EVENTS

From the Therapy chapter for the 3rd edition of Clinical Epidemiology, by DL Sackett
17 April 2004 (day 108)

Events Check List:

1 ❍	Based on your question, decide which events you want to capture to determine both benefit and harm.
2 ❍	Generate objective and precise criteria for these events.
3 ❍	Decide how, when and by whom to ascertain the occurrence of these events.
4 ❍	Consider event-hierarchies
5 ❍	Decide how and to whom these events will be reported.
6 ❍	If required for validity or credibility, set up a system for adjudicating these events.
7 ❍	Based on your question, decide whether any events will be ineligible for the analysis, and develop a plan for their bias-free removal.

Glossary Note: I will use the term "event" to describe an occurrence, incident or experience of a trial patient that is important in answering the study question about benefit and harm. You will find other words used for this purpose elsewhere. I won't use the term "endpoint" because of its connotation that the study is over when it occurs. I won't use the term "outcome" because it suggests a final, fixed state that cannot change, especially for the worse. Finally, I will stretch the term "event" to include continuous measures (such as functional capacity or symptom severity) when they provide important answers to questions about efficacy or safety.

We examined follow-up data for the three events of TIA, stroke and death, and for harm that might have been caused by our treatment.
TIAS: We reviewed each follow-up examination for the presence and number of attacks over each period of follow-up. The occurrence of any ischemic attacks during a follow-up period constituted an event. Since patients who go on to stroke may not, and patients who die cannot, have further ischemic attacks, the last two events were included in this analysis of continuing attacks. Among subjects with multiple events, the date of the first relevant event was used for the analysis.
Stroke. From follow-up data, we identified two groups of patients with possible strokes and submitted their "blinded" records for adjudication. The first group were patients in whom neurologists reported a stroke in a follow-up narrative summary. The second group had complained of one or more ischemic events with a residual neurologic deficit lasting for more than 24 hours. Since dead patients cannot proceed to stroke, death was included in these analyses.
Death. Finally, we documented each death and classified its underlying cause (cerebrovascular, coronary, other vascular or nonvascular).
Harm. In addition, at each follow-up visit we carried out standardized searches for side effects and toxicity (vomiting, nausea, upper abdominal pain, hematemesis, melena, heartburn, renal colic or urinary stone, hematuria, hypoglycemic reaction, skin rash, and an open-ended question about any other complications).

Events in clinical trials help us document the success of our experimental treatment in achieving 1 or more of the "5 Ps": Preventing or Postponing bad events, Promoting good events, Palliating the course of a relentless disorder, or Poisoning the patient with an adverse effect[1]. The events

[1] Alliterations (repeating the same starting sound in several words in a sequence) are so commonly employed as memory aids in clinical medicine that I couldn't resist inventing and inserting one here. A student of alliteration in medicine has noted that the letter "P" predominates, and that "five-P" sequences are not uncommon (e.g., the "Pink, Puffy, Painful,

prevented (or at least postponed) in the RRPCE study were of the yes/no ("discrete") variety. Other "events" in other trials, such as blood pressure, number of days of diarrhea in the past month, or quality of life, are of a matter of degree ("continuous") variety. I'll discuss both sorts in this section.

On to the check-list:

1. <u>Based on your question, decide which events you want to capture to determine both benefit and harm.</u>

As in every other step in your trial, the "right way" to decide on the events you want to capture is determined by the precise question you hope to answer. Was it about prevention, promotion, postponement, palliation, or poison? You could collect dozens of different sorts of events, but the effort and cost of ascertaining each one of them with accuracy and precision can be huge. In addition, as you'll learn in the next section (if you don't already know it), your risk of drawing a false-positive conclusion (that your treatment works when, in fact, it doesn't) rises with the number of different sorts of events you analyze.

You should therefore keep the number of different sorts of events to a minimum. One way to do this is to ask two blunt questions: First, which events do you need to capture to answer your key questions about benefit and harm? Second, which events will be decisive to a clinician and patient as they discuss whether to use your treatment? If answering the first question requires a substantially longer list of events than answering the second, you are overdoing it.

Composite events

In order to answer your study question, you may need to create a "composite" event made up of several disparate events, any one of which constitutes a bad result for patients. Sometimes, the components of these composite events are part of the same pathogenetic mechanism. For example, I chaired the Data Safety and Monitoring Committee for a trial of an ACE-inhibitor (ramipril) in a mix of patients who, for one reason or several, were at increased risk of cardiovascular morbidity and mortality[1]. The composite event in that trial was myocardial infarction, stroke, or cardiovascular death. As it happened, the drug was efficacious enough, and the trial large enough, to generate statistically significant reductions for each of these 3 events considered individually.

In other trials, the components of a composite event might be quite independent from each other, resulting from entirely different pathogenetic mechanisms. For example, in the NASCET trial, patients who underwent endarterectomy risked stroke and death both from their underlying disease and from mechanical complications of their carotid artery surgery. For another example, in carrying out our trials of anticoagulants and antithrombotics among patients with vascular disease, we and our patients were equally concerned about the events we hoped to prevent (e.g., pulmonary embolism) and the "poisonous" adverse events we might cause with our interventions (e.g., life-threatening hemorrhage).

The point here is that the components of the compound event must make sense to readers, and must not simply be a ploy for increasing the numbers of events in a trial that is too small.

Composite events pose problems in interpretation[2]. Just because an experimental treatment reduces the overall occurrence of a composite event, that doesn't mean (and shouldn't be interpreted to mean) that the treatment benefits every individual component of that composite

---

Perspiring, and Peeling" skin of a child with mercury-induced acrodynia) [Hayden GF. Alliteration in medicine: a puzzling profusion of p's. BMJ 1999;319:1605-8.].

event. The latter conclusion has to wait until several, similar trials have been reported. Only then can we meta-analyze the individual components of a composite event to see whether they benefit from treatment.

There's a special case of composite events in which each succeeding component constitutes a progressively worse manifestation of the same pathogenetic process. Such a "hierarchy" appeared in the scenario as TIA or stroke or death. Hierarchies are important enough to deserve their own entry on the check-list, and I discuss them as item #4.

You may want to beef up your composite event by substituting a "biomarker" or "risk-factor" for one or more of your "hard" events. The creation and use of such "surrogate events" is tricky business, as you'll discover in the following paragraphs.

Surrogate events

Sometimes (always?) you are short on time and patients for your trial. Events, even composite ones, may be so rare, and your treatment may take so long to affect them, that you question the wisdom of even attempting a trial. In that situation, you might be tempted to change your "event" from the hard, clinical event itself to a "marker" or "surrogate" which, when measured after a few weeks of treatment, predicts the clinical event that won't occur for months or years. Alternatively, you might want to add the surrogate to the clinical event, forming a composite event. For example, in a coronary prevention trial, you might want to add the coronary risk factor levels such as LDL cholesterol or blood pressure (that every study patient has) to the few clinical coronary events (that only a few patients have).

Surrogate events (sometimes called "surrogate biomarkers") possess some real advantages. You can measure surrogates early in your trial. Moreover, they are typically "continuous" events so that every trial patient has a surrogate "event" (remember the "simple truth" back on page xx that the important number in an RCT is the number of events, not the number of patients?). As a result, the power of your trial (the probability that you'll detect the minimally important difference, if it really exists) soars. No wonder, then, that you find surrogate markers such as tumor size in cancer trials, CD-4 lymphocyte counts in HIV-AIDS trials, and bone mineral density in osteoporotic fracture trials.

The problem is that a favorable change in your surrogate marker may not, in fact, guarantee a favorable change in the occurrence of later events. This can arise in two ways. First, the link between the surrogate marker and the event may be only statistical, not biological. In that case, changing the surrogate "risk-factor" will not change risk. For example, a group of allergists performed an RCT in which they added an oral leukotriene receptor antagonist to a drug regimen for asthma[3]. They reported that altho the new drug had a favorable effect on some surrogate inflammatory biomarkers, it did not improve lung function. For a second example, various treatments for osteoporosis can increase bone mineral density, a surrogate for fractures. However, a meta-analysis of several trials found no relation between changes in bone mineral density (the surrogate marker) and the risk of non-vertebral fractures (the clinical event)[4].

There is a second way that a favorable change in your surrogate marker may not guarantee a favorable change in later events. This occurs when your experimental treatment produces some other, harmful effect by a mechanism that is unrelated to your surrogate marker. In that case, the harmful effect could swamp any favorable effect predicted by improvements in your surrogate marker. There are lots of examples (and, tragically, lots of dead patients) that attest to this major drawback of surrogate markers. Several anti-arrhythmic drugs have been shown to reduce serious arrhythmias following heart attacks, and several other drugs improve exercise tolerance in patients with heart failure. Unfortunately, however, both sets of drugs increase mortality. In another "head-to-head" trial, although the newer, sexier antihypertensive drug doxazosin reduced

diastolic blood pressure as well as "old-fashioned" chlorthalidone, patients randomized to doxazosin were twice as likely to wind up in heart failure[5].

Given these problems with surrogate events, it's wise to avoid relying on them (except perhaps in early Phase II trials). For a more detailed discussion of surrogate events, see an excellent discussion led by Heiner Bucher[6].

Harm

We often don't search hard enough for those events that constitute harm ("Poison"). There are two explanations (but no excuses) for this lack of attention. First, we never start an RCT hoping to demonstrate that our experimental treatment does more harm than good. We focus on the other, "positive P's," and devote the majority of our brains, time, and resources to their ascertainment.

The second explanation follows from the fact that most "serious unanticipated adverse events" (SUAEs) are rare. By their very nature, Phase III RCTs are insensitive in validating them. Mike Gent and I named the statistical explanation for this, which you may already have encountered on page xx, the "inverse rule of 3.[7]" It simply states that, to be 95% sure of observing at least one adverse event that occurs once per x-many patients, you have to observe 3x patients. That is, if your treatment causes one adverse event in every 1,000 patients who receive it, you'll have to follow 3,000 patients to be 95% confident that you'll observe even one of them. Moreover, to conclude that they are occurring significantly more frequently among experimental than control patients, you need to follow several times 3x patients.

Phase III trials are just too small to detect the awful but rare unanticipated adverse events. That's why we've had to employ Phase IV "post-marketing surveillance" studies to detect them and to convince both their manufacturers and regulators to withdraw them.

There are benefits and risks in "lumping" all adverse events together and in "splitting" them into several small categories. Lumping makes for easy summaries, but will hide serious, rare adverse events. Similarly, splitting will provide great detail (e.g., dyspepsia, heart burn, abdominal pain, cramp, etc), but (like a series of small RCTs) can obscure an overall effect of treatment.

Nonetheless, the fact that serious unanticipated adverse events are rare is no excuse for ignoring them. As you read in the previous section on intervention and follow-up, and as you'll read in the later section on monitoring, you must both respond clinically to them at once and document their occurrence for later investigation.

2. <u>Generate objective and precise criteria for these events</u>.

For each type of event, you need to develop criteria that are replicable, not only by your study clinicians, but also by readers of your subsequent trial report. Hammering these criteria out with potential clinical collaborators will not only improve their specification and reproducibility, but also strengthen your collaborators' sense of ownership and cooperation with your trial.

For example, the events in our venous thromboembolism trials included positive venograms as reported by one of two radiologists[8]. Before beginning this series of trials, we asked these two experts to read the same set of venograms, and found substantial disagreement (their kappa was only slightly greater than 0.6). We then shut them in a room with a set of venograms and a view-box, and told them not to come out until they resolved their disagreement. To everyone's pleasant surprise, they discovered that they agreed well on the presence or absence of specific venographic features, but disagreed on how they should be combined into an overall

interpretation. Once they emerged with their mutually satisfactory set of venographic criteria, their agreement soared (their new kappa was almost 0.9).

In a similar fashion, you might want to run workshops in which volunteer patients with and without the target events are examined (and re-examined) by potential study clinicians. In this way you can speed the generation of both common criteria and high precision in ascertaining them.

When events are continuous measures such as blood pressure, functional status, or quality of life, you need to decide the appropriate "instruments" with which to measure them. You also need to decide who should apply these instruments. For example, should study clinicians be the ones who measure study patients' blood pressure? If so, do they need hearing tests, training sessions for accuracy and precision (including whether to report muffling or disappearance for diastolic blood pressure, whether to record the nearest lowest 2 mm Hg or the closest 5 mm Hg, and the like) and periodic retesting? Or do you want to measure blood pressure automatically by machine? If so, how often do you want to compare its results with those obtained by a person? Gordon Guyatt and Peter Tugwell discuss the development and validation of questionnaires and interviews for continuous measures in chapter 6.

These tactics apply equally to composite events and surrogate markers. Furthermore, when a composite event includes clinical interventions, such as the decision to offer revascularization or to hospitalize for heart failure, you need to be sure that study clinicians remain blind to patients' treatments.

3. Decide how, when and by whom to ascertain the occurrence of these events.

Some events guarantee ascertainment, such as disabling strokes or deaths, and you need not schedule special appointments to ferret them out. Other events (such as recurrent mild TIAs or upset stomachs from study drugs) may go unrecognized without standardized inquiries during repeated follow-up visits. Regardless of whether events shout for recognition or only murmur, you must first decide how to ascertain each of them. Can you wait for patients to spontaneously call or visit your study center? Or, should you carry out standardized searches for them at regularly scheduled follow-up visits? As elsewhere, the right answer to these queries resides in your study question and, in particular, on its location along the explanatory – management axis.

Next you must decide when you will ascertain events. Your objective here is to schedule follow-up visits often enough to keep your patients on their trial regimens, ascertain transient events, detect important adverse effects, and keep track of their whereabouts. On the other hand, you don't want to wear out your study patients and their clinicians by insisting on unnecessarily frequent visits. In the RRPCE trial, we needed to ascertain the frequency of continuing TIAs (whose manifestations are transient by definition), so we saw those patients every 3 months. In the HOPE trial of ramipril among patients at high-risk for dramatic and permanent cardiovascular events, we needed to see patients only twice a year[9].

Finally, you need to decide who will ascertain these events. Will clinical laboratories report them as a matter of course? Or, does their recognition require considerable clinical skill, in which case your study clinicians must ascertain them? Or, are they continuous measures from a quality of life questionnaire-interview, in which case trained lay interviewers will out-perform most clinicians by a wide margin? Finally, will study patients record them in personal diaries, in which case a friendly, conscientious study clerk could collect them or obtain their contents by telephone? Who does what has major consequences for your trial budget.

Should you forewarn trial patients about mild side-effects?

A special set of issues surrounds ascertaining mild side-effects of trial treatments. We discovered them in our 3-center trial of aspirin and sulfinpyrazone for patients with unstable angina (a life-threatening illness)[10]. Half the patients at each center received aspirin (with or without additional sulfinpyrazone). At two centers (I'll call them A and B), our co-investigators listed "occasional gastrointestinal irritation and skin rash" as potential side-effects in their consent forms. Our co-investigator did not mention these mild side-effects in the consent forms at the third center, C. (Yes, local ethics committees had approved both versions of the consent forms.)

During the trial, "informed" patients in centers A and B were far more likely to report mild gastrointestinal side-effects (e.g., nausea, indigestion, heartburn) than "uninformed" patients in center C. Interestingly, however, these side-effects were not associated with aspirin; only 56% of symptomatic patients were, in fact, receiving aspirin. Of real concern was the finding that "informed" patients were 6 times as likely as "uninformed" patients to stop their study drugs because of these mild gastrointestinal side-effects. We couldn't blame this huge effect on other differences between "informed" and "uninformed" patients. They were equally likely to develop major side-effects such as frank gastrointestinal bleeding. Moreover, none of the patients in centers A and B stopped taking their study drugs because of minor side-effects arising outside the g-I tract (e.g., weakness, vertigo, tinnitus). We concluded that we had "sensitized" our study patients, not only to attribute the mild g-I complaints we all encounter to the study drugs, but also to stop taking them as a result.

Of course you should always inform your study patients to be on the look-out for severe side-effects and to take immediate action when they occur. Just remember that, if you "sensitize" your study patients to potential mild side-effects as they enter your trial, they may be both more likely to report them and more likely to stop your study drugs "because" of them.

4. <u>Consider event-hierarchies</u>

As you can see from the RRPCE scenario, we sought 3 different events that shared 3 interesting properties. First, their severity ranged from minor (TIA) to the supremely severe (death). Second, all were manifestations of the same biologic process, the progressive atherosclerotic deterioration of an artery serving part of the brain. Third, the occurrence of a more severe event in the group made it impossible for the affected patient to subsequently display a less severe event. That is, patients with a total loss of sensation and movement in an arm could no longer display the transient sensory or motor deficits in that same arm that we call TIAs, and you can't have a stroke if you're already dead. We had to recognize "event-hierarchies" in which a more severe event along the hierarchy precluded the occurrence of a lesser event. As a result, we could never report the lesser event in isolation. Thus, in describing the efficacy of aspirin it would have been nonsensical for us to report the frequency of recurring TIAs all by by themselves, because aspirin could have stopped our patients' recurring TIAs by killing them. Similarly, in our subsequent trial of these same two drugs among patients with unstable angina, we never reported myocardial infarctions by themselves, but always as "non-fatal myocardial infarction or cardiac death.[11]"

I hope you noted in that last example that "cardiac death" is cause-specific mortality, not total mortality. Is that restriction appropriate? After all, patients who get killed by a bus or die from stroke can't have non-fatal myocardial infarctions either. I think that it is appropriate for you to designate cause-specific mortality as an event when two conditions are met. First, individuals who are blind to treatment must adjudicate each death and decide its cause. Second, the specific cause should account for most of the deaths that occur among your study patients. In our unstable angina trial, the adjudicators attributed 39 of the 44 deaths (89%) to be cardiac in origin. Moreover, this proportion of cause-specific mortality was so high that including all deaths in the analysis led to the same conclusion (that aspirin helped reduce them). We'll talk more about this in the section on analysis and interpretation of trials.

5. Decide how and to whom these events will be reported.

Your follow-up forms need to capture and document all relevant features of the events in your trial. For the first 4 P's (Prevention, Promotion, Postponement and Palliation) it is usually appropriate to simply document the event on the follow-up form and submit it with ordinary speed for adjudication and eventual incorporation into the analysis. However, for the 5th P, Poisoning, any major or life-threatening event, especially when it is unexpected, requires immediate action (often including notification of licensing bodies). This is another reason why most RCTs should recruit one or more outsiders (not otherwise involved in the trial) who monitor its progress and respond to just such events. Severe unanticipated adverse events (SUAEs) can then be reported immediately to the trial's monitor, who breaks the code and decides whether urgent decisions need to be made about modifying the experimental regimen or even stopping the trial.

6. If required for validity or credibility, set up a system for adjudicating these events.

This sort of adjudication has become standard practice for any trial in which knowing the patient's treatment group could influence (consciously or subconsciously) the reporting of the occurrence or severity of important events. In the RRPCE trial, we took the records of patients who died or whom we suspected had suffered strokes, and purged them of any information about their study drugs. We then had these purged records reviewed independently by two senior neurologist-adjudicators who were blind to their treatment. Our adjudicators then compared notes, and resolved any disagreements by discussing them in the presence of one of the directors (also blind) of the Methods Center. These adjudicators also ruled whether the stroke was minor (no impairment in activities of daily living), moderate (impairment in activities of daily living but residing at home and out of bed for all or part of the day) or severe (bedfast or institutionalized for reasons of disability). Although our adjudicators usually agreed with the diagnoses made in study centers, there were important exceptions.

Event adjudication is a lengthy process, and in RCTs in chronic diseases it is not unusual for it to lag behind the occurrence of events by 6 months or more. However, there are 2 good reasons for striving to shorten this lag time. First, early in your trial, the frequency of true events can be compared with your pre-trial estimates. This comparison will tell you whether your sample size is adequate for answering your study question (and, if not, will provide a starting point for re-estimating your sample size needs). Second, later in your trial the need to know the number of true events becomes urgent as you apply the statistical warning rules that inform your decisions about stopping or continuing your trial.

7. Based on your question, decide whether any events will be ineligible for the analysis, and develop a plan for their bias-free removal.

We've already described the exclusion of ineligible patients, both before and soon after randomization. But what about the exclusion of events that occur well into the trial? Here is another excerpt from the RRPCE trial, this time about the eligibility of study events for the efficacy analysis at the explanatory pole:

> Because it was believed that sulfinpyrazone took one week to produce a biologically appreciable effect, we decided to exclude any events occurring in the first week of therapy with any of the four regimens. Furthermore, since the withdrawal of patients from the trial might be precipitated by a deterioration in their neurologic status (and thus their exclusion from subsequent analyses might bias the results in favor of their study regimen), we charged any events occurring within the first six months after withdrawal against the corresponding study regimen even if the patient stopped taking the study medication at the time of withdrawal. Any bias resulting from this maneuver should be against showing a benefit of treatment.

This strategy was consistent with the explanatory nature of the question we asked in that trial: Can sulfinpyrazone work under ideal circumstances? We derived it from what we knew about the pharmacodynamics of sulfinpyrazone (it took 7 days to exert its effect on platelet function). If sulfinpyrazone were effective, the inclusion, in the efficacy analysis, of events before 7 days of treatment and more than 6 months after its withdrawal would unfairly blame it for events it couldn't control. This inclusion would raise the sulfinpyrazone event rate and decrease both the relative and absolute risk reductions attributed to it. In our primary analyses, we removed these "ineligible" events from all arms of the trial to prevent biased comparisons between treatment groups.

But you risk the credibility, if not the validity, of your trial's conclusion if you remove (or "censor") any events that occur after randomization to eligible study patients. Such post-randomization exclusions do (and usually should) increase skepticism about your trial's conclusions. Furthermore, in carrying out post-randomization exclusion of events you might move your interpretation of a positive result so far away from its pragmatic application to destroy its clinical usefulness. In the RRPCE trial, we also performed an "intention-to-treat" analysis that included all the "ineligible" events. As it happened, sulfinpyrazone remained ineffective in this pragmatic analysis, but the benefit of aspirin became even greater.

Designating any events that occur after randomization "ineligible" is a very risky strategy, and I don't recommend it. Instead, I urge you to recruit enough patients to swamp the negative consequences of including such events in your analysis.

REFERENCES

[1] Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 2000;342:145-53.

[2] Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003;289:2554-9.

[3] Currie GP, Lee DK, Haggart K, Bates CE, Lipworth BJ. Effects of montelukast on surrogate inflammatory markers in corticosteroid-treated patients with asthma. Am J Respir Crit Care Med. 2003;167: 1232-8. Epub 2002 Nov 27.

[4] Guyatt GH, Cranney A, Griffith L, Walter S, Krolicki N, Favus M, Rosen C. Summary of meta-analyses of therapies for postmenopausal osteoporosis and the relationship between bone density and fractures. Endocrinol Metab Clin North Am 2002;31:659-80.

[5] D'Agostino RB. Debate: The slippery slope of surrogate outcomes. Curr Control Trials Cardiovasc Med 2000;1:76-8.

[6] Bucher H, Guyatt G, Cook D, Holbrook A, McAlister F. Therapy and applying the results. Surrogate Outcomes. In: Guyatt G, Rennie D, eds. Users Guides to the Medical Literature. AMA Press, 2002.

[7] Sackett DL, Haynes RB, Gent M, Taylor DW. Compliance. In: WHW Inman, ed. Monitoring for Drug Safety. Lancaster: MTP Press, 1980.

[8] Hull RD, Raskob GE, Hirsh J, Jay RM, Leclerc JR, Geerts WH, Rosenbloom D, Sackett DL, Anderson C, Harrison L, et al.

Continuous intravenous heparin compared with intermittent subcutaneous heparin in the initial treatment of proximal-vein thrombosis.
N Engl J Med. 1986;315:1109-14.

[9] The Heart Outcomes Prevention Evaluation study investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. N Engl J Med 2000; 342:145 – 53.

[10] Myers MG, Cairns JA, Singer J. The consent form as a possible cause of side effects. Clin Pharmacol Ther 1987;42:250-3

[11] Cairns JA, Gent M, Singer J, Finnie KJ, Froggatt GM, Holder DA, Jablonsky G, Kostuk WJ, Melendez LJ, Myers MG, Sackett DL, Sealey BJ, Tanser PH. Aspirin, sulfinpyrazone, or both in unstable angina. Results of a Canadian multicenter trial. N Engl J Med 1985;313:1369-75