Changes in Statistical Significance

One of our two quantitative signals consisted of a change in statistical significance: the original review reported a statistically significant result and the addition of new trials resulted in loss of that significance or, conversely, a previously non-significant result became statistically significant.

In order for this signal to achieve its desired goal of capturing changes in evidence that would clearly warrant updating the original review, we wanted to avoid labeling trivial changes in p-values as signals for updating. For example, if a systematic review reported a borderline p-value of 0.049 and addition of data from new eligible trials resulted in an updated p-value of 0.051, this would have no face validity as an important signal for updating.

To avoid labeling such trivial or borderline changes as signals for updating, we excluded all nominal changes in statistical significance in which the original and updated meta-analytic results both had p-values between 0.04 and 0.06. We did not exclude changes in which just one of the p-values was 'borderline.' For instance, if the original p-value were 0.02 and the updated result were 0.059, we would count this as a change in statistical significance, because, even if one wanted to call the updated value of p=0.059 a 'borderline significant result, this clearly represents an important change from the previous, apparently robust result of p=0.02.

The specific choice of 0.04 to 0.06 was arbitrary, but changing these numbers would not substantially have affected the results, as those that we did count as changes in statistical significance fell well outside this range. In other words, there were no cases in which both the original and the updated values would both have been labeled 'borderline', using a slightly different range of p-values, such as p=0.03 to p=0.07.

The table below lists all of the specific p-values that generated quantitative signals for a change in statistical significance. Note that, in many cases, these changes in the meta-analytic were accompanied by confirmatory qualitative signals. For example, the second row reflects the results of a trial that was published in a high impact journal [Finfer S et al. A comparison of albumin and saline for fluid resuscitation in the intensive care unit *N Engl J Med.* 2004] and having almost 5 times the sample size of previous trials combined, which showed that using albumin in critically ill patients produced no increase in mortality, contrary to increase in mortality reported in the original review. [Alderson P et al. Human albumin solution for resuscitation and volume expansion in critically ill patients. *Cochrane Database Syst Rev.* 2002]

P value of original review	P value after adding new trials	Outcome and comparison involved
0.020	0.197	All-cause mortality: Amphotercin B versus no fungal prophylaxis in cancer patients
0.000	0.387	All-cause mortality: albumin vs. crystalloids in critically ill patients
0.286	0.003	Major bleeding: moderate to high intensity oral anticoagulation plus aspirin versus aspirin alone
0.297	0.044	Serious vascular events (vascular death or death from unknown cause or MI (with or without clearly non-vascular death) or stroke (with or without clearly non-vascular death): Aspirin plus dipyridamole versus Aspirin alone
0.056	0.009	All-cause mortality: beta-blockers versus placebo in CHF
0.020	0.148	Risk of contrast nephropathy: acetylcysteine vs placebo
0.351	0.045	All-cause mortality: Abciximab vs Placebo
0.016	0.106	Need for coronary artery bypass grafting: angioplasty vs medical therapy for non-acute coronary artery disease
0.196	0.039	30-Day Death or stroke: carotid endarterectomy vs endovascular treatment
0.112	0.008	All-cause mortality: local thrombolysis vs. transfer for primary angioplasty
0.111	0.043	asthma exacerbation requiring systemic steroids: anti-leukotrienes versus inhaled steroids in adults and children
0.143	0.019	10-year all-cause mortality: radiotherapy vs surgery for breast cancer
0.404	0.000	abdominal pain and discomfort: tegaserod vs placebo for irritable bowel syndrome
0.297	0.004	Cardiovascular and thromboembolic events: post-

		menopausal hormone replacement therapy vs placebo
0.071	0.037	Progression to any primary end point: early vs deferred AZT in HIV infected patients without AIDS defining illnesses
0.161	0.001	Mortality following percutaneous coronary intervention (PCI) at 30 days and 6 months: platelet glycoprotein IIb/IIIa inhibitors vs placebo
0.185	0.010	Clinical treatment success at 1 month: treatment of Helicobacter pylori infection vs control for patients with nonulcer dyspesia
0.000	0.062	Ovulation rate: metformin and clomifene vs clomifene alone for women with polycystic ovary syndrome