

# Poor Agreement between Meta-analysis and Subsequent Large Randomised Controlled Trials in Peri-operative Medicine

Philip Peyton<sup>1</sup>, Harry Sivakumar<sup>2</sup>

1. A/Prof. Austin Hospital & University of Melbourne 2. Dept Anaesthesia, Southern Health, Melbourne, Australia



**AIMS:** To assess the reliability of meta-analysis (MA) in peri-operative medicine and anaesthesia, by comparing the findings of MAs with those of subsequent large randomized controlled trials (RCTs) of the same interventions and endpoint outcomes.

**METHODS:** Using Medline & PubMed, large RCTs ( $n > 1000$ ), published since 2000 in the anaesthesia & peri-operative medicine/critical care literature were identified. All previous MAs investigating the same intervention and population were sourced. For all reported major morbid endpoints common to each RCT and its related MA(s), results (significant or non-significant at  $p < 0.05$ ) were compared on a 2X2 table. The positive & negative predictive value (PPV & NPV) & Kappa were calculated. In a sub-analysis results of all RCTs were compared with only the largest recent MA. Significance of agreement was tested by McNemar's  $X^2$  with Bonferroni adjustment for multiple endpoint comparisons. Study size was also compared.

**RESULTS:** 18 large RCTs and 48 MAs investigating the effects of 15 interventions were identified. 150 endpoint outcomes were found suitable for comparison (Table 1). PPV was 8.3%, NPV 88.5%, Kappa -0.0331 indicating agreement worse than expected by chance alone.

In the sub-analysis of 18 large RCTs and 18 largest recent MAs, 51 endpoint outcomes were compared (Table 2). PPV was 11.1%, NPV 79.2%, Kappa -0.0931 indicating agreement worse than expected by chance alone. The Exact McNemar's  $X^2$  was 12.5, adjusted  $p = 0.0014$ .

Table 1 Sig = Significant Result for Endpoint		Large RCT	
		Sig	Non-Sig
All MAs	Sig	6	66
	Non-Sig	9	69

Table 2 Sig = Significant Result for Endpoint		Large RCT	
		Sig	Non-Sig
Largest recent MA	Sig	3	24
	Non-Sig	5	19

The median study size (interquartile range) for large RCTs was  $n = 4,263$  (2,046–6,911). This was not different to that of the MAs,  $n = 2,340$  (1,121–5,360) ( $p = 0.1798$  on the Wilcoxon rank sum test).

**CONCLUSIONS:** The outcomes of the 18 large RCTs were predicted correctly on 50% of occasions by the preceding MAs. There was a strong tendency towards positive findings in MA which were not substantiated by subsequent large RCTs, which is likely to be a result of positive publication bias in the literature in our field.<sup>1,2</sup> This finding questions reliance on MA in clinical decision making in anaesthesia and peri-operative medicine.

**References:** De Oliveira *et al.* Publication bias in the anesthesiology literature. *Anesth Analg.* 2012 May;114(5):1042-8  
Peyton P, Chong S, Collins N. Publication bias in anaesthesia literature – A meta-analysis & review. *ANZCA ASM* 2014