16.77 per 1000 patients versus 16.06 in 1999). A difference in the types of admissions between the two periods was also recorded and adjusted for in analysis: "We could control for all major characteristics associated with the hospital...". The definition of cardiac arrest in this study was "unusual". We adopted a working definition of cardiac arrest-namely, that a staff member was so concerned about a patient that they made a cardiac arrest call, regardless of whether the patient was actually having a cardiac arrest. That the incidence of "unexpected cardiac arrest calls" and subsequent mortality fell between the two periods should not necessarily be taken at face-value:

i) the widespread "education and audit process" in 1999 could presumably have resulted, as the authors conceded, in "improved ward management of patients with clinical instability" and/or a heightened awareness of such instability, leading to an overall increase use of the cardiac arrest/MET, as found also in the MERIT study. The call out rate of the 1996 "cardiac arrest team" was not given nor was information about the resuscitation status of the patients; such would be of interest as cardiac arrest teams may operate as METs in some environments; the MERIT study recording 50% of calls to cardiac arrest teams in control hospitals not for a cardiac arrest or unexpected death.

ii) by virtue of using an operational definition of "cardiac arrest", there was potential for misclassification bias, although there was apparently "no significant difference in the actual reasons for the call between the two periods". It is thus difficult to accept that the "early intervention by a medical team reduced the incidence of unexpected cardiac arrest in hospital by about half"; rather, as canvassed by the authors, the effect may have been due to "a system of support, ongoing education, and performance feedback". It is also hard to reconcile the dismissal, by the authors, of the alternative explanation for the reduction in "cardiac arrest calls"; that is a "...natural regression" due to medical progress, or at worst, random fluctuation". The decline in cardiac arrest and in-hospital mortality both began before the "gradual" introduction of the MET in 1997 (Figure 1, page 2). The "formal" education and audit process, begun in 1999, was associated with an increase in the MET call-out, but no subsequent decline (1999 to 2000) in the rate of cardiac arrest or hospital mortality.

As previously noted, "natural regression due to medical progress" does confound the interpretation of
improvements in mortality outcomes due to uncontrolled interventions; in particular:
i) the Cleveland Health Quality Choice program, which noted improved mortality subsequent to the introduction of an initiative to profile hospital performance. This improvement over time was shown to exist in the rest of the state of Ohio where such profiling did not occur and
ii) the improved mortality of the New York Cardiac Surgery Program's initiatives were mirrored in regions where state-wide outcome reporting, introduced by the Program, was not operative. The MERIT investigators also canvassed "...increased awareness of patient safety, causing systematic changes in delivery of health care...".1

Bellomo et al.,6 conducted a prospective "before-after trial of a medical emergency team" comparing mortality outcomes in two 4 month periods, separated by a preparation and education and a run-in period of the MET (1999 - 2001). Again, a decrease in the number of cardiac arrests and deaths from cardiac arrests was noted, "before" versus "after", and inpatient deaths also declined (risk ratio (RR) 0.74; 95%CI: 0.70 - 0.79). This latter decrease was also evident when inpatient deaths were discounted for the number of deaths due to cardiac arrest (RR 0.78; 95% confidence intervals [CI]: 0.65 - 0.94). Unlike the studies of Buist et al. and Bristow et al.,4 no formal adjustment for case-mix was reported; for the "Major surgical procedures" there was a significant difference in the distribution of the categories, "before" versus "after" (χ² test: P = 0.006), but no detail of major medical categories (for example, disease related groups [DRGs]) was provided. The authors provided data on patient and procedure numbers with 95% CI, but the overlap method (of 95% CI) is known to be conservative and lack power compared with standard methods, and 83% CI have been recommended for assessing whether or not two means are significantly different from one another at α = 0.05. In a subsequent Reply-to-the-Editor,19 the notion was reiterated that "...introducing the MET was effective, not the MET per se", which sentiment appeared consistent with that of Buist et al.6 above.

DeVita et al., in a retrospective analysis over a period of 6.8 years (1996-2002), found a decrease in the incidence of cardiopulmonary arrests (6.5 to 5.4 per 1000 admissions), but no change (defined as death during the calendar day of the arrest) "before and after the increased use of MET". An increase in the hospital case mix severity was recorded over this time. Of interest, there also appeared to be no change in the proportion of fatal arrests in the Buist et al (P = 0.017, significance set by authors at 0.001) and Bellomo et al studies (P = 0.31).6 Ball et al.,6 using historical controls, reported an increased survival to hospital discharge of ICU discharged patients (RR 1.08; 95% CI: 1.00 - 1.18) over a two year period (2000 - 2002) with the introduction of a critical care outreach team. The above studies stand in contrast to a one year study (2000 - 2001), using historical controls, where the reduction in the cardiac arrest rate and overall mortality did not achieve statistical significance.10

Problems with before-after studies
Besides the obvious lack of concurrent randomised controls, what were the problems which may have confounded the interpretation of these "before-after" studies? Firstly, as noted above, there may be a "natural" change in the outcome of interest over time, which may also be paralleled by, or be a function of, quality of care. Secondly, the Hawthorne effect12 of change in outcome(s) resultant upon "studying an intervention".11 In other medical disciplines, such effects have been noted.22-23 as have means for controlling the Hawthorne effect in randomised controlled trials.24 Thirdly, regression to the mean,25-26 which is a function of sampling variability and measurement error (or reliability). Although traditionally described with respect to continuous outcomes,26 binary events (for example, mortality) may also demonstrate regression to the mean.21-13 Thus clinicians should be mindful of the cautions of Greimeder et al.: "In before-and-after studies such as those typically conducted during continuous quality improvement projects, which typically do not have control groups, investigators need to consider control group effects when they assess results of intervention.... A substantial proportion of the reductions reported ....may be caused by the Hawthorne (placebo) effect ... or regression to the mean."26 This is not to say that the authors of the above four studies did not consider alternative explanations for the observed treatment effects; what we are concerned with is the interpretation of these effects compared with those of the two studies where control groups were utilised.

The MERIT study
Despite its size and sophisticated design, the MERIT study was unable to demonstrate an effect, treatment versus control, in any of the 4 tabulated outcome rates (Table 3, page 2095; P ≥ 0.31). The most illuminating result of the MERIT study, from the perspective of the positive results of the comparator "before-after" MET studies, was the substantial (up to 37%) and statistically significant decreases (P ≤ 0.03) in outcome rates in the control hospital group over the study period (see Table 4, page 2095: "Control hospitals"). For instance, the control hospitals sustained decreases in cardiac arrest rates from a baseline of 2.61 to 1.64 per 1000 admissions. This baseline rate was somewhat less than
that recorded in the “before” time periods of Bristow et al (3.3 to 4.8 per 1000; 1996 admissions); Buist et al (2.90 per 1000 admissions; 1996)8 and Bellomo et al (2.98 per 1000 admissions; 1999),12 although the modest time-trend of these study baseline rates (1996 to the MERIT study) was non-significant (P = 0.2). That control hospital outcome reductions of this magnitude could occur over a relatively short period (6 months), compared with the long study periods (2 to 6 years) of the comparator trials, suggests that the MET effect “per se” may remain elusive. The MERIT investigators noted that “The fact that both control and MET hospitals improved their adverse outcome rate during the study could have reduced our ability to record a positive treatment effect”.13 However, this begs the question of why, despite increased MET call-out, outcome events in the MET hospitals did not decrease as anticipated. The relatively short time course of events was also alluded to in comparison with the efficacy of trauma systems which “…have taken up to 10 years before any effect on mortality has been detected”. That effectiveness is not necessarily time-lagged is illustrated by the experience of the Major Trauma Services in South Australia (1997-2000),32 where mortality decreased within 1 year of the introduction of the State Trauma System in 1997 (statistically significant by 1998), although, ironically, similar cautions apply to the interpretation of the success of this initiative as with the MET system.

One aspect of the uncertainty of the effect of METs is that of the sensitivity, or lack thereof, of call-out criteria, as mentioned by Bristow et al, above. All MET studies used a fairly common constellation of clinical signs as call-out criteria, which were derived from, or have been subject to, observational study.36-41 In only two of these studies37,38 were the clinical signs subject to formal predictive analysis using logistic regression and there was no use of cross-validation. The point estimates (as odds ratios [OR]) for individual clinical predictors from these two studies suggested substantial associated “risk”, with OR ranging from 2.4-14.4,41 or in combination, 6.3-37.2.38 However, CI were extremely wide with upper OR limits of 8041 and 300,38 indicating substantial estimate uncertainty, due to “perfect” prediction of small numbers of events and/or multicollinearity.44,45 Of interest, 67% of abnormal observations resolved spontaneously in the Buist et al study.43 Thus, the introduction of the MET system may increase call-out rates, but not necessarily influence the outcome course of (irreversible) underlying pathophysiological processes.

Overview and conclusions

It is difficult to believe that the “comprehensive educational strategy” associated with the introduction of a system such as the MET does not influence process of care at the general hospital ward level, directly or by a Hawthorne-like effect (“contamination” in the trial setting).1 In the current climate, where “Health care policy is intensely focused on patient safety and medical error, evidence-based practice, and quality indicators”,46 the MET system appears to be “…a rational and reasonable change that may improve patient care, and is unlikely to make things worse”.1 However, an inappropriately resourced “administrative” solution to the perception that patients are dying “while we wait for evidence”,2 may indeed have an unintended adverse effect; both in respect of acute and non-acute care. In the United Kingdom, the “Outreach” program was reported to have been funded at £142 million.15

The MERIT investigators suggested that the MERIT study would be able to “assist the design and management of other such studies”,1 but as “ineffectiveness” is hard to prove and accept47 and, as doctors appear unable to stand idle and do nothing,48 “the motivation for opposition to the change…(to MET)…” has been called into question.3 It is indeed true that “Absence of evidence is not evidence of absence”,49 under conditions of uncertainty, confidence intervals may illuminate this “absence of evidence”.50 However, in the case of the MERIT trial, the 95% CI were “…consistent with anything between large positive effect and a large negative effect”. If level I evidence may never become available, should we accept “cash for no questions”,51 that is “…get on with reducing errors rather than worry about describing the problem?52 To do so would presumably entail abandoning the scientific requirement for rigorous uncertainty.52 As reminded by Vincent et al: “It is hard to think of any other major public health problem for which this argument would be taken seriously.”51

Dr. J. L. Moran
Department of Intensive Care Medicine,
The Queen Elizabeth Hospital, Woodville
SOUTH AUSTRALIA 5011

Associate Professor P. J. Solomon
School of Applied Mathematics,
University of Adelaide, Adelaide
SOUTH AUSTRALIA 5005

REFERENCES


Clough JD, Engler D, Canuto P E. Cleveland Health Quality Choice was a failure, not a martyr. Qual Saf Health Care 2002;1:391.


Goldhill DR, White S A, Sumner A. Physiological values and procedures in the 24 hours before ICU admission from the ward. Anaesthesia 1999;54:529-534.


"You see but you do not observe, Watson"

An oration delivered by Dr. G. M. Clarke on the occasion of his receiving the inaugural Joint Faculty of Intensive Care (JFCIM) Medal at the JFCIM ASM held in Sydney, June 2005

As mentioned in the citation given by Dr. van Heerden, I was the inaugural Dean of the original Faculty of Intensive Care of ANZCA. This prompted a former chairman of examinations, Richard Lee, to comment: "Old deans never die, they just lose their faculties!" It seems to me unfortunate that the first part of this statement is pretty unlikely. As to the second part, I am sure you will all be in a position to judge it's veracity within a short time from the beginning of this talk.

The formation of the Joint Faculty of Intensive Care Medicine of ANZCA and the RACP has been a further step forward in the continuing evolution of solidly based Intensive Care medical training, certification and practice in Australia and New Zealand. I encourage Fellows who have wondered about helping in Faculty and College affairs to offer their services. Faculty and College life is far from boring. You meet wonderful people, many of extremely high intellect; there is camaraderie, humour, excitement and even danger!

Mind you, exceptional intellect and an offbeat sense of humour are probably not the best combination of qualities in a person who is to represent the Faculty on a Government committee. We appointed such a person to a Victorian Government committee. As I respect his privacy, I will only tell you that his initials are GP and he works at Monash. During the first meeting he departed prematurely, never to return again. During his departure it is rumoured that he stated "I do not know why you are all wasting your time discussing the meaning of life, the universe and everything, when everyone knows the answer is 42!"

What of excitement and danger?

On the occasion of a formal dinner at ANZCA I arrived resplendent, and I thought probably magnificent, in my dinner suit. I had bought that suit when I was seventeen years old with money hard earned on the wheat bins of rural Western Australia. Joan Sheales, the ultimate boss of ANZCA, on spotting me said, "Geoff, I want you to follow me up to my office." I immediately had this dreadful sense of foreboding, like when the headmaster tells you to report to his study, now! (As a schoolboy, for reasons which were quite beyond me at the time, this was not an infrequent experience). Joan and I entered the office and she closed the door behind me. She then moved behind her desk, opened the left hand top drawer and produced the biggest pair of glittering scissors I had ever seen. Good G-d, I thought, you could both mules and dock a ram with those. Never in my life, I thought, have I ever done anything to deserve this. Joan then descended upon me, scissors open. To my amazement and relief she used the scissors to cut away all the bits of frayed cotton, silk and other fluff that were dangling from the lapels of my old drinking (correction "dinner") jacket. On returning home my beloved dinner suit went into the bin, much to Susan's delight and I went out and bought a new one.

The quote from Conan Doyle's Sherlock Holmes, "You see but you do not observe, Watson," refers to the fact that although Watson may have seen something, he often did not observe the finer detail and thus its significance. To some extent this may apply to all of us in our every day lives.

My interest tonight is in exploring what others may see and or observe from our actions as specialists in Intensive Care Medicine. This includes the way we manage patients, deal with their relatives and interact with other health care workers. My biggest concern is that, when busy or stressed, we are tempted to take short