



OPEN ACCESS

STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts

J r mie F Cohen,^{1,2} Dani l A Korevaar,¹ Constantine A Gatsonis,³ Paul P Glasziou,⁴ Lotty Hooft,⁵ David Moher,^{6,7} Johannes B Reitsma,⁸ Henrica CW de Vet,⁹ Patrick M Bossuyt¹

For numbered affiliations see end of article.

Correspondence to: J F Cohen jeremie.cohen@inserm.fr

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2017;358:j3751 <http://dx.doi.org/10.1136/bmj.j3751>

Many abstracts of diagnostic accuracy studies are currently insufficiently informative. We extended the STARD (Standards for Reporting Diagnostic Accuracy) statement by developing a list of essential items that authors should consider when reporting diagnostic accuracy studies in journal or conference abstracts. After a literature review of published guidance for reporting biomedical studies, we identified 39 items potentially relevant to report in an abstract. We then selected essential items through a two round web based survey among the 85 members of the STARD Group, followed by discussions within an executive committee. Seventy three STARD Group members responded (86%), with 100% completion rate. STARD for Abstracts is a list of 11 quintessential items, to be reported in every abstract of a diagnostic accuracy study. We provide examples of complete reporting, and developed template text for writing informative abstracts.

Abstracts play a critical role in the use of research. Clinicians and researchers use abstracts to decide whether they should read the full journal article, attend the conference presentation, or contact the authors for more information. Systematic reviewers screen large amounts of abstracts to assess study

eligibility. In some cases, study abstracts may be the only information available to clinicians, researchers, reviewers, guideline developers, or policy makers.¹ In evaluations, the proportion of diagnostic accuracy studies presented as conference abstracts that are eventually reported in articles was found to be as low as 39%.²⁻⁴

We recently evaluated the quality of reporting of abstracts of diagnostic accuracy studies published in several high impact journals and abstracts presented at a major ophthalmology conference.^{5,6} In line with previous authors,^{3,7} we found that many of these abstracts were insufficiently informative. Key items, such as eligibility criteria, study setting, patient sampling procedures, and confidence intervals around accuracy estimates were reported in less than half of the abstracts.^{5,6} This makes it difficult for readers to assess the validity and applicability of the study findings.

Ideally, studies should be free from deficiencies, and the results of the study should reflect the “true” accuracy of the test under evaluation. Major sources of bias in diagnostic accuracy studies include methodological flaws in participant recruitment, data collection, test execution and interpretation, and data analysis.^{8,9} Even when free of bias, study findings are not necessarily generalisable to all applications. Diagnostic accuracy can vary across studies because of variations in study setting, participant characteristics, disease prevalence and severity, and aspects of test execution and interpretation.¹⁰ Risk of bias and concerns about applicability can only be evaluated if study reports are sufficiently informative.

Aim and scope

The Standards for Reporting Diagnostic Accuracy (STARD) initiative was developed in response to increasing evidence of suboptimal reporting of diagnostic accuracy studies in scientific journals.^{8,9} The STARD Group developed a list of essential items that should be presented in all full reports of diagnostic accuracy studies.¹¹ Since its launch in 2003, STARD has been endorsed by more than 200 journals, including *The BMJ*. Study reports of diagnostic accuracy studies have become more complete since then, although there is still room for further improvement.¹² STARD 2015, an update of the original STARD statement, was recently published by *The BMJ*, *Radiology*, and *Clinical Chemistry*.¹³⁻¹⁵

Unlike some other reporting guidelines, such as CONSORT (Consolidated Standards Of Reporting Trials) for randomised controlled trials¹ and PRISMA (Preferred Reporting Items for Systematic reviews and Meta-analyses) for systematic reviews,¹⁶ STARD so far has not provided guidance for writing abstracts. Here

SUMMARY POINTS

- The STARD statement has become an internationally accepted reporting guideline for diagnostic accuracy studies
- STARD for Abstracts is intended to improve the completeness and informativeness of journal and conference abstracts of diagnostic accuracy studies
- STARD for Abstracts presents a minimal set of 11 items (specific journals or organisations could ask for additional information); whenever space restrictions allow it, authors may also incorporate other STARD 2015 elements in their abstract

Table 1 | STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts

Section	Item
	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
Background and Objectives	Study objectives
Methods	Data collection: whether this was a prospective or retrospective study
	Eligibility criteria for participants and settings where the data were collected
	Whether participants formed a consecutive, random, or convenience series
	Description of the index test and reference standard
Results	Number of participants with and without the target condition included in the analysis
	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
Discussion	General interpretation of the results
	Implications for practice, including the intended use of the index test
Registration	Registration number and name of registry

AUC=area under the curve.

we present a separate reporting guideline that can help to improve the informativeness of abstracts of diagnostic accuracy studies, both for journals and for conferences.

The guiding principle in the development of the checklist was to identify essential items that should be reported in all abstracts of diagnostic accuracy studies, considering the usual 200 to 300 word limit. The items can assist authors in presenting informative abstracts and help readers in deciding whether to invest time in reading the full report, attending the conference presentation, or contacting the authors for more information.

Methods for developing STARD for Abstracts

Detailed survey methods and results are presented in supplementary eAppendix 1 and eTables 1 and 2. We relied on standard processes for developing reporting guidelines.¹⁷ Initially we formed an executive committee (DAK and JFC, clinicians, respectively, doctoral and postdoctoral research fellows; PMB, CAG, JBR, and LH, respectively, professor in clinical epidemiology, professor in biostatistics, associate professor in clinical epidemiology, and co-director of the Dutch Cochrane Centre) and developed a protocol.¹⁸ We then conducted a literature review, which focused on previously published guidance for reporting biomedical studies (full texts and abstracts), including STARD 2015, and on studies of the methodological quality of diagnostic accuracy studies.⁵ Thereafter we listed 39 items judged potentially relevant to report in abstracts (see supplementary eAppendix 2).

We then invited the STARD Group, which consists of 85 clinical epidemiologists, statisticians, journal editors, and other stakeholders, to participate in a two round web based survey, aiming at obtaining consensus on which of these items were deemed essential.

Seventy three STARD Group members responded in both rounds (86%), with 100% completion rate. In the first round, participants were asked to rate to what extent each candidate item would be essential for abstracts. Consensus, defined as a positive response by at least two thirds of the respondents, was reached for 10 items. We then developed a draft STARD for Abstracts checklist and circulated it within the STARD Steering Committee. That list was fine tuned until the executive committee agreed.

In the second round, STARD Group members were asked whether they thought any of the remaining candidate items, apart from the 10 selected in the first round, should be added to the list. No consensus was reached about adding any other item. In both rounds of the survey, participants had the option to provide comments in open comment boxes.

After the survey, a revised draft STARD for Abstracts checklist was established. During a teleconference in August 2015, the executive committee agreed on incorporating two additional elements, merging these into the already selected items. This was based on concerns expressed in comments by STARD Group members during the survey. The draft list of 10 items was then circulated to members of the STARD Steering Committee to provide feedback. Before the final list was agreed upon, the executive committee decided to add an 11th item about study registration (see supplementary eAppendix 3 for a description of the flow of items through the process), to ensure consistency with another STARD initiative promoting the prospective registration of diagnostic accuracy studies.¹⁹

STARD for Abstracts

STARD for Abstracts presents a checklist of 11 essential items, to be considered in every abstract

Table 2 | Key STARD terminology

Term	Explanation
Index test	The test under evaluation
Target condition	The disease, disease stage, event, or condition that the index test is expected to detect
Reference standard	The test or procedure used for establishing the presence or absence of the target condition
Intended use of the test	Whether the index test is used for diagnosis, screening, staging, monitoring, surveillance, prediction, prognosis, or other reasons

Table 3 | STARD for Abstracts template text

STARD for Abstracts item	Template text
Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	[diagnostic accuracy/sensitivity and specificity/predictive value . . .] of [index test] for diagnosing [target condition]
Study objectives	to evaluate the [diagnostic accuracy/sensitivity and specificity/predictive value . . .] of [index test] in patients with suspected [target condition]
Data collection: whether this was a prospective or retrospective study	<ul style="list-style-type: none"> • In this [prospective/retrospective] study . . . • We conducted a [prospective/retrospective] study . . . • Data were collected [prospectively/retrospectively] . . .
Eligibility criteria for participants and the settings where the data were collected	<p>eligible for inclusion were [adults/children/men/women] [age X to Y years] with suspected [target condition] . . .</p> <ul style="list-style-type: none"> • based on [presenting signs and symptoms] • who underwent [index test] and [reference standard] <p>[mono/multi]centre study in [primary/secondary/tertiary] care in [country] . . .</p>
Whether participants formed a consecutive, random, or convenience series	a [consecutive series/random sample/convenience sample] of patients with . . .
Description of the index test and reference standard	<ul style="list-style-type: none"> • all patients underwent [index test with key elements of description] . . . • [reference standard with key elements of description] was used as the reference standard . . .
Number of participants with and without the target condition included in the analysis	Of [X] patients included in the analysis, the diagnosis of [target condition] was confirmed in [Y] and excluded in [Z]
Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	The [sensitivity and specificity/positive and negative predictive values/positive and negative likelihood ratios] of [index test] were [A (95% CI B to C)] and [D (95% CI E to F)], respectively
General interpretation of the results	[Index test] showed [high/low/insufficient . . .] [accuracy/sensitivity/specificity . . .] for detecting/diagnosing [target condition]
Implications for practice, including the intended use of the index test	[Index test] [should/should not/could/could not . . .] be used for [diagnosis/screening/staging/monitoring/surveillance . . .] in [patients/adults/children] suspected of [target condition] seen in [primary/secondary/tertiary] care
Registration number and name of registry	[Name of registry]: [Registration number]

AUC=area under the curve

that reports on a diagnostic accuracy study (see table 1 for the checklist and table 2 for key terminology). The structure of STARD for Abstracts follows that of a typical biomedical abstract, with headings pertaining to Background and Objectives, Methods, Results, and Discussion sections.

Because 10 out of 11 STARD for Abstracts items are similar to those from STARD 2015 (see supplementary eTable 3), we did not develop a separate explanation and elaboration document; instructions can be found in STARD 2015.²⁰ To illustrate the information that corresponds to each item, we collected examples of complete reporting (see supplementary eAppendices 4-7). To further assist authors in writing abstracts, we developed template text for each item and an example abstract (see table 3 for template text and the box for an example of application).

Applicability and implementation

In developing STARD for Abstracts, we aimed at identifying items that would apply to any abstract of a diagnostic accuracy study. The list presents a minimum, and specific journals or organisations could ask for additional information, such as variability across readers in imaging studies or analytical performance in laboratory tests studies. Whenever space restrictions allow it, authors may incorporate other elements from STARD 2015 in their abstract.

Based on our evaluations, we believe it is possible to address all 11 items within the 200 to 300 word limit that typically applies for abstracts (supplementary eAppendices 4-7 illustrate real abstracts, with 237 to 339 words, which comply well with the checklist). We do not make recommendations about how abstracts should be structured but only recommend that this

Box 1: Example of an application of STARD for Abstracts template text

Point-of-care D-Dimer testing for diagnosing pulmonary embolism in primary care

- **Objective:** To evaluate the negative and positive predictive value of D-Dimer testing in patients with suspected pulmonary embolism in primary care.
- **Methods:** We conducted a prospective study among 70 general practitioners in the UK. Eligible for inclusion were consecutive adults, age 18 to 70 years, with suspected pulmonary embolism based on presenting signs and symptoms. All consenting patients underwent a qualitative point-of-care D-Dimer test (with a positivity cut-off of 80 ng/mL) performed by the general practitioner. Patients with a positive D-Dimer test result were referred to secondary care for further management according to national guidelines. Three months' clinical follow-up was used as the reference standard in patients with a negative D-Dimer test result.
- **Results:** Of 500 patients included in the analysis, the diagnosis of pulmonary embolism was confirmed in 50 and excluded in 450. Three cases of pulmonary embolism were observed among the 273 patients with negative D-Dimer test results. The negative predictive value of point-of-care D-dimer testing was 98.9% (95% confidence interval 96.8% to 99.8%) and the positive predictive value was 20.7% (15.6% to 26.6%).
- **Discussion:** With a high negative predictive value, point-of-care D-Dimer testing could be used for the triage of adults suspected of pulmonary embolism seen in primary care.
- **Registration:** ClinicalTrials.gov: NCT02593219.
- **Word count:** 205

This abstract was created for illustrative purposes only. It is based on a virtual study.

minimal set of information should be reported within every abstract. Some conferences invite authors to provide a figure with their abstract. If so, we recommend considering submission of a diagram reflecting the design and flow of participants through the study.¹³

To improve the completeness of reporting, simply developing a list of items is insufficient; dissemination, endorsement, and implementation are also critical.^{17 21} We invite journal editors and conference organisers to endorse STARD for Abstracts, by drawing attention to this list of items in their instructions to authors and conference websites. The template texts may also facilitate writing abstracts of diagnostic accuracy studies (see table 3 and the box).

Conclusions

We acknowledge that an important share of the burden of improving reporting and reducing waste in research is currently put on journal editors and peer reviewers, as they play a major role in the final stages of the publication process. Authors should also take action, as should other stakeholders, such as funders and academic institutions.²² We need to convince scientific institutions and universities that complete reporting forms an irrefutably indispensable element of good research practice and should be taught as such in academic training programmes—for example, as part of scientific writing courses.^{21 23}

We believe that STARD for Abstracts can help to improve the quality of reporting of diagnostic accuracy studies through the inclusion of essential study information in every abstract, thereby increasing the value of such studies to the clinical and research community.

AUTHOR AFFILIATIONS

¹Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands

²INSERM UMR 1153 and Department of Pediatrics, Necker-Enfants malades Hospital, AP-HP, Paris Descartes University, Paris, France

³Department of Biostatistics, Brown University School of Public Health, Providence, Rhode Island, USA

⁴Centre for Research in Evidence-Based Practice, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia

⁵Dutch Cochrane Centre, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands

⁶Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

⁷School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Canada

⁸Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands

⁹Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, Netherlands

STARD Group collaborators: Todd Alonzo, Douglas G Altman, Augusto Azuara-Blanco, Lucas Bachmann, Jeffrey Blume, Patrick M Bossuyt, Isabelle Boutron, David Bruns, Harry Büller, Frank Buntinx, Sarah Byron, Stephanie Chang, Jérémie F Cohen, Richelle Cooper, Joris de Groot, Henrica CW de Vet, Jon Deeks, Nandini Dendukuri, Jac Dinnes, Kenneth Fleming, Constantine A Gatsonis, Paul P Glasziou, Robert M Golub, Gordon Guyatt, Carl Heneghan, Jørgen Hilden, Lotty

Hoof, Rita Horvath, Myriam Hunink, Chris Hyde, John Ioannidis, Les Irwig, Holly Janes, Jos Kleijnen, André Knottnerus, Daniël A Korevaar, Herbert Y Kressel, Stefan Lange, Mariska Leeftang, Jeroen G Lijmer, Sally Lord, Blanca Lumbreras, Petra Macaskill, Erik Magid, Susan Mallett, Matthew McInnes, Barbara McNeil, Matthew McQueen, David Moher, Karel Moons, Katie Morris, Reem Mustafa, Nancy Obuchowski, Eleanor Ochodo, Andrew Onderdonk, John Overbeke, Nitika Pai, Rosanna Peeling, Margaret Pepe, Steffen Petersen, Christopher Price, Philippe Ravaut, Johannes B Reitsma, Drummond Rennie, Nader Rifai, Anne Rutjes, Holger Schunemann, David Simel, Iveta Simera, Nynke Smidt, Ewout Steyerberg, Sharon Straus, William Summerskill, Yemisi Takwoingi, Matthew Thompson, Ann van den Bruel, Hans van Maanen, Andrew Vickers, Gianni Virgili, Stephen Walter, Wim Weber, Marie Westwood, Penny Whiting, Nancy Wilczynski, and Andreas Ziegler.

STARD for Abstracts is released under a Creative Commons license (CC BY-NC license <http://creativecommons.org/licenses/by-nc/4.0>) that allows others to use and distribute the list if unmodified and if they acknowledge the source. All material related to STARD for Abstracts is available at the EQUATOR website (www.equator-network.org/reporting-guidelines/stard/).

The protocol of this study was published on the EQUATOR website (www.equator-network.org/wp-content/uploads/2009/02/STARD-for-Abstracts-protocol.pdf).

Contributors: PMB had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis; he is the guarantor. JFC analysed the web based survey, drafted the STARD for Abstracts list, and wrote the first draft of the manuscript. PMB, JFC, CAG, DAK, CAG, LH, and JBR coordinated the development of STARD for Abstracts, wrote the protocol, designed the web based survey, and finalised the list (executive committee). All authors, who are members of the STARD Steering Committee, revised the manuscript for important intellectual content and approved the final version submitted for publication. Members of the STARD Group participated in the web based survey. PMB supervised the development of STARD for Abstracts.

Funding: No external funding.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

Transparency: The manuscript's guarantor (PMB) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is noncommercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- Hopewell S, Clarke M, Moher D, et al. CONSORT Group. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008;371:281-3. doi:10.1016/S0140-6736(07)61835-2
- Wilson C, Kerr D, Noel-Storr A, Quinn TJ. Associations with publication and assessing publication bias in dementia diagnostic test accuracy studies. *Int J Geriatr Psychiatry* 2015;30:1250-6. doi:10.1002/gps.4283
- Brazzelli M, Lewis SC, Deeks JJ, Sandercock PA. No evidence of bias in the process of publication of diagnostic accuracy studies in stroke submitted as abstracts. *J Clin Epidemiol* 2009;62:425-30. doi:10.1016/j.jclinepi.2008.06.018
- Korevaar DA, Cohen JF, Spijker R, et al. Reported estimates of diagnostic accuracy in ophthalmology conference abstracts were not associated with full-text publication. *J Clin Epidemiol* 2016;79:96-103. doi:10.1016/j.jclinepi.2016.06.002
- Korevaar DA, Cohen JF, Hoof L, Bossuyt PM. Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. *J Clin Epidemiol* 2015;68:708-15. doi:10.1016/j.jclinepi.2015.01.014
- Korevaar DA, Cohen JF, de Ronde MW, Virgili G, Dickersin K, Bossuyt PM. Reporting weaknesses in conference abstracts of diagnostic accuracy studies in ophthalmology.

- JAMA Ophthalmol* 2015;133:1464-7. doi:10.1001/jamaophthalmol.2015.3577
- 7 Estrada CA, Bloch RM, Antonacci D, et al. Reporting and concordance of methodologic criteria between abstracts and articles in diagnostic test studies. *J Gen Intern Med* 2000;15:183-7. doi:10.1046/j.1525-1497.2000.03189.x
 - 8 Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-51. doi:10.1001/jama.1995.03530080061042
 - 9 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6. doi:10.1001/jama.282.11.1061
 - 10 Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189-202. doi:10.7326/0003-4819-140-3-200402030-00010
 - 11 Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003;49:1-6. doi:10.1373/49.1.1
 - 12 Korevaar DA, Wang J, van Enst WA, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology* 2015;274:781-9. doi:10.1148/radiol.14141160
 - 13 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527. doi:10.1136/bmj.h5527
 - 14 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Clin Chem* 2015;61:1446-52. doi:10.1373/clinchem.2015.246280
 - 15 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology* 2015;277:826-32. doi:10.1148/radiol.2015151516
 - 16 Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts Group. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* 2013;10:e1001419. doi:10.1371/journal.pmed.1001419
 - 17 Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217. doi:10.1371/journal.pmed.1000217
 - 18 Cohen JF, Korevaar DA, Hooft L, Reitsma JB, Bossuyt PM. Development of STARD for Abstracts (Protocol) 2015. www.equator-network.org/wp-content/uploads/2009/02/STARD-for-Abstracts-protocol.pdf.
 - 19 Korevaar DA, Hooft L, Askie LM, et al. Facilitating Prospective Registration of Diagnostic Accuracy Studies: A STARD Initiative. *Clin Chem* 2017;63:1331-41. doi:10.1373/clinchem.2017.272765
 - 20 Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799. doi:10.1136/bmjopen-2016-012799
 - 21 Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267-76. doi:10.1016/S0140-6736(13)62228-X
 - 22 Begley CG, Buchan AM, Dirnagl U. Robust research: Institutions must do their part for reproducibility. *Nature* 2015;525:25-7. doi:10.1038/525025a
 - 23 Moher D, Altman DG. Four proposals to help improve the medical research literature. *PLoS Med* 2015;12:e1001864.

Supplementary information: eAppendices 1-7 and eTables 1-3