

Test-treatment RCTs are sheep in wolves' clothing

Ferrante di Ruffano, Lavinia; Deeks, Jonathan J.

DOI:

[10.1016/j.jclinepi.2015.06.013](https://doi.org/10.1016/j.jclinepi.2015.06.013)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Ferrante di Ruffano, L & Deeks, JJ 2015, 'Test-treatment RCTs are sheep in wolves' clothing', *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2015.06.013>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Eligibility for repository: Checked on 14/09/2015

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Accepted Manuscript

Test-treatment RCTs – sheep in wolves' clothing

Lavinia Ferrante di Ruffano, Jonathan J. Deeks

PII: S0895-4356(15)00320-0

DOI: [10.1016/j.jclinepi.2015.06.013](https://doi.org/10.1016/j.jclinepi.2015.06.013)

Reference: JCE 8924

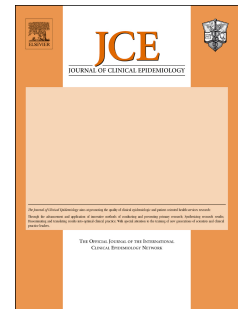
To appear in: *Journal of Clinical Epidemiology*

Received Date: 26 May 2015

Accepted Date: 21 June 2015

Please cite this article as: di Ruffano LF, Deeks JJ, Test-treatment RCTs – sheep in wolves' clothing, *Journal of Clinical Epidemiology* (2015), doi: 10.1016/j.jclinepi.2015.06.013.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Letter 689wds:

Test-treatment RCTs – sheep in wolves’ clothing

Lavinia Ferrante di Ruffano and Jonathan J Deeks, University of Birmingham, UK.

We thank Siontis and colleagues for their interesting review of diagnostic RCTs [1]. Assuming that changes to patient health are caused by improvements in test accuracy, the authors expected health effects observed in their cohort of RCTs to be complemented by changes in downstream diagnostic and treatment decision-making. Yet the authors found no overall correlation. We would argue their findings are based on a flawed assumption: that diagnostic accuracy is the only mechanism driving change to health outcomes. Although this reflects previous thought [2], recent research suggests this to be an incomplete model with tests shown to change patient health through 3 other causal chains of effect: by the direct effects of the testing process, by changing timeframes of management, and by altering patient and clinician perceptions [3]. Using this complete model, we offer alternative explanations for the authors’ findings and challenge their call for more diagnostic RCTs.

First, the authors selected two process outcomes (further test use and treatment use) that are not always fair surrogate markers for more appropriate decision-making. More accurate tests improve patient health through their ability to enhance decisions, which does not necessarily alter the frequency of subsequent test or treatment use. Equally, while more accurate tests could increase testing by identifying more patients who need further investigation, less able tests might also increase further testing if they reduce a clinician’s diagnostic confidence. In other words, the meaning of observed changes in further test use or treatment use, and whether one would expect any correlation with health effects, is entirely dependent on the clinical context being evaluated.

By focussing on accuracy-driven change, the authors did not always analyse the processes actually driving changes to patient health in each trial. This is clearly shown in the example of point-of-care ultrasonography (PLUS) for ER trauma cases given by the authors (Box 2). In this add-on comparison, observed improvements to patient health (reduced probability of experiencing serious complications) are largely driven by PLUS’s ability to expedite urgent treatment; this is reflected in the trial’s primary finding that PLUS significantly reduced time-to-operative care (on average 57 minutes vs. 166 minutes) [4]. The reduction in further testing (use of CT) is therefore unrelated to changes in downstream patient health in this trial, and is instead likely to reflect physicians’ confidence in the ability of PLUS. The lack of correlation observed in this study between health effects and markers of accuracy is therefore also explained because health effects are not always driven by improved accuracy in this cohort of 140 trials.

It is highly probably that many of the trials are underpowered to detect differences in patient outcomes, which could in part explain both the rarity of significant health effects and lack of correlation with process outcomes. In an RCT comparing tests, treatment effects are diluted by the subgroup of patients receiving no change in diagnosis and treatment. Small differences in the sensitivity of comparative tests (say 20%) means the proportion of diseased patients (in a group with 20% prevalence) benefiting from different treatment will also be a small subset of the study population [3,5] (20% x 20%, or 4%); thus to detect differences in patient outcomes sample sizes must be inflated to account for this 'dilution effect' (by 25 times, or 1/4%). The 49 RCTs evaluating diagnostic tests (a third of the cohort) had a median size of 372, which is smaller than found in RCTs of treatments [6].

We suggest caution before embarking on a test-treatment RCT. There are major barriers to conducting valid trials that can deliver: competing test-treat pathways must be described, interventions and complex decision-making protocolized, sample sizes must be large enough, and the web of mechanisms driving change to patient health (accuracy and otherwise) must be identified and measured to ensure that diagnostic processes are operating as intended. Often this is not achievable within acceptable budgets and timeframes, and the RCT cannot deliver. We would rather highlight an excellent alternative: using modelling to link evidence for all important mechanisms that are identified, on a case-by-case basis, to cause changes to patient health.

References

- [1] Siontis KC, Siontis GCM, Contopoulos-Ioannidis DG, Ioannidis JPA. Diagnostic tests often fail to lead to changes in patient outcomes. *J Clin Epi* 2014;67:612–621.
- [2] Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88–94.
- [3] Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests – A framework for designing and evaluating trials. *BMJ* 2012;344:e686.
- [4] Melniker LA, Leibner E, McKenney MG, Lopez P, Briggs WM, Mancuso CA. Randomized controlled clinical trial of point-of-care, limited ultrasonography for trauma in the emergency department: the first sonography outcomes assessment program trial. *Ann Emerg Med* 2006;48:227–235.
- [5] Pletcher MJ, Pignone M. Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. *Circulation* 2011;123:1116–1124.
- [6] Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.