Tim Peters
Professor of Primary Care Health Services Research
Head of the School of Clinical Sciences

Design, analysis and reporting of 2×2 factorial trials

Structure of this talk

- Factorial RCTs – definition and potted history
- Rationale, indications and contra-indications
- Key issues in design
- Key issues in analysis
- Reporting
- Complexities
- Conclusions

Design of simple 2x2 factorial trial

		Treatment A		
		Yes	**No**	**Margin**
Treatment B	**Yes**	Both A & B (AB)	B alone (OB)	All B
	No	A alone (AO)	Neither A nor B (OO)	All non-B
Margin		All A	All non-A	

THE THERAPEUTIC VALUE OF DIGITALIS IN PNEUMONIA*

JOHN WYCKOFF, M.D.
EUGENE F. DU BOIS, M.D.
AND
I. OGDEN WOODRUFF, M.D.
NEW YORK

In the present study it was decided that patients should be selected for the digitalis or nondigitalis groups in the following way: Patients were received into the "pneumonia series" according to the date and hour of admission, and alternate patients were treated with serum. The combination of serum and digitalis therapy led to the grouping of patients into four classes, selected only by the time of admission, and termed arbitrarily A, B, C and D. The treatment of the patients in these four classes was as follows:

Class A received neither serum nor digitalis.
Class B received serum only.
Class C received digitalis only.
Class D received both serum and digitalis.

This system of classification operated in each ward independent of other wards, so that factors of general care and nursing might be the same for each class of treatment.

Wyckoff J, et al (1930). The therapeutic value of digitalis in pneumonia. JAMA 95:1243-9.

THE LANCET

Volume 247, Issue 6407, 15 June 1946, Pages 881-883

ORIGINAL ARTICLES

DIET IN THE TREATMENT OF INFECTIVE HEPATITIS: THERAPEUTIC TRIAL OF CYSTEINE AND VARIATION OF FAT-CONTENT

Clifford Wilson D.M. Oxfd,, M. R. Pollock M.B. Camb. and A. D. Harris M.R.C.S.

(1): Cysteine supplementation / Low fat diet
(2): Cysteine supplementation / High fat diet
(3): No cysteine supplementation / Low fat diet
(4): No cysteine supplementation / High fat diet

Total N=103

Cite this as: *BMJ* 2010;340:c869
doi: 10.1136/bmj.c869

RESEARCH METHODS & REPORTING

CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials

David Moher,[1] Sally Hopewell,[2] Kenneth F Schulz,[3] Victor Montori,[4] Peter C Gøtzsche,[5] P J Devereaux,[6] Diana Elbourne,[7] Matthias Egger,[8] Douglas G Altman[2]

"Most RCTs have [paral
substantial minority do not: 45% of RCTs published in December 2000, and 39% in December 2006."

JAMA®

Analysis and Reporting of Factorial Trials: A Systematic Review

Finlay A. McAlister; Sharon E. Straus; David L. Sackett; et al.

JAMA. 2003;289(19):2545-2553 (doi:10.1001/jama.289.19.2545)

http://jama.ama-assn.org/cgi/content/full/289/19/2545

Analysis and Reporting of Factorial Trials
A Systematic Review

Finlay A. McAlister, MD, MSc

Sharon E. Straus, MD, MSc

David L. Sackett, MD, MSc

Douglas G. Altman, DSc

Context Although factorial trials have become common, standards for the analysis and reporting of such trials have not been established and, despite concerns about the possibility of unrecognized interactions between therapies in factorial trials, the magnitude of this potential problem is unknown.

Objective To examine the rationale, methods, and analysis of randomized factorial trials.

Data Sources and Study Selection We searched MEDLINE, EMBASE, and the Cochrane Controlled Trials Register using the terms *factorial*, *interaction*, *2×2*, *2 by 2*, and *incremental* to identify factorial randomized trials published from January 2000 to July 2002. To identify trials missed by the electronic search, we performed a hand search of English-language trials in a defined topic area (using the term *myocardial*

FACTORIAL RANDOMIZED TRIALS permit investigators to evaluate 2 (or more) interventions in a single experiment. In its simplest 2×2 form, shown in TABLE 1, par-

BMC Medical Research Methodology

BioMed Central

Debate

Open Access

Design, analysis and presentation of factorial randomised controlled trials

Alan A Montgomery*[1], Tim J Peters[1] and Paul Little[2]

Address: [1]Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6JL, UK and [2]Community Clinical Sciences Division (Primary Medical Care Group), Faculty of Medicine, Health and Biological Sciences, Southampton University, Aldermoor Health Centre, Southampton SO16 5ST, UK

Email: Alan A Montgomery* - alan.a.montgomery@bristol.ac.uk; Tim J Peters - tim.peters@bristol.ac.uk; Paul Little - p.little@soton.ac.uk

* Corresponding author

Published: 24 November 2003

BMC Medical Research Methodology 2003, 3:26

Received: 31 July 2003
Accepted: 24 November 2003

This article is available from: http://www.biomedcentral.com/1471-2288/3/26

Review in Trials 2011

Reporting of factorial trials of complex interventions in community settings: a systematic review

Alan Montgomery, Margaret Astin, Tim Peters

Montgomery *et al. Trials* 2011, **12**:179

Design of simple 2x2 factorial trial

		Treatment A		
		Yes	**No**	**Margin**
Treatment B	**Yes**	Both A & B (AB)	B alone (OB)	All B
	No	A alone (AO)	Neither A nor B (OO)	All non-B
Margin		All A	All non-A	

To 'factorialise' or – key issues

- Nature of interventions
- Study context including participants
- Comparisons of interest
- Sample size constraints
- Outcome measures (scale/model)

Indications – 1

- Both interventions can be used in conjunction with each other
- Unit(s) of randomisation and timing of allocations means the design is feasible
- Intended model for (especially binary) outcome measure is not likely to in itself introduce an interaction

Indications – 2

- When treatments act ***independently***
  - effect of A is same in presence/absence of B
  - additive effects when used together
- When ***interaction*** is of primary interest
  - if combined effect less than or greater than additive
  - usually results in a much larger sample size

Contraindications

- When interaction is suspected, and either:
  - not of primary interest
  - not detectable within available sample size/resources

Then require either (effectively) separate trials, and certainly separate analyses, for A & B

Sheikh et al *BJGP 2002;52:746-751.*

“ … there is need for either prior knowledge of interaction between treatments or the ability to reliably detect or exclude such an interaction in the new study are limiting factors in the use of this trial design. ”

Agreed, but ...

- Potential for efficiency gains should not be overlooked
- Even under-powered investigation of interaction effects may be of interest and value long-term

- So is this the go-to design for >1 intervention or a BOGOF mirage?

[a question for debate ...]

Factorial RCTs – examples from my experience

- Montgomery et al [BJGP 2003;53:446-453] Decision Analysis and video/leaflet for newly diagnosed hypertension

- Richards/Bankhead et al [J Med Screen 2001;8:91-98/99-105] flag in notes and GP letter for breast screening attendance

Design issues – sample size

- Usually powered to detect *main effects*:
  - perform separate calculations for each intervention and use the larger of the two (or more)

- For the *interaction*:
  - based on magnitude of interaction considered important, often very difficult to specify
  - usually results in much larger study size (akin to a subgroup analysis in a parallel groups RCT)
  - in practice, main problem then is inability to rule out an interaction with any degree of confidence

Analytical issues

- Additional to standard CONSORT guidelines
- Type of outcome and, for binary measures, nature of the model
  - what is multiplicative on one scale can be additive on another
- Primary hypothesis
  - ‘average’ effects
  - differential effect/interaction
- Analysis ‘at the margins’

Analysis ‘at the margin

		Decision Analysis		
		Yes	**No**	
Information video/leaflet	**Yes**	Both	VL alone	All VL
	No	DA alone	Neither	All non-VL
		All DA	All non-DA	

Analysis 'inside the ta

		Decision Analysis		
		Yes	**No**	
Information video/leaflet	**Yes**	Both	VL alone	All VL
	No	DA alone	Neither	All non-VL
		All DA	All non-DA	

Was there any interaction?

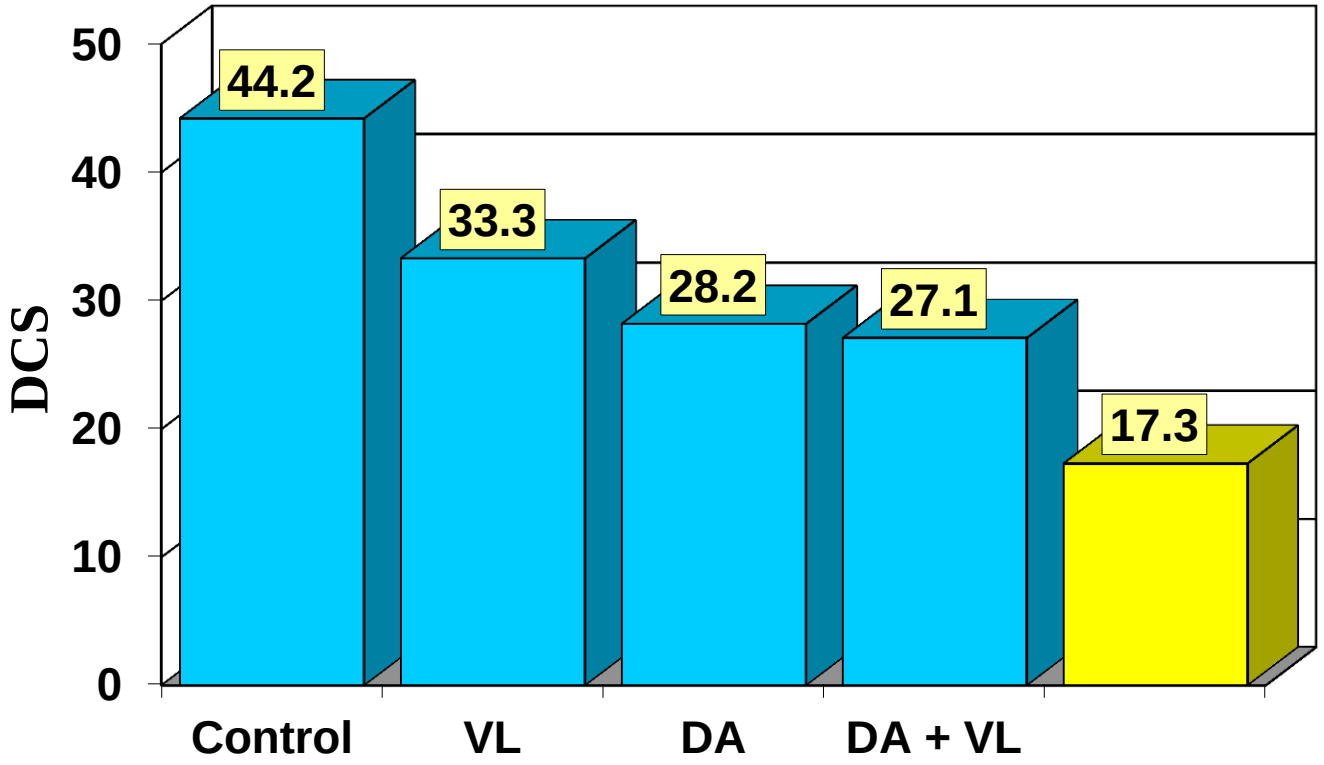
		Decision Analysis	
		Yes	**No**
Information video/leaflet	**Yes**	27	33
	No	28	44

Outcome is Total Decisional Conflict score, with lower values representing less uncertainty about anti-hypertensive treatment

So less a case of BOGOF and more one of GONOFT?

Interaction: continuous outcome

Decisional Conflict Score:



Interaction = 12.5 (5.4 to 19.5), p=0.001

Analysis of a 2×2 factorial RCT

- *Main effects*:
  - usual basis for the *a priori* power calculation
  - each intervention adjusted for the other (as well as other design variables) in a regression model
  - analysis realises full precision benefits

- *Interaction*:
  - term in regression model for interaction – ‘extra’ effect of receivi
  - usually underpowered

Reporting of results

- Descriptive statistics for outcome at follow-up
  - for each factorial cell
  - at the margins
- Primary comparative analyses
  - estimate with 95%CI and p-value *at the margins*
  - one for each intervention in the trial
- Secondary analyses should include
  - *interaction* estimate with 95%CI and p-value

Reporting – how did we do (in the past!)

- RCT of DA vs. VL
  - followed main hypotheses by focussing on the margins
  - noted magnitude and p-value for (antagonistic) interaction
  - BUT, perhaps should have given more prominence to outcomes inside the table, and acknowledged consequent limitations w.r.t. precision
  - impact was to underestimate the separate effects of the interventions while overestimating (at least, potentially over-stating) their combined effect

Reporting – how did we do (in the past!)

- Individually randomised breast screening trial
  - presented both at the margins and inside the table results
  - no evidence of interaction and could rule out antagonistic effect so no concerns about overstating effects or power
- Cluster randomised breast screening trial
  - presented both at the margins and inside the table results
  - no evidence of interaction b
    effects in either direction, especially antagonistic
  - Concomitant findings for main effects to some extent got us off the hook (one was more clearly effective and cost-effective) BUT we may have under-appreciated the potential impact on (and of) precision here too
- Economic evaluation raise

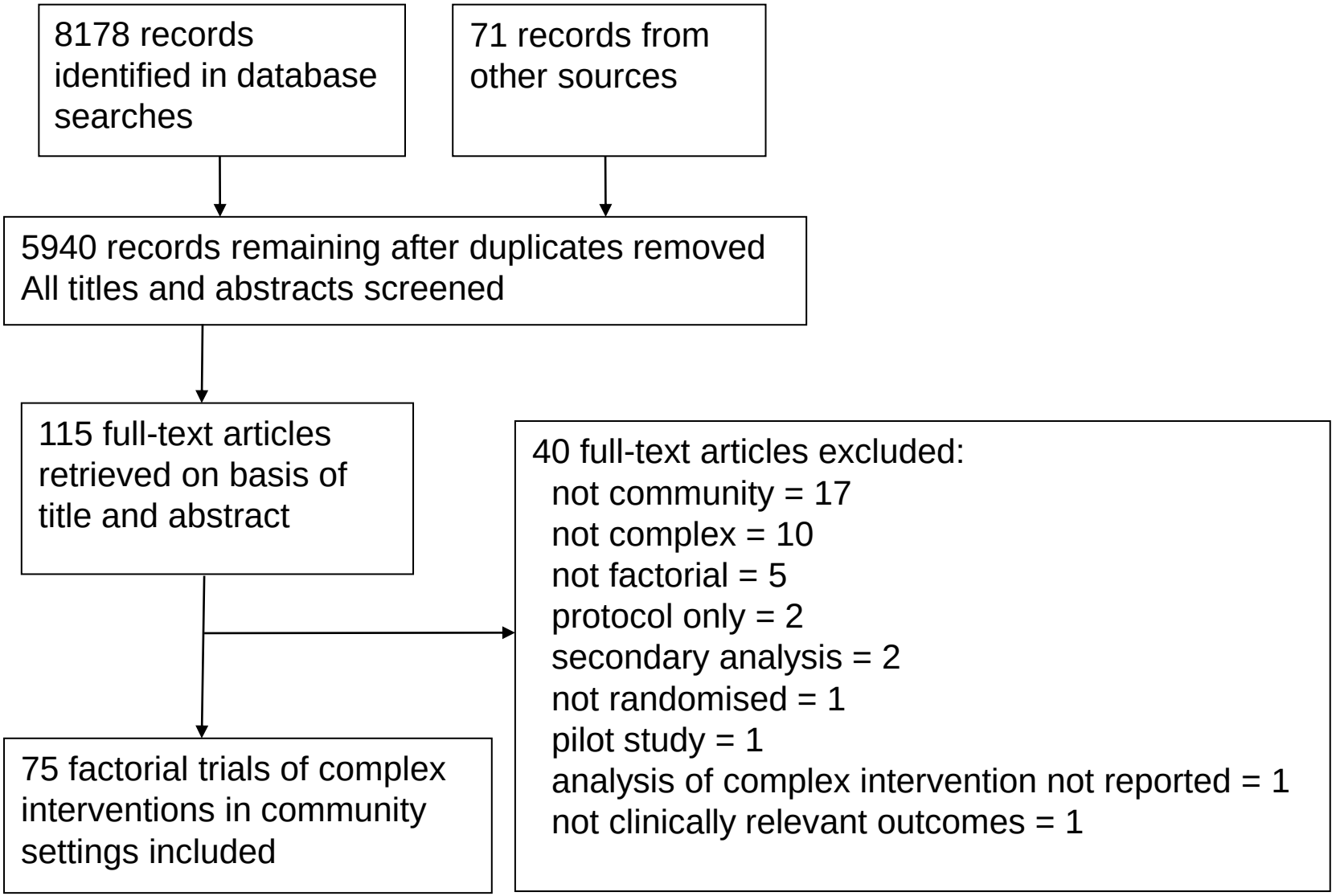
Complexities

- Economic evaluations
  - in practice, margins are unrealistic
  - intervention (combinations) to be compared are thus arguably ‘inside the table’
- Cluster trials
  - interventions can be at different levels
  - antagonistic interactions more common (especially for behavioural interventions), and these give more power problems
- Complex interventions
  - recently published systematic review on the reporting of factorial trials of complex interventions in community settings
    [Montgomery *et al. Trials* 2011,12:179]

What makes an intervention complex?

- Number of interacting components
- Number and difficulty of behaviours required
- Number of groups targeted
- Number and variability of outcomes
- Degree of tailoring permitted

Results (1) – Information flow



Results (2) – Description of studies

Year	N (%)
2000-2003	27 (36)
2004-2009	48 (64)

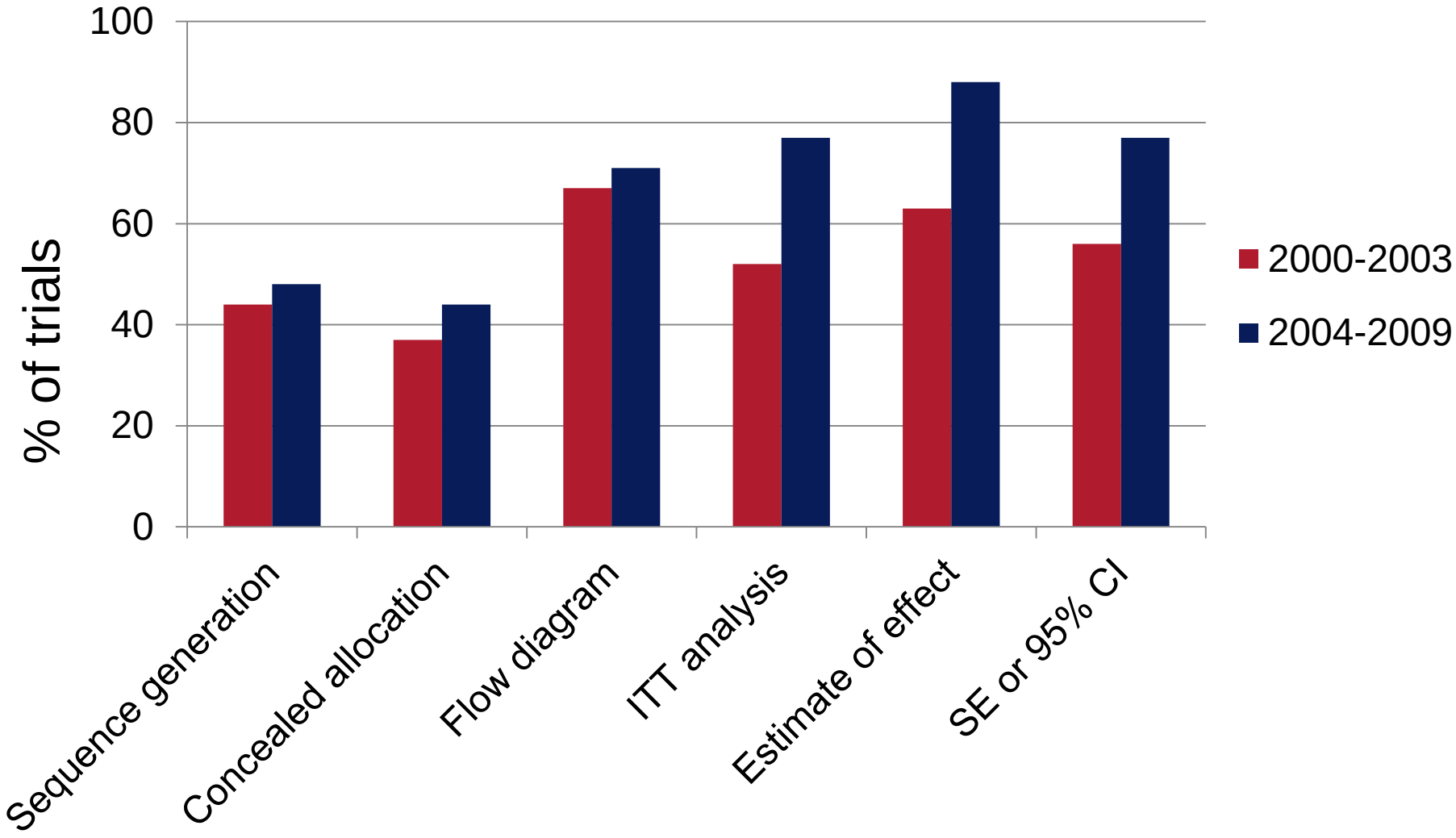
Age Group	N (%)
Adults	68 (85)
Children	4 (5)
Mixed	3 (4)

Country	N (%)
USA	27 (36)
UK	26 (35)
Other Europe	10 (13)
Canada	5 (7)
Other	7 (9)

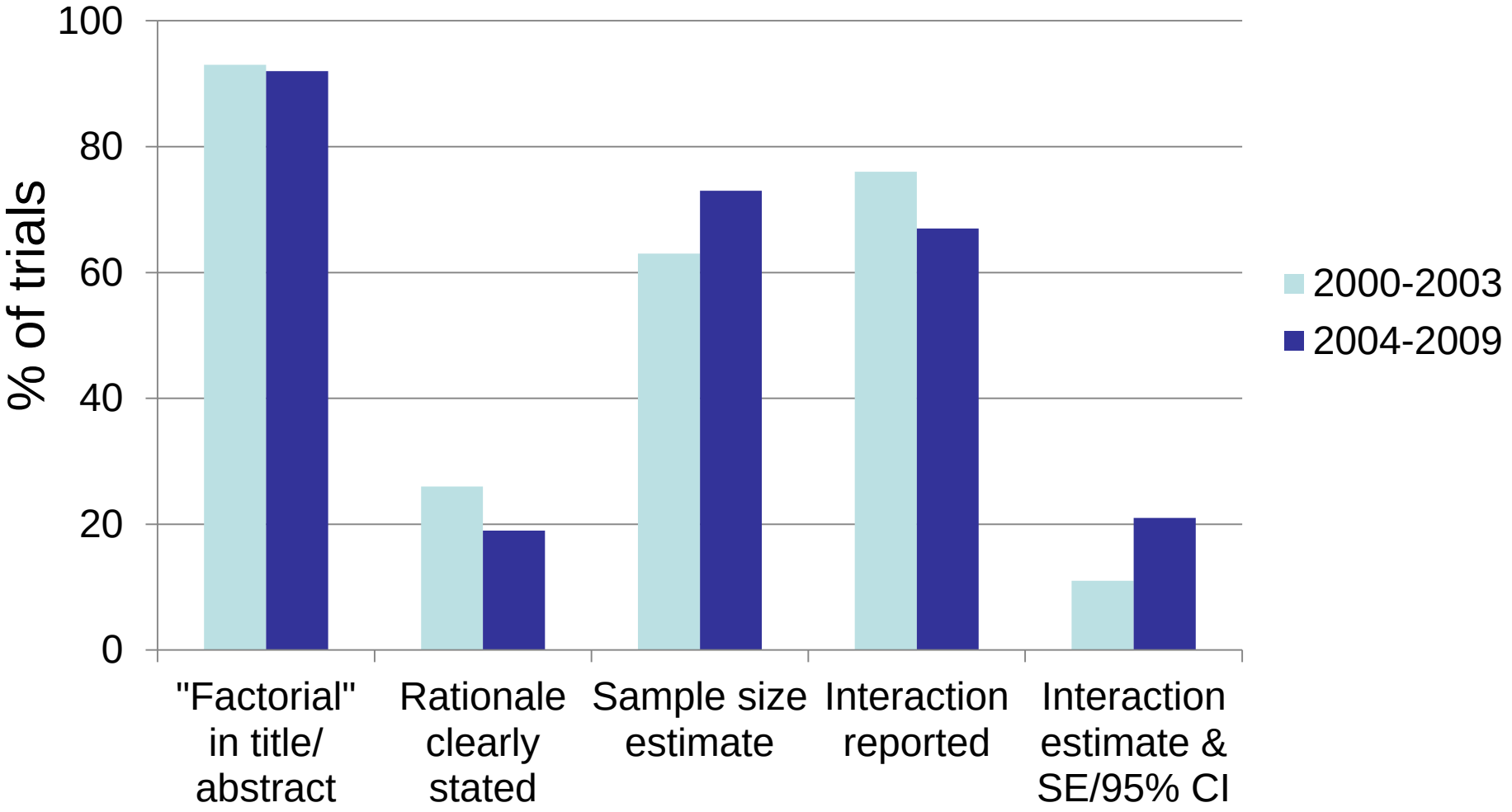
Number of target conditions/behaviours = 51

- acute respiratory infection
- cancer screening
- smoking cessation
- weight management/diet/physical activity

Results (3)


100
80
60
40
20
0
% of trials
Sequence generation
Concealed allocation
Flow diagram
ITT analysis
Estimate of effect
SE or 95% CI
2000-2003
2004-2009

Results (4)


100
80
60
40
20
0
% of trials
"Factorial" in title/ abstract
Rationale clearly stated
Sample size estimate
Interaction reported
Interaction estimate & SE/95% CI
2000-2003
2004-2009

Conclusions – overall

- Design & analytical issues not always fully understood
- Independence of effects uncertain in advance but usually low power to detect interactions
  - attraction of 2-for-1 is strong, but …
  - if interaction present, 'inside t consequences for precision …
- Reporting of factorial trials could be better
  - randomisation; sample size calculation; participant flow; quantifying interaction
- Poor reporting hampers judgement of quality
- CONSORT extension

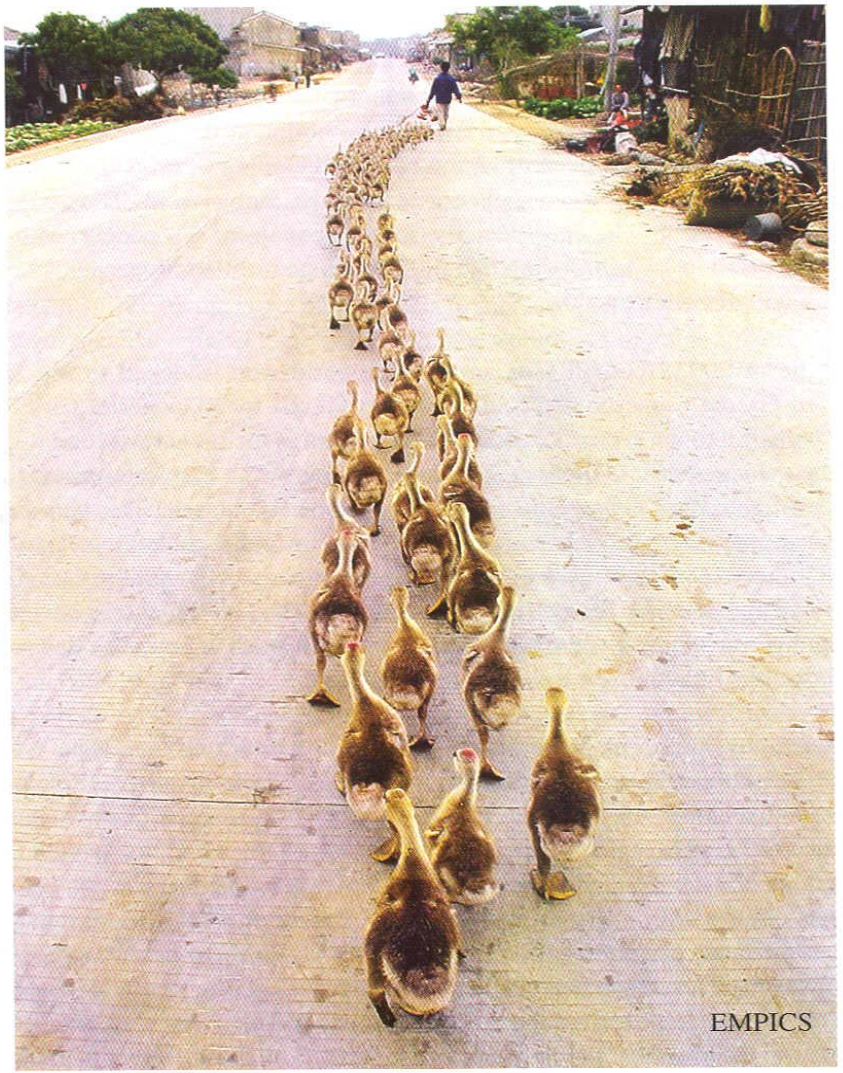
Cite this as: *BMJ* 2010;340:c869
doi: 10.1136/bmj.c869

RESEARCH METHODS & REPORTING

CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials

David Moher,[1] Sally Hopewell,[2] Kenneth F Schulz,[3] Victor Montori,[4] Peter C Gøtzsche,[5] P J Devereaux,[6] Diana Elbourne,[7] Matthias Egger,[8] Douglas G Altman[2]

"...other planned extens
been completed; they will cover trials with the following designs: multiarm parallel, **factorial**…. "



A man takes his goose chicks for a walk in Taiping, about 1,100 miles south of Beijing, (AP Photo/Eugene Hoshiko)

Acknowledgements, thanks

- NIHR Senior Investigator award
- RSS
- Clare Bankhead, Suzanne Richards, Paul Little, Margaret Astin
- Alan Montgomery