RESEARCH ARTICLE

An attempt to reproduce a previous meta-analysis and a new analysis regarding the impact of directly observed therapy on tuberculosis treatment outcomes

Brian McKay *, **Maria Castellanos**, **Mark Ebell**, **Christopher C. Whalen**, **Andreas Handel***

Department of Epidemiology and Biostatistics, The University of Georgia, Athens, Georgia, United States of America

* bmckay52@uga.edu (BM); ahandel@uga.edu (AH)
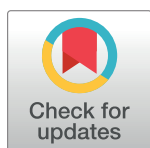
Abstract

Directly observed therapy (DOT) is almost universally used for the treatment of TB. Several meta-analyses using different methods have assessed the effectiveness of DOT compared to self-administered therapy (SAT). The results of these meta-analyses often conflict with some concluding DOT is superior and others that there is little or no difference. Meta-analyses can guide policymaking, but such analyses must be reliable. To assess the validity of a previous meta-analysis, we tried to reproduce it. We encountered problems with the previous analysis that did not allow for a meaningful reproduction. We describe the issues we encountered here. We then performed a new meta-analysis comparing the treatment outcomes of adults given treatment with SAT versus DOT. Outcomes in the new analysis are loss to follow-up, treatment failure, cure, treatment completed, and all-cause mortality. All data, documentation, and code used to generate our results is provided. Our new analysis included four randomized and three observational studies with 1603 and 1626 individuals respectively. The pooled relative risks (RR) are as follows: Lost to follow-up (RR = 1.2, 95% CI 0.9, 1.7), Treatment Failure (RR = 1.1, 95% CI 0.6, 2), Cure (RR = 0.9, 95% CI 0.8, 1.1), Treatment Completion (RR = 1, 95% CI 0.9, 1.1), Mortality (RR = 0.9, 95% CI 0.6, 1.3). Based on data from our new meta-analysis, the magnitude of the difference between DOT and SAT for all reported outcomes is small, and none of the differences are statistically significant.

Introduction

Directly observed therapy (DOT) for the treatment of tuberculosis (TB) is a component of the World Health Organization's (WHO) recommended DOTS program. While the DOTS program has been successful in reducing tuberculosis incidence, the average annual decrease is far short of the needed reduction to meet the goals of the WHO End TB strategy [1]. DOT requires patients with TB to be observed by another individual while taking their medicine.

Evidence to support DOT's effect on cure rates [2,3], adherence [2,3], or any other measure of treatment success is not very compelling [4,5]. Leading to an intense debate about the use and effectiveness of DOT [6–9].

A meta-analysis of data from six randomized controlled trials (RCT) suggested that DOT did not increase the cure or treatment completion significantly [3]. We believe that due to the small number of RCT and the high likelihood that there will not be any new ones addressing this issue, it is necessary also to consider the evidence from observational studies. Tian and others reported the findings from eight RCTs and fifteen observational studies [10]. They found DOT to be more successful than SAT when the analysis was restricted to observational studies, but no difference was observed among RCT. Their analysis included retrospective cohort and cross-sectional studies, which are generally considered to provide lower quality evidence. To ensure the best quality evidence, we believe only prospective studies should be considered [11].

One previous meta-analysis included both RCT and prospective cohort studies (PCS) [12]. This study by Pasipanodya and Gumbo (which will be referred to as PG), included five RCT and five PCS. The inclusion of prospective studies helps provide valuable information without the high risk of bias associated with retrospective studies. However, PG's study was criticized due to methodological concerns [13–15]. Thus, calls to gather more evidence to define the effectiveness of DOT versus SAT are still being made [8].

Since the question of DOT vs. SAT is a significant public health issue, we deemed it valuable to try and reproduce the results found by PG, taking into account and addressing previously raised criticisms as needed. In doing so, our investigation found other inconsistencies in PG beyond those already raised [13–15]. This prevented us from reproducing the results presented in PG. We instead performed an entirely new analysis that is reproducible and based on what we consider to be the most appropriate data and analysis methodology. Our analysis confirms results for RCT studies that there is little evidence for the superiority of DOT versus SAT in regards to any of the WHO-defined treatment outcomes.

Methods

Standards

We followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines [16].

Search strategy

We performed a systematic search of the literature from January 1st, 1960 to July 7th, 2016. The search included Pub Med and gray literature such as The Cochrane library. The studies found were reviewed and assessed by two researchers independently for inclusion (BM and MC), and consensus resolved any conflict. We attempted to contact authors if the published data were not complete or were reported in a manner that results in the desired population were not available. Further details can be found in the supplementary material (SM).

Study selection criteria

Included studies met the following criteria: patients were diagnosed by microscopic examination of sputum smear or culture, assigned to either DOT or SAT, evaluated for treatment failure during the treatment, and were treated with short-course chemotherapy regimen including isoniazid, rifampin, and pyrazinamide. Study designs were limited to RCT and PCS. Studies were excluded based on the following criteria: study population includes children

(< 15 years), retrospective design, included specialized re-treatment regimens or biased assignment of patients receiving re-treatment. Studies were not excluded based on the language unless reasonable attempts to have them translated failed.

Exposure definitions

We used DOT definitions created by the WHO, which defines DOT as "an element of the DOTS program and refers to the direct observation of patients swallowing their medication" [17]. The observers in some studies are family members, community members or lay medical personnel. The training level of these individuals is not always clear, but studies have shown the impact of training level on outcomes is negligible [2,18–21]. When a patient's treatment involves little or no supervision, it is considered SAT.

Outcomes

Outcomes are defined by the WHO [17]. Included in the final analysis are lost to follow-up (LTFU), treatment failure, treatment completion, death, and cure. Lost to follow-up is used in place of the term "default" and is defined as an individual that does not complete treatment.

Data abstraction and quality assessment of included studies

Data were abstracted independently by two researchers (BM and MC), and any disagreements were settled by consensus. We contacted the authors of studies that include children, retreatment, or extrapulmonary TB in an attempt to get the data excluding these individuals. Similarly, the authors of papers that only report some of the treatment outcomes or do not report the data in a usable format were also contacted.

To assess study quality, we used the "Cochrane Risk of Bias Tool for Cohort Studies" (CRBCS) to evaluate bias and quality of the PCS [22], and the Jadad score for RCT [23]. Two investigators (BM and MC) assessed study quality independently, and any disagreements were resolved by consensus. We compared the results of our data re-abstraction, study selection, and study quality to those obtained in PG. A detailed examination of differences in study inclusion between PG and our analysis is included in the SM.

Data analysis

All analyses were done in R [24] using the "metafor" package [25]. Relative risk (RR) was used as the primary outcome. RR is preferred over risk difference (RD) because RR accounts for the baseline risk and provides a more consistent summary of effects [26].

Relative risk was calculated as SAT risk over DOT risk. A RR less than 1 for a positive outcome or a RR greater than 1 for an adverse outcome indicates that DOT is favored.

Since some of the studies had no events for outcomes, we used a continuity correction of adding .5 to all the cells in any 2x2 table with a single 0. We did not exclude studies with zero outcomes from the RR analysis [27]. PG reported effect size using the DerSimonian Laird (DL) method. In a meta-analysis of only a few studies, with each study having a diverse sample and effect size, the DL method can lead to biased estimates of overall effect and underestimate the amount of heterogeneity [28–31]. We believe the use of restricted maximum likely hood is more appropriate and used it for all the random effects models [28,32].

To decide on the statistical model to use, PG used fixed effects if $I^2 < 30\%$ and random effects if $I^2 > 30\%$. While this approach is common, we contend that it is not entirely appropriate. Instead, the model choice should be determined based on the assumptions about the underlying biology and distribution of the included studies [33]. Since the studies will likely be

from different settings with different populations, there is no reason to assume *a priori* that effect sizes should be the same. Therefore, a random effects model is more appropriate irrespective of the amount of heterogeneity, and it was used for all effect size calculations [33].

PG performed a meta-regression to investigate and explain the observed heterogeneity between studies. We decided that due to the small sample size it was not appropriate. Instead, we examined the effects of study design via subgroup analysis. As done by PG, publication bias is assessed using visual inspection and the Egger test of funnel plot symmetry [34]. Since the overall sample of studies is small, these tests may lack sufficient power. We therefore also included Orwin's method of "N Fail Safe" and the "Trim and Fill" methods to provide additional information when judging publication bias [35].

We also performed a systematic sensitivity analysis to identify studies that may be outliers and studies with a significant level of influence on the estimated effect size.

Results

Reproducibility of PG

Study selection. Our set of studies included four RCTs and two PCSs. These studies differed from studies reported in PG: one RCT [36] and 4 PCS [37–40] included in PG should not have been included by PG based on their stated exclusion and inclusion criteria. We found two additional observational studies that met the inclusion criteria [41,42] which were not included in PG. With the information provided by PG, it is not possible to tell whether the studies were not found by PG or if they were included in the search results but not included in the analysis. Further details are provided in the SM.

Quality assessment of studies. We re-evaluated the quality of all the included studies from PG with the Jadad score for comparison. Assessing the quality of the RCT, we found a low risk of bias and overall high quality for three RCT which agreed with PG. There was disagreement for one RCT that we found to have a high risk of bias [36]. Assessing the quality of PCS with the Jadad score, we did not agree with PG that there was a low risk of bias. Using the Jadad score all of the PCS are given a score of 1, but when re-evaluated with the CRBCS all but one of the studies was shown to be of good quality. Further details are provided in the SM.

Treatment outcomes. We were unable to reproduce PG's results for lost to follow-up (Fig 1 and Table 2 in [12]), even when we tried to use the same underlying studies, data, and methods. We provide detailed information on our inability to reproduce those results (Figures P-R and Tables Z-AA in SM). PG analyzed the outcomes of relapse and acquired drug resistance. Since none of the studies which reported these outcomes meet the inclusion criteria (either ours or those in PG), and since these studies were not designed to assess the differences between DOT and SAT for these outcomes [13], we decided that those analyses were not meaningful and did not try to perform them.

New meta-analysis of DOT vs. SAT

Study Selection and Characteristics. Our systematic review generated 166 unique results, 11 [39–49] studies compared DOT and SAT treatment outcomes in a prospective manner (Fig 1). We contacted the authors of 4 papers in an attempt to get the data on the population of interest; specifically, data on adults only [39,40,42,48] and data stratified by new or re-treatment patients [42] for certain outcomes. We received responses from authors of 2 studies, but neither was able to provide the data. We received no response from the corresponding authors of the other papers. In total 7 [41–44,46,47,49] studies met the inclusion criteria for at least one of the outcomes of interest.
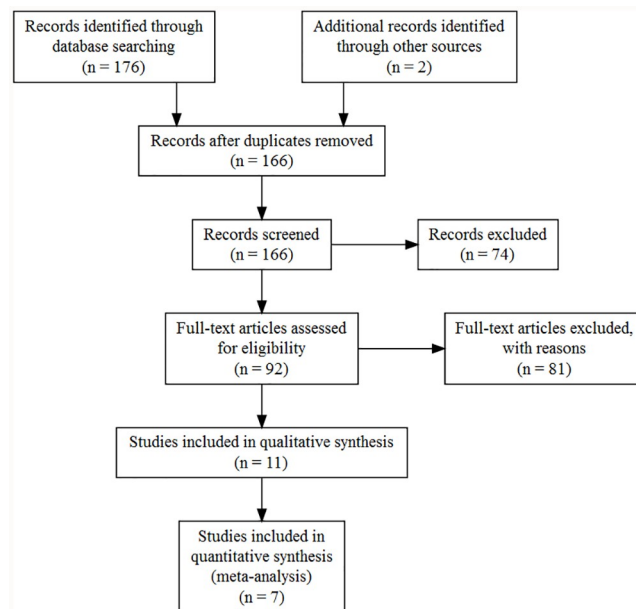
Fig 1. PRISMA diagram. Results of systematic literature review.

https://doi.org/10.1371/journal.pone.0217219.g001

The total number of individuals across all included studies is 3229. This number is less than the 13751 individuals reported in PG due to erroneous inclusion of several additional studies, as described above. Of those 3229 individuals, 1603 were randomized to interventions, 914to DOT, and 689 to SAT. The remaining 1626 patients came from PCS and were not randomly assigned to a treatment group, 998 received DOT, and 628 received SAT. A spreadsheet which provides detailed information on the systematic search and data abstraction is provided in the SM.

Quality assessment of included studies. Assessing the quality of the RCT [43,44,47,49], we found a low risk of bias and overall high quality. Assessing the quality of the PCS, we found one study to be of high quality with a low risk of bias [46]. The other two studies were found to be at an increased risk of bias [41,42]. A spreadsheet detailing the quality assessment of all the studies is included in the SM.

Lost to follow-up. The pooled risk of LTFU in the SAT group was 19.7% (95% CI 10.2, 29.1) vs. 16.1% (95% CI 6.5, 25.7) for the DOT group (Fig 2). The calculated risks for individual studies are shown in the supplementary materials (Table A in S1 Text). The pooled RR was 1.2 (95% CI 0.9, 1.7). The pooled RD (RD = SAT—DOT) was 3.7% (95% CI -1.2, 8.7). Figure for the RD is in the SM (Figure A in S1 Text). Both measures indicate that DOT is favored in reducing the risk of LTFU, but the results are not statistically significant. This result disagrees with the significant difference favoring DOT reported in PG [12].

Treatment failure. The pooled risk of treatment failure in the SAT group was 2.1% (95% CI 0.3, 3.8) and 1.3 (95% CI 0.4, 2.1) for the DOT group (Fig 3). The calculated risks for individual studies are shown in the supplementary materials (Table B in S1 Text). The pooled RR was 1.1 (95% CI 0.6, 2). The pooled RD was 0.5% (95% CI -1, 1.9). Figure for the RD is in the SM (Figure B in S1 Text). Both measures indicate that DOT is slightly favored over SAT, but the results are not statistically significant.

Completion of treatment. The pooled risk of completing treatment in the SAT group was 71.5% (95% CI 65, 78) vs. 72.3% (95% CI 63.4, 81.2) for the DOT group (Fig 4). The
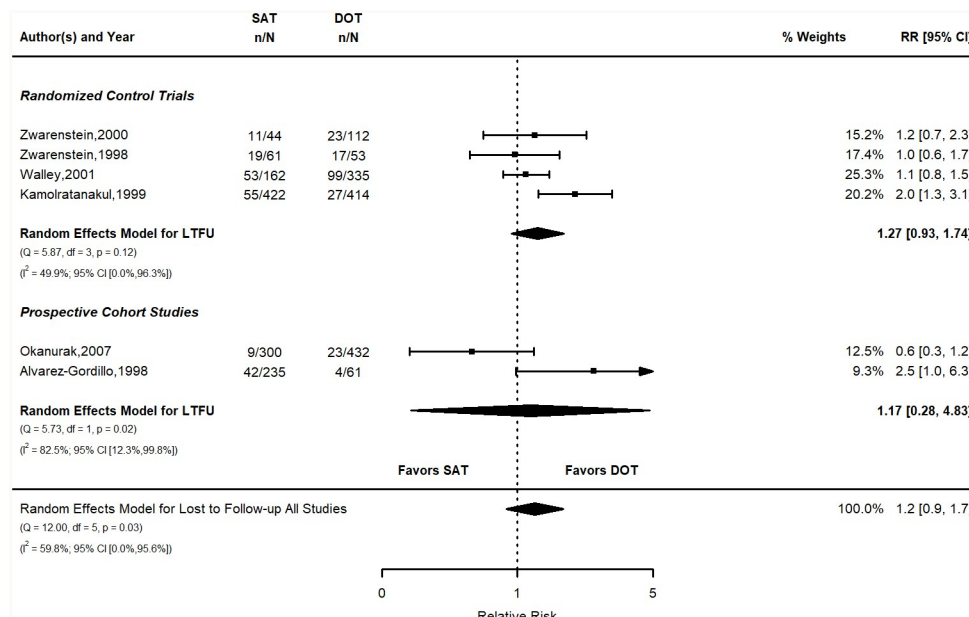
Fig 2. LTFU relative risk stratified by study design. Q = Cochrane's Q; df = Degrees of Freedom; p = P-value associated with Q; I^2 = Proportion of variation due to heterogeneity and corresponding 95% confidence interval.

calculated risks for individual studies are shown in the supplementary materials (Table C in S1 Text). The pooled RR was 1 (95% CI 0.9, 1.1). The pooled RD was -2.1% (95% CI -8, 3.9). Figure for the RD is in the SM (Figure C in S1 Text). Both measures show no evidence of any difference between DOT and SAT.

Death. The pooled risk of death in the SAT group was 3.1% (95% CI 1.6, 4.6) vs. 3.2% (95% CI 1.6, 4.7) for the DOT group (Fig 5). The calculated risks for individual studies are shown in the supplementary materials (Table D in S1 Text). The pooled RR was 0.9 (95% CI 0.6, 1.3). The pooled RD was -0.1% (95% CI -2.1, 1.9). Figure for the RD is in the SM (Figure D in S1 Text). The overall RR indicates that SAT is slightly favored over DOT, but the results are not statistically significant or consistent between study design type.

Cure. The pooled risk of cure in the SAT group was 56.7% (95% CI 47.6, 65.8) vs. 59.1% (95% CI 43, 75.1) for the DOT group (Fig 6). The calculated risks for individual studies are shown in the supplementary materials (Table E in S1 Text). The pooled RR was 0.9 (95% CI 0.8, 1.1). The pooled RD was -4% (95% CI -13.1, 5.1). Figure for the RD is in the SM (Figure E in S1 Text). Both measures indicate that DOT is slightly favored over SAT, but the results are not statistically significant.

Sensitivity and influence. Lost to follow-up was the only outcome that was sensitive to the absence of a single study [46]. When the study is removed, the point estimated risk difference becomes statistically significant, and the estimated heterogeneity goes to zero. Detailed results of the sensitivity and influence analysis for each outcome are included in the SM (Table F-Y in S1 Text).

Publication bias. None of the statistical tests or visual inspection of the funnel plots indicated the presence of publication bias for any of the included outcomes. It is important to note that the power to detect any publication bias is small with so few studies. Detailed results are shown in the SM (Figure F-O in S1 Text).
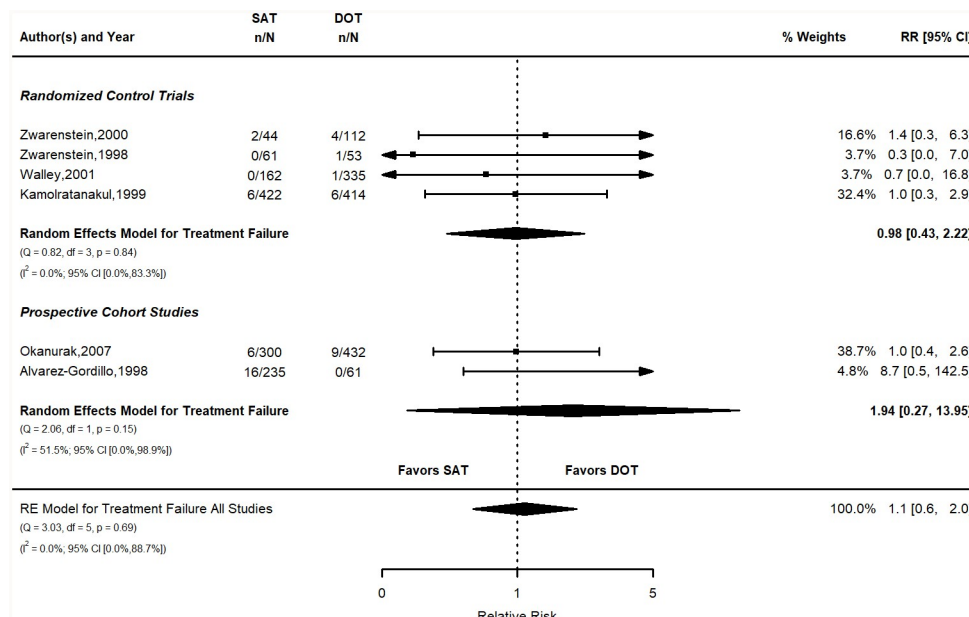
Fig 3. Treatment failure relative risk stratified by study design. Q = Cochrane's Q; df = Degrees of Freedom; p = P-value associated with Q; I² = Proportion of variation due to heterogeneity and corresponding 95% confidence interval.

https://doi.org/10.1371/journal.pone.0217219.g003

Discussion

History has demonstrated that TB control is required to protect public health [50], but the DOT component of TB control has been more controversial. There are a number of meta-
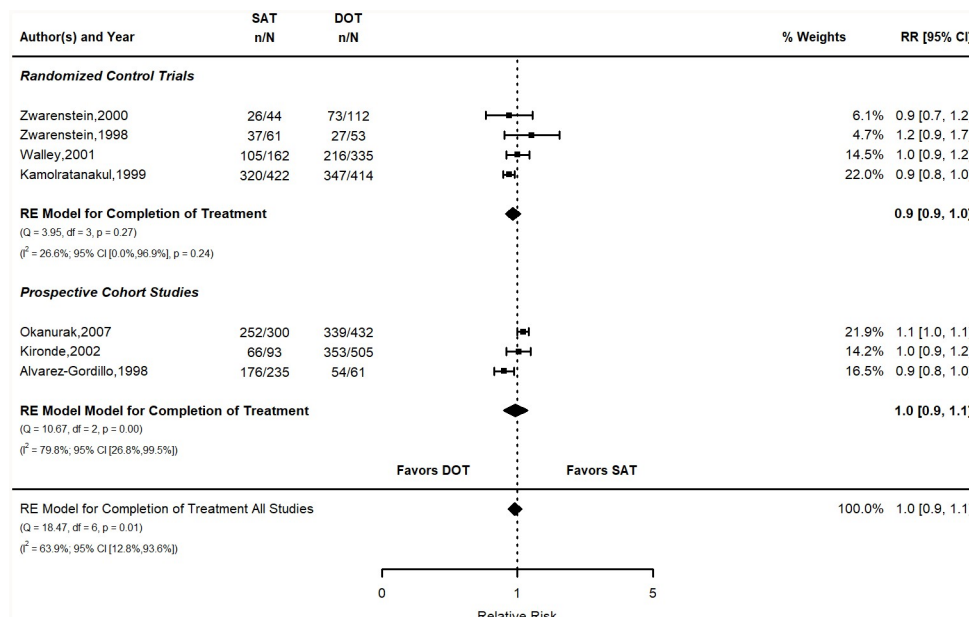

Fig 4. Completion of treatment relative risk stratified by study design. Q = Cochrane's Q; df = Degrees of Freedom; p = P-value associated with Q; I² = Proportion of variation due to heterogeneity and corresponding 95% confidence interval.

https://doi.org/10.1371/journal.pone.0217219.g004

Fig 5. Death relative risk stratified by study design. Q = Cochrane's Q; df = Degrees of Freedom; p = P-value associated with Q; I^2 = Proportion of variation due to heterogeneity and corresponding 95% confidence interval.
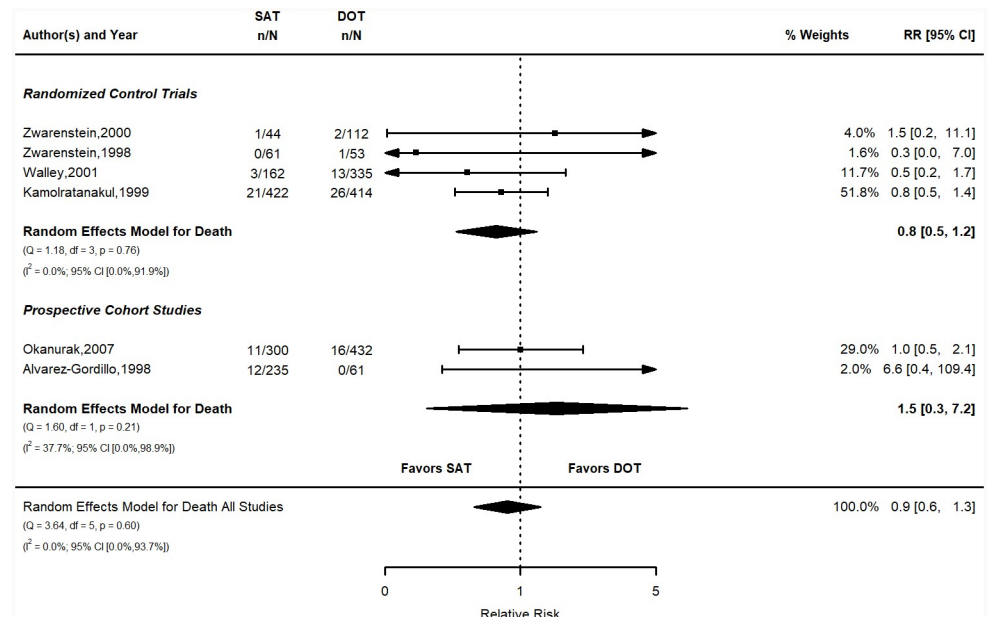
analyses that have been published, and the methodology used for the systematic search of the literature appears to impact the results. Specifically, the types of study designs included seems to determine if the findings are statistically significant or not. Studies that include retrospective studies favor DOT and show a statistically significant difference when comparing DOT and SAT [10,51,52].

Additionally, the definition for DOT and SAT groups can vary significantly within and between meta-analyses. This leads to differences in the studies included as well as how study groups are divided. These issues confirm the need for transparent and complete reporting of the methods used to generate the results for a meta-analysis. Recent Cochrane meta-analyses addressed the effectiveness of DOT, but only included RCT and concluded that there is little evidence in favor of DOT [2,3]. Due to the small number of RCT and the high likelihood that there will not be any new ones addressing this issue, it is useful to include data from PCS to provide a more comprehensive assessment of treatment outcomes for DOT and SAT. While the inclusion of PCS is controversial, it provides additional information and is useful when only a few RCT are available.

A previous meta-analysis by PG included both RCT and PCS. However, as described here, their analysis had methodologic issues rendering their results unreliable [13–15]. Our new meta-analysis corrects problems found in PG and expands the scope of the Cochrane review by including not only RCT but also PCS. We have attempted to make our study easy to reproduce. To that end, all the data and code required to reproduce all our results, as well as, documentation of the entire systematic review process, are provided in the SM. In contrast to PG, we did not find a statistically significant difference between DOT and SAT as regards to LTFU. Similar to past Cochrane reviews[2,3] and PG, we found no statistically significant difference between DOT and SAT for treatment failure. Our findings are in agreement with meta-analyses including only RCT [2,3]. Despite the inclusion of both RCT and PCS, the amount of data available for analysis is still limited, which constrains the power of the results and leads to at
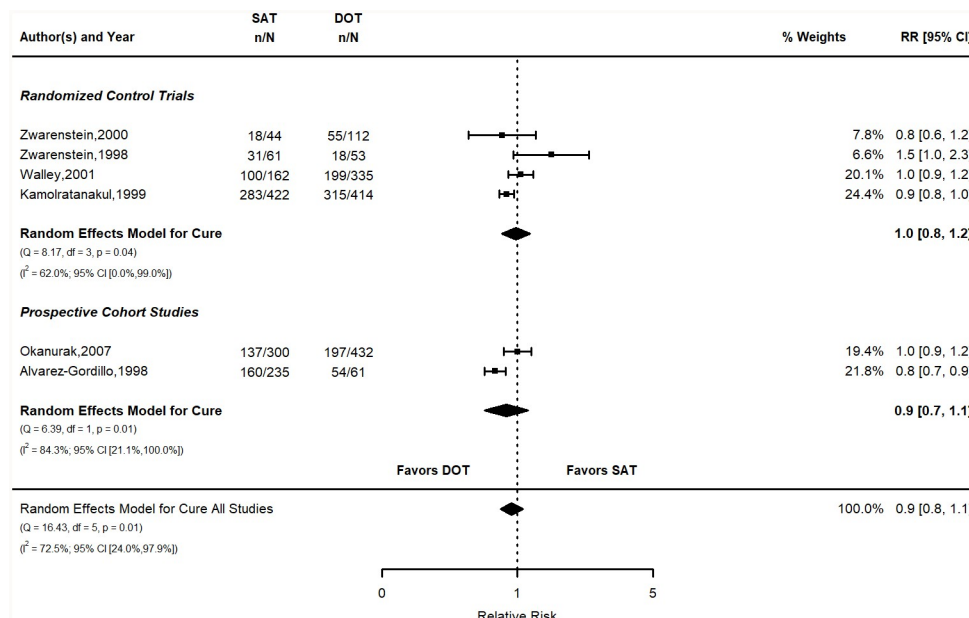
Fig 6. Cure relative risk stratified by study design. Q = Cochrane's Q; df = Degrees of Freedom; p = P-value associated with Q; I^2 = Proportion of variation due to heterogeneity and corresponding 95% confidence interval.

https://doi.org/10.1371/journal.pone.0217219.g006

times wide confidence intervals. Attempts to obtain additional data from published studies were unsuccessful. The inclusion of these data could change the results of our analysis.

We did not address the effects of DOT on the development of secondary drug resistance because of the paucity of the information. The emergence of multi-drug resistant TB and extensively drug-resistant TB threatens TB control now and in the future. In theory, direct observation of effective medications may reduce the likelihood of acquired drug resistance. Evaluating acquired drug resistance as the primary outcome in a meta-analysis may require greater dependence on observational data, since extensive clinical trials may not be feasible.

The results of our study suggest that in adult patients with smear or culture-positive tuberculosis, there is little evidence that DOT is better than SAT. While there often seems to be some indication that DOT is better than SAT, the observed difference is small. Our analysis found no statistically significant differences between DOT and SAT for the primary indicators of treatment success when considering both RCT and PCS. Only in the stratified analysis of RCT studies is there any evidence of a significant difference between DOT and SAT for the RD of LTFU. We believe our findings further support the current lack of evidence that DOT is superior to SAT for improving cure rates, adherence [2,3], or any other measure of treatment success [4,5]. It is important to note that despite the importance of TB we agree there is a lack of high-quality data available in the literature [8], and while we believe traditional RCTs are unlikely to occur in the future there is a need for more high-quality evidence in the form of prospective cohort and cluster randomized trials [8]. Such studies could provide a clear understanding of the effectiveness of DOT. Additionally, the ability to compare novel means of treatment monitoring should be conducted based on the needs and consideration of the populations being served[53].

In summary, we could not reproduce a previous meta-analysis, due to inconsistencies regarding study inclusion and insufficient information about the analysis methods. Further improvements toward robust reproducibility of meta-analyses, especially those directly related to interventions are needed [54–57]. Our new analysis found that for adult patients 15 years or

older with drug-susceptible pulmonary TB, there is no statistically significant difference between DOT and SAT and the magnitude of the difference between DOT and SAT for all reported outcomes is small. The role of DOT in the DOTS strategy remains controversial and we agree with McLaren et al. that further efforts are needed to evaluate its role [8].

Data sharing statement

Documentation and data for reproducibility are freely available in the supplementary materials.

Supporting information

S1 Text. Supplementary Documentation. This document contains information about all the supplementary materials, explains how to reproduce the results of our analysis, and has additional results not shown in the main text.
(DOCX)

S1 Folder. Zip folder. This folder contains all of the files described in the Supplementary Documentation.
(ZIP)

Author Contributions

Conceptualization: Christopher C. Whalen, Andreas Handel.

Data curation: Brian McKay, Maria Castellanos.

Formal analysis: Brian McKay, Andreas Handel.

Investigation: Brian McKay, Maria Castellanos, Andreas Handel.

Methodology: Brian McKay, Mark Ebell, Christopher C. Whalen, Andreas Handel.

Project administration: Brian McKay.

Software: Brian McKay.

Supervision: Mark Ebell, Christopher C. Whalen, Andreas Handel.

Validation: Mark Ebell.

Visualization: Brian McKay.

Writing – original draft: Brian McKay, Maria Castellanos.

Writing – review & editing: Brian McKay, Maria Castellanos, Mark Ebell, Christopher C. Whalen, Andreas Handel.

References

1. Glaziou P, Falzon D, Floyd K, Raviglione M. Global epidemiology of tuberculosis. Semin Respir Crit Care Med. 2013; 34(1):3–16. https://doi.org/10.1055/s-0032-1333467 PMID: 23460002

2. Volmink J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev. 2007; 4.

3. Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev. 2015;(5).

4. de Steenwinkel JEM, Marian T, de Knegt GJ, Verbrugh HA, Aarnoutse RE, Boeree MJ, et al. Consequences of noncompliance for therapy efficacy and emergence of resistance in murine tuberculosis caused by the beijing genotype of mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2012; 56(9):4937–44. https://doi.org/10.1128/AAC.00124-12 PMID: 22802244

5. Srivastava S, Pasipanodya JG, Meek C, Leff R, Gumbo T. Multidrug-resistant tuberculosis not due to noncompliance but to between-patient pharmacokinetic variability. Journal of Infectious Diseases. 2011; 204(12):1951–9. https://doi.org/10.1093/infdis/jir658 PMID: 22021624

6. Garner P, Volmink J. Directly observed treatment for tuberculosis-less faith, more science would be helpful [editorial]. British Medical Journal. 2003; 327(7419):823–4.

7. Macq JC, Theobald S, Dick J, Dembele M, others. An exploration of the concept of directly observed treatment (dot) for tuberculosis patients: From a uniform to a customised approach. The International Journal of Tuberculosis and Lung Disease. 2003; 7(2):103–9.

8. McLaren ZM, Milliken AA, Meyer AJ, Sharp AR. Does directly observed therapy improve tuberculosis treatment? More evidence is needed to guide tuberculosis policy. BMC Infectious Diseases. 2016; 16 (1):537. https://doi.org/10.1186/s12879-016-1862-y PMID: 27716104

9. Denkinger C, Grenier J, Stratis A, Akkihal A, Pant-Pai N, Pai M. Mobile health to improve tuberculosis care and control: A call worth making. The International Journal of Tuberculosis and Lung Disease. 2013; 17(6):719–27. https://doi.org/10.5588/ijtld.12.0638 PMID: 23541232

10. Tian J-H, Lu ZX, Bachmann MO, Song FJ. Effectiveness of directly observed treatment of tuberculosis: A systematic review of controlled studies. The International Journal of Tuberculosis and Lung Disease. 2014; 18(9):1092–8. https://doi.org/10.5588/ijtld.13.0867 PMID: 25189558

11. Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing risk of bias and confounding in obser- vational studies of interventions or exposures: Further development of the rti item bank. Agency for Healthcare Research; Quality (US); 2013.

12. Pasipanodya JG, Gumbo T. A meta-analysis of self-administered vs directly observed therapy effect on microbiologic failure, relapse, and acquired drug resistance in tuberculosis patients (vol 57, pg 21, 2013). Clinical Infectious Diseases. 2013; 57(8):1223–3.

13. Nunn AJ, Phillips PPJ. An analysis with serious flaws. Clinical Infectious Diseases. 2013; 57(7):1064–5. https://doi.org/10.1093/cid/cit433 PMID: 23811419

14. Moonan PK, Weis SE. Caveat emptor? Meta-analysis of studies comparing self-observed therapy and directly observed therapy for tuberculosis. Clinical Infectious Diseases. 2013; 57(7):1062–3. https://doi.org/10.1093/cid/cit431 PMID: 23811420

15. Seaworth BJ, Armitige LY, Griffith DE. First do no harm: Adverse events, drug intolerance, and hepato- toxicity: How can we not justify directly observed therapy for treating tuberculosis? Clinical Infectious Diseases. 2013; 57(7):1063–4. https://doi.org/10.1093/cid/cit432

16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta- analyses: The prisma statement. Journal of Clinical Epidemiology. 2009; 62(10):1006–12. https://doi.org/10.1016/j.jclinepi.2009.06.005 PMID: 19631508

17. WHO. What is dots?: A guide to understanding the who-recommended tb control strategy known as dots. World Health Organization. 1999;

18. Lwilla F, Schellenberg D, Masanja H, Acosta C, Galindo C, Aponte J, et al. Evaluation of efficacy of community-based vs. institutional-based direct observed short-course treatment for the control of tuber- culosis in kilombero district, tanzania. Tropical Medicine & International Health: TM & IH. 2003; 8 (3):204–10.

19. MacIntyre CR, Goebel K, Brown GV, Skull S, Starr M, Fullinfaw RO. A randomised controlled clinical trial of the efficacy of family-based direct observation of anti-tuberculosis treatment in an urban, devel- oped-country setting. The International Journal of Tuberculosis and Lung Disease. 2003; 7(9):848–54. PMID: 12971668

20. Wandwalo E, Kapalata N, Egwaga S, Morkve O. Effectiveness of community-based directly observed treatment for tuberculosis in an urban setting in tanzania: A randomised controlled trial. The Interna- tional Journal of Tuberculosis and Lung Disease. 2004; 8(10):1248–54. PMID: 15527158

21. Wright J, Walley J, Philip A, Pushpananthan S, Dlamini E, Newell J, et al. Direct observation of treat- ment for tuberculosis: A randomized controlled trial of community health workers versus family mem- bers. Tropical Medicine & International Health: TM & IH. 2004; 9(5):559–65.

22. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. Vol. 5. Wiley Online Library; 2008.

23. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the qual- ity of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials. 1996; 17 (1):1–12. PMID: 8721797

24. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. https://www.R-project.org/

25. Viechtbauer W. Metafor: Meta-analysis package for r [Internet]. 2017. https://CRAN.R-project.org/package=metafor

26. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Statistics in medicine. 2002; 21(11):1575–600. https://doi.org/10.1002/sim.1188

27. Friedrich JO, Adhikari NK, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. BMC medical research methodology. 2007; 7(1):5.

28. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. Journal of Educational and Behavioral Statistics. 2005; 30(3):261–93.

29. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. Statistics in medicine. 2007; 26(9):1964–81. https://doi.org/10.1002/sim.2688 PMID: 16955539

30. Thorlund K, Wetterslev J, Awad T, Thabane L, Gluud C. Comparison of statistical inferences from the dersimonian–Laird and alternative random-effects model meta-analyses–an empirical assessment of 920 cochrane primary outcome meta-analyses. Research synthesis methods. 2011; 2(4):238–53. https://doi.org/10.1002/jrsm.53

31. Jackson D, Bowden J, Baker R. How does the dersimonian and laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? Journal of Statistical Planning and Inference. 2010; 140(4):961–70.

32. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. Statistics in medicine. 1999; 18(20):2693–708. PMID: 10521860

33. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Random-effects model. In: Introduction to meta-analysis. John Wiley & Sons, Ltd; 2009. pp. 69–75.

34. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Bmj. 1997; 315(7109):629–34. https://doi.org/10.1136/bmj.315.7109.629

35. Orwin RG. A fail-safe n for effect size in meta-analysis. Journal of educational statistics. 1983; 8 (2):157–9.

36. Tuberculosis Research C. A controlled clinical trial of oral short-course regimens in the treatment of sputum-positive pulmonary tuberculosis. The International Journal of Tuberculosis and Lung Disease. 1997; 1(6):509–17. PMID: 9487448

37. Jasmer RM, Seaman CB, Gonzalez LC, Kawamura LM, Osmond DH, Daley CL. Tuberculosis treatment outcomes: Directly observed therapy compared with self-administered therapy. American journal of respiratory and critical care medicine. 2004; 170(5):561–6. https://doi.org/10.1164/rccm.200401-095OC

38. Ormerod L, Horsfield N, Green R, others. Tuberculosis treatment outcome monitoring: Blackburn 1988–2000. The International Journal of Tuberculosis and Lung Disease. 2002; 6(8):662–5. PMID: 12150476

39. Anuwatnonthakate A, Limsomboon P, Nateniyom S, Wattanaamornkiat W, Komsakorn S, Moolphate S, et al. Directly observed therapy and improved tuberculosis treatment outcomes in thailand. PLoS One. 2008; 3(8):e3089. https://doi.org/10.1371/journal.pone.0003089 PMID: 18769479

40. Pungrassami P, Johnsen SP, Chongsuvivatwong V, Olsen J. Has directly observed treatment improved outcomes for patients with tuberculosis in southern thailand? Tropical Medicine & International Health. 2002; 7(3):271–9.

41. Del Carmen Alvarez-Gordillo G, Dorantes-Jimenez J. Shortened directly observed treatment applied to the tuberculosis control program. Salud Publica de Mexico. 1998; 40(3):272–5.

42. Kironde S, Kahirimbanyib M. Community participation in primary health care (phc) programmes: Lessons from tuberculosis treatment delivery in south africa. African Health Sciences. 2002; 2(1):16–23. PMID: 12789110

43. Zwarenstein M, Schoeman JH, Vundule C, Lombard CJ, Tatley M. Randomised controlled trial of self-supervised and directly observed treatment of tuberculosis. The Lancet. 1998; 352(9137):1340–3.

44. Zwarenstein M, Schoeman JH, Vundule C, Lombard CJ, Tatley M. A randomised controlled trial of lay health workers as direct observers for treatment of tuberculosis. The International Journal of Tuberculosis and Lung Disease. 2000; 4(6):550–4. PMID: 10864186

45. Tandon M, Gupta M, Tandon S, Gupta KB. DOTS versus self administered therapy (sat) for patients of pulmonary tuberculosis: A randomised trial at a tertiary care hospital. Indian journal of medical sciences. 2002; 56(1):19–21. PMID: 12508627

46. Okanurak K, Kitayaporn D, Wanarangsikul W, Koompong C. Effectiveness of dot for tuberculosis treatment outcomes: A prospective cohort study in bangkok, thailand. The International Journal Of Tuberculosis And Lung Disease. 2007; 11(7):762–8. PMID: 17609051

47. Walley JD, Khan MA, Newell JN, Khan MH. Effectiveness of the direct observation component of dots for tuberculosis: A randomised controlled trial in pakistan. The Lancet. 2001; 357(9257):664–9.

48. Mathema B, Pande S, Jochem K, Houston R, Smith I, Bam D, et al. Tuberculosis treatment in nepal: A rapid assessment of government centers using different types of patient supervision. The International Journal of Tuberculosis and Lung Disease. 2001; 5(10):912–9. PMID: 11605884

49. Kamolratanakul P, Sawert H, Lertmaharit S, Kasetjaroen Y, Akksilp S, Tulaporn C, et al. Randomized controlled trial of directly observed treatment (dot) for patients with pulmonary tuberculosis in thailand. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1999; 93(5):552–7. https://doi.org/10.1016/s0035-9203(99)90379-6 PMID: 10696421

50. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in new york city: Turning the tide. New England Journal of Medicine. 1995; 333(4):229–33. https://doi.org/10.1056/NEJM199507273330406

51. Alipanah N, Jarlsberg L, Miller C, Linh NN, Falzon D, Jaramillo E, et al. Adherence interventions and outcomes of tuberculosis treatment: A systematic review and meta-analysis of trials and observational studies. PLoS medicine. 2018; 15(7):e1002595. https://doi.org/10.1371/journal.pmed.1002595 PMID: 29969463

52. Müller A, Osório C, Silva D, Sbruzzi G, Tarso P de, Dalcin R. Interventions to improve adherence to tuberculosis treatment: Systematic review and meta-analysis. The International Journal of Tuberculosis and Lung Disease. 2018; 22(7):731–40. https://doi.org/10.5588/ijtld.17.0596

53. WHO, others. Guidelines for the treatment of drug-susceptible tuberculosis and patient care. World Health Organization. 2017;

54. Peng RD. Reproducible research in computational science. Science. 2011; 334(6060):1226–7. https://doi.org/10.1126/science.1213847 PMID: 22144613

55. Schwab M, Karrenbach M, Claerbout J. Making scientific computations reproducible. Computing in Science & Engineering. 2000; 2(6):61–7.

56. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. PLoS Biol. 2015; 13(6):e1002165. https://doi.org/10.1371/journal.pbio.1002165 PMID: 26057340

57. Mesirov JP. Accessible reproducible research. Science. 2010; 327(5964):415–6. https://doi.org/10.1126/science.1179653 PMID: 20093459