Comparing the efficacy of drug regimens for pulmonary tuberculosis: meta-analysis of endpoints in early phase clinical trials

Laura J Bonnett – Department of Biostatistics, University of Liverpool, Liverpool UK

Gie Ken-Dror – Department of Biostatistics, University of Liverpool, Liverpool, UK

Gavin Koh – Clinical Development, GlaxoSmithKline, Middlesex, UK

Geraint R Davies – Department of Clinical Infection, Microbiology & Immunology, University of Liverpool, Liverpool, UK

Key points: This meta-analysis, which systematically appraises the performance of single drugs and combination regimens across early clinical endpoints in trials of treatment of pulmonary tuberculosis, demonstrates that the existing evidence base supporting Phase II methodology in tuberculosis is highly incomplete.

Key words: Tuberculosis, Efficacy, Meta-Analysis

Running title: Efficacy of drugs for pulmonary TB

Corresponding Author: Dr Geraint Davies - Department of Clinical Infection, Microbiology & Immunology, University of Liverpool, 8 West Derby Street, Liverpool, L69 7BE

+44 (0)151 7959685

gdavies@liverpool.ac.uk

Alternative Author: Dr Laura Bonnett - Department of Biostatistics, University of Liverpool, Waterhouse Buildings, Block F, 1-5 Brownlow Street, Liverpool, L69 3GL

+44 (0)151 7959686

l.j.bonnett@liverpool.ac.uk

Abstract

A systematic review of early clinical outcomes in tuberculosis was undertaken to determine ranking of efficacy of drugs and combinations, define variability of these measures on different endpoints, and to establish the relationships between them. Studies were identified by searching PubMed, MEDLINE, EMBASE, LILACs and reference lists of included studies. Outcomes were early bactericidal activity results over two, seven and 14 days, and the proportion of patients with negative culture at eight weeks. 133 trials reporting Phase IIA (early bactericidal activity) and IIB (culture-conversion at two months) outcomes were identified. Only nine drug combinations were assessed on more than one Phase IIA endpoint and only three were assessed in both Phase IIA and IIB trials. The existing evidence base supporting Phase II methodology in tuberculosis is highly incomplete. In future, a broader range of drugs and combinations should be more consistently studied across a greater range of Phase II endpoints.

Introduction

First-line tuberculosis therapy has remained unchanged for forty years. While "short-course" treatment is effective in clinical trials, in practice the six months required for successful cure is burdensome for patients and tuberculosis programmes. Identifying new and ultra-short regimens will require identification of suitable surrogate outcomes to facilitate progression of novel treatment regimens through Phase II to Phase III trials and de-risk drug development [1].

The current "gold standard" Phase III endpoint is a composite of treatment failure and relapse up to 24 months following treatment completion. Use of this binary outcome which is rare in the comparator arm (<5% with standard short course regimens) mandates large sample sizes to adequately power clinical trials. The prolonged follow-up required further adds to trial costs, making the definitive outcome unsuitable for extensive evaluation of drug combinations or dose-finding.

Numerous surrogate outcomes have been used for these purposes in Phase II. Phase IIA studies of "early bactericidal activity" (EBA) based on quantitative sputum bacteriology enrol small patient numbers for up to two weeks [2]. While the original rationale for such studies was dose-finding for single agents, more recent studies have evaluated drug combinations [3, 4]. This concept has been extended into larger Phase IIB studies with combination therapy lasting up to two months [5]. The most studied Phase IIB outcome has been sputum culture conversion at fixed time-points, usually two months [6-8]. This endpoint is supported by regulators for conditional approval of novel drugs [9], but there remains a lack of consensus amongst trialists as to the utility of EBA studies and of other approaches to intermediate bacteriological data such as time-to-event and regression modelling [10].

A complete understanding of the performance of TB treatment regimens in early phase clinical trials is critical to understanding their usefulness in predicting Phase III trial results and in

calibrating preclinical models of treatment. While the goals of historical Phase IIA and IIB regimens are distinct, with the former focussing on proof-of-concept for individual drugs and the latter on identifying the best combinations of drugs, it seems important to understand whether this information can be transmitted rationally through these phases. We undertook a systematic review of early clinical outcomes in tuberculosis (within the first two months of treatment), focusing on the key drugs comprising modern and historical first-line treatment regimens, to determine the overall ranking of efficacy of drugs and combinations, to define the variability of these measures of effect on the different endpoints used, and to establish the relationships between them.

Methods

The review included randomised clinical trials (RCTs) including patients with smear- and culture-positive pulmonary tuberculosis, being treated for the first time or with known isoniazid mono-resistant organisms, and including regimens containing any combination of historic or novel drugs used or proposed for use in first-line treatment regimens. Pre-defined outcomes of interest were EBA over two (EBA_{0-2}), seven (EBA_{0-7}) and 14 days (EBA_{0-14}), and the proportion of patients with negative culture results at eight weeks. A systematic search of databases was conducted on 12th December 2016 (see Appendix). Risk of bias was considered.[11] Pooled estimates of each outcome for each drug or combination were obtained. Meta-regression was used to examine the impact of clinical covariates on the effect size of culture results at eight weeks. Analyses were performed using R version 2.14.1 [12]. Full methodology is detailed in the Appendix.

Results

Included Studies

Figure 1 shows the number of studies included at each stage of the review. The main reasons for exclusion were failure to meet the inclusion criteria, specifically previously-treated or drug-resistant patients, and study design other than RCT. 133 relevant studies were identified which reported outcomes of interest. Of these, 37 were Phase II studies and 96 were Phase III studies reporting intermediate bacteriological outcomes. All 96 Phase III studies contributed to the Phase IIB outcome only – no Phase III studies contributed data to the Phase IIA outcomes. Together these studies provide data relating to 37,173 patients and 67 drug combinations.

Figure 2 summarises numbers of studies and patients pertaining to each drug combination across all the outcomes of interest. Since only combinations including drugs of interest to the review are summarised, in some cases only data concerning the control arms of trials are presented. Additionally, studies with two or more trial arms were analysed separately. The composition of each drug combination refers only to the period preceding the endpoint of interest. Therefore, for the eight week culture outcome usually only drugs used in the initiation phase of treatment are reported without the associated continuation phase drugs. Where the initiation phase was less than two months, however, continuation drugs are also listed. The regimens for which most data were available were HRZE, SHRZ and HRZ combination therapy.

Twenty-four studies reported Phase IIA outcomes. $EBA_{0\text{-}2}$ was reported in all 24 studies (141 trial arms, 35 drug combinations, 1424 patients). In some cases, studies considered the same drug combination but different treatment strategies and dosing intervals. Others considered single formulation treatments versus combined formulations, and some considered multiple dosages of a drug. $EBA_{0\text{-}7}$ and $EBA_{0\text{-}14}$ were reported in only six (23 trial arms, 14 drug

combinations, 296 patients) and eight studies (46 trial arms, 27 drug combinations, 449 patients) respectively.

The proportion of patients culture negative by eight weeks was reported in 104 studies considering Phase IIB outcomes. These studies investigated 45 different drug combinations in 34,418 patients. One study reported both Phase IIA and Phase IIB outcomes [13].

Forest plots for each drug combination and outcome and associated numerical results are presented in the Appendix. Results shown in Figures 3 to 8 are graphical summaries based on standard doses recommended in treatment guidelines in the case of historic drugs, or doses going into Phase III trials in the case of novel drugs.

Risk of Bias

87 (65%) studies provided information on sequence generation. In most cases (94%) patients were "randomly allocated" so studies were classified as unclear risk of bias. Some publications mentioned stratifying by factors such as severity, or used permuted block designs, random tables, or similar and were classified as low risk of bias. Five studies referred to quasi-randomisation and were therefore classified as high risk of bias.

Only 30 (23%) studies mentioned allocation concealment. Of these, 28 (93%) studies used sealed envelopes and were classified as low risk. 97 (73%) studies either reported that the study was un-blinded, or did not specify blinding procedures and were classified as high risk of bias. 15 included studies were of a double-blinded nature including the use of telephone randomisation, or prearranged lists, although one stated that it was double-blind during the maintenance phase of treatment only [14]. Most other blinded studies mentioned that radiographers or laboratory staff were blinded to treatment - these were considered as single-blinded designs and classified as low risk of bias.

97 (73%) of studies were published pre-CONSORT when selective reporting had not been raised as a possible source of bias. In all studies published post-CONSORT, the risk of bias is unclear as there is insufficient information to determine whether the published reports include all expected outcomes, including those pre-specified. 85 (64%) studies reported reasons for exclusions, or numbers lost to follow-up.

Due to the limited number of high quality studies, sensitivity analyses assessing the impact of risk of bias was not performed.

Phase IIA Studies

EBA_{0-2}

Pooled results for the EBA_{0-2} outcome can be seen in Figure 3. Of the 32 drugs and combinations, only five were studied in more than 30 patients. Hence the confidence intervals on pooled estimates of effect are wide and frequently overlap. Some drugs (Rb, Z, J and S) do not demonstrate any significant efficacy on this endpoint and can also clearly be distinguished from H, the most commonly studied and precisely estimated drug. However, even quite commonly studied combinations containing H such as HRZ, HRZE and SHRZ do not appear significantly different from H monotherapy using the EBA_{0-2} endpoint. Similarly, it does not appear possible to separate the effect of HRZE from any of its component drugs, with the exception of Z.

EBA_{0-7} and EBA_{0-14}

Pooled results for the EBA_{0-7} and EBA_{0-14} outcomes can be seen in Figures 4 and 5 respectively. Although the number of studies reporting these endpoints was fewer, variability of these endpoints appeared lower than for EBA_{0-2}. Even so, it did not appear possible to distinguish

statistically between the drugs and regimens studied, including combinations such as HRZE and drugs as diverse as H, Pa and J.

Phase IIB Studies

Figures 6 to 8 summarise the pooled estimates of the proportion of patients culture negative at eight weeks, grouped by culture method. The overall point estimates of culture conversion for the most effective rifampicin-containing regimens on this endpoint (HRZ, SHRZ and HRZE) exceeded 85% while for most non-rifampicin regimens this estimate was no better than 50%. Though relatively precise estimates were obtained for frequently studied modern short-course regimens such as HRZ, SHRZ and HRZE, the performance of these regimens was not statistically distinguishable from historical regimens comprising similar numbers of trials and patients such as HS, HPS or SHR. This appeared to reflect high inter-trial variability within regimens as measured by I^2 and τ^2 estimated using two approaches (see Appendix).

Fewer data were available relating to culture conversion as measured by liquid culture. Although variability appeared lower than with solid culture, this could be due to the smaller number of studies included in this analysis. Variability increased when pooled results based on solid and liquid culture methods were reported and confidence intervals for all the regimens tested using liquid culture results overlapped.

Meta-regression

The results of meta-regression analyses can be seen in Table 1. For the selected drug combinations HRZE, HRZ and SHRZ neither year of publication nor geographical location were statistically significant. HIV co-infection could only be examined for the drug combination HRZE as there was insufficient data for HRZ and SHRZ. It was also not significantly associated with the proportion of patients who were culture negative at either

weeks. The inclusion of R and Z in a regimen independently explained significant heterogeneity among drug combinations (see Appendix).

Ranking

Since it was difficult to discriminate between regimens in terms of formal statistical inference, we evaluated whether the rank order of regimens was consistent between different endpoints. The ranking however was highly constrained by the limited number of drugs and regimens studied in both phases, principally because of the ethical unacceptability of prolonged monotherapy in two month studies and the lack of historical combination EBA studies.

Nine distinct regimens (HRZE, H, J, JPa, JZ, Pa, PaMZ, PaZ, and R) were considered on at least two Phase IIA endpoints. All nine were considered when examining $EBA_{0\text{-}2}$, and all except Pa were considered when examining $EBA_{0\text{-}7}$. Seven were considered (H, HPa, HZ, Pa, PaMZ, and PaZ) when examining $EBA_{0\text{-}14}$. Only three of these regimens (HRZE, H, and R) were considered for two month culture conversion on solid media.

Using the relative order among the drugs in the common sets and basing results only on solid culture, the rankings are shown in Table 2. Though qualitative rankings for the available regimens were reasonably consistent the dataset was too small to be able to draw conclusions about their usefulness for decision making.

Discussion

This review is the first to systematically appraise the performance of single drugs and combination regimens across early clinical endpoints in trials of treatment of pulmonary tuberculosis. Though we focussed only on the set of drugs most relevant to historical and modern first-line therapy, we identified 133 trials reporting Phase IIA and IIB outcomes

comprising more than 37,000 patients and 67 drug combinations. However, the diversity of treatment regimens represented in 14 day Phase IIA studies was much lower, with only nine drug combinations assessed on more than one Phase IIA endpoint and only three of these combinations assessed in both Phase IIA and IIB trials. While these findings partly reflect the history, development and differing goals of such trials, the narrowness of this evidence base is concerning and suggests a potentially serious gap in rational translation between these two critical phases of development. While rankings of the efficacy of treatment appeared reasonably consistent on different Phase IIA endpoints, the existing dataset does not provide convincing support to current practices and inter-trial variability was high in many cases.

We selected four outcome measures for our review based on those most commonly reported in the included studies [15]. However, there was large variation in reporting, particularly of EBA measures, with many unique to a single study. Overall the quality of reporting, particularly of Phase IIA studies made data extraction and synthesis challenging and imposes limitations in interpretation of the data. The striking feature of the available dataset is the variability of pooled estimates of effect for all the endpoints examined. For $EBA_{0\text{-}2}$ and two month culture conversion, this variability was particularly marked, with overlapping confidence intervals for the majority of regimens. Though there were appreciable differences between the best-performing regimens on these endpoints (H, HRZ, HRZE and SHRZ) and the worst (Z, S, SH and SHP), this suggests that such trials may lack the power to formally discriminate between regimens where differences in treatment effect are more modest but still clinically relevant. The reasons for this variability were difficult to explore using the data available, given the quality and consistency of reporting.

We used the data as reported – some studies adopted an intention to treat approach to analysis and included patients with missing or contaminated culture results, while others used a per protocol approach and excluded these patients. Poor quality reporting meant it was mostly

impossible to distinguish these situations. This may account for some of the observed heterogeneity.

The variance of the pooled estimates for EBA endpoints may be inflated by the regression coefficients being based on different numbers of observations. There were no such methodological problems for the two month culture-conversion results, suggesting that the observed heterogeneity is likely to be a real clinical effect. Among the most likely sources of this within-regimen heterogeneity are pharmacological (ethnic differences in absorption, elimination and clearance), bacteriological (differences in initial bacterial burden, and virulence) and patient factors (disease stage such as presence of cavities, co-morbidities such as malnutrition, diabetes, and HIV).Our meta-regression analysis was able to explore a very limited subset of such variables for a few of the most common regimens on a single endpoint.

Since the review incorporated all reported trials over the last sixty years, evolution in the efficiency and standardisation even of solid culture methods may have contributed to variation in two month culture results, with older studies tending to produce numerically higher rates (not statistically significant) of culture conversion due to lower assay sensitivity compared to more modern methods. We tested this via meta-regression for drug combinations HRZE, HRZ and SHRZ and in all three cases year of publication was not significant. Isoniazid resistance, whether known or undetected, may have tended to increase heterogeneity in outcomes, but this is difficult to assess due to small numbers of patients. While it is known that at least for modern regimens such as HRZE the risk of poor outcome for patients with H-resistant organisms is only modestly increased [16, 17] this could be more important for older regimens although there was insufficient data to test this in a meta-regression. Finally, patient factors particularly chronicity of disease, geographical location and HIV co-infection could also have increased inter-trial variability within regimens although there was insufficient data to test this for chronicity of disease. It has been observed that culture conversion at two months may vary

widely even between study sites in individual trials [18] and may be influenced by the lower sputum bacillary load observed in HIV+ patients.

We propose three approaches which could help to overcome some of the limitations this review identifies in the existing evidence base for Phase II trials in TB. Firstly, assembly of a database of individual patient data relating to the trials identified would facilitate re-analysis of the trials and also enable computation of endpoints not reported in the original study publication, which may help to address the lack of diversity of regimens on each endpoint. Such an effort is currently in progress by the PreDiCT-TB consortium and we anticipate an update of this review based on individual patient data when that process has been completed.

Second, development of a core outcome set for tuberculosis trials which could be applied to new studies in the field would likely assist both investigators and systematic reviewers in choosing and reporting endpoints in such a way that the contribution of each trial to the overall evidence base is maximised [19]. This would provide a minimum but not exhaustive set of clearly defined outcomes to be reported in each study.

Finally, we also suggest wider use of novel, more efficient, adaptive screening trial designs which would enable a broader range of regimens to be studied in Phase II than was previously possible. However, such trials may also impact any meta-analysis to which they contribute in terms of the endpoints that they prioritise and any bias they might introduce due to early termination.

Our review shows that the existing evidence base supporting Phase II methodology in tuberculosis is highly incomplete. To truly understand and improve drug development in tuberculosis, it is desirable that a broader range of drugs and combinations be more consistently studied across a greater range of Phase II endpoints than is currently available and that these regimens be rigorously compared in a cumulative meta-analytic framework. Though this

review forms an initial contribution, achieving this goal will require a coordinated and multidisciplinary effort by the TB trials community.

Funding

The research leading to these results has received funding from the Innovative Medicines Initiative Joint Undertaking under grant agreement n°115337, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

Conflict of Interest

All authors declare no conflicts of interest.

Acknowledgements

The authors wish to thank Sarah Donegan for her review of this manuscript.

References

1. Lienhardt C, Raviglione M, Spigelman M, et al. New Drugs for the Treatment of Tuberculosis: Needs, Challenges, Promise, and Prospects for the Future. Journal of Infectious Diseases **2012**; 205(suppl 2): S241-S9.
2. Donald PR, Diacon AH. The early bactericidal activity of anti-tuberculosis drugs: a literature review. Tuberculosis **2008**; 88, Supplement 1: S75-S83.
3. Jindani A, Aber VR, Edwards EA, Mitchison DA. The early bactericidal activity of drugs in patients with pulmonary tuberculosis. The American review of respiratory disease **1980**; 121(6): 939-49.
4. Diacon AH, Dawson R, von Groote-Bidlingmaier F, et al. 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: a randomised trial. The Lancet 380(9846): 986-93.
5. Rustomjee R, Lienhardt C, Kanyok T, et al. A Phase II study of the sterilising activities of ofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease **2008**; 12(2): 128-38.
6. Phillips PP, Fielding K, Nunn AJ. An evaluation of culture results during treatment for tuberculosis as surrogate endpoints for treatment failure and relapse. PloS one **2013**; 8(5): e63840.
7. Wallis RS, Wang C, Meyer D, Thomas N. Month 2 Culture Status and Treatment Duration as Predictors of Tuberculosis Relapse Risk in a Meta-Regression Model. PloS one **2013**; 8(8): e71116.

8. Wallis RS, Peppard T, Hermann D. Month 2 Culture Status and Treatment Duration as Predictors of Recurrence in Pulmonary Tuberculosis: Model Validation and Update. PloS one **2015**; 10(4): e0125403.
9. Center for Drug Evaluation and Research. Pulmonary Tuberculosis: Developing drugs for treatment. In: Services US Department of Health and Human Services: Food and Drug Administration, **2013**.
10. Davies GR, Phillips PPJ, Jaki T. Adaptive clinical trials in tuberculosis: applications, challenges and solutions. The International Journal of Tuberculosis and Lung Disease **2015**; 19(6): 626-34.
11. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. www.cochrane-handbook.org: The Cochrane Collaboration, **2011**.
12. R Cran Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, **2015**.
13. Kennedy N, Fox R, Kisyombe GM, et al. Early bactericidal and sterilizing activities of ciprofloxacin in pulmonary tuberculosis. The American review of respiratory disease **1993**; 148(6 Pt 1): 1547-51.
14. Snider DE, Gracyk J, Bek E, Rogowski J. Supervised six-months treatment of newly diagnosed pulmonary tuberculosis using isoniazid, rifampin, and pyrazinamide with and without streptomycin. American Review of Respiratory Disease **1984**; 130(6): 1091-4.
15. Bonnett LJ, Davies GR. Quality of outcome reporting in phase II studies in pulmonary tuberculosis. Trials **2015**; 16(1): 518.
16. Aber VR, Nunn AJ. Factors affecting relapse following short-course chemotherapy. Bulletin of the International Union Against Tuberculosis **1978**; 53: 260-4.

17. Espinal MA, Kim S, Suarez PG, et al. Standard short-course chemotherapy for drug-resistant tuberculosis: Treatment outcomes in 6 countries. JAMA **2000**; 283(19): 2537-45.

18. Burman WJ, Goldberg S, Johnson JL, et al. Moxifloxacin versus ethambutol in the first 2 months of treatment for pulmonary tuberculosis. American journal of respiratory and critical care medicine **2006**; 174(3): 331-8.

19. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials **2007**; 8: 39.

Tables

Table 1: Meta-regression results for the impact of selected variables on the proportion of patients who were culture negative at eight weeks – beta coefficients describe how the outcome variable changes with a unit increase in the explanatory variable

	Covariate – beta regression coefficient (p-values)		
Drugs	Year of Publication	Proportion of patients with HIV co-infection	Geographical location (Africa or not)
HRZE	-0·00 (0·27)	0·00 (0·79)	0·01 (0·87)
HRZ	0·01 (0·24)	Insufficient data*	-0·12 (0·45)
SHRZ	0·00 (0·57)	Insufficient data**	-0·08 (0·17)

* Only one study presented proportion of patients with HIV co-infection

**No studies presented proportion of patients with HIV co-infection

Table 2: Ranking of Drugs Across Outcomes based on a subset of regimens for which at least two of the EBA results were available.

Drug(s)	EBA0-2: Ranking Patients (Regimens)	EBA0-7: Ranking Patients (Regimens)	EBA0-14: Ranking Patients (Regimens)	2 month: Ranking Patients (Regimens)
H	1 149 (16)	1 36 (4)	-	3 533 (6)
HRZE	2 51 (6)	4 21 (2)	2 50 (6)	1 1618 (8)
PaMZ	3 15 (1)	3 12 (1)	1 13 (1)	-
R	4 28 (3)	2 13 (1)	-	2 77 (2)
PaZ	5 15 (1)	5 14 (1)	3 14 (1)	-
Pa	6 29 (2)	-	6 26 (2)	-
JPa	7 14 (1)	6 14 (1)	5 14 (1)	-
JZ	8 15 (1)	7 15 (1)	4 15 (1)	-
J	9 41 (3)	8 26 (2)	7 28 (2)	-

Figure Legends

Figure 1: Literature Review process

Figure 2: Drug Combinations of Included Studies

Figure 3: EBA_{0-2} Day Results (fixed effects – Generalised Inverse Variance Method)

Figure 4: EBA_{0-7} Days (fixed effects – Generalised Inverse Variance Method)

Figure 5: EBA_{0-14} Days (fixed effects – Generalised Inverse Variance Method)

Figure 6: Culture Negativity at eight weeks - Solid Culture (random effects – DerSimonian and Laird Method)

Figure 7: Culture Negativity at eight weeks - Liquid Culture (random effects – DerSimonian and Laird Method)

Figure 8: Culture Negativity at eight weeks – Solid & Liquid Culture (random effects – DerSimonian and Laird Method)