Purpose of review
Four randomized trials have investigated the combination of clopidogrel plus aspirin for secondary prevention of vascular outcomes in 54,949 patients. Here we argue that attempts to translate the results of these trials into clinical practice have proven frustrating because of the following statistical considerations: differences in study populations and study design make comparisons difficult (comparisons of ‘apples and oranges’), incomplete factorial designs prevent proper contrasts (examining ‘bits and pieces’ of a larger picture), results concern widely different vascular diseases (‘puzzling subgroups’), and negative results are easily misinterpreted.

Recent findings
Between 1996 and 2004 three major randomized trials assessed combinations of aspirin and clopidogrel, finding: Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events (CAPRIE) in favor of clopidogrel alone versus aspirin alone, Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) in favor of clopidogrel plus aspirin versus aspirin alone, and Management of Atherothrombosis with Clopidogrel in High-risk Patients (MATCH) in favor of clopidogrel plus aspirin versus clopidogrel alone. A recently completed fourth trial (Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management and Avoidance; CHARISMA) was a ‘negative study’ comparing aspirin alone to aspirin plus clopidogrel.

Summary
Even after four large randomized trials we still do not know the optimal treatment for secondary prevention of stroke. We suggest that subsequent trials should focus on a particular vascular disease and test hypotheses that relate to a specific mechanism.

Keywords
clinical trials, clopidogrel

Abbreviations
CAPRIE Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events
CHARISMA Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management and Avoidance
CURE Clopidogrel in Unstable Angina to Prevent Recurrent Events
MATCH Management of Atherothrombosis with Clopidogrel in High-risk Patients
MI myocardial infarction

Introduction
The relative efficacy of different combinations of aspirin and clopidogrel to prevent cardiovascular events have been contrasted in several large randomized clinical trials including Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events (CAPRIE)[1], Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) [2] and Management of Atherothrombosis with Clopidogrel in High-risk Patients (MATCH) [3]. Together these trials have randomized 39,346 patients, but as yet the appropriate treatment for primary and secondary prevention of stroke remains a matter of debate [4,5]. While individually each of these studies was conducted with the highest level of quality, they have failed to provide conclusive guidance for secondary prevention because they provide only comparisons between ‘apples and oranges’, include only ‘bits and pieces’ of the information needed, include puzzling subgroups, and are subject to misinterpretation in the case of a negative study. The surrounding confusion is a product of these shortcomings in information. We will first review the three studies CAPRIE, CURE and MATCH. We then see if a more recent fourth study clarified concerns and uncertainties about the first three.

A brief overview of the first three studies is necessary to appreciate the confusion resulting from their interpretation.

CAPRIE randomized 19,185 symptomatic patients to either clopidogrel (75 mg) or aspirin (325 mg). By design, one-third of the patients had a previous stroke, one third had a previous myocardial infarction (MI) and one-third had peripheral vascular disease. Patients were followed for between one and 3 years (median follow-up of 1.91 years). The study showed an 8.7% (95% CI 0.3%–16.5%;
Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
the dose of aspirin or clopidogrel used? Frankly, who knows: making comparisons between apples and oranges is challenging.

**Bits and pieces?**

The European Stroke Prevention Study 2 (ESPS2) used a two-by-two factorial design to assess the efficacy of aspirin and dipyridamole individually or in combination (Fig. 1a) [6]. This powerful design allows for the assessment of the following: the effect of aspirin, a comparison of the pooled data from those receiving neither aspirin nor dipyridamole plus those on dipyridamole only compared to the pooled data of those receiving aspirin only and those receiving both aspirin and dipyridamole; the effect of dipyridamole, a comparison of the pooled data from those receiving neither aspirin nor dipyridamole and those receiving aspirin only versus the pooled data of those receiving dipyridamole only and those receiving both aspirin and dipyridamole; and the potential of a synergistic (or nonadditive) effect of aspirin and dipyridamole, i.e. if the efficacy gains among those receiving both aspirin and dipyridamole represents the sum of the gains from taking aspirin and the gains from taking dipyridamole.

In contrast, Fig. 1b shows the comparisons that can be made on the basis of studies conducted on aspirin and clopidogrel. Interestingly, each of these major studies shares one treatment with each of the other two major studies. Thus, instead of having the complete factorial design we have only bits and pieces from which to draw inferences.

The transitive property in mathematics states that if A is greater than B, and if B is greater than C, then A is greater than C. It is tempting to use this type of logic to make decisions as to which of the three alternative active treatments (aspirin, clopidogrel or aspirin plus clopidogrel) is superior. The uncertainty associated with the comparisons within each of the studies makes this logic problematic. A sports analogy, where uncertainty introduces similar problems, may make this point clear. On November 19th, 2004, the University of North Carolina Tar Heel basketball team was ranked number two in the country, when they lost to Santa Clara University (Santa Clara 77, UNC 66). This established Santa Clara as a better team than UNC. Subsequently, on December 29th, Yale defeated Santa Clara 90 to 84, establishing Yale as superior to Santa Clara. Since: UNC is the number two team in the nation, and Yale is superior to Santa Clara, who is superior to UNC, then Yale is better than the second best team in the nation, thus establishing Yale as the number one team in the nation. In addition, New Mexico, Pacific, Cal Poly, UC Irvine, Arizona State, Central Connecticut State, Brigham Young, Gonzaga, and San Francisco had also all defeated Santa Clara during the year, hence they should also have been the number one team in the nation. This ludicrous example of logic is faulty because of the uncertainty involved in sporting events: a victory on one night does not establish a team as ‘globally’ superior to another. In much the same way, however, there is uncertainty in research studies, and differences between treatments are also estimated with error. Few people are comfortable drawing transitive relationships in sports (go Yale!); perhaps we should be equally cautious drawing such transitive conclusions from bits and pieces of incomplete study designs (as we have in Fig. 1b). Unfortunately, the illogic of using the transitive property to compare studies means that physicians are left in a situation where large studies have been performed but their results cannot be combined in a coherent way.

**Confusing subgroups?**

As the largest study and the study providing a direct comparison of monotherapy with aspirin versus clopidogrel, an understanding of the results of CAPRIE is central to the decision of prescribing treatments to prevent stroke. Overall, clopidogrel was superior to aspirin in CAPRIE, showing an 8.7% (95% CI 0.3–16.5%) benefit (Fig. 2); however, this effect barely reached the traditional level of significance ($P = 0.043$). As discussed...
above, patients in CAPRIE were drawn from three patient sub-populations in approximately equal representation: those having had a stroke, MI or peripheral vascular disease (PVD). The interpretation of CAPRIE, however, is substantially complicated by the presence of a significant interaction within the study population \( (P = 0.042) \) indicating that it is likely that the relative benefit of clopidogrel differs depending on the study sub-population (Fig. 2). As is evident in Fig. 2, the overall treatment effect is largely affected by a 23.8% benefit in the clopidogrel group for the PVD subpopulation. The large effect among PVD patients was substantially dampened by a 3.7% benefit for aspirin in the MI subpopulation (i.e. nonsignificantly and in the opposite direction). The 7.3% effect in favor of clopidogrel in the stroke subpopulation was intermediate and similar to the overall effect; however, this difference did not reach a level of statistical significance \( (P = 0.26) \).

When faced with the decision of whether to prescribe clopidogrel to a stroke patient, the information from CAPRIE is confusing at best. It is not surprising that statistical significance was not achieved in the stroke subpopulation since the study was not powered to achieve significance in the subpopulations. Had the effects across the subpopulations been consistent, then it would be reasonable to proceed with treatment of the stroke patient, based on the results from the CAPRIE study. The evidence of a substantial difference between the patient subpopulations implies that a general ‘average’ effect across the patient population as a whole is not a reasonable interpretation of the results. The lack of a significant association in stroke patients raises the question of whether they are more similar to MI patients (for whom no effect was observed), more similar to PVD patients (for whom a substantial effect was observed), or intermediate between the two. When treating a stroke patient, this presents the clinician with the dilemma of interpreting the results based on confusing subgroups.

A final point should be made here about MATCH. In this study, approximately 70% of the patients had diabetes and 54% of the recurrent strokes were small vessel in nature. Thus, it is likely that a very different mechanism is in play when compared to coronary artery disease. Thus we do not know if the outcome would have been different if only patients with stroke from large artery disease had been included.

**Interpretation of ‘negative’ studies?**

A study is considered ‘negative’ when it fails to reject the null hypothesis of no difference between treatment groups. Frequently, results from negative studies (such as MATCH) are described thus: ‘there are no differences between treatment groups’. This interpretation is simply incorrect. There are two major reasons why a trial may fail to show a significant difference: the difference is in fact small or does not exist, or the study was not adequately powered to document differences that truly do exist. As such, rather than ‘there are no differences between treatment groups’, the correct interpretation is that there is no evidence in these data that indicates which treatment is superior: a substantially different statement. As an example, consider the statement ‘all swans are white’. To examine this statement, a sample of swans is drawn. There are two possible outcomes: all the swans in the sample are white; at least one swan in the sample is not white. The latter outcome establishes the falsehood of the statement (i.e., rejects the null hypothesis that all swans are white). The first outcome, however, does not prove the statement since, had we taken a different sample, we may have found a non-white swan. Obviously, although the statement can never be proven, a larger sample of swans provides more definitive evidence that the probability that the statement is false is likely to be very small.

Hypothesis testing in clinical trials follows very similar principles. In hypothesis testing, there are two types of
errors that can be made. A type I error is made if we incorrectly reject the null hypothesis. In the swan analogy this would be equivalent to saying that all swans are white, when in fact they are not (perhaps we had an unusual sample with a ‘dirty’ swan). Similarly, a type II error is made if we fail to reject the null hypothesis when it is false. In the swan analogy, this would be equivalent to having a sample of all white swans and failing to reject the assumption that all are white. Statistical power, which is defined as one (1.0) minus the probability of a type II error, is therefore defined as the probability that the study will reject the null hypothesis of no treatment difference if there truly is a difference. Thus, failing to reject hypotheses for sufficiently powered tests provides more definitive evidence that the likelihood that the null hypothesis is false is very small. Hence, in interpreting the evidence from a negative study, one should examine whether the study was adequately powered to detect important differences of interests.

So, what difference was the MATCH study designed to detect, and to what degree was the study powered to detect this difference? The MATCH study was designed to detect a 14% treatment difference, with 80% power. In MATCH, this means that even if a 14% treatment difference did exist there was a 20% chance that the study would fail to reject the null hypothesis of no difference. Just as 5% is the standard for rejecting the null hypothesis (i.e., $P < 0.05$), studies are frequently designed with either 80% or 90% power to detect a treatment difference. Interestingly, 5%, 80%, and 90% are simply all arbitrary numbers: why do we require $P < 0.05$ to reject a hypothesis, yet we are comfortable with a study with only 80% power?

We should all reflect on the meaning of a negative study with 80% power. First, the MATCH trial did nothing incorrect; 80% power is an ‘industry standard’ number (much like $P < 0.05$). It is surprising (at least to this author), however, that as a scientific community, in partnership with industry, we are comfortable committing the financial resources required to mount a trial of the scope of MATCH and exposing patients to randomly assigned treatments under a situation where there is a one in five chance of not detecting a treatment effect even if the hypothesized one does exist.

MATCH was designed to detect a 14% difference. Is this a reasonable effect? Perhaps the most reasonable comparison is the observed 13% treatment effect for combined treatment (dipyridamole plus aspirin) versus single treatment in ESPS2 [6]. This implies a difference of 14% is achievable, but it is still slightly larger than achieved in ESPS2. It is relatively small, however, compared to the 20% difference between combination treatment and treatment with aspirin observed in CURE. So, does MATCH confirm or refute an effect of combination therapy? The observed effect was a 6.4% (95% CI –4.6% to 16.3%) event reduction for combination therapy as compared to monotherapy with clopidogrel. The old way of thinking is that since this does not differ significantly from a 0% difference, this is a negative study and thus does not support combination therapy. A new way of thinking is (hopefully) emerging, however, by which we recognize that the information available from the study resides in the estimated difference and its confidence limits. Taking this approach combination therapy showed a trend in favor of being superior; it is true that it is not significantly different from a 0% effect, and so we should interpret this with caution; but it is also not significantly different than the a priori difference of 14%, which was selected as being a clinically meaningful difference. That is, while we cannot say that there was clearly a treatment difference, we also cannot say that the observed difference was less than the a priori ‘clinically meaningful’ difference that the study was powered to detect.

How should we interpret a negative study such as MATCH? It is clear that we should not say that based on this study, there is no evidence of difference between treatments. It is also clear that we cannot say that this establishes combination therapy as the better treatment. If one had to comment in this uncertain situation, it is reasonable to say that our best guess is that combination therapy results in a 6.4% reduction in risk; however, this is an uncertain guess and it is not unreasonable to expect to see as much as a 4.6% harm, or as much as a 16.3% benefit. In summary, we should interpret these findings with logic and caution.

Does performing more studies clarify our understanding?

One could only hope that adding new information from a more recent study would clarify our understanding of the clinical usefulness of clopidogrel. The recently reported Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management and Avoidance (CHARISMA) trial [7], however, has, if anything, done the opposite. CHARISMA recruited 15 603 patients with clinically evident cardiovascular disease (‘symptomatic’ patients) or multiple risk factors (‘asymptomatic’ patients). Like the CURE study, CHARISMA randomized patients to either aspirin/clopidogrel or aspirin/placebo. While all patients recruited into CURE had unstable angina, the CHARISMA population included approximately 11% with ‘angina with documented multi-vessel coronary disease’, 26% with other evidence of coronary disease (interventions or myocardial infarctions), 28% with stroke, 18% with peripheral vascular disease, and 21% asymptomatic patients with multiple risk factors (percentages sum to more than 100% because groups are not mutually exclusive).
It is an informative exercise to see if we could have predicted the results of CHARISMA. We would surely have been heavily influenced by the CURE results, showing a striking 20% (95% CI 10–28%) benefit for treatment with clopidogrel. We would also have also been influenced by the CAPRIE results where there was a nonsignificant harm for treatment with clopidogrel among patients with coronary disease, but a significant benefit for patients with peripheral vascular disease and a nonsignificant benefit among patients with stroke. Unlike CURE (which included only coronary patients), CHARISMA was ‘enriched’ by patients with stroke and peripheral vascular disease, and as such one could speculate that the benefit of clopidogrel in CHARISMA would actually be larger than CURE. The exact opposite was found.

CHARISMA failed to find a benefit for the addition of clopidogrel to aspirin with a relative risk of 0.93 (95% CI 0.85–1.05; P = 0.22), basically a finding that leads to the response ‘now what?’ This result is close to a perfect teaching example of the dangers of generalizing findings in subgroups (confusing subgroups), attempting to make interpretations between different drug contrasts (bits and pieces), and how to interpret a negative study. As such, rather than clarifying our understanding, it is hard not to conclude that CHARISMA has only further clouded the situation. To make matters worse, CHARISMA reported an a priori hypothesized subgroup analysis, which showed a borderline significant (P = 0.045) interaction between treatment and symptomatic status with significant protection in the symptomatic population (relative risk 0.88; 95% CI 0.77–0.998; P = 0.046) but nonsignificant harm in the asymptomatic population (relative risk 1.2; 95% CI 0.91–1.59). As acknowledged by the authors, this finding was the only one of only several subgroups where effect modification was found; however, the paper suggests (by showing results for 12 subgroup analyses) that at least 12 such analyses were conducted, and in this case under the null hypothesis that clopidogrel has no effect there is a 46% (1 – (1 – 0.05)^12) chance that at least one subgroup analysis indicating significant effect modification would be reported. Depressingly, after a remarkable 54,949 patients randomized (the 39,246 in CAPRIE, CURE and MATCH plus the 15,603 in CHARISMA), the authors of CHARISMA concluded that the finding in symptomatic patients ‘requires further study’.

**Conclusion**

It is hard not to be pessimistic about whether the piece-meal approach employed to date will ever answer the dual antiplatelet therapy question definitively, even if the number of patients were to soar into six digits (not to mention active comparison trials with alternative promising treatments such as dipyridomole) [8]. Thus, what do we tell our patients: aspirin, clopidogrel, or both? What does the information from randomizing 54,949 patients tell us? All it tells us is that we can be sure that we are unsure of the appropriate approach. CAPRIE tells us that monotherapy with clopidogrel is slightly superior to aspirin, but not really for treating stroke patients, and the benefit of clopidogrel apparently is dependent on the medical history of the patient. CURE tells us combination therapy is clearly superior to monotherapy with aspirin, but this was conducted in a population of almost exclusively coronary patients. MATCH, a study actually performed in stroke patients, suggests a benefit for combination therapy but fails to reach a level where definitive statements can be made. CHARISMA tells us to avoid dual antiplatelet therapy altogether. Go figure.

This sorry state of affairs suggests that we need to do something differently. Although we may not be at the point where we can substitute ‘one size fits all’ megatrials for personalized patient management based on genotype [9], surely a more homogenous disease mechanism approach should be employed. After all, in the field of secondary stroke prevention, it has been those trials that have focused on a particular stroke mechanism, carotid stenosis and atrial fibrillation, that have yielded significant and large effect sizes. Subgroup analysis is not a good alternative. It is an approach, as we have seen, which is fraught with difficulties and suggests differences that then go on to be proven wrong when tested directly [10]. Without a change, it can be envisaged that the next trial will be called MIASMA.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (p. 85).


CHARISMA recruited 15 603 patients with clinically evident cardiovascular disease (‘symptomatic’ patients) or multiple risk factors (‘asymptomatic’ patients), and failed to show a difference by treatment assignment with a relative risk of 0.93 (95% CI 0.83–1.05; P = 0.22) for aspirin plus clopidogrel versus aspirin alone.

