Epidemiology series

Sample size slippages in randomised trials: exclusions and the lost and wayward

Kenneth F Schulz, David A Grimes

Proper randomisation means little if investigators cannot include all randomised participants in the primary analysis. Participants might ignore follow-up, leave town, or take aspartame when instructed to take aspirin. Exclusions before randomisation do not bias the treatment comparison, but they can hurt generalisability. Eligibility criteria for a trial should be clear, specific, and applied before randomisation. Readers should assess whether any of the criteria make the trial sample atypical or unrepresentative of the people in which they are interested. In principle, assessment of exclusions after randomisation is simple: none are allowed. For the primary analysis, all participants enrolled should be included and analysed as part of the original group assigned (an intent-to-treat analysis). In reality, however, losses frequently occur. Investigators should, therefore, commit adequate resources to develop and implement procedures to maximise retention of participants. Moreover, researchers should provide clear, explicit information on the progress of all randomised participants through the trial by use of, for instance, a trial profile. Investigators can also do secondary analyses on, for instance, per-protocol or as-treated participants. Such analyses should be described as secondary and non-randomised comparisons. Mishandling of exclusions causes serious methodological difficulties. Unfortunately, some explanations for mishandling exclusions intuitively appeal to readers, disguising the seriousness of the issues. Creative mismanagement of exclusions can undermine trial validity.

Proper randomisation[1,2] means little if investigators cannot include all randomly assigned participants in their primary analysis. Hence, a crucial aspect of assessing a randomised controlled trial pertains to exclusions, withdrawals, losses, and protocol deviations. How should investigators handle participants who refuse entry, ignore follow-up, leave town, or take aspartame when they were instructed to take aspirin? Unfortunately, many inappropriate approaches to dealing with these types of problem actually seem logical and falsely appealing. Therein lies their insidious nature, because such inappropriate approaches can result in serious biases. Here, we address the effect of exclusions made before and after randomisation.

Exclusions before randomisation

Investigators can exclude participants before randomisation. The eventual randomised treatment comparison will remain unbiased (good internal validity), irrespective of whether researchers have well-founded or whimsical reasons for exclusion of particular individuals. However, exclusions at this stage can hurt extrapolation, the generalisability, of the results (external validity). For most investigations, we therefore recommend that eligibility criteria be kept to a minimum, in the spirit of the large and simple trial.[3,4] However, some valid reasons exist for exclusion of certain participants. Individuals could, for example, have a condition for which an intervention is contraindicated, or they could be judged likely to be lost to follow-up. The trial question should guide the approach.[5] Sometimes, however, investigators impose so many eligibility criteria that their trial infers to a population of little apparent interest to anyone, and, in addition, recruitment becomes difficult. If investigators exclude too many participants, or the wrong participants, their results might not represent the people of interest, even though the randomised controlled trial might have been meticulously done—ie, the results could be true but potentially irrelevant.

What to look for in exclusions before randomisation

The eligibility criteria should indicate the population to which the investigators wish to infer. When judging the results of a trial, readers should make sure that the eligibility criteria are clear and specific. Most importantly, the criteria should have been applied before randomisation. Readers should also assess whether any of the criteria make the study sample atypical, unrepresentative, or irrelevant to the people of interest. In practice, however, results from a trial will infrequently be totally irrelevant: "most differences between our patients and those in trials tend to be quantitative (they have different ages or social classes or different degrees of risk of the outcome event or of responsiveness to therapy) rather than qualitative (total absence of responsiveness or no risk of the event)."[6] Such qualitative differences in response are rare; thus, trials tend to have rather robust external validity.[6]

Exclusions after randomisation

Exclusions made after randomisation threaten to bias treatment comparisons. Randomisation itself configures unbiased comparison groups at baseline. Any erosion, however, over the course of the trial from those initially unbiased groups produces bias, unless, of course, that erosion is random, which is unlikely. Consequently, for the primary analysis, methodologists suggest that results for all patients who are randomly assigned should be analysed, and, furthermore, should be analysed as part of the group to which they were initially assigned.[3,7] Trialists refer to such an approach as an intent-to-treat analysis. Simply put: once randomised, always analysed as assigned.

Intent-to-treat principles underlie the primary analysis in a randomised controlled trial to avoid biases associated with non-random loss of participants.[8–10] Investigators can

Lancet 2002; **359:** 781–85

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA (K F Schulz PhD, D A Grimes MD)

Correspondence to: Dr Kenneth F Schulz
(e-mail: KSchulz@fhi.org)

also do secondary analyses, preferably preplanned, based on only those participants, for example, who fully comply with the trial protocol (per protocol) or who receive the treatment irrespective of randomised assignment (on-treatment or as-treated). Secondary analyses are acceptable as long as researchers label them as secondary and non-randomised comparisons. Trouble brews, however, when investigators exclude participants and, in effect, present a secondary, non-randomised comparison as the primary randomised comparison from a trial. In reality, this analysis represents a cohort study masquerading as a randomised controlled trial. Exclusion of participants from an analysis can lead to misleading conclusions (panel 1).[11–14]

Researchers often do not provide adequate information on excluded participants.[7,15,16] Furthermore, in a review of 249 randomised controlled trials published in major general medical journals in 1997, only 2% (five of 249) of reports explicitly stated that all randomly assigned participants were analysed according to the randomised group assignment.[17] About half of the reports (119 of 249) noted an intent-to-treat analysis, but many provided no details to support this claim.

Additionally, researchers frequently do not report anything with respect to exclusions.[7] Left in this information void, many readers deduce that certain trials used intent-to-treat principles and had no exclusions. We call this scenario no apparent exclusions. Readers commonly view trials with no apparent exclusions as less biased, when in fact unreported exclusions probably occurred in many of them. Indeed, trials with no apparent exclusions were methodologically weaker than those reporting at least some exclusions.[7] In other words, some of the more biased trials might be mistakenly interpreted as unbiased, and many of the less biased trials as biased; we call this inconsistency the exclusion paradox. Until researchers comprehensively report exclusions after randomisation, readers should be aware of this unsettling irony.

What to look for in exclusions after randomisation

Before we launch into attributes of proper handling of exclusions after randomisation, we should acknowledge the tenuous ground on which any such discussion rests. Reporting on exclusions is poor, with the exclusion paradox misleading readers. Investigators should provide clear, explicit information on the progress through the trial of all randomised participants, and when such information is absent, readers should be sceptical. The flow diagrams specified in the CONSORT statement provide appropriate guidelines.[18,19]

Optimally, of course, investigators would have no exclusions after randomisation and use an intent-to-treat analysis. Assessment of exclusions after randomisation is simple: none are allowed. All participants enrolled should be analysed as part of the original group assigned. Clinical research is not normally that simple, but the principle holds. One pragmatic hint for minimising exclusions after randomisation involves randomly assigning individuals at the last possible moment. If randomisation takes place when the participant is first identified, but before treatment is initiated, then any exclusions arising before treatment still become exclusions after randomisation. Investigators can address this potential difficulty by delaying randomisation until immediately before treatment begins.[20]

If investigators report exclusions after randomisation, those exclusions should be carefully scrutinised because they could bias comparisons. Exclusions arise after randomisation for several reasons, including discovery of patient ineligibility, postrandomisation-pretreatment outcome, deviation from protocol, and losses to follow-up.

Discovery of participant ineligibility

In some trials, participants are enrolled and later discovered not to have met the eligibility criteria. Exclusions at this point could seriously bias the results, since discovery is probably not random. For example, participants least responsive to treatment or who have side-effects might draw more attention and, therefore, might be more likely to be judged ineligible than other study participants. Alternatively, a physician who had treatment preferences for certain participants might withdraw individuals from the trial if they were randomly assigned to what he believes to be the wrong group.

Participants discovered to be ineligible should remain in the trial. An exception could be made if establishment of eligibility criteria is difficult. In such instances, investigators could obtain the same information from each patient at time of randomisation and have it centrally

Panel 1: **A randomised controlled trial of sulfinpyrazone versus placebo for prevention of repeat myocardial infarction**

For this trial, the researchers reported a primary analysis that compared rates of death from cardiac causes rather than from all cardiac deaths.[11,12] In their analysis, inappropriate exclusions due to eventual discovery of patient ineligibility caused a problem:[13] the investigators withdrew as ineligible seven patients who had received treatment—six in the treatment group and one in the placebo group—resulting in more patients who died being withdrawn from the treatment group than from the placebo group.

Moreover, results of a detailed audit of this trial by the US Food and Drugs Administration (FDA) indicate that additional patients from the placebo group could have been declared ineligible on the basis of similar criteria, but were not.[13] Furthermore, the trial protocol did not mention exclusion of ineligible patients after entry, particularly patients who had died. The researchers also excluded two deaths in the sulfinpyrazone group and one death in the placebo group as non-analysable because of poor compliance. However, the trial protocol did not include plans to exclude patients because of poor compliance.

Additionally, the investigators used a 7-day rule. They declared as non-analysable any death of a patient who had not received treatment for at least 7 days or who died more than 7 days after termination of treatment. The FDA review committee did not criticise this practice strongly, principally because the protocol described the 7-day rule, and also because the rule had little overall effect on the results.

Overall, these inappropriate exclusions did, however, affect the results of the study.[13] Although the researchers initially reported a 32% reduction (p=0·058) in rates of death from cardiac causes for participants who took the drug, a reanalysis showed a weaker result. When individuals judged ineligible or non-analysable were included in the originally assigned groups, the reduction was only 21% (p=0·16). It is noteworthy that only p values were provided. We urge the use of confidence intervals in reporting results.[14]

Moreover, the fallout from inappropriate exclusions, as ascertained by the FDA, cast doubt over the trial. The FDA advisory committee announced that sulfinpyrazone could not be labelled and advertised as a drug to prevent death in the critical months after a heart attack because, on close examination, the data were not as convincing as they seemed at first glance.

reviewed by an outside source, blinded to the assigned treatment. That source, whether a person or group, could then withdraw patients who did not satisfy the eligibility criteria, presumably in an unbiased way.

Postrandomisation, pretreatment outcome

Researchers sometimes report exclusion of participants on the basis of outcomes that happen before treatment has begun or before the treatment could have had an effect. For example, in a clinical trial of a specific drug's effect on death rates, investigators withdrew as non-analysable data on all patients who died after randomisation but before treatment began or before they had received at least 7 days of treatment. This winnowing seems intuitively attractive, because none of the deaths can then be attributable to treatment. But the same argument could be made for excluding data on all patients in a placebo group who died during the entire study interval, because, theoretically, none of these deaths could have been related to treatment. This example illustrates the potential for capriciousness in addressing postrandomisation, pretreatment outcomes.

Randomisation tends to balance the non-attributable deaths in the long run. Any tinkering after randomisation, even if done in the most scientific and impartial manner, cannot improve upon that attribute, but can hurt it. More importantly, this meddling sometimes serves as a post hoc rationale for inappropriate exclusions.

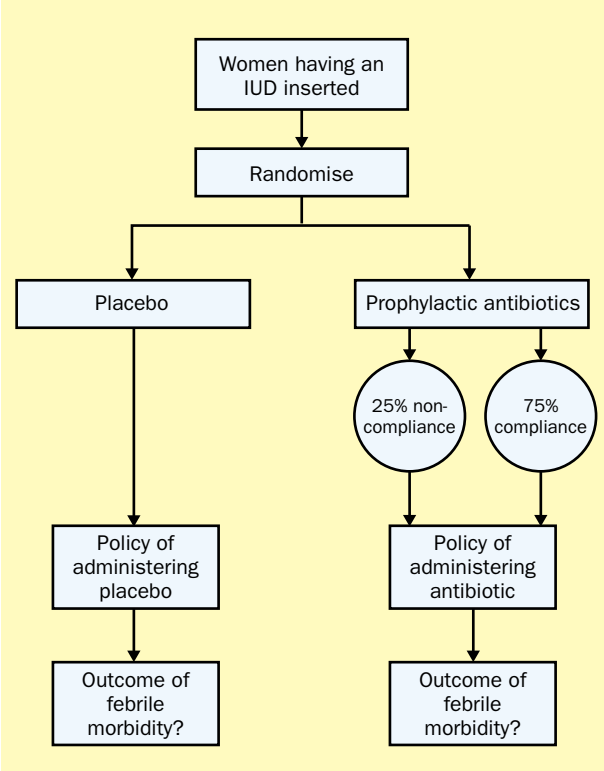
Post hoc rationalisation arises when investigators observe the results and then frame rules that favour their hypotheses. Assume that an investigator postulates that a drug used for treatment reduced the death rate associated with a particular condition. After analysis of data, however, the investigator notes that 14 deaths in the treatment group and two deaths in the placebo group arose before treatment had begun or before the drug had been taken for at least 7 days. She then rationalises the deaths as unrelated to treatment, and withdraws them from analysis. Such a response would seriously bias her results, even though her reasoning in the report would likely seem logical.

Imposed a priori, such rules only complicate trial implementation; imposed a posteriori, they lead to biased and invalid results. In assessment of randomised controlled trials, identification of when researchers stipulated rules usually proves impossible. We prefer to find, in reports of randomised controlled trials, that investigators did not allow any withdrawal of participants after randomisation. The data of all randomised patients should be analysed. Planned or unplanned, the exclusion of non-analysable outcomes on grounds of efficiency is not a generally accepted practice in the analysis of a randomised clinical trial.[21]

Protocol deviations

Deviations from assigned treatment happen in many trials. Some investigators suggest that participants who deviate substantially from the allotted treatment should be excluded in the final analysis, or should be included only up to the point of deviation. Although this approach seems attractive, it has a serious flaw: "the group which deviates from one protocol and the group which deviates from the other protocol may be so different [. . .] that the treatment comparison in the remaining patients will be severely biased."[3]

For example, suppose investigators want to know if prophylactic antibiotics reduce febrile morbidity associated with insertion of an intrauterine device (IUD). Investigators randomly allocate participants to receive antibiotics or placebo (figure). Unfortunately, 25% of the patients in the antibiotic group deviate from the protocol and do not take their antibiotics. In effect, these deviates receive the same treatment—that is, nothing—as the placebo group. Should the investigators exclude them from analysis? Alternatively, should investigators merge them with the placebo group and compare them with the compliant patients in the antibiotic group who adhered to the protocol? Some investigators opt for one of these speciously attractive options.



Schematic of randomised IUD patients, accounting for their compliance with treatment during the trial
IUD=intrauterine device.

For the primary analysis, however, neither option proves acceptable. The two treatment groups would no longer be comparable. The participants who did not take antibiotics might have been in better health or might have better tolerated insertion of the IUD. In either instance, they were probably less susceptible to febrile morbidity. If investigators exclude the deviates, the antibiotic group will contain only the more susceptible: the treatment comparison would be biased. If investigators include the deviates in the placebo group, then not only will those left in the antibiotic group be more susceptible to febrile morbidity, but the placebo group will have been infiltrated with less susceptible patients: the treatment comparison would be even more biased. Those who deviated could be sicker rather than healthier—it does not matter. The point remains that the treatment comparison would be systematically biased.

All protocol deviations should be followed up, and their data should be analysed with the group to which they were originally assigned. In our example, the deviates from the antibiotic group should remain with the antibiotic group. Similarly, any deviates in the placebo group should remain in that group. Despite what happened during the course of the trial, investigators should compare the group randomly allocated to antibiotics with the group allocated

to placebo. This approach, in addition to being unbiased, will provide a pragmatic answer to the question of primary clinical interest—eg, does the policy of giving prophylactic antibiotics for IUD insertion prevent febrile morbidity? Thus, if researchers report excluding protocol deviates, or if they report moving protocol deviates from one group to another group, the resultant treatment comparison should be considered biased, analogous to an observational study.

Loss to follow-up

Losses to follow-up are perhaps the most vexing of the proffered reasons for exclusions after randomisation. Participants might move or might refuse to continue participating in the trial. Participants lost to follow-up could still be included in the analysis if outcome information could be obtained from another source, such as gathering data from a national death registry. Such opportunities, however, rarely arise. Without outcomes from those lost to follow-up, investigators have little choice but to exclude them from the analysis. Any losses damage internal validity, but differential rates of loss among comparison groups cause major damage. Hence, investigators must minimise their losses to follow-up.

Minimisation of loss in some trials exudes difficulties. Investigators should commit adequate attention and resources to develop and implement procedures to minimise losses.[10] For example, investigators might exclude patients before randomisation if deemed likely to be lost to follow-up. Alternatively, they could obtain contact information to locate lost participants or hire special follow-up personnel who visit unresponsive participants, or both.

Some investigators add innovative twists that cultivate high follow-up rates. One approach uses a large number of conveniently placed follow-up clinics. Too often investigators expect participants to visit a single, inconvenient location. Shortening the data collection instrument to a manageable size caters to the participants' wishes and needs. Investigators foster follow-up by not overburdening participants. Such instruments might not only promote higher follow-up rates, but might also engender higher quality data on the main items of interest. Elimination of loss completely could be impossible, but investigators too frequently profess insurmountable difficulties. Many investigators could work harder than they do to obtain higher follow-up rates (panel 2).

What is an acceptable rate of loss to follow-up? Only one answer, 0%, ensures the benefits of randomisation. Obviously, this is unrealistic at times. Some researchers suggest a simple five-and-20 rule of thumb, with fewer than 5% loss probably leading to little bias, greater than 20% loss potentially posing serious threats to validity, and in-between levels leading to intermediate levels of problems.[22] Indeed, in their experience with sensitivity analyses, use of the worst case scenario, they opine, and we agree, that a trial would be unlikely to successfully withstand challenges to its validity with losses of more than 20%.[6] Indeed, some journals refuse to publish trials with losses greater than 20%.[6]

Although the five-and-20 rule is useful, it can oversimplify the problem in situations with infrequent outcomes.[22] Expectations for loss to follow-up depend on various factors, such as the topic examined, the outcome event rate, and the length of follow-up. For example, if researchers examined outcomes during the first day after birth to women delivering in hospitals, we would expect no losses. If the researchers examined use of microbicides by women in Africa (who usually have no phones and sometimes lack street addresses) to prevent HIV-1 transmission over a 1-year follow-up period, however, we would expect perhaps 5–15% loss to follow-up, although hoping for lower. Actually, most investigators have done much worse under such circumstances, but recent exhaustive efforts have yielded loss to follow-up rates of about 1·5%.[23] Another useful general rule of thumb suggests not allowing the loss to follow-up rate to exceed the outcome event rate.

Perhaps more important than the absolute overall loss to follow-up rate is the comparative loss rates in the groups. Researchers should analyse the data for differential rates of loss in the groups. Bias could arise when losses are related to differences in unpleasantness, toxicity, or efficacy of the treatments. In any case, investigators should have recorded and analysed the outcomes from those participants lost, at least up to the point of loss.

Panel 2: **Approaches to maximisation of participant follow-up**

Hire a person to manage and encourage follow-up

Hire personnel to call participants or to visit participants at their homes or place of work, if participants are not returning for follow-up

Exclude before randomisation those likely to be unwilling to return

Exclude before randomisation those likely to move

Obtain contact information to prompt participants to return for follow-up and to facilitate location of participants if they do not return—eg, mail, telephone, and e-mail for enrolled participants, for close friends or relatives who do not live with the participant, and for the participant's family doctor

Obtain an identification number, such as a national health-care number

Establish follow-up venues suited to participants rather than to investigators and trial implementers—eg, more locations than just the central clinic or hospital, close to where participants live, convenient to access, and sensitive to waiting time

Streamline trial procedures to move participants quickly through a follow-up visit

Keep data collection instrument short so as to not overburden the participant

Provide excellent and free medical care

Provide monetary subsidies, primarily for time and travel costs incurred by participants

Conclusion

Trialists should endeavour to minimise exclusions after randomisation and to do intent-to-treat analyses. They should also follow the CONSORT statement for reporting.[18,19] The flow diagram (trial profile) helps particularly to track the progress of participants through a trial.

For readers, non-reporting of exclusions results in interpretation difficulties, such as the exclusion paradox, which misleads readers about trial quality. Moreover, mishandling of exclusions causes serious methodological difficulties. Unfortunately, some explanations provided in reports for such difficulties intuitively appeal to readers, which disguises the seriousness of the issues. Readers must battle both inadequate reporting and their intuition to discover potential threats to validity.

We thank Willard Cates and David L Sackett for their helpful comments on an earlier version of this report. Much of the material stems from our 15 years of teaching the Berlex Foundation Faculty Development Course.

References

1 Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002; **359:** 515–19.
2 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002; **359:** 614–18.
3 Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, I: introduction and design. *Br J Cancer* 1976; **34:** 585–612.
4 Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, II: analysis and examples. *Br J Cancer* 1977; **35:** 1–39.
5 Sackett DL. On some prerequisites for a successful clinical trial. In: Shapiro SH, Louis TA, eds. Clinical trials: issues and approaches. New York: Marcel Dekker, 1983: 65–79.
6 Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingstone, 2000.
7 Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996; **312:** 742–44.
8 Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991; **10:** 1595–605.
9 Lewis JA, Machin D. Intention to treat: who should use ITT? *Br J Cancer* 1993; **68:** 647–50.
10 Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000; **21:** 167–89.
11 The Anturane Reinfarction Trial Research Group. Sulfinpyrazone in the prevention of cardiac death after myocardial infarction. *N Engl J Med* 1978; **298:** 289–95.
12 The Anturane Reinfarction Trial Research Group. Sulfinpyrazone in the prevention of sudden death after myocardial infarction. *N Engl J Med* 1980; **302:** 250–56.
13 Temple R, Pledger GW. The FDA's critique of the anturane reinfarction trial. *N Engl J Med* 1980; **303:** 1488–92.
14 Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2002; **359:** 57–61.
15 DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; **306:** 1332–37.
16 Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Control Clin Trials* 1984; **5:** 328–47.
17 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; **319:** 670–74.
18 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports or parallel-group trials. *Lancet* 2001; **357:** 1191–94.
19 Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; **134:** 663–94.
20 Friedman L, Furberg C, DeMets D. Fundamentals of clinical trials. St Louis: Mosby, 1996.
21 Meier P. Anturane reinfarction trial. *N Engl J Med* 1981; **304:** 730.
22 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone, 1997.
23 Roddy RE, Zekeng L, Ryan KA, Tamoufé U, Tweedy KG. The effect of Nonoxynol-9 gel on urogenital gonorrhea and chlamydial infection: randomized controlled trial. *JAMA* (in press).

Uses of error

No problem

Paolo Gallinaro

25 years ago, I started another orthopaedic night shift on duty. My colleague, finishing his, told me that a young man with a pelvic fracture had been admitted an hour or so earlier. "There is no problem with him", he said, "his blood pressure is 120/80 mm Hg, and he has an intravenous line with saline running in." Radiographs showed an open book pelvis: a fractured anterior ring and slight disruption of the sacroiliac joints. Immediate action did not seem necessary, so I took care of other patients with minor injuries. Only later did I have a look at the "no problem" patient. Although his blood pressure had not changed, he was pale and his extremities were cold. Clearly, he had haemorrhagic shock. I called a general surgeon. He suspected an arterial retroperitoneal bleed, and decided to operate. Opening the large retroperitoneal haematoma just made the bleeding worse—and he found no major source. The patient died. Patients with similar injuries still die as a result of similar mistakes, but we have since changed the management, and prognosis, of closed pelvic fractures by using aggressive fluid resuscitation, stabilisation, and intensive care.

Dipartimento di Traumatologia, Ortopedia e Medicina del Lavoro, Universita' di Torino, 10126 Torino, Italy (Prof P Gallinaro MD)