

Research article

Open Access

The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C?

Stuart G Baker*¹ and Barnett S Kramer²

Address: ¹Division of Cancer Prevention, National Cancer Institute and ²Offices of Disease Prevention and Medical Applications of Research, National Institutes of Health

E-mail: Stuart G Baker* - sb16i@nih.gov; Barnett S Kramer - KramerB@OD.NIH.GOV

*Corresponding author

Published: 13 November 2002

BMC Medical Research Methodology 2002, 2:13

This article is available from: <http://www.biomedcentral.com/1471-2288/2/13>

© 2002 Baker and Kramer; licensee BioMed Central Ltd. This article is published in Open Access: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Received: 13 September 2002

Accepted: 13 November 2002

Abstract

Background: If intervention A bests B in one randomized trial, and B bests C in another randomized trial, can one conclude that A is better than C? The problem was motivated by the planning of a randomized trial, where A is spiral-CT screening, B is x-ray screening, and C is no screening. On its surface, this would appear to be a straightforward application of the transitive principle of logic.

Methods: We extended the graphical approach for omitted binary variables that was originally developed to illustrate Simpson's paradox, applying it to hypothetical, but plausible scenarios involving lung cancer screening, treatment for gastric cancer, and antibiotic therapy for clinical pneumonia.

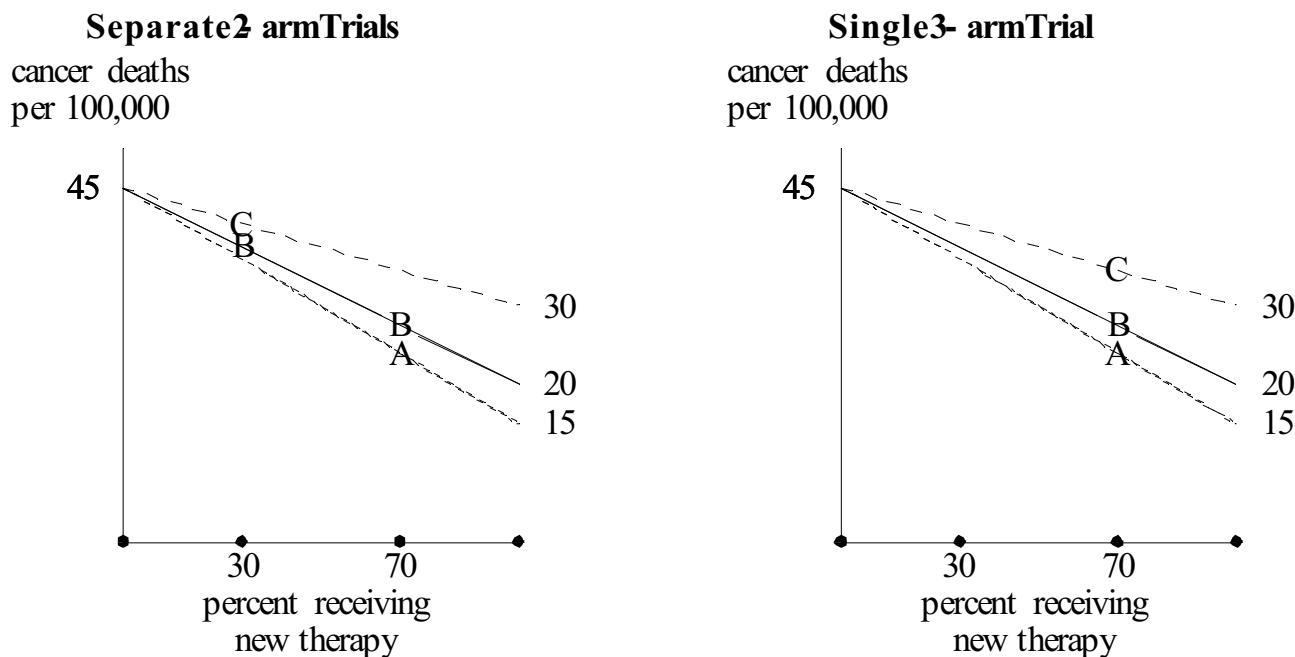
Results: Graphical illustrations of the three examples show different ways the transitive fallacy for randomized trials can arise due to changes in an unobserved or unadjusted binary variable. In the most dramatic scenario, B bests C in the first trial, A bests B in the second trial, but C bests A at the time of the second trial.

Conclusion: Even with large sample sizes, combining results from a previous randomized trial of B versus C with results from a new randomized trial of A versus B will not guarantee correct inference about A versus C. A three-arm trial of A, B, and C would protect against this problem and should be considered when the sequential trials are performed in the context of changing secular trends in important omitted variables such as therapy in cancer screening trials.

Background

Consider three statements: A, B, and C. In formal logic, if A implies B, and B implies C, then A implies C. This can be illustrated by a Venn diagram in which set A lies entirely within set B and set B lies entirely within set C. This implies set A lies entirely within set C. This is an example of transitivity. Extending this logical construct to the design and interpretation of clinical trials, one might conclude

that (even with large sample sizes) if in a randomized trial, intervention A is shown superior to intervention B, and in another randomized trial intervention B is shown superior to C, then A will be shown superior to C in another randomized trial.

**Figure 1**

Screening for lung cancer. Hypothetical results are shown for A = spiral CT, B = x-ray, and C = no screening. In the sequential study, results for B and C are similar in the first trial (when 30% of the subjects receive the new therapy), and results for A and B are similar in the second trial (when 70% of the subjects receive the new therapy). However, as shown with three-arm trial, it is incorrect to make the transitive inference that when 70% of the subjects receive the new therapy, the results for A and C would be similar.

However, statistical association is not generally transitive. Consider three random variables: A, B, C. If A is positively correlated with B, and B is positively correlated with C, then A may or may not be positively correlated with C [2].

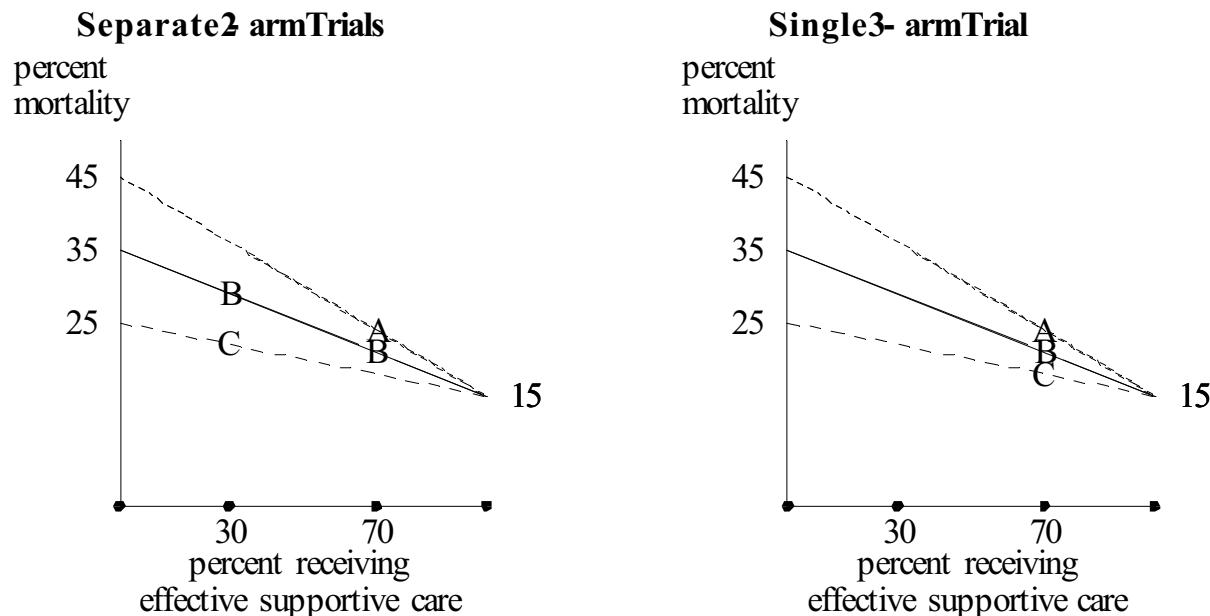
To our knowledge no one has investigated transitivity of results from separate randomized trials. In fact, for the sake of perceived efficiency and limited resources, the principle of transitivity is sometimes assumed in clinical trial design. The possibility of a transitive fallacy surfaced in discussions about a planned randomized clinical trial to assess efficacy of low dose spiral computed tomography (CT) for lung cancer screening.

The aim of the investigation is the use of graphical methods to explore conditions when transitive inference for randomized trials does not hold. As the mathematician, John Allen Paulos writes, "It's odd that logical acuity, rather than helping one to clarify statements, often reveals hidden ambiguities within them. Instead of leading one to form more conclusions, it makes clear that fewer conclusions are justified." [3]

Methods

We extended the graphic in [1,4] for illustrating Simpson's paradox. In the Simpson's paradox graphic, the horizontal axis is the fraction of subjects with an unobserved variable and the vertical axis is the probability of a binary outcome. One diagonal line in the plot represents the effect of unobserved variable on the probability of outcome, given treatment A. A second diagonal line, which is lower and parallel to the first, represents the effect of the unobserved variable on the probability of outcome, given treatment B. The graphic shows that if the fraction with the unobserved variable differs between groups receiving A and B (as in an observational study), then subjects receiving treatment B could have a higher probability of outcome than subjects receiving A, even though the line for A is higher than the line for B.

For this investigation, we added a third treatment to the graphic and investigated three hypothetical, but plausible, scenarios, one involving lung cancer screening, one involving treatment for gastric cancer, and one involving antibiotics. Unlike the Simpson's paradox example, in all the scenarios there is an interaction between an unobserved

**Figure 2**

Treatment for gastric cancer. Hypothetical results are shown for A = radical gastrectomy / splenectomy, B = "simple" gastrectomy, and C = radiation. In the sequential study, results for B and C differ in the first trial (when 30% of the subjects receive effective supportive care), and results for A and B are similar in the second trial (when 70% of the subjects receive effective supportive care). However, as shown with three-arm trial, it is incorrect to make the transitive inference that, when 70% of the subjects receive effective supportive care, the results for A and C would differ.

binary variable and the intervention. For each scenario we considered two study designs: (1) separate randomized trials of B versus C and A versus B, and (2) a randomized three-arm trial of A, B, and C. In the discussion below, we emphasize that the trial sizes would be large enough to eliminate considerations of simple statistical variation or imprecision in measurement of outcome variables. We also emphasize that the results would hold if there were no bias from contamination or noncompliance.

Results

We present graphical illustrations of the transitive fallacy in three hypothetical, but plausible, scenarios.

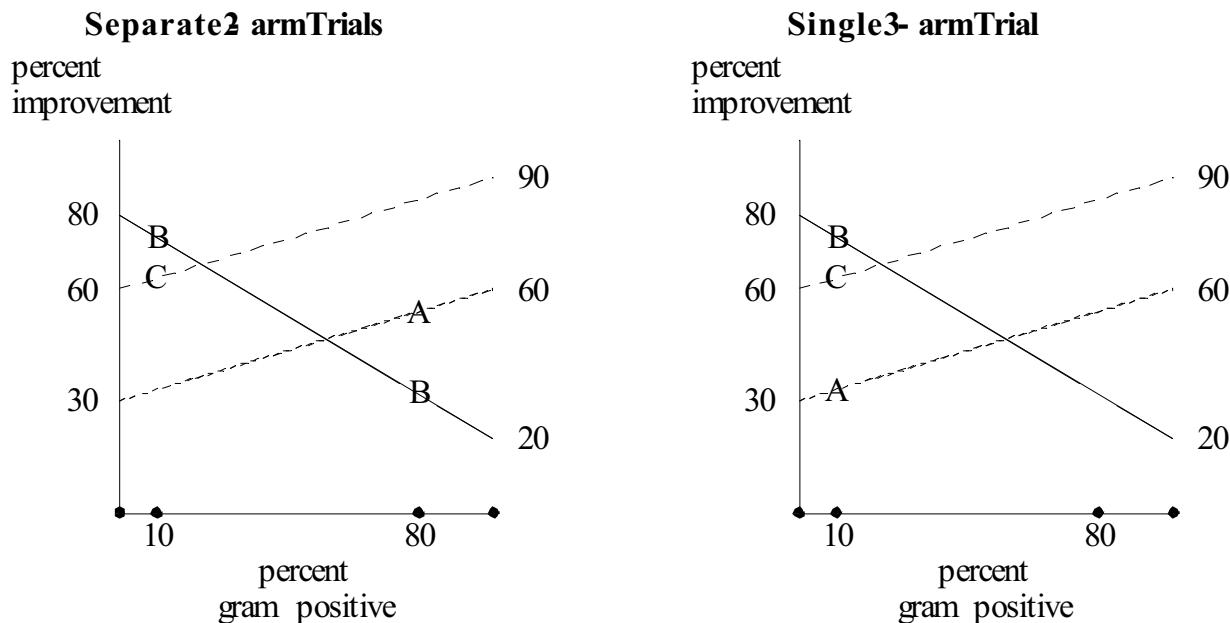
Lung cancer screening with improved therapy (Figure 1)

Let A denote spiral-CT screening, B denote chest x-ray screening, and C denote the control group of no screening or "usual care." The endpoint is lung cancer mortality rate. A previous randomized trial of B versus C found similar lung cancer mortality rates for the two interventions [5,6]. Currently there is discussion of a new randomized trial to compare A and B. Suppose the new randomized trial shows similar cancer mortality rates for A and B. Would

that constitute proof of similar cancer mortality rates for A and C?

The unobserved binary variable is an indicator of whether or not subjects received a (relatively) new (and effective) therapy after early detection. Although type of therapy is observable, it is not generally analyzed or reported in papers summarizing the results of screening trials. We suppose that the new therapy decreases the lung cancer mortality rates for A, B, and C, but at different rates (i.e. a quantitative interaction). In addition, unless there is substantial overdiagnosis, it is reasonable that the greatest decrease would occur with A due to earliest detection. Also, unless the therapy is effective at all stages of cancer, it is reasonable that the smallest decrease would occur with C due to late detection and greater total tumor burden.

We realistically assume that the percent of subjects who receive the new therapy has increased over time as randomized treatment trials establish its worth. For purposes of illustration, we specify that in the first trial of B versus C, 30 percent of the subjects receive the new therapy, and in the planned second trial (either A versus B, or A versus B versus C) 70 percent of the subjects will receive the new therapy.

**Figure 3**

Antibiotic treatment for clinical pneumonia. Hypothetical results are shown for A = antibiotic for gram-positive, B = antibiotic for gram-negative, C = antibiotic for gram-positive that is more effective than A. In the sequential study, A bests B in the first trial (when 80% of the subjects are gram positive), and B bests C in the second trial (when 10% of the subjects are gram positive). However, as shown with the three-arm trial, it would be incorrect to make the transitive inference that when 10% of the subjects are gram-positive, A is better than C.

Figure 1 shows a realistic set of outcomes. In the first trial the cancer mortality rate is similar for B and C. A second trial of A versus B also indicates similar cancer mortality rates (Figure 1, left). Under the transitive fallacy, one would incorrectly conclude that the cancer mortality rates for A and C are similar at the time of the second trial. However, when 70% of the subjects receive the new therapy, the three-arm trial correctly shows that A is substantially better than C (Figure 1, right).

Cancer treatment with improved supportive care (Figure 2)

A second hypothetical example involves treatment for gastric cancer. Let A denote, radical gastrectomy/splenectomy, B denote "simple" gastrectomy, and C denote radiation. Suppose the endpoint is percent mortality over some time period and the unobserved covariate is effective supportive care. It is plausible that supportive care improves over time with better intensive care and better antibiotics for any infections that arise.

In an earlier period when a small percentage of subjects receive effective supportive care, the more aggressive treatments carry substantially more treatment-related

mortality. As shown in Figure 2 (left side), a randomized trial of B versus C during this earlier period demonstrates considerably higher mortality for B than C. In a later period, when a larger percentage of subjects receive effective supportive care, the mortality rates converge. As shown in Figure 2 (left side) a randomized trial of A versus B during the later period indicates similar mortality rates. Under the transitive fallacy, one would incorrectly conclude that the cancer mortality rates for A and C differ substantially at the time of the second trial. However, when a large percentage of subjects receive effective supportive care, the three-arm trial correctly shows that mortality rates for A and C are similar (Figure 2, right).

Antibiotic treatment with change in percent gram positive (Figure 3)

A third hypothetical example involves the use of empiric antibiotics to treat clinical pneumonia. Suppose A is an antibiotic that treats gram positive organisms but not gram negative; B is an antibiotic that treats gram negative but not gram positive pneumonias; and C is an antibiotic that treats gram positive pneumonias better than A. The endpoint is fraction successfully treated. Suppose the percent of organisms that are gram positive is an unmeasured

covariate. This is a realistic scenario given that the spectrum of bacterial infections can shift over time, or can differ from hospital to hospital at the same time.

Suppose a randomized trial of A versus B has been completed and investigators are considering a randomized trial of B versus C. (In this scenario, A versus B occurs prior to B versus C, even though it is depicted farther to the right on the graph.) We realistically assume the percent of organisms that are gram positive has decreased by the time of the second trial. For purposes of illustration suppose 80 percent of the subjects are gram positive in the first trial and only 10% are gram positive in the second trial.

In this situation A bests B in a randomized trial of patients who mainly have gram-positive infections, and B bests C in a randomized trial of patients with mainly gram-negative infections. (Figure 2 left). Under the transitive fallacy, one would incorrectly conclude that when 10% of the subjects have gram-negative infections, A would best C. However, the three-arm study correctly shows that when 10% of the subjects have gram-negative infections, C would best A (Figure 2 right).

Conclusion

Given only a previous randomized trial of B versus C and a new randomized trial of A versus B, inference about A versus C can be misleading. In contrast a three-arm randomized trial of A, B, and C, will yield appropriate inference about both A versus B and A versus C. Validity of the sequential studies strategy (B versus C, and A versus B) rests on the assumption that there is no intervening important covariate that could confound the implied principle of transitivity. Given the amount of resources that are often invested in large "definitive" clinical trials, the possibility of such covariates should be an explicit part of the discussion in designing the trials and interpreting the results.

Authors' contributions

In discussions about designing a planned randomized trial of lung cancer screening with one of the trial investigators, the authors pointed out that the concept of transitivity might not hold and that there was therefore some risk in conducting a two-arm study. After further discussion, they decided to develop this concept into a manuscript. SGB had the general idea of using the graphic from Simpson's paradox to explain lack of transitivity. BSK worked out crucial details of how it applies in practice. SGB wrote the initial draft and BSK made substantial improvements. Both authors read and approved the final manuscript.

Competing interests

None declared.

Disclaimer

This works represents the opinions of the authors and does not necessarily represent the opinions of the federal government or of the Department of Health and Human Services.

References

1. Baker SG, Kramer BS: **Good for women, good for men, bad for people: Simpson's paradox and the importance of sex-specific analysis in observational studies.** *Journal of Women's Health & Gender-Based Medicine* 2001, **10**:867-872.
2. Langford E, Schwertman N, Owens M: **Is the property of being positively correlated transitive?** *The American Statistician* 2001, **55**:322-325.
3. Paulos JA: **I Think Therefore I Laugh: The Flip Side of Philosophy.** New York: Columbia University Press 2000, 23-29.
4. Wainer H: **The BK-Plot: Making Simpson's paradox clear to the masses,** *Chance* **15**:60-62.
5. Fontana RS, Sanderson DR, Woolner LB, Taylor WF, Miller WE, Muham JR, et al: **Screening for lung cancer: a critique of the Mayo Lung Project.** *Cancer* 1991, **67**(4 Suppl):1155-64.
6. Marcus PM, Bergstrahl EJ, Fagerstrom RM, Williams DE, Fontana R, Taylor WF, Prorok PC: **Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up.** *Journal of the National Cancer Institute* 2000, **92**:1308-1316.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/2/13/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedCentral will be the most significant development for disseminating the results of biomedical research in our lifetime."
Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright

Submit your manuscript here:
<http://www.biomedcentral.com/manuscript/>



editorial@biomedcentral.com