*Acta Anaesthesiol Scand 2005; **49**: 927–929*
Printed in UK. All rights reserved

ACTA ANAESTHESIOLOGICA SCANDINAVICA
doi: 10.1111/j.1399-6576.2004.00622.x

Factorial design provides evidence to guide practice of anaesthesia

K. KORTTILA[1] and C. C. APFEL[2]
[1]*Department of Anaesthesia and Intensive Care, Helsinki University Hospital, Helsinki, Finland,* [2]*Outcomes Research™ Institute and Department of Anaesthesiology and Perioperative Medicine, University of Louisville, Louisville, KT*

Many scientific articles are written merely to get something published, neglecting the clinician who would like the medical literature to guide their practice. Evidence-based medicine is expected to help in clinical decision-making. Systematic reviews of the literature followed by a meta-analysis of randomized, controlled trials (RCT) have claimed to represent the highest strength of evidence. However, the results published in meta-analyses have not always been confirmed in subsequent large RCTs. An analysis of 12 large RCTs and 19 meta-analyses addressing the same questions found that the outcomes of these large RCTs were not predicted accurately 35% of the time by previously published meta-analyses. Therefore, meta-analyses of several small RCTs do not obviate the need for large, multicentre RCTs, which can still be considered as a gold standard for the development of clinical guidelines or practice plans. Moreover, large RCTs using a factorial design can be highly efficient because they can answer several clinical questions at the same time and offer the only systematic approach to investigate an interaction of combinations in multimodal approaches.

Accepted for publication 20 October 2004

Key words: Controlled trial; factorial design; meta-analysis; randomized.

A RECENT viewpoint published in *Lancet* (1) addressed the concern that medicine is 'moving from the mystic certainty towards scientific uncertainty'. Many articles are written just to get something published, neglecting the clinician reader who would like the literature to guide their practice. A few years ago, a UK newspaper wrote that only 5% of papers published in the *British Medical Journal* reached the minimum standards of scientific soundness and clinical relevance and that in most journals the figure was less than 1% (2). How much does, for example, the following conclusion published in a clinical journal aimed at practitioners guide their practice: 'In this pilot study, the null hypothesis that both treatments will show equal results cannot be confirmed or rejected because of the small number of participants (3)?'. Is it not important for scientific papers to serve the needs of their readers as well as those of their writers?

Evidence-based medicine

Today evidence-based medicine is expected to help in clinical decision-making. It is believed that the strength of evidence from lowest to strongest is as follows: clinical experience, basic research (*in vitro* research and animal studies), observational studies, individual randomized controlled trials (RCTs), and systematic reviews of RCTs (of which a 'quantitative systematic review' is often called a meta-analysis). Large randomized, controlled trials are also generally considered as the gold standard in evaluations of the efficacy of clinical practice plans. Actually, RCTs have been used for a long time to obtain data and evidence for clinicians. Now meta-analyses are widely used to provide evidence to support clinical strategies. As of June 17, 2004, PubMed identifies 16,379 items for 'meta-analysis' alone and 159 items for 'meta-analysis and anaesthesia'; the respective figures for 'RCT' alone and 'RCT and anaesthesia' being 214,833 and 12,661 items, respectively.

Since large RCTs are not always available, clinicians may have to rely on a meta-analysis to support their clinical strategies. However, papers of great variety and different quality have been published under the title 'meta-analysis.' A meta-analysis of, e.g. five RCTs not chosen systemically, does not guide clinicians towards sound patient care. However a thorough,

quantitative, systematic review of the medical literature carried out by two independent investigators, followed by a meta-analysis with point estimates with confidence intervals for the benefits and risks, is likely to give clinicians valuable guidance. The principles applied and taught by the Cochrane Collaboration are very helpful in allowing investigators to perform high-standard, systematic reviews followed by a meta-analysis. However, critics raised many intrinsic weaknesses of meta-analysis (4, 5). For example, pooled results incorporate biases of the individual studies, and despite systematic reviews a selection of studies is not always easy as there is usually an inevitable heterogeneity among them. There is also a publication bias, which refers to the tendency of authors to submit studies with positive results for publication and the tendency of journals to accept them. Many manuscripts with negative results may remain in investigators' drawers and thus are not considered in meta-analyses.

Discrepancies between meta-analyses and large RCTs

LeLorier et al. (6) reported that there are discrepancies between meta-analyses and subsequent large randomized trials. They compared the results of large randomized trials of 1000 patients or more that were published in four major peer-reviewed journals – *New England Journal of Medicine, Lancet, Annals of Internal Medicine* and *Journal of American Medical Association* – with the results of meta-analyses published earlier on the same topics. They identified 12 large RCTs and 19 meta-analyses addressing the same questions and found that the outcomes of these large RCTs were not predicted accurately 35% of the time by previously published meta-analyses. If there had been no subsequent large RCT, the meta-analysis would have led to adoption of an ineffective treatment and to rejection of a useful treatment in 32% of cases. Publication bias and heterogeneity of the trials included in the meta-analysis may explain such discrepancies. Furthermore, concomitant therapies may have changed after some of the smaller RCTs included in meta-analysis were published. The authors recommend that if large, good RCTs have been conducted, practice guidelines should strongly be influenced by their results (6).

Factorial trial design

Well-conducted, large, multicentre RCTs require major efforts and are therefore expensive. However, using a factorial design one large RCT can answer several questions simultaneously and allows assessment of the effect of drug combinations that will help clinicians. To a limited extent, factorial designs have been used in medicine before. However, a recently published study, IMPACT, is the only one to date in which patients were simultaneously randomized to more than three interventions (patients were, in fact, randomized to six interventions) (7, 8). Moreover, it is the first study to date which was adequately powered to allow for a systematic and complete investigation of interactions between up to three factors; a question that is essential if the best combination from a multimodal approach needs to be identified. But, what is the factorial design?

Table 1

Design of a 2×2 factorial design. Note that Factor B can be compared for states 0 and 1 by combining groups 1 and 2 (Factor B, state 0) and groups 3 and 4 (Factor B, State 1).

Group	Factor A	Factor B
1	0	0
2	1	0
3	0	1
4	1	1

The simplest example of a factorial trial design is a so-called 2×2 factorial design (Table 1). It has two factors (A and B) with two different states (0 or 1). The design has several advantages. First, it is possible to compare the combined results of patients who received factor B (groups 3 and 4) with the combined results of patients who did not receive factor B (groups 1 and 2), because factor A is equally distributed between the combined groups. At the same time, the groups in the table can be rearranged so that the effect of factor A can be determined by comparing the combined result of patients who received factor A (groups 2 and 4) with the combined results of patients who did not receive factor A (groups 1 and 3), because factor B is now equally distributed between the combined groups (Table 2).

The effect of factor A and the effect of factor B could have been determined in a study with only *three* groups (Table 3); however, by adding one additional

Table 2

Regrouping in a 2×2 factorial design. Note that Factor A can be compared for states 0 and 1 by combining groups 1 and 3 (Factor A, state 0) and groups 2 and 4 (Factor A, State 1).

Group	Factor A	Factor B
1	0	0
3	0	1
2	1	0
4	1	1

Table 3

Comparison of two interventions in three groups without a factorial design.

Group	Intervention A	Intervention B
1	0	0
2	0	1
3	1	0

group, i.e. using a factorial design, the sample size that can be analyzed for both comparisons was doubled and the combined treatment of A and B can be assessed. As a consequence, when using a factorial design, the size of the groups can be halved (compared to a simple RCT) if the primary intention is to determine only the 'main effects'. What makes the factorial design so highly efficient is that two factors are studied at the same time and each patient is randomized to both interventions.

Despite advantages, factorial designs have rarely been used in clinical trials. The presence of an additional factor makes the group appear more heterogenous, but it is actually not the case, as the distribution of the other factor is completely controlled by the study design. In addition to being able to control several factors affecting the end-points, the factorial design is also the only systematic approach to identify interactions in clinical trials and there is no more efficient way to do so. The fact that a considerably higher sample size is required to analyze the interaction is probably the main reason why factorial designs are rarely performed. However, if the main factors are to be quantified, without investigating interactions, the number of patients needed are far less than those with a conventional approach because each patient provides information for each of the factors. A more detailed explanation with an example of a $2 \times 2 \times 2$ (2^3) factorial design is presented on the internet (*http://www.ponv.org*).

When assessing the single and combined benefits of six interventions for the prevention of postoperative nausea and vomiting, Apfel et al. used a 2^6 factorial design (IMPACT) (7, 8). Each of the 5199 patients was randomly assigned to six interventions: ondansetron (4 mg i.v.) or no ondansetron; dexamethasone (4 mg i.v.) or no dexamethasone; droperidol (1.25 mg i.v.) or no droperidol; propofol or a volatile anaesthetic (isoflurane, desflurane, or sevoflurane); nitrogen or nitrous oxide; and remifentanil or fentanyl. These six treatments led to 64 possible (i.e. 2^6) treatment combinations. The large enrolment and the multifactorial design of the IMPACT study allowed simultaneous evaluation of the antiemetic efficacy of three antiemetic interventions and three anaesthetic interventions and of all possible combinations of two or three interventions. Ondansetron, dexamethasone, and droperidol each reduced the risk of postoperative nausea and vomiting by about 26%; propofol reduced it by 19%; and nitrogen (i.e. omission of nitrous oxide) by 12%. Thus total intravenous anaesthesia (combination of propofol and nitrogen) was as effective as any of the antiemetics alone. All the interventions acted independently of one another and independently of the patient's baseline risk for postoperative nausea and vomiting. Since relative risk reduction was independent from other interventions or factors, the absolute risk reduction was critically dependent on the patient's baseline risk.

So far, IMPACT is the only RCT in medicine that has allowed for the analysis of three-factor interactions, and it provided reliable evidence to clinicians to guide the practice of anaesthesia. In the future, RCTs with a factorial design may help us to get more information for clinical decision making.

References

1. O'Donnell M. Evidence-based illiteracy: time to rescue 'the literature'. *Lancet* 2000; **355**: 489–91.
2. Boseley S. News Report. The Guardian, June 24, 1998.
3. Lamers HJ, Jamin RH, Zaat JO *et al.* Dietary advice for acute diarrhoea in general practice: a pilot study. *Br J Gen Pract* 1998; **48**: 1819–23.
4. Feinstein AR. Meta-analysis. statistical alchemy for the 21st century. *J Clin Epidemiol* 1995; **48**: 71–9.
5. Bailar JC III The practice of meta-analysis. *J Clin Epidemiol* 1995; **48**: 149–57.
6. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomised controlled trials. *N Engl J Med* 1997; **337**: 536–42.
7. Apfel CC, Korttila K, Abdalla M *et al*. An international multicenter protocol to assess the single and combined benefits of antiemetic interventions in a controlled clinical trials of a $2 \times 2 \times 2 \times 2 \times 2 \times 2$ factorial design (IMPACT). *Contr Clin Trials* 2003; **24**: 736–51.
8. Apfel CC, Korttila K, Abdalla M *et al.* for the IMPACT Investigators: a factorial trail of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med* 2004; **350**: 2441–51.

Address:
Prof. Kari Korttila
Department of Anaesthesia and Intensive Care
Helsinki University Central Hospital
PO Box 140
FIN-00029 HUS
Helsinki
Finland
e-mail: kari.korttila@hus.fi