

## Supplementary File 1: Additional methods

### REGIMEN AND ADHERENCE DATA EXTRACTION

A standardised form was used for data extraction from clinical notes. Data collection was completed on 15<sup>th</sup> December 2017. Drug names, start and end dates, dosing, and frequency of administration were collected, as well as any notation in a patient's notes by hospital staff of issues with adherence (including dates and the number of doses missed, where possible) and patient outcomes. Use of directly observed therapy (DOT) was also recorded. In the UK DOT is generally provided to patients deemed to be at especial risk of non-adherence, either at the start of treatment or during treatment, although some hospitals routinely DOT all patients for the first two months. It frequently is used with thrice weekly dosing. Duration calculations omitted gaps in treatment.

### PHENOTYPIC LEVELS OF DRUG RESISTANCE

607/626 (97.0%) of samples were phenotyped within a single reference laboratory, one sample at a second site external to London, and the rest within a second London reference laboratory. Drug sensitivity tests were performed to standard operating procedures across all sites.[1, 2] Resistance ratios were used to determine which strains were resistant to H and which highly resistant. The growth of test strains across three test slopes is compared to wild type strains and so, depending on the controls, the cut-off concentration threshold can vary. High levels of resistance are usually called when there is growth on all three slopes, including at the highest concentration (0.2mg/l H). Resistant, but not highly resistant, strains usually grown on two of the three slopes, up to 0.1mg/l H.

### OTHER EXPOSURE VARIABLES

Site of disease was combined with smear status to generate a single variable with four strata: pulmonary with or without extrapulmonary site(s) smear positive, pulmonary with or without extrapulmonary site(s) smear negative or smear status missing, meningeal or other central nervous system (CNS) sites, other extrapulmonary sites only. The separate meningeal/CNS grouping was due to the difficulty of treating TB in these sites.

The presence of one or more social risk factors (homelessness, problematic drug use, problematic alcohol use, and imprisonment) and whether or not they were a current risk- was coded into a single variable.

'Severe' non-adherence to treatment was classed as treatment gaps of two months or more, or any period of taking less than 80% of prescribed doses.

### MODEL BUILDING

Our knowledge of the literature and previous studies was used to decide on the *a priori* confounders age, sex, phenotype, thrice weekly dosing and adherence. Additional potential confounders were then identified through causal frameworks.[3] The model-building process to generate the final multivariable model has been described before.[4, 5] Briefly, we started with a model containing all *a priori* and potential confounders and undertook a step-by-step backwards deletion strategy that sequentially removed potential confounders that were not determined to fulfil the three rules of confounding, whilst retaining the *a priori* confounders. A *priori* it was decided that model fit using linear and categorical variables for age, year and time before Hr was known would be compared in the model containing the final covariate set. Subsequently, thrice weekly dosing, adherence, Hr phenotype, and Hr genotype were assessed for effect modification within this model. All p-values quoted are from likelihood ratio tests. We undertook a complete case analysis.

Thrice weekly dosing and adherence were collinear, thus only thrice weekly dosing was included in the baseline model.

## REFERENCES

1. Collins CH, Grange JM, Yates MD. Tuberculosis Bacteriology: Organization and Practice. 2nd edition ed: CRC Press; 1997.
2. Parte A, Whitman WB, Goodfellow M, et al. Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria. 2nd edition ed: Springer; 2012.
3. Victora CG, Huttly SR, Fuchs SC, et al. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol.* 1997;26(1):224-7.
4. Greenland S, Pearce N. Statistical foundations for model-based adjustments. *Annu Rev Public Health.* 2015;36:89-108.
5. Stagg HR, Abubakar I, Brown J, et al. Towards better guidance on caseload thresholds to promote positive tuberculosis treatment outcomes: a cohort study. *BMC Med.* 2016;14:52.