ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

The balanced incomplete block design is not suitable for the evaluation of complex interventions

Jasper Trietsch[a,*], Pieter Leffers[b], Ben van Steenkiste[a], Richard Grol[c], Trudy van der Weijden[a]

[a]*Department of General Practice, School for Public health and Primary Care (CAPHRI), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands*

[b]*Department of Epidemology, Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands*

[c]*IQ Healthcare, Radbout University Nijmegen, PO Box 9101 (144), 6500HB Nijmegen, The Netherlands*

Accepted 8 July 2014; Published online xxxx

Abstract

Objectives: In quality of care research, the balanced incomplete block (BIB) design is regularly claimed to have been used when evaluating complex interventions. In this article, we reflect on the appropriateness of using this design for evaluating complex interventions.

Study Design and Setting: Literature study using PubMed and handbooks.

Results: After studying various articles on health services research that claim to have applied the BIB and the original methodological literature on this design, it became clear that the applied method is in fact not a BIB design.

Conclusion: We conclude that the use of this design is not suited for evaluating complex interventions. We stress that, to prevent improper use of terms, more attention should be paid to proper referencing of the original methodological literature.

Keywords: Research methodology; Evaluation studies; Primary care; Health care evaluation mechanisms; Quality of health care; Epidemiologic research design

1. Introduction

Incorrect citing in scientific literature can lead to a chain of erroneous interpretations or the use of incorrect methods or terms. In this short report, we describe the case of the balanced incomplete block (BIB) design. In quality of care research, the BIB design is regularly mentioned in the sections reporting on the applied study design. Study of the literature on the BIB design led us to original work by Cochran and Cox from the 1950s describing a design that appeared to be very different from the designs that were actually used in the quality of care research articles claiming to have applied the BIB. This raised questions about possible consequences of improper reporting of the BIB design for our own research with respect to the interpretation of study results.

In this article, we describe the features of the original BIB design, and we compare it with the design that we used in our earlier research projects. We illustrate where we failed in citing of the literature on the BIB design, and we describe the implications of this error for the validity of conclusions of our research.

1.1. Features of the original BIB design

Comparing the outcomes of multiple interventions under various conditions is a well-known challenge in experimental research. It is often impossible to carry out such studies because of limitations on the number of available research subjects and because of limited resources. In agriculture, a solution for this problem was developed for crop-optimization studies by statisticians involved in combinatorial problems research. These methods were thoroughly described by Cochran and Cox [1]. For instance, when testing several new genetic varieties of corn under different growth conditions, vast areas of land would be needed. The BIB design enables researchers to compare harvest returns of the varieties using plots of land (blocks), which each have different conditions between but uniform conditions within these plots. Not all varieties of corn will be grown on each of these plots: incomplete testing. By balancing the allocation of the varieties over the different

Funding: The authors received funding for this article from The Netherlands Organisation for Health Research and Development (grant number 945-17-102) (ZonMw); project number 94517102 and OWM Centrale Zorgverzekeraars group (grant number H05SU21), Zorgverzekeraar UA (CZ insurance company).

* Corresponding author. Tel.: +31 43 3882302; fax: +31 43 3619344.
E-mail address: Jasper.Trietsch@maastrichtuniversity.nl (J. Trietsch).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2014.07.006

What is new?

Key findings

- Contrary to what has been claimed in various publications reporting on quality of care research, the balanced incomplete block (BIB) design has not been used. We show that the use of this design is not suited for this type of research.

What this adds to what was known?

- The BIB design dates from the first half of the last century. Because of its efficiency increasing potential, it can be an attractive design when resources are limited and when certain prerequisites are met. We show that indirect referencing has led to improper use of the name of the design in quality of care research and that the design should also not be applied in the evaluation of complex interventions.

What is the implication and what should change now?

- To increase transparency and enable critical appraisal of evidence, authors and peer reviewers should pay more attention to appropriate use of terms such as BIB design.

Box 1 General theorem for BIB designs

$bk = vr$ and
$r(k-1) = \lambda(v-1)$
When,
$v > k > 0$
$r > 0$
$\lambda > 0$

v = the number of compared (genetic) varieties
b = the number of plots of land (blocks)
r = the number of plots in which each variety is present
k = the number of varieties per plot
λ = the number of plots in which each pair of varieties is present (pair replication)

plots, a comparison of outcome (eg, returns in harvest) can be made between varieties that were never really compared under the same conditions. The simplest example is that of difference in harvest between varieties A and C under condition I can be estimated, although they were not directly tested against each other under condition I. Under the assumption of the absence of effect modification by the condition, the harvest difference between A and C can be calculated by comparing varieties A and B under condition I and varieties B and C under condition II. Statistical testing is done using analysis of variance [1,2].

To speak of a BIB design, several conditions have to be met. A theorem is available to test whether a design meets the requirements of a BIB design. The notation of the parameters of a BIB design is $\{b,r,k,v,\lambda\}$ (Box 1). Fig. 1 shows a {3,2,2,3,1} design: a design of three blocks with a variation replication rate of 2, two varieties per block, three different varieties, and a pair replication rate of 1, as described previously.

1.2. Features of the BIB design as applied in quality improvement research

According to Cochran and Cox, the BIB design is suitable for situations in which repeated testing of varieties will lead to the same result, as can be expected when conditions can be well controlled such as in agricultural or laboratory sciences. Unfortunately, in most types of clinical research, patients will be permanently influenced by the intervention that is being evaluated and therefore repeated testing cannot be expected to lead to the same result. As a consequence, the BIB design cannot be used for patient-centered research. However, several publications within quality of care research on complex health care interventions report on the use of the BIB [3–6], and it was also advocated as appropriate for complex guideline implementation trials [7].

Testing all components of the complex interventions separately is generally not possible because of limited resources or limitations on the number of available research subjects, let alone that all components can be tested under the various conditions. The applied design is claimed to overcome these limitations while it is also considered to be attractive because it controls for the Hawthorne effect [8]. However, from the published reports, it can be concluded that the BIB design was not applied at all.

Our research group has used the same design in studies on complex quality improvement interventions [9–11]. An example is the work of Verstappen et al. They performed a cluster randomized controlled trial and claimed

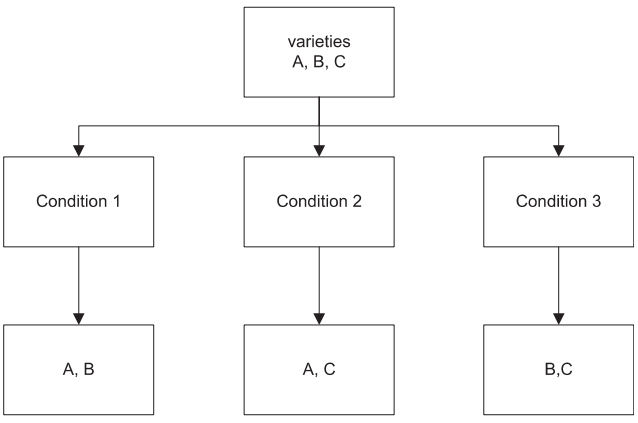

Fig. 1. Example of a balanced incomplete block design with {3,2,2,3,1} design.

to have applied the BIB design. The strategy under study was the "small group peer review," and the aim was reducing inappropriate test ordering by general practitioners. Thirteen general practitioners (GP) groups participated as units of research. They were randomized to receive feedback on one of two sets of clinical topics relating to the treatment of their patients (incomplete testing). Both arms acted as "placebo comparator" for the other arm with the treatment and the placebo aspect of each treatment completely linked. Therefore, this was a cluster randomized trial with only two conditions. Fig. 2 shows the typical design used in such trials. Put in terms of the BIB design, there are two varieties (GP groups with their allocated treatments; $v = 2$) under evaluation. There are only two blocks (two variations of the intervention; $b = 2$), each variety appears only once ($r = 1$), the number of varieties per block is 1 ($k = 1$), and the number of blocks in which each pair of varieties appear is 2 ($\lambda = 2$). From checking these numbers with the conditions in Box 1, it is immediately clear that these studies do not fulfill the criteria for a BIB design.

Recently, we completed the evaluation of a complex implementation strategy based on audit, provision of feedback, and educational materials with peer group discussion guided by clinical guidelines, claiming to use the BIB design [12]. This strategy was implemented in groups of GPs from the same region and was moderated by community pharmacists and test ordering experts (local opinion leaders). We integrated the small group peer review strategy in routine health care, which resulted in an extremely pragmatic trial [13]. With hindsight, we now conclude that our strategy also did not meet the criteria of a proper BIB design.

Figures 1 and 2 show schematic representations of the simplest possible BIB design and of a typical trial in quality improvement research. It is immediately clear that a two-armed trial cannot be a BIB design.

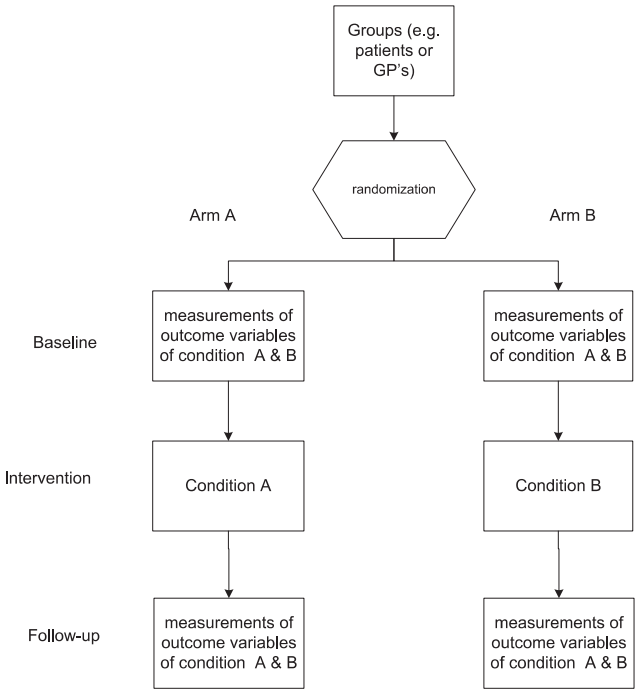

Fig. 2. A typical design used in quality improvement research.

1.3. What went wrong and what is the impact?

Eccles and Grimshaw described using the BIB design in a similar setting in the area of quality improvement research as we did [3]. When checking the references of their published studies, we were struck by the fact that authors regularly refer to earlier work from colleagues in the field and not to the original publications on the BIB design. The BIB design can be traced back in time to the work of Cochran and Cox and further back to the work of Bose and that of Yates and even to the famous "Kirkman's schoolgirl problem" [1,14,15]. It became clear that the used design does not match with the description of the BIB design by Cochran and Cox at all. The fact that the description is difficult to understand for non-statisticians makes it understandable that authors tend not to refer to it or refer to it without having read and understood the contents. This way, an interpretation error made by one group of authors has led to systematic inappropriate use of a methodological term. We think that the smaller the number of researchers in the field the less probable it seems that others will notice such an error.

The question remains what the implications of the improper use of the BIB design are. Fortunately, we found that the authors performed appropriate statistical analyses for the design that they actually used. The validity of the conclusions of the publications should be judged on the assumptions that underlie the comparisons in these studies. The main assumption is that the feedback strategies can act as placebo strategies for each other. The improper use of the term "BIB design" has no bearing on this.

2. Conclusions

The consequences of the improper citing to the BIB design have been very limited. However, authors, reviewers, and editors should be alert when indirect citations are used maybe especially when complex methodological or statistical methods are involved that are not commonly used. In the present case, indirect referencing has only led to improper use of a beautiful and complicated name for a very standard study design. As admitting not being acquainted with a design may feel uncomfortable for many, we fear that the unjustified use of a complicated term will impede critical debate about the validity of methods that study people. It certainly leads to failure to honor developers of research methodology for their contributions to science.

Acknowledgments

There are no contributors to this article besides the authors.

Author contributions: J.T. conceived of the article and drafted the manuscript. J.T. and P.L. researched literature on the subject. B.v.S., T.v.d.W., and R.G. read and corrected previous versions of the manuscript. All authors read and approved the final manuscript.

References

[1] Cochran WG, Cox GM. Experimental designs. New York, NY: Wiley; 1950.

[2] Wakeling IN, Buck D. Balanced incomplete block designs useful for consumer experimentation. Food Qual Prefer 2001;12(4):265–8.

[3] Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. BMJ 2002;325:941.

[4] Kasje WN, Denig P, Stewart RE, de Graeff PA, Haaijer-Ruskamp FM. An educational programme for peer review groups to improve treatment of chronic heart failure and diabetes mellitus type 2 in general practice. J Eval Clin Pract 2006;12:613–21.

[5] Campbell BF, Sengupta S, Santos C, Lorig KR. Balanced incomplete block design: description, case study, and implications for practice. Health Educ Q 1995;22(2):201–10.

[6] Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D. An evaluation of computerised guidelines for the management of two chronic conditions. Newcastle, UK: Centre for Health Services Research; 2002.

[7] Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Fam Pract 2000;17(Suppl 1):S11–6.

[8] Verstappen WHJM, van der Weijden T, ter Riet G, Grimshaw J, Winkens R, Grol RPTM. Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. J Clin Epidemiol 2004;57:1119–23.

[9] Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, Grimshaw J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial. JAMA 2003;289:2407–12.

[10] Verstappen WHJM, van der Weijden T, Dubois WI, Smeele I, Hermsen J, Tan FES, et al. Improving test ordering in primary care: the added value of a small-group quality improvement strategy compared with classic feedback only. Ann Fam Med 2004;2(6):569–75.

[11] Bindels R, Hasman A, Kester AD, Talmon JL, de Clercq PA, Winkens RAG. The efficacy of an automated feedback system for general practitioners. Int J Med Inform 2003;11:69–74.

[12] Trietsch J, van der Weijden T, Verstappen W, Janknegt R, Muijrers P, Winkens R, et al. A cluster randomized controlled trial aimed at implementation of local quality improvement collaboratives to improve prescribing and test ordering performance of general practitioners: study Protocol. Implement Sci 2009;4:6.

[13] Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464–75.

[14] Bose RC. On the construction of balanced incomplete block designs. Ann Hum Genet 1939;9(4):353–99.

[15] Yates F. Incomplete randomized blocks. Ann Hum Genet 1936;7(2):121–40.