BMJ Open

Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR)

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005636
Article Type:	Research
Date Submitted by the Author:	12-May-2014
Complete List of Authors:	Mears, Jessica; University College London, Department of Infection and Population Health Abubakar, Ibrahim; University College London, Department of Infection and Population Health; Public Health England, Centre for Infectious Disease Surveillance and Control Cohen, Ted; Harvard School of Public Health, Harvard University, Division of Global Health Equity, Brigham and Women's Hospital and Department of Epidemiology McHugh, Timothy; Centre for Clinical Microbiology, Research Department of Infection, Royal Free Campus, University College London Sonnenberg, Pamela; University College London, Department of Infection and Population Health
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Tuberculosis < INFECTIOUS DISEASES, MOLECULAR BIOLOGY

Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR)

Jessica Mears [1], Ibrahim Abubakar [1,2], Theodore Cohen [3], Timothy D McHugh [4] & Pam Sonnenberg [1,*]

[1] Department of Infection and Population Health, University College London

[2] Centre for Infectious Disease Surveillance and Control, Public Health England

[3] Division of Global Health Equity, Brigham and Women's Hospital and Department of Epidemiology, Harvard School of Public Health, Harvard University

[4] Centre for Clinical Microbiology, Department of Infection, University College London

*Corresponding author:

Dr Pam Sonnenberg

3rd Floor Mortimer Market Centre

University College London

London WC1E 6JB

p.sonnenberg@ucl.ac.uk

Word count: 1,733

Abstract

Objectives: To systematically review the evidence for the impact of study design and setting on the interpretation of TB transmission using clustering derived from Mycobacterial Interspersed Repetitive Units – Variable Number Tandem Repeats (MIRU-VNTR) strain typing.

Data sources: Medline, Embase, CINHAL, Web of Science and Scopus were searched for articles published before November 2012

Review methods: Studies in humans that reported the proportion of clustering of TB isolates by MIRU-VNTR were included in the analysis. Univariable meta-regression analyses were conducted to assess the influence of study design and setting on the proportion of clustering.

Results: The search identified 14 eligible articles reporting clustering between 22.1% and 61.2%. The proportion of culture positive isolates and the number of MIRU-VNTR loci typed explained 49% and 34% of the between study variation, respectively, and had a significant association with the proportion of clustering.

Conclusions: Although MIRU-VNTR typing is being adopted worldwide there is a paucity of data on how study design and setting may influence estimates of clustering. We have highlighted study design variables for consideration in the design and interpretation of future studies.

Strengths and Limitations of Study

- This is a timely evaluation of the impact of study design on estimates of TB clustering using MIRU-VNTR strain typing because it has been incorporated into national typing services globally.
- There were insufficient data available to fully explore the impact of study design and setting on estimates of clustering.

Introduction

The introduction of molecular typing methods has improved our understanding of *Mycobacterium tuberculosis* (TB) transmission and has changed local and national control policies [1–5]. The proportion of cases that are clustered is often used to estimate the amount of ongoing transmission within the population, based on the assumption that cases with indistinguishable strain types are part of a chain of transmission. TB molecular typing methodology is changing rapidly and it is important that we better understand how to interpret the outputs and thus act.

TB molecular typing methods include Spoligotyping [6], insertion sequence *6110* (IS*6110*) restriction fragment length polymorphism (RFLP) analysis (the recent gold standard) [7], mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing [8], and whole genome sequencing [9–11]. Published reviews have identified factors that might influence or bias clustering by IS*6110* RFLP [12,13]. No study has repeated this analysis using more up-to-date typing methods, which is important for understanding of the epidemiology of TB and to shape the application of molecular typing to improve TB control.

Published meta-analyses and modelling studies using IS*6110* RFLP data show that the proportion of clustering observed can be affected by 1) study design (affecting the proportion of eligible cases that are included in the study); 2) features of the typing method (such as the ability to type isolates with low copy numbers); and 3) study setting (such as characteristics of the study population). For example, the proportion of clustering increases when the fraction of the total data sampled increases [13–15] and when study duration increases [16].

MIRU-VNTR is currently the preferred method of molecular typing [17–21], and can be used together with Spoligotyping [8]. Relative to IS*6110* RFLP, MIRU-VNTR does not have to exclude isolates with a low IS*6110* copy number, has a faster turnaround time, is high throughput and the numeric strain types are more easily compared. MIRU-VNTR strain typing is increasingly being adopted worldwide [1,22–27], yet unlike IS*6110* RFLP, the evidence for the interpretation of the findings such as the impact of study design and setting on clustering have not been reviewed. Although the two typing methods have been shown to have a similar discriminatory value, the markers evolve independently and at different rates, resulting in a difference in clustering between the two methods [28]. This suggests that there could be differences in the way study design, typing method and setting affects clustering by the two methods. We conducted a systematic review to assess the evidence for the impact of study design and setting on the interpretation of TB

transmission using clustering derived from MIRU-VNTR strain typing – as has been shown using IS*6110* RFLP typing.

Methods

Five electronic databases were searched (EMBASE, ISI Web of Science, CINHAL, Scopus and Medline (Ovid)) up to 1 November 2012. The search strategy combined the following terms with Boolean operators: Tuberculosis, strain typing, and transmission. The search was limited to studies using the standard MIRU-VNTR method [8], in humans only, and in English.

All titles and abstracts from each of the searches were examined. The full text of each paper was obtained and reviewed if the study reported MIRU-VNTR strain typing of *M.tuberculosis* complex isolates with at least 15 of the standardised 24 loci [8,29,30].

Studies using fewer than 15 loci were not included because the level of discrimination is inadequate for epidemiological use (n=97) [8]. Studies that used loci different to the standardised 15 and 24 set were not included in the analysis in order to reduce the heterogeneity between studies (n=11). All publication types were included in this first screen to ensure that no relevant data were missed.

Reviews, letters, editorials, outbreaks or case reports (n=99) were excluded in the second screen. Studies that used incomplete sampling (e.g. random samples, studies using subsets of populations such as MDR patients) (n=30) and studies that had a sample size of less than 50 (n=2) were also excluded.

A reviewer extracted the following data items from all included studies using a form developed in Excel (Microsoft 2010): publication details (year, authors, study country), study details (study duration, loci typed, secondary typing method, study population), the proportion of total TB isolates clustered by MIRU-VNTR strain typing, and the covariates of interest: the number of clustered and unique isolates; the maximum size of clusters; the proportion of clusters containing two cases; the prevalence of culture-positivity among TB patients included in the study; the proportion of culture positive isolates typed; risk factors for clustering; and the Hunter Gaston Discriminatory Index (HGDI) [31]).

Authors were contacted if TB incidence rate was not reported. Where no response was received WHO country estimates of TB incidence for the study year were used [32].

Data were analysed in Stata 12. Where studies reported data from more than one set of loci, the method with the highest discriminatory value was included (i.e. MIRU-VNTR 24 would be chosen

over MIRU-VNTR 15, and MIRU-VNTR 15 plus Spoligotyping would be chosen over MIRU-VNTR 15 alone) (n=5). This review was not concerned with summary measures of clustering, but factors that influenced clustering; therefore articles must have included at least one of the covariates. Continuous variables were transformed where the distribution was skewed. The proportion clustered was transformed using the Freeman Tukey transformation [33]. Univariable meta-regression analyses were carried out to determine the effect of the study design covariates on the proportion of clustered isolates. All covariates in the analysis were hypothesised to influence the proportion clustered *a priori*.

Results

The search identified 5607 references resulting in 12 journal articles and 2 conference abstracts included after deduplication and title/abstract/full text screening (Figure 1). The main characteristics of the included studies are shown in Table 1. Studies were published between 2007 and 2011 and the clustering reported varied from 22.1% [34] to 61.2% [35].

The univariable meta-regression shows evidence for the proportion of clustering to decrease as the prevalence of culture-positivity among TB patients included in the study increases (p=0.03; Table 2), accounting for 49% of the between study variation. There was also evidence for the proportion of clustering to decrease as the number of MIRU-VNTR loci typed increased from 15 to 24 (p=0.02), explaining 34% of the between study variation. There was no evidence of the other study design or study setting variables significantly influencing the proportion clustered. Though non-significant (p>0.05), the size of the study and the maximum cluster size explained 15% and 27% of the between study variation, respectively.

Discussion

This review identified 14 studies that met the inclusion criteria. We illustrate that the interpretation of studies using MIRU-VNTR to estimate clustering is subject to bias relating to study design; however, there were insufficient data available to fully explore the impact of study design and setting on estimates of clustering.

As expected, we found that the proportion of clustering decreased with a greater number of MIRU-VNTR loci typed. Our finding that the prevalence of culture-positivity among TB patients included in the study influences the estimates of transmission within a population is counterintuitive and not consistent with estimates of the influence of sampling on the proportion of clustering using *IS*6110 RFLP typing [36]. This may reflect the relationship between TB burden and resource poor/rich

settings and the consequent availability of culture diagnostic laboratory services; i.e. in resource poor settings where there is a high burden of TB (and, therefore, high rates of clustering) the prevalence of culture positive TB cases is low. The finding may also be due to chance, with only 8 studies included in the analysis of this variable.

The other study design variables included in this analysis, such as study duration, did not significantly influence the proportion of isolates that were clustered, contrary to previous findings [12]. This is likely to be because of a lack of good quality evidence: only 14 studies met the inclusion criteria for the review and of those only three reported all the variables of interest, reducing the power of the analysis and precluding multivariable meta-regression. In addition, the range of the variables may have been too limited to show any impact on clustering estimates. For example, the proportion of culture positive isolates typed had a narrow range from 81.9% to 100%. Furthermore, most of the studies were from low TB burden settings and therefore may be reflecting the rate at which imported cases have matching strain types by chance, rather than rates of recent transmission.

This study is a timely evaluation of the impact of study design on estimates of TB clustering using MIRU-VNTR strain typing because it has been incorporated into national typing services globally [23,37]. The findings are relevant where strain typing is used to evaluate TB control systems across different settings because the proportion of clustering is influenced by the prevalence of culture positive TB cases in the study setting. Given that strain typing methods are advancing beyond MIRU-VNTR typing and that the application of whole genome sequencing to TB control and public health strategies has been demonstrated [9–11,38], it is important that the biases in the analysis of such methods are explored and compared. Understanding how to design and compare research studies for public health will greatly improve the benefit gained from newer technologies.

This review has highlighted the need for better quality reporting in primary studies to enable future reviews to be more robust. A lack of standards for the molecular epidemiology of infectious diseases may explain the poor quality of reporting; this field would benefit from the introduction of such standards (STROBE-ID, submitted).

The use of TB strain typing as a public health tool in TB control programmes is increasing globally. We have identified a lack of good quality studies that can contribute to our understanding in interpreting the molecular typing of TB. We have also shown that the proportion of clustering derived from MIRU-VTNR typing is influenced by the number of loci typed and the prevalence of culture-positivity among TB patients included in the study, highlighting these as important considerations in the design and interpretation of future studies.

Conflict of interest

Nothing to declare.

Acknowledgements

We would like to acknowledge Ross Harris from the Statistics Unit at Public Health England for his advice on meta-regression.

Author contributions

All authors made substantial contributions to the conception and design of the review, and the analysis and interpretation of data. JM drafted the article and PS, IA, TM and TC revised it critically for important intellectual content. All authors approved the final version for publication.

Funding

JM is funded through a Public Health England and University College London Impact Studentship. IA is funded through a NIHR Senior Research Fellowship.

Ethics

Ethical approval was not required as this review analyses data that is in the public domain.

Data sharing

No additional data are available

References

1. Lambregts-van Weezenbeek CSB, Sebek MMGG, Van Gerven PJHJ, De Vries G, Verver S, et al. (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. Int J Tuberc Lung Dis 7: S463–470.

2. Borgdorff MW, Van den Hof S, Kremer K, Verhagen L, Kalisvaart N, et al. (2010) Progress towards tuberculosis elimination: secular trend, immigration and transmission. Eur Respir J 36: 339–347. doi:10.1183/09031936.00155409.

3. Kik SV, Verver S, Van Soolingen D, De Haas PEW, Cobelens FG, et al. (2008) Tuberculosis Outbreaks Predicted by Characteristics of First Patients in a DNA Fingerprint Cluster. Am J Respir Crit Care Med 178: 96–104. doi:10.1164/rccm.200708-1256OC.

4. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology

laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 31: 1677–1682.

5. De Vries G, Van Hest RAH, Richardus JH (2007) Impact of mobile radiographic screening on tuberculosis among drug users and homeless persons. Am J Respir Crit Care Med 176: 201–207. doi:10.1164/rccm.200612-1877OC.

6. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35: 907–914.

7. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31: 406–409.

8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 44: 4498–4510. doi:10.1128/JCM.01392-06.

9. Schürch AC, Van Soolingen D (2011) DNA fingerprinting of Mycobacterium tuberculosis: From phage typing to whole-genome sequencing. Infect Genet Evol. Available: http://www.ncbi.nlm.nih.gov/pubmed/22067515. Accessed 13 March 2012.

10. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364: 730–739. doi:10.1056/NEJMoa1003176.

11. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. (2013) Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13: 137–146. doi:10.1016/S1473-3099(12)70277-3.

12. Houben RMGJ, Glynn JR (2009) A systematic review and meta-analysis of molecular epidemiological studies of tuberculosis: development of a new tool to aid interpretation. Tropical Medicine & International Health 14: 892–909. doi:10.1111/j.1365-3156.2009.02316.x.

13. Fok A, Numata Y, Schulzer M, FitzGerald MJ (May) Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies [Review Article]. The International Journal of Tuberculosis and Lung Disease 12: 480–492.

14. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, Van Soolingen D (2011) Influence of Sampling on Clustering and Associations With Risk Factors in the Molecular Epidemiology of Tuberculosis. Am J Epidemiol. Available: http://aje.oxfordjournals.org/content/early/2011/05/23/aje.kwr061. Accessed 29 March 2012.

15. Glynn JR, Bauer J, De Boer AS, Borgdorff MW, Fine PE, et al. (1999) Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis 3: 1055–1060.

16. Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD, et al. (2005) The Importance of Recent Infection with Mycobacterium tuberculosis in an Area with High HIV Prevalence: A

Long-Term Molecular Epidemiological Study in Northern Malawi. J Infect Dis 192: 480–487. doi:10.1086/431517.

17. De Beer JL, Kremer K, Ködmön C, Supply P, Van Soolingen D (2012) First Worldwide Proficiency Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex Strains. Journal of Clinical Microbiology 50: 662–669. doi:10.1128/JCM.00607-11.

18. Maes M, Kremer K, Van Soolingen D, Takiff H, De Waard JH (2008) 24-locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis (Edinb) 88: 490–494. doi:10.1016/j.tube.2008.04.003.

19. Sougakoff W (2011) Molecular epidemiology of multidrug-resistant strains of Mycobacterium tuberculosis. Clinical Microbiology and Infection 17: 800–805. doi:10.1111/j.1469-0691.2011.03577.x.

20. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res 38: W326–331. doi:10.1093/nar/gkq351.

21. Supply P (2010) MIRU-VNTR typing: the new international standard for TB molecular epidemiology Symposium of the Institut Pasteur de Tunisia.

22. Van Soolingen D, Borgdorff MW, De Haas PE, Sebek MM, Veen J, et al. (1999) Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J Infect Dis 180: 726–736. doi:10.1086/314930.

23. Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step approach for large-scale, prospective genotyping of Mycobacterium tuberculosis isolates in the United States. J Clin Microbiol 43: 688–695. doi:10.1128/JCM.43.2.688-695.2005.

24. New CDC Program for Rapid Genotyping of Mycobacterium Tuberculosis Isolates (2005). JAMA 293: 2086–2086. doi:10.1001/jama.293.17.2086.

25. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, et al. (2000) A prospective evaluation of the clinical value of nation-wide DNA fingerprinting of tuberculosis isolates in Denmark. Int J Tuberc Lung Dis 4: 295–299.

26. Bauer J, Yang Z, Poulsen S, Andersen AB (1998) Results from 5 years of nationwide DNA fingerprinting of Mycobacterium tuberculosis complex isolates in a country with a low incidence of M. tuberculosis infection. J Clin Microbiol 36: 305–308.

27. Zolnir-Dovc M, Poljak M, Erzen D, Sorli J (2003) Molecular epidemiology of tuberculosis in Slovenia: results of a one-year (2001) nation-wide study. Scand J Infect Dis 35: 863–868.

28. Hanekom M, Van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Hoek KGP, et al. (2008) Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of Mycobacterium tuberculosis Beijing strains in a setting of high incidence of tuberculosis. J Clin Microbiol 46: 3338–3345. doi:10.1128/JCM.00770-08.

29. Supply P, Lesjean S, Savine E, Kremer K, Van Soolingen D, et al. (2001) Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based

on mycobacterial interspersed repetitive units. J Clin Microbiol 39: 3563–3571. doi:10.1128/JCM.39.10.3563-3571.2001.

30. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniewski FA (2006) Progression toward an improved DNA amplification-based typing technique in the study of Mycobacterium tuberculosis epidemiology. J Clin Microbiol 44: 2492–2498. doi:10.1128/JCM.01428-05.

31. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465–2466.

32. WHO | TB data (n.d.). WHO. Available: http://www.who.int/tb/country/en/index.html. Accessed 12 December 2012.

33. Freeman MF, Tukey JW (1950) Transformations Related to the Angular and the Square Root. Ann Math Statist 21: 607–611. doi:10.1214/aoms/1177729756.

34. Oelemann MC, Diel R, Vatin V, Haas W, Rüsch-Gerdes S, et al. (2007) Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol 45: 691–697. doi:10.1128/JCM.01393-06.

35. Evans J (2010) Analysis of prevalent Mycobacterium tuberculosis strains in the United Kingdom: detection, distribution and expansion of MIRU-VNTR profiles containing high numbers of isolates. European Society of Clinical Microbiology and Infectious Diseases. Vienna, Austria.

36. Glynn JR, Vyonycky E, Fine PEM (1999) Influence of Sampling on Estimates of Clustering and Recent Transmission of Mycobacterium tuberculosis Derived from DNA Fingerprinting Techniques. American Journal of Epidemiology 149: 366 –371.

37. TB Strain Typing Project Board HPA (2011) TB Strain Typing Cluster Investigation Handbook for Health Protection Units. Available: https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealthProtectionProgrammes/Respiratory/TB/StrainTyping/. Accessed 30 November 2011.

38. Walker TM, Monk P, Grace Smith E, Peto TEA (2013) Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect. doi:10.1111/1469-0691.12183.

39. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR (2011) Tuberculosis transmission in Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. Infect Genet Evol 11: 124–131. doi:10.1016/j.meegid.2010.09.013.

40. Allix-Béguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008) Standardised PCR-based molecular epidemiology of tuberculosis. Eur Respir J 31: 1077–1084. doi:10.1183/09031936.00053307.

41. Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 46: 1398–1406. doi:10.1128/JCM.02089-07.

42. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, et al. (2011) Evaluation of Mycobacterium tuberculosis typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. J Clin Microbiol 49: 4173–4178. doi:10.1128/JCM.05293-11.

43. Ojo OO, Sheehan S, Corcoran DG, Nikolayevsky V, Brown T, et al. (2010) Molecular epidemiology of Mycobacterium tuberculosis clinical isolates in Southwest Ireland. Infect Genet Evol 10: 1110–1116. doi:10.1016/j.meegid.2010.07.008.

44. Dymova MA, Liashenko OO, Poteiko PI, Krutko VS, Khrapov EA, et al. (2011) Genetic variation of Mycobacterium tuberculosis circulating in Kharkiv Oblast, Ukraine. BMC Infect Dis 11: 77. doi:10.1186/1471-2334-11-77.

45. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med 105 Suppl 1: S67–73. doi:10.1016/S0954-6111(11)70014-2.

46. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, et al. (2009) Prospective universal application of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for fast identification of clustered and orphan cases. J Clin Microbiol 47: 2026–2032. doi:10.1128/JCM.02308-08.

47. Hamblion EL, Wynne-Edwards E, Anderson C, Anderson SR (2011) A summary of strain typing and clustering of TB in London in 2010 and an analysis of the associated risk factors. Thorax 66: A88–A89. doi:10.1136/thoraxjnl-2011-201054c.50.

48. Mandal S, Bradshaw L, Anderson LF, Brown T, Evans JT, et al. (2011) Investigating Transmission of Mycobacterium Bovis in the UK, 2005-2008. J Clin Microbiol. Available: http://jcm.asm.org/content/early/2011/03/23/JCM.02299-10. Accessed 24 April 2012.

49. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, et al. (2011) Molecular epidemiology of Mycobacterium tuberculosis in East Lancashire 2001-2009. Thorax 66: 709–713. doi:10.1136/thx.2011.158881.

50. Nikolayevskyy VV, Brown TJ, Bazhora YI, Asmolov AA, Balabanova YM, et al. (2007) Molecular epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in Mycobacterium tuberculosis strains from the southern Ukraine. Clin Microbiol Infect 13: 129–138. doi:10.1111/j.1469-0691.2006.01583.x.

Tables

Table 1: Studies included in the analysis

Reference	Author	Country	Studysite [a]	Method [b]	Loci [c]	Study duration (months)	Clustered + unique isolates	TB incidence (per 100,000)	TB/HIV co-infection [d]	Prevalence of culture positivity	% culture positive typed	No. clusters	Max cluster size	HGDI	Proportion clustering	Recent transmission (%)
[39]	Asgharzadeh, M	Azerbaijan	r	15	o	12	156	26.0		94.6	98.7	22	5	0.9966	32.7	18.6
[40]	Allix-Beguec, C	Belgium	r	24	n	24	530	35.2	5.1	86.1	87.9	53	23		29.6	19.6
[41]	Allix-Beguec, C	Belgium	r	24,S	n	39	802	35.2	5.1	81.8	84.7	82			28.8	19.6
[34]	Oelemann, M	Germany	ci	24,S	n	12	154	12.7			100	11			22.1	14.9
[42]	Roetzer, A	Germany	r	24,S	n	48	277	3.2	0.09		100	18	22		27.1	20.6
[43]	Ojo, OO	Ireland	r	24,S	n	36	171	15.3	3.3	79.5	96.1	15	12	0.9996	27.5	18.7
[44]	Dymova, MA	Russia	r	15	o	3	98	94.0	3.8		100	8		0.9900	31.6	23.5
[45]	Bidovec-Stojkovic, U	Slovenia	co	24,S	n	12	196	10.6	0.04		100	29	6	0.9965	36.2	21.4
[46]	Alonso-Rodriguez, N	Spain	r	15	n	27	281	26.0	6		81.9		8		43.1	24.4
[35]	Evans, J	UK	r	15	o	48	4207	15.0	8.2	58.3	100	439			61.2	50.8
[47]	Hamblion, E	UK	r	24	n	9	964	44.9	8.2		100				37.0	
[48]	Mandal, S	UK	co	15	o	48	102		8.2	90.7	87.2	8	12		30.4	22.6
[49]	Sails, A	UK	r	15	o	102	332	18.3	8.2	33.9	100	42	13		42.8	30.1
[50]	Nikolayevsky, V	Ukraine	r	15	o	4	225	80.4	3.9	39.2	97.4	31		0.9700	60.4	46.7

[a] ci=city, r=region, co=country

[b] 15=15 MIRU-VNTR loci, 24=24 MIRU-VNTR loci, S=with Spoligotyping

[c] o= old 12 MIRU loci (MIRU 2, 4, 10, 16, 20, 23, 24, 26, 27,30, 31, 39, 40), n=new 12 MIRU loci (MIRU 10, 16, 26, 31, 40 + Mtub 04, 21, 39 + ETR A C + QUB 11b, 26)

[d] estimates from the literature of the prevalence of TB/HIV co-infection reported in the study area

Table 2: Univariable metaregression showing the coefficients for change in the proportion of clustering and the percentage of between-study variation explained by variables describing the study design and setting.

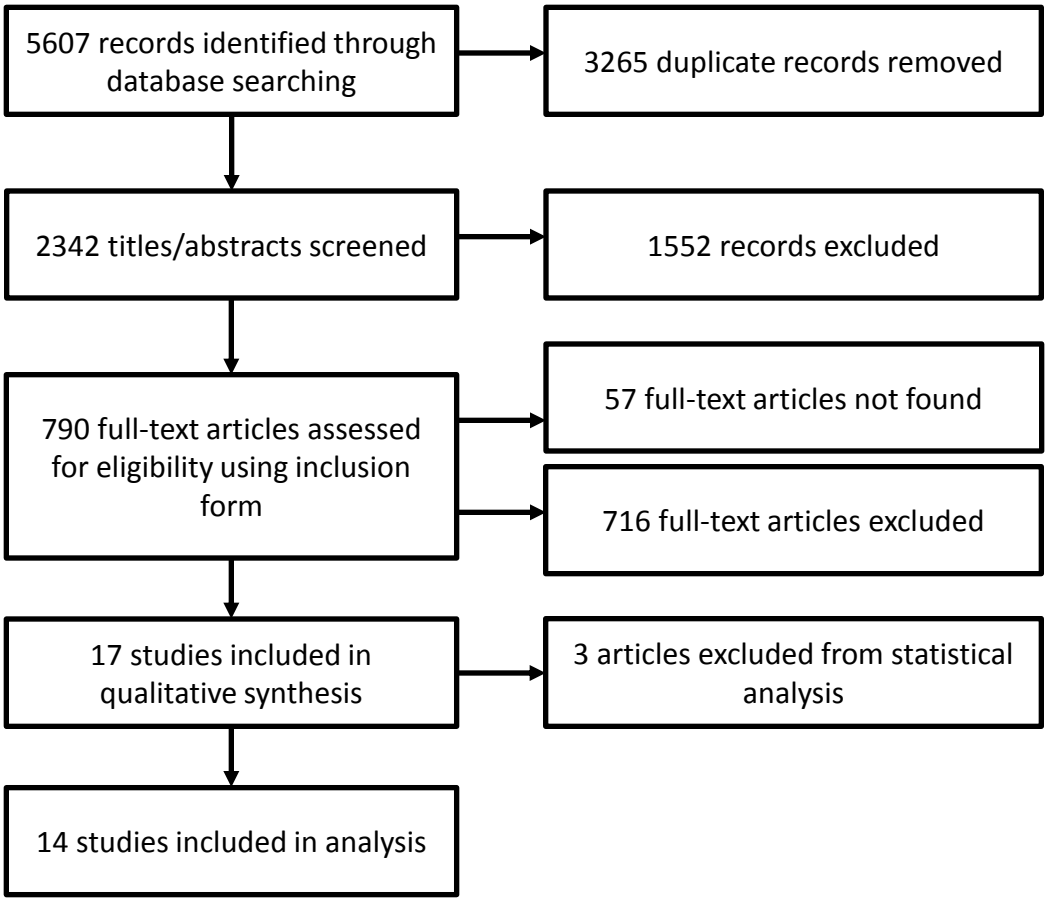
	n	Coefficient [π]	CI	p	Adj R^{2} [¥]
Study design					
Study duration (months)	14	0.003	-0.063, 0.069	0.919	-8.47
Prevalence of culture positivity	8	-0.913	-1.732, -0.094	0.034	49.36
% culture positive typed	14	0.161	-0.731, 1.053	0.701	-6.99
Study size	14	-4.462	-10.000, 1.076	0.105	14.89
Number of loci (ref 15 loci)					
24 loci	14	-0.282	-0.519, -0.045	0.023	34.1
Study setting					
TB incidence	13	0.082	-0.097, 0.22	0.334	0.04
TB/HIV co-infection	12	0.088	-0.087, 0.263	0.288	3.28
Maximum cluster size	9	0.137	-0.035, 0.309	0.101	26.91
% clusters with 2 cases	7	0.004	-0.007, 0.016	0.396	-2.39

[π] Coefficients for the change in the proportion of clustering for each covariate. E.g. for a one-month increase in study duration, the proportion of clustering increases by 0.003.

[¥] The proportion of between-study variation explained by the univariate metaregression.

Figure Caption

Figure 1: Results of systematic search, screening and data extraction.


5607 records identified through database searching
3265 duplicate records removed
2342 titles/abstracts screened
1552 records excluded
790 full-text articles assessed for eligibility using inclusion form
57 full-text articles not found
716 full-text articles excluded
17 studies included in qualitative synthesis
3 articles excluded from statistical analysis
14 studies included in analysis

Appendix: Medline/Embase search strategy

1. (tubercle adj3 (bacillus or bacilli)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

2. ((mycobacterium or mycobacteria) adj3 (bovis or africanum or microti or canetti)).mp.

3. exp tuberculosis/ or mycobacterium tuberculosis/ or tuberculosis.mp. or tb.mp. or Mtb.mp. or "M tuberculosis complex".mp.

4. or/1-3

5. Minisatellite Repeats/ or Genotype/ or Interspersed Repetitive Sequences/ or DNA Fingerprinting/ or Bacterial Typing Techniques/

6. "miru".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

7. "vntr".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

8. (miru adj3 vntr).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

9. (mycobacterial adj3 interspersed adj3 repetitive adj3 units).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

10. (dna adj3 fingerprinting).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

11. ((strain adj3 type) or (strain adj3 typing) or (strain adj3 types)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

12. ((molecular adj3 typing) or (molecular adj3 strain adj3 typ*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

13. (genotype or genotyping or genotypes).ti,ab.

14. (minisatellite adj3 repeat*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

15. molecular epidemiology/mt or (molecular adj3 epidemiology).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

16. or/5-15

17. exp disease outbreaks/ or (outbreak adj3 analysis).mp. or (outbreak adj3 investigation).mp. or (outbreak adj3 management).mp. or (tuberculosis adj3 outbreak).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

18. exp contact tracing/ or (contact adj3 tracing).mp. or (contact* adj3 traced).mp. or (contact adj3 screen*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

19. exp case management/ or (case adj3 management).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

20. exp Risk Factors/

21. (risk adj3 factor*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

22. exp Epidemiologic Factors/

23. infectious disease transmission.mp. or exp Disease Transmission, Infectious/

24. exp case management/ or (case adj3 management).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

25. program evaluation/ or evaluation studies as topic/ or (program adj3 evaluation).mp. or (programme adj3 evaluation).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

26. public health practice/ or (public adj3 health).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

27. ((tuberculosis adj3 control) or (tb adj3 control)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

28. (molecular adj3 surveillance).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

29. exp cluster analysis/ or (cluster* adj3 rate*).mp. or (cluster* adj3 growth).mp. or (cluster* adj3 analysis).mp. or (cluster adj3 investigation).mp. or (proportion adj3 cluster*).mp. or (molecular adj3 cluster*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

30. ((recent adj3 transmission) or (transmission adj3 event*) or (transmission adj3 rate*) or (chain adj3 transmission) or (transmission adj3 setting*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

31. or/17-30

32. 4 and 16

33. 32 and 31

34. limit 33 to yr="1998-Current"

35. limit 34 to english language

36. animals/

37. humans/

38. 36 not 37

39. 35 not 38

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	14
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	4
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	5
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	5
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	12
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	12
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	14
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	12
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	14
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	5
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	6
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	7

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

BMJ Open

Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): A systematic review

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005636.R1
Article Type:	Research
Date Submitted by the Author:	25-Nov-2014
Complete List of Authors:	Mears, Jessica; University College London, Department of Infection and Population Health Abubakar, Ibrahim; University College London, Department of Infection and Population Health; Public Health England, Centre for Infectious Disease Surveillance and Control Cohen, Ted; Harvard School of Public Health, Harvard University, Division of Global Health Equity, Brigham and Women's Hospital and Department of Epidemiology McHugh, Timothy; Centre for Clinical Microbiology, Research Department of Infection, Royal Free Campus, University College London Sonnenberg, Pamela; University College London, Department of Infection and Population Health
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Tuberculosis < INFECTIOUS DISEASES, MOLECULAR BIOLOGY

Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): A systematic review

Jessica Mears[1], Ibrahim Abubakar[1,2,3], Theodore Cohen[4], Timothy D McHugh[5] & Pam Sonnenberg[1,*]

[1] Department of Infection and Population Health, University College London

[2] Centre for Infectious Disease Surveillance and Control, Public Health England

[3] Clinical Trials Unit, Medical Research Council, London

[4]Division of Global Health Equity, Brigham and Women's Hospital and Department of Epidemiology, Harvard School of Public Health, Harvard University

[5] Centre for Clinical Microbiology, Department of Infection, University College London

*Corresponding author:

Dr Pam Sonnenberg

3rd Floor Mortimer Market Centre

University College London

London WC1E 6JB

p.sonnenberg@ucl.ac.uk

Word count: 2439

Abstract

Objectives: To systematically review the evidence for the impact of study design and setting on the interpretation of TB transmission using clustering derived from Mycobacterial Interspersed Repetitive Units – Variable Number Tandem Repeats (MIRU-VNTR) strain typing.

Data sources: Medline, Embase, CINHAL, Web of Science and Scopus were searched for articles published before 21st October 2014.

Review methods: Studies in humans that reported the proportion of clustering of TB isolates by MIRU-VNTR were included in the analysis. Univariable meta-regression analyses were conducted to assess the influence of study design and setting on the proportion of clustering.

Results: The search identified 27 eligible articles reporting clustering between 0% and 63%. The number of MIRU-VNTR loci typed, requiring consent to type patient isolates (as a proxy for sampling fraction), the TB incidence and the maximum cluster size explained 14%, 14%, 27% and 48%, respectively, and had a significant association with the proportion of clustering .

Conclusions: Although MIRU-VNTR typing is being adopted worldwide there is a paucity of data on how study design and setting may influence estimates of clustering. We have highlighted study design variables for consideration in the design and interpretation of future studies.

Strengths and Limitations of Study

- This is a timely evaluation of the impact of study design on estimates of TB clustering using MIRU-VNTR strain typing because it has been incorporated into national typing services globally.
- The strength of this meta-analysis was limited by the lack of detail reported by the included studies, highlighting the need for better quality reporting in primary studies.

Introduction

The introduction of molecular typing methods has improved our understanding of *Mycobacterium tuberculosis* (TB) transmission and has changed local and national control policies [1–5]. The proportion of cases that are clustered is often used to estimate the amount of ongoing transmission within the population, based on the assumption that cases with indistinguishable strain types are part of a chain of transmission. TB molecular typing methodology is changing rapidly and it is important that we better understand how to interpret the outputs and thus act.

TB molecular typing methods include Spoligotyping [6], insertion sequence *6110* (IS*6110*) restriction fragment length polymorphism (RFLP) analysis (the recent gold standard) [7], mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing [8], and whole genome sequencing [9–11]. Published reviews have identified factors that might influence or bias clustering by IS*6110* RFLP [12,13]. No study has repeated this analysis using more up-to-date typing methods, which is important for understanding of the epidemiology of TB and to shape the application of molecular typing to improve TB control.

Published meta-analyses and modelling studies using IS*6110* RFLP data show that the proportion of clustering observed can be affected by 1) study design (affecting the proportion of eligible cases that are included in the study); 2) features of the typing method (such as the ability to type isolates with low copy numbers); and 3) study setting (such as characteristics of the study population). For example, the proportion of clustering increases when the fraction of the total data sampled increases [13–15] and when study duration increases [16].

MIRU-VNTR is currently the preferred method of molecular typing [17–21], and can be used together with Spoligotyping [8]. Relative to IS*6110* RFLP, MIRU-VNTR does not have to exclude isolates with a low IS*6110* copy number, has a faster turnaround time, is high throughput and the numeric strain types are more easily compared. MIRU-VNTR strain typing is increasingly being adopted worldwide [1,22–27], yet unlike IS*6110* RFLP, the evidence for the interpretation of the findings such as the impact of study design and setting on clustering have not been reviewed. Although the two typing methods have been shown to have a similar discriminatory value, the markers evolve independently and at different rates, resulting in a difference in clustering between the two methods [28]. This suggests that there could be differences in the way study design, typing method and setting affects clustering by the two methods. We conducted a systematic review to assess the evidence for the impact of study design and setting on the interpretation of TB

transmission using clustering derived from MIRU-VNTR strain typing – as has been shown using IS*6110* RFLP typing.

Methods

Five electronic databases were searched (EMBASE, ISI Web of Science, CINHAL, Scopus and Medline (Ovid)) up to 20th October 2014. The search strategy combined the following terms with Boolean operators: Tuberculosis, strain typing, and transmission. The search was limited to studies using the standard MIRU-VNTR method [8], in humans only, and in English.

All titles and abstracts from each of the searches were examined. The full text of each paper was obtained and reviewed if the study reported MIRU-VNTR strain typing of *M.tuberculosis* complex isolates with at least 15 of the standardised 24 loci (ETR A, B, C, D, E; MIRU 2, 10, 16, 20, 23, 24, 26, 27, 39, 40; VNTR 424, 1955, 2163b, 2347, 2401, 3171, 3690, 4052, 4156) [8,29,30].

Studies using fewer than 15 loci were not included because the level of discrimination is inadequate for epidemiological use (n=121) [8]. Studies that used loci different to the standardised 15 and 24 set were not included in the analysis in order to reduce the heterogeneity between studies (n=19). All publication types were included in this first screen to ensure that no relevant data were missed.

Reviews, letters, editorials, outbreaks or case reports (n=103) were excluded in the second screen. Studies that used incomplete sampling (e.g. random samples, studies using subsets of populations such as multidrug-resistant patients) (n=47) and studies that had a sample size of less than 50 (n=4) were also excluded.

A reviewer (JM) extracted the following data items from all included studies using a form developed in Excel (Microsoft 2010): publication details (year, authors, study country), study details (study duration, loci typed, secondary typing method, study population, whether participant consent was required (a characteristic of the study design that was used as proxy for sampling fraction, assuming that where consent was required the sampling fraction was low)), the number of clustered and unique isolates, and the covariates of interest: the maximum size of clusters; the proportion of clusters containing two cases; the proportion of the population that was culture positive; the proportion of culture positive isolates typed; risk factors for clustering; and the Hunter Gaston Discriminatory Index (HGDI) [31]). IA extracted data from 10% of the papers for external validity, disagreements were discussed and a consensus agreed upon.

The main outcome measure – the proportion of TB isolates clustered by MIRU-VNTR strain typing – was calculated as the number of clustered isolates/number of clustered+unique isolates. Where there were uncertainties JM consulted with IA

Authors were contacted if TB incidence rate was not reported. Where no response was received WHO country estimates of TB incidence for the study year were used [32]. As so few studies reported the proportion coinfected with TB/HIV, these estimates for the study country were taken from an EU-wide survey and WHO country profiles.[33,34] Due to poor recording of the sampling fraction (the number of isolates typed/ the total number of culture positive TB cases diagnosed during the study period (n=19)), whether the study required the consent of participants (yes/no) was included as a proxy for (high/low) sampling fraction. The risk of bias within each study was assessed using the STROME-ID checklist. [35]

Data were analysed in Stata 12. Where studies reported data from more than one set of loci, the method with the highest discriminatory value was included (i.e. MIRU-VNTR 24 would be chosen over MIRU-VNTR 15, and MIRU-VNTR 15 plus Spoligotyping would be chosen over MIRU-VNTR 15 alone) (n=8). This review was not concerned with summary measures of clustering, but factors that influenced clustering; therefore articles must have included at least one of the covariates. Continuous variables were transformed where the distribution was skewed. The proportion clustered was transformed using the Freeman Tukey transformation [36]. Study heterogeneity was assessed using a forest plot and the chi^2 test of heterogeneity. Univariable meta-regression analyses were carried out to determine the effect of the study design covariates on the proportion of clustered isolates. All covariates in the analysis were hypothesised to influence the proportion clustered *a priori*.

Sensitivity analyses were conducted to see the effect of removing studies reporting 0% clustering, with only extra-pulmonary TB cases, only *M.bovis* cases, studies using the 'old 12' MIRU loci as part of their 15 loci, and studies assessed as having a high likelihood of bias (STROME-ID score less than 20).

Results

The search identified 7274 references resulting in 27 studies (25 journal articles and 2 conference abstracts) included after deduplication and title/abstract/full text screening (Figure 1). The main characteristics of the included studies are shown in Table 1. Studies were published between 2007 and 2014 and the clustering reported varied from 0% [37] to 62.8% [38]. In all studies, clustered

isolates were defined as having identical strain types based on the MIRU-VNTR loci typed, with or without Spoligotyping. 17 studies included isolates from newly diagnosed TB cases, three studies reported including isolates from new and chronic cases of TB, and seven did not report this information. In addition, ten studies did not include repeat isolates from the same patient, one study included a repeat isolate from one patient, and the remaining 17 did not report whether repeat isolates were included or not. Furthermore, four studies included isolates with missing loci in the cluster analysis, whereas four excluded isolates with missing loci, and the remaining 20 did not report how they dealt with missing loci. The number of studies reporting each variable of interest is shown in Table 2.

A forest plot shows the spread of clustering reported by number of loci and additional typing method (Figure 2). Significant heterogeneity was identified between the studies ($p<0.001$), suggesting that a meta-regression would be an appropriate analysis.

The univariable meta-regression shows evidence for the proportion of clustering to decrease as the number of MIRU-VNTR loci typed increased from 15 to 24 ($p=0.04$; Table 3), accounting for 14% of the between study variation, and to increase when the study participants consented to being included in the study ($p=0.03$), accounting for 14% of the between study variation. The proportion of clustering increased as the TB incidence in the population increased ($p=0.007$, Adj $R^2 = 26.7$). There was also evidence for the proportion of clustering to increase as the maximum cluster size increased ($p=0.001$), accounting for 48% of between study variation. There was no evidence of the other study design or study setting variables significantly influencing the proportion clustered. Though non-significant ($p>0.05$), the TB/HIV coinfection rate in the population explained 2% of the between study variation. Too few studies included information on the proportion of clusters containing two cases, proportion of the study sample with previous TB or with pulmonary TB, and the proportion of the population with culture positive TB, so these could not be included in the analysis (Table 2).

Sensitivity analyses to examine the effect of excluding studies reporting 0% clustering,[37] only *M.bovis* cases,[39] studies using the 'old 12' MIRU loci,[39–44] and studies assessed as having a high risk of bias,[37,45–48] did not generally change the results. The proportion of culture positive TB in the population remained insignificant but explained 2.6% of the between study variation when excluding 0% clustering ($p=0.278$ and Adj $R^2=2.62$). Similarly, the proportion of culture positive TB in the population remained insignificant but explained 2.6% of the between study variation when excluding studies with the highest risk of bias ($p=0.278$ and Adj $R^2=2.62$). The number of loci typed became non-significant, but explained 9.6% and 10.5% of the between study variation when

excluding studies using the 'old 12' loci and the highest risk of bias, respectively (p=0.106, Adj R^2=9.63; p=0.111, Adj R^2=10.51, respectively).

Discussion

This review identified 27 studies that met the inclusion criteria. We illustrate that the interpretation of studies using MIRU-VNTR to estimate clustering is subject to bias relating to study design and setting; however, there were insufficient data available to fully explore this impact.

As expected, we found that the proportion of clustering decreased with a greater number of MIRU-VNTR loci typed, with increasing TB incidence and with increasing maximum cluster size. We found that requiring consent to type patient isolates reduced the proportion of clustering, which is expected, given that the sampling fraction would be lower in these studies.

The other study design variables included in this analysis, such as study duration, did not significantly influence the proportion of isolates that were clustered, contrary to previous findings [12]. This is likely to be because of a lack of good quality evidence: of the 27 studies that met the inclusion criteria for the review, none reported all the variables of interest, reducing the power of the analysis and precluding multivariable meta-regression (Table 2). Importantly, key details of cluster analyses were not reported consistently across the studies, such as whether repeat isolates from the same patients were included, or typing profiles with missing loci were included, introducing new, unmeasured biases. In addition, the range of the variables may have been too limited to show any impact on clustering estimates. For example, the proportion of culture positive isolates typed ranged from 34.5% to 100%, with 17 of the 19 studies reporting this variable from 81.9% to 100%. Furthermore, most of the studies (17/27=63%) were from low TB burden settings and therefore may be reflecting the rate at which imported cases have matching strain types by chance, rather than rates of recent transmission.

The sensitivity analysis suggested that, when excluding the studies with the greatest risk of bias, the culture-positivity in the population might explain a small amount of the between study variation. This is counterintuitive and not consistent with estimates of the influence of sampling on the proportion of clustering using *IS*6110 RFLP typing [49]. This may reflect the relationship between TB burden and resource poor/rich settings and the consequent availability of culture diagnostic laboratory services; i.e. in resource poor settings where there is a high burden of TB (and, therefore, high rates of clustering) the prevalence of culture positive TB cases is low. The finding may also be due to chance, with only 14 studies included in the analysis of this variable. In the sensitivity analysis

excluding studies that used the 'old 12' loci, the effect of the number of loci typed becomes non-significant. This is likely because studies using the 'old 12' accounted for six out of ten studies reporting 15 loci, reducing the number of studies and the power of the model.

This study is a timely evaluation of the impact of study design on estimates of TB clustering using MIRU-VNTR strain typing because it has been incorporated into national typing services globally [23,50]. The findings are relevant where strain typing is used to evaluate TB control systems across different settings because the proportion of clustering is influenced by the number of loci typed, the TB incidence and the maximum cluster size. Given that strain typing methods are advancing beyond MIRU-VNTR typing and that the application of whole genome sequencing to TB control and public health strategies has been demonstrated [9–11,51], it is important that the biases in the analysis of such methods are explored and compared. Understanding how to design and compare research studies for public health will greatly improve the benefit gained from newer technologies.

The strength of this meta-analysis was limited by (a lack of) detail reported by the included studies. This review has highlighted the need for better quality reporting in primary studies to enable future reviews to be more robust. Recently published standards for reporting of molecular epidemiology for infectious diseases should improve the quality of reporting.[35] This review is further limited by our inability to access 58 of the title/abstract screened articles for full text screening.

The use of TB strain typing as a public health tool in TB control programmes is increasing globally. We have identified a lack of good quality studies that can contribute to our understanding in interpreting the molecular typing of TB. We have also shown that the proportion of clustering derived from MIRU-VTNR typing is influenced by the number of loci typed, whether consent is required to type isolates, TB incidence in the study setting, and the maximum cluster size, highlighting these as important considerations in the design and interpretation of future studies.

Conflict of interest

Nothing to declare.

Acknowledgements

We would like to acknowledge Ross Harris from the Statistics Unit at Public Health England for his advice on meta-regression.

Author contributions

All authors made substantial contributions to the conception and design of the review, and the analysis and interpretation of data. JM drafted the article and PS, IA, TM and TC revised it critically for important intellectual content. All authors approved the final version for publication.

Funding

JM is funded through a Public Health England and University College London Impact Studentship. IA is funded through a NIHR Senior Research Fellowship.

Ethics

Ethical approval was not required as this review analyses data that is in the public domain.

Data sharing

No additional data are available

References

1. Lambregts-van Weezenbeek CSB, Sebek MMGG, van Gerven PJHJ, de Vries G, Verver S, et al. (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. Int J Tuberc Lung Dis 7: S463–470.

2. Borgdorff MW, van den Hof S, Kremer K, Verhagen L, Kalisvaart N, et al. (2010) Progress towards tuberculosis elimination: secular trend, immigration and transmission. Eur Respir J 36: 339–347. doi:10.1183/09031936.00155409.

3. Kik SV, Verver S, Van Soolingen D, De Haas PEW, Cobelens FG, et al. (2008) Tuberculosis Outbreaks Predicted by Characteristics of First Patients in a DNA Fingerprint Cluster. Am J Respir Crit Care Med 178: 96–104. doi:10.1164/rccm.200708-1256OC.

4. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 31: 1677–1682.

5. De Vries G, van Hest RAH, Richardus JH (2007) Impact of mobile radiographic screening on tuberculosis among drug users and homeless persons. Am J Respir Crit Care Med 176: 201–207. doi:10.1164/rccm.200612-1877OC.

6. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35: 907–914.

7. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31: 406–409.

8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 44: 4498–4510. doi:10.1128/JCM.01392-06.

9. Schürch AC, van Soolingen D (2011) DNA fingerprinting of Mycobacterium tuberculosis: From phage typing to whole-genome sequencing. Infect Genet Evol. Available: http://www.ncbi.nlm.nih.gov/pubmed/22067515. Accessed 13 March 2012.

10. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364: 730–739. doi:10.1056/NEJMoa1003176.

11. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. (2013) Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13: 137–146. doi:10.1016/S1473-3099(12)70277-3.

12. Houben RMGJ, Glynn JR (2009) A systematic review and meta-analysis of molecular epidemiological studies of tuberculosis: development of a new tool to aid interpretation. Tropical Medicine & International Health 14: 892–909. doi:10.1111/j.1365-3156.2009.02316.x.

13. Fok A, Numata Y, Schulzer M, FitzGerald MJ (May) Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies [Review Article]. The International Journal of Tuberculosis and Lung Disease 12: 480–492.

14. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, Van Soolingen D (2011) Influence of Sampling on Clustering and Associations With Risk Factors in the Molecular Epidemiology of Tuberculosis. Am J Epidemiol. Available: http://aje.oxfordjournals.org/content/early/2011/05/23/aje.kwr061. Accessed 29 March 2012.

15. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, et al. (1999) Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis 3: 1055–1060.

16. Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD, et al. (2005) The Importance of Recent Infection with Mycobacterium tuberculosis in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study in Northern Malawi. J Infect Dis 192: 480–487. doi:10.1086/431517.

17. De Beer JL, Kremer K, Ködmön C, Supply P, van Soolingen D (2012) First Worldwide Proficiency Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex Strains. Journal of Clinical Microbiology 50: 662–669. doi:10.1128/JCM.00607-11.

18. Maes M, Kremer K, van Soolingen D, Takiff H, de Waard JH (2008) 24-locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis (Edinb) 88: 490–494. doi:10.1016/j.tube.2008.04.003.

19. Sougakoff W (2011) Molecular epidemiology of multidrug-resistant strains of Mycobacterium tuberculosis. Clinical Microbiology and Infection 17: 800–805. doi:10.1111/j.1469-0691.2011.03577.x.

20. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res 38: W326–331. doi:10.1093/nar/gkq351.

21. Supply P (2010) MIRU-VNTR typing: the new international standard for TB molecular epidemiology Symposium of the Institut Pasteur de Tunisia.

22. Van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, et al. (1999) Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J Infect Dis 180: 726–736. doi:10.1086/314930.

23. Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step approach for large-scale, prospective genotyping of Mycobacterium tuberculosis isolates in the United States. J Clin Microbiol 43: 688–695. doi:10.1128/JCM.43.2.688-695.2005.

24. Centers for Disease Control and Prevention (2005) New CDC Program for Rapid Genotyping of Mycobacterium Tuberculosis Isolates. JAMA 293: 2086–2086. doi:10.1001/jama.293.17.2086.

25. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, et al. (2000) A prospective evaluation of the clinical value of nation-wide DNA fingerprinting of tuberculosis isolates in Denmark. Int J Tuberc Lung Dis 4: 295–299.

26. Bauer J, Yang Z, Poulsen S, Andersen AB (1998) Results from 5 years of nationwide DNA fingerprinting of Mycobacterium tuberculosis complex isolates in a country with a low incidence of M. tuberculosis infection. J Clin Microbiol 36: 305–308.

27. Zolnir-Dovc M, Poljak M, Erzen D, Sorli J (2003) Molecular epidemiology of tuberculosis in Slovenia: results of a one-year (2001) nation-wide study. Scand J Infect Dis 35: 863–868.

28. Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Hoek KGP, et al. (2008) Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of Mycobacterium tuberculosis Beijing strains in a setting of high incidence of tuberculosis. J Clin Microbiol 46: 3338–3345. doi:10.1128/JCM.00770-08.

29. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, et al. (2001) Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. J Clin Microbiol 39: 3563–3571. doi:10.1128/JCM.39.10.3563-3571.2001.

30. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniewski FA (2006) Progression toward an improved DNA amplification-based typing technique in the study of Mycobacterium tuberculosis epidemiology. J Clin Microbiol 44: 2492–2498. doi:10.1128/JCM.01428-05.

31. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465–2466.

32. WHO | TB data (n.d.). WHO. Available: http://www.who.int/tb/country/en/index.html. Accessed 12 December 2012.

33. Kruijshaar ME, Pimpin L, Abubakar I, Rice B, Delpech V, et al. (2011) The burden of TB-HIV in the EU: how much do we know? A survey of surveillance practices and results. Eur Respir J 38: 1374–1381. doi:10.1183/09031936.00198310.

34. World Health Organization (n.d.) WHO Tuberculosis Country Profiles. Available: http://www.who.int/tb/country/data/profiles/en/.

35. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, et al. (2014) Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. The Lancet Infectious Diseases 14: 341–352. doi:10.1016/S1473-3099(13)70324-4.

36. Freeman MF, Tukey JW (1950) Transformations Related to the Angular and the Square Root. Ann Math Statist 21: 607–611. doi:10.1214/aoms/1177729756.

37. Guang-ming D, Zhi-guo Z, Peng-ju D, Qian Z, Li W, et al. (2013) Differences in the population of genetics of Mycobacterium tuberculosis between urban migrants and local residents in Beijing, China. Chinese Medical Journal 126: 4066–4071. doi:10.3760/cma.j.issn.0366-6999.20130216.

38. Zmak L, Obrovac M, Katalinic Jankovic V (2014) First insights into the molecular epidemiology of tuberculosis in Croatia during a three-year period, 2009 to 2011. Scandinavian Journal of Infectious Diseases 46: 123–129.

39. Mandal S, Bradshaw L, Anderson LF, Brown T, Evans JT, et al. (2011) Investigating transmission of Mycobacterium bovis in the United Kingdom in 2005 to 2008. J Clin Microbiol 49: 1943–1950. doi:10.1128/JCM.02299-10.

40. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR (2011) Tuberculosis transmission in Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. Infect Genet Evol 11: 124–131. doi:10.1016/j.meegid.2010.09.013.

41. Dymova MA, Liashenko OO, Poteiko PI, Krutko VS, Khrapov EA, et al. (2011) Genetic variation of Mycobacterium tuberculosis circulating in Kharkiv Oblast, Ukraine. BMC Infect Dis 11: 77. doi:10.1186/1471-2334-11-77.

42. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, et al. (2011) Molecular epidemiology of Mycobacterium tuberculosis in East Lancashire 2001-2009. Thorax 66: 709–713. doi:10.1136/thx.2011.158881.

43. Evans J (2010) Analysis of prevalent Mycobacterium tuberculosis strains in the United Kingdom: detection, distribution and expansion of MIRU-VNTR profiles containing high numbers of isolates. European Society of Clinical Microbiology and Infectious Diseases. Vienna, Austria.

44. Nikolayevskyy VV, Brown TJ, Bazhora YI, Asmolov AA, Balabanova YM, et al. (2007) Molecular epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in Mycobacterium tuberculosis strains from the southern Ukraine. Clin Microbiol Infect 13: 129–138. doi:10.1111/j.1469-0691.2006.01583.x.

45. L. Z, M. O, V. KJ (2014) First insights into the molecular epidemiology of tuberculosis in Croatia during a three-year period, 2009 to 2011. Scandinavian Journal of Infectious Diseases.

46. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, et al. (2011) Evaluation of Mycobacterium tuberculosis typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. J Clin Microbiol 49: 4173–4178. doi:10.1128/JCM.05293-11.

47. Dymova MA, Kinsht VN, Cherednichenko AG, Khrapov EA, Svistelnik AV, et al. (2011) Highest prevalence of the Mycobacterium tuberculosis Beijing genotype isolates in patients newly diagnosed with tuberculosis in the Novosibirsk oblast, Russian Federation. J Med Microbiol 60: 1003–1009. doi:10.1099/jmm.0.027995-0.

48. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, et al. (2009) Prospective universal application of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for fast identification of clustered and orphan cases. J Clin Microbiol 47: 2026–2032. doi:10.1128/JCM.02308-08.

49. Glynn JR, Vyonycky E, Fine PEM (1999) Influence of Sampling on Estimates of Clustering and Recent Transmission of Mycobacterium tuberculosis Derived from DNA Fingerprinting Techniques. American Journal of Epidemiology 149: 366 –371.

50. TB Strain Typing Project Board HPA (2011) TB Strain Typing Cluster Investigation Handbook for Health Protection Units 1st Edition. Available: https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealthProtectionProgrammes/Respiratory/TB/StrainTyping/. Accessed 30 November 2011.

51. Walker TM, Monk P, Grace Smith E, Peto TEA (2013) Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect. doi:10.1111/1469-0691.12183.

52. Gurjav U, Jelfs P, McCallum N, Marais BJ, Sintchenko V (2014) Temporal dynamics of Mycobacterium tuberculosis genotypes in New South Wales, Australia. BMC infectious diseases 14: 455–455. doi:10.1186/1471-2334-14-455.

53. Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 46: 1398–1406. doi:10.1128/JCM.02089-07.

54. Allix-Béguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008) Standardised PCR-based molecular epidemiology of tuberculosis. Eur Respir J 31: 1077–1084. doi:10.1183/09031936.00053307.

55. Tuite AR, Guthrie JL, Alexander DC, Whelan MS, Lee B, et al. (2013) Epidemiological evaluation of spatiotemporal and genotypic clustering of mycobacterium tuberculosis in Ontario, Canada. International Journal of Tuberculosis and Lung Disease 17: 1322–1327.

56. Tessema B, Beer J, Merker M, Emmrich F, Sack U, et al. (2013) Molecular epidemiology and transmission dynamics of Mycobacterium tuberculosis in Northwest Ethiopia: new phylogenetic lineages found in Northwest Ethiopia. Bmc Infectious Diseases 13: 131. doi:10.1186/1471-2334-13-131.

57. Smit PW, Haanpera M, Rantala P, Couvin D, Lyytikainen O, et al. (2013) Molecular Epidemiology of Tuberculosis in Finland, 2008-2011. Plos One 8: e85027. doi:10.1371/journal.pone.0085027.

58. Oelemann MC, Diel R, Vatin V, Haas W, Rüsch-Gerdes S, et al. (2007) Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol 45: 691–697. doi:10.1128/JCM.01393-06.

59. Ojo OO, Sheehan S, Corcoran DG, Nikolayevsky V, Brown T, et al. (2010) Molecular epidemiology of Mycobacterium tuberculosis clinical isolates in Southwest Ireland. Infect Genet Evol 10: 1110–1116. doi:10.1016/j.meegid.2010.07.008.

60. Aleksic E, Merker M, Cox H, Reiher B, Sekawi Z, et al. (2013) First Molecular Epidemiology Study of Mycobacterium tuberculosis in Kiribati. PLoS ONE 8. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8.

61. De Beer JL, van Ingen J, de Vries G, Erkens C, Sebek M, et al. (2013) Comparative study of IS6110 restriction fragment length polymorphism and variable-number tandem-repeat typing of Mycobacterium tuberculosis isolates in the Netherlands, based on a 5-year nationwide survey. J Clin Microbiol 51: 1193–1198. doi:10.1128/JCM.03061-12.

62. Varghese B, Supply P, Shoukri M, Allix-Beguec C, Memish Z, et al. (2013) Tuberculosis Transmission among Immigrants and Autochthonous Populations of the Eastern Province of Saudi Arabia. PLoS ONE 8. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2.

63. Lim LK-Y, Sng LH, Win W, Chee CB-E, Hsu LY, et al. (2013) Molecular Epidemiology of Mycobacterium tuberculosis Complex in Singapore, 2006-2012. Plos One 8: e84487. doi:10.1371/journal.pone.0084487.

64. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med 105 Suppl 1: S67–73. doi:10.1016/S0954-6111(11)70014-2.

65. Jonsson J., Hoffner S., Berggren I., Bruchfeld J., Ghebremichael S., et al. (2014) Comparison between RFLP and MIRU-VNTR genotyping of mycobacterium tuberculosis strains isolated in stockholm 2009 to 2011. PLoS ONE. Available: http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0095159&representation=PDF.

66. Muwonge A, Malama S, Johansen TB, Kankya C, Biffa D, et al. (2013) Molecular Epidemiology, Drug Susceptibility and Economic Aspects of Tuberculosis in Mubende District, Uganda. PLoS ONE 8. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697.

67. Hamblion EL, Wynne-Edwards E, Anderson C, Anderson SR (2011) A summary of strain typing and clustering of TB in London in 2010 and an analysis of the associated risk factors. Thorax 66: A88–A89. doi:10.1136/thoraxjnl-2011-201054c.50.

68. Hang NTL, Maeda S, Lien LT, Thuong PH, Hung NV, et al. (2013) Primary Drug-Resistant Tuberculosis in Hanoi, Viet Nam: Present Status and Risk Factors. Plos One 8: UNSP e71867. doi:10.1371/journal.pone.0071867.

Tables

Table 1: The study setting and design characteristics of the included articles

Ref	Study setting							Study design							Risk of bias [d]	Clustering (%)[e]
	Study area and country	TB incidence (per 100,000)	TB/HIV (per 100,000)[a]	Previous TB treatment (%)	Pulmonary TB (%)	Maximum cluster size	Clusters of size 2 (%)	Study duration (months)	Study size (clustered + unique isolates)	Culture positive in study population (%)	Culture positive isolates typed (%)	Typing method [b]	Loci typed [c]	Consent required		
[52]	New South Wales, Australia	6.7	0.2	0.0	63.7	.	.	36	1128	.	.	m24	N	no	low	20.1
[40]	Tabriz and Orumieh, Azarbaijan	26.0	.	5.2	87.0	5	81.8	12	156	.	94.5	m15	O	no	low	32.7
[53]	Brussels-Capital Region, Belgium	35.2	5.1	10.8	.	23	64.2	24	530	86.1	87.9	m24	N	no	low	29.6
[54]	Brussels-Capital Region, Belgium	35.2	5.1	.	100	.	.	39	802	81.8	84.7	m24s	N	no	low	28.8
[55]	Ontario, Canada	4.8	0.4	.	.	18	58.8	65	2016	.	.	m24s	N	no	low	23.1
[37]	Changping District, Beijing, China	.	0.3	.	100	0	.	30	318	31.5	94.6	m24	N	no	high	0.0
[38]	Croatia	19.0	0.1	.	.	45	48.3	36	1587	.	.	m15	N	no	high	62.8
[56]	Amhara region, Northwest Ethiopia	.	24.0	17.6	100	13	.	5	244	.	.	m24	N	yes	low	45.1
[57]	Finland	5.0	0.0	.	.	20	.	48	1048	75.4	99.4	m15s	.	no	low	33.9
[58]	Hamburg, Germany	12.7	.	.	.	.	45.5	12	154	78.2	91.1	m24s	N	no	low	22.1
[46]	Schleswig-Holstein, Germany	3.2	0.1	.	.	22	44.4	48	277	.	.	m24s	N	no	high	27.1
[59]	South West Ireland	15.3	3.3	.	82.7	12	.	36	171	79.5	96.1	m24s	N	no	low	27.5
[60]	South Tawara, Kiribati	370.0	.	4.1	100	25	55.6	24	73	45.4	98.6	m24s	N	yes	low	75.3
[61]	Netherlands	6.5	0.2	.	.	.	57.2	60	3978	.	100.1	m24	N	no	low	46.7
[41]	Kharkiv, Russia	94.0	3.8	63.3	100	10	50.0	3	98	.	100	m15	O	yes	high	31.6
[62]	Eastern province, Saudi Arabia	4.0	.	.	73.1	24	19.0	24	522	.	.	m24s	N	no	low	40.2

[63]	Singapore	40.5	1.2	.	.	21	48.0	24	1128	82.0	34.5	m24s	N	no	low	30.8
[64]	Slovenia	10.6	0.0	.	.	6	.	12	196	94.4	97.5	m24s	N	no	low	36.2
[48]	Almeria, Spain	26.0	6.0	.	.	8	.	27	281	.	81.9	m15	N	no	high	43.1
[65]	Sweden	4.8	0.1	.	.	10	.	36	406	.	.	m24s	N	no	low	21.2
[66]	Mubende, Uganda	.	86.0	31.1	87.8	11	70.0	6	67	21.5	90.5	m15s	N	yes	low	35.8
[42]	East Lancashire, UK	18.3	8.2	.	.	13	58.3	102	332	48.5	69.9	m15	O	no	low	42.8
[39]	UK	.	8.2	.	42.3	12	50.0	48	102	90.7	87.2	m15	O	no	low	30.4
[67]	London, UK	44.9	8.2	.	.	.	.	9	964	36.0	100	m24	N	no	.	37.0
[43]	Midlands, UK	15.0	8.2	.	.	.	.	48	4207	58.3	100	m15	O	no	.	61.2
[44]	Odessa and Nikolaev, Ukraine	80.4	3.9	34.2	100	.	.	4	225	.	.	m15	O	yes[f]	low	60.4
[68]	Hanoi, Vietnam	146.0	10.0	0.0	100	.	.	20	465	92.7	91.9	m15s	N	yes	low	55.3

[a] Estimates from of the prevalence of TB/HIV co-infection in the study country [33,34]

[b] 15=15 MIRU-VNTR loci (made up of the 'old 12' or 'new 12' defined in the footnote below), 24=24 MIRU-VNTR loci (ETR A, B, C, D, E; MIRU 2, 10, 16, 20, 23, 24, 26, 27, 39, 40; VNTR 424, 1955, 2163b, 2347, 2401, 3171, 3690, 4052, 4156), S=with Spoligotyping

[c] O= old 12 MIRU loci (MIRU 2, 4, 10, 16, 20, 23, 24, 26, 27,30, 31, 39, 40), N=new 12 MIRU loci (MIRU 10, 16, 26, 31, 40 + Mtub 04, 21, 39 + ETR A C + QUB 11b, 26)

[d] Risk of bias was assessed using the STROME-ID checklist. Studies scoring <20 were categorised as have a high risk of bias

[e] The proportion of clustering was calculated as the number of clustered isolates/number of clustered + unique isolates

[f] 11.3% did not consent to being part of the study. The other studies that required consent for isolates to be typed did not report the refusal rate

Table 2: The number of studies that reported the variables of interest

	Reported	Missing
Study setting		
TB incidence	8	15
TB/HIV co-infection	5	22
Previous TB treatment	9	18
Proportion pulmonary TB	14	13
Maximum cluster size	19	8
% clusters with 2 cases	14	13
Study design		
Study duration	27	0
Study size	27	0
% population that is culture positive	15	12
% culture positive typed	19	8
24 loci (compared to 15)	27	0
Repeat isolates	12	15
Missing loci	8	19
Double alleles	1	26
Consent required	6[a]	21
Epidemiological information	6	21

[a] Only one study reported the consent rate

Table 3: Univariable metaregression showing the coefficients for change in the proportion of clustering and the percentage of between-study variation explained by variables describing the study design and setting.

	n	Coefficient [a]	CI	p	Adj R^{2} [b]
Study setting					
TB incidence	23	0.14	0.04-0.24	0.007	26.74
TB/HIV co-infection	23	0.04	-0.03-0.11	0.246	2.00
Maximum cluster size	19	0.20	0.09-0.30	0.001	48.20
Study design					
Study duration	27	-0.02	-0.09-0.06	0.677	-3.37
% population that is culture positive	15	0.34	-1.23-1.96	0.661	-5.92
% culture positive typed	19	0.22	-1.08-1.52	0.725	-5.41
Study size	27	0.03	-0.11-0.16	0.702	-3.31
24 loci (compared to 15)	27	-0.30	-0.59--0.01	0.04	13.58
Consent required	27	0.38	0.04-0.72	0.029	14.41

[a] Coefficients for the change in the proportion of clustering for each covariate. E.g. for a one-month increase in study duration, the proportion of clustering increases by 0.003.

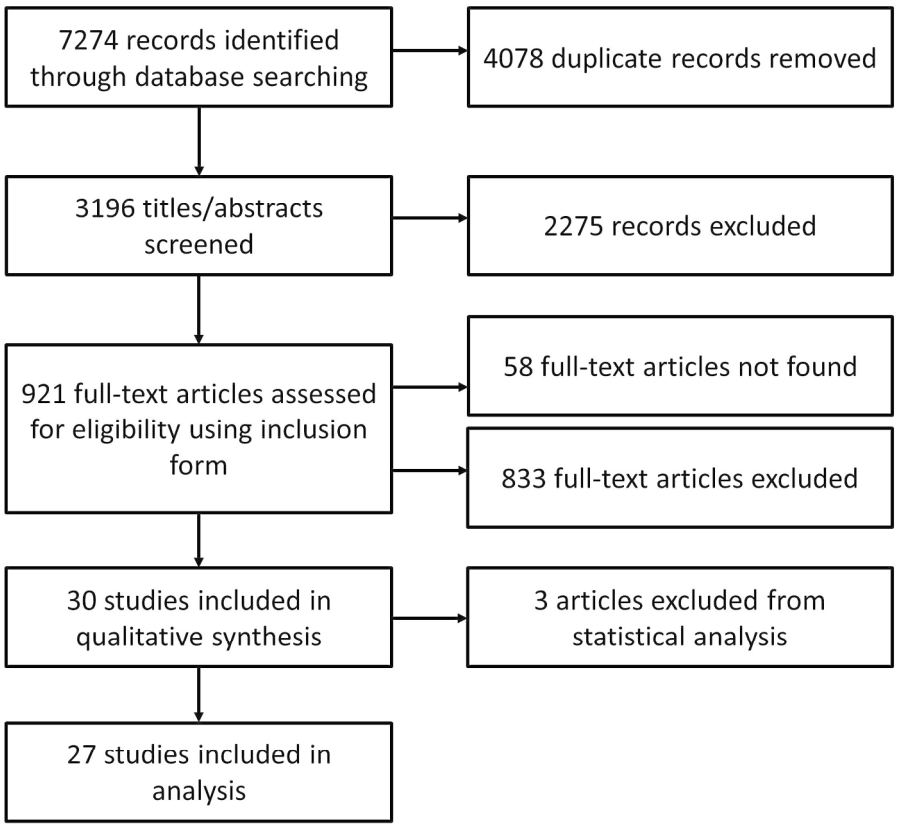
[b] The proportion of between-study variation explained by the univariate meta-regression.

Figure Caption

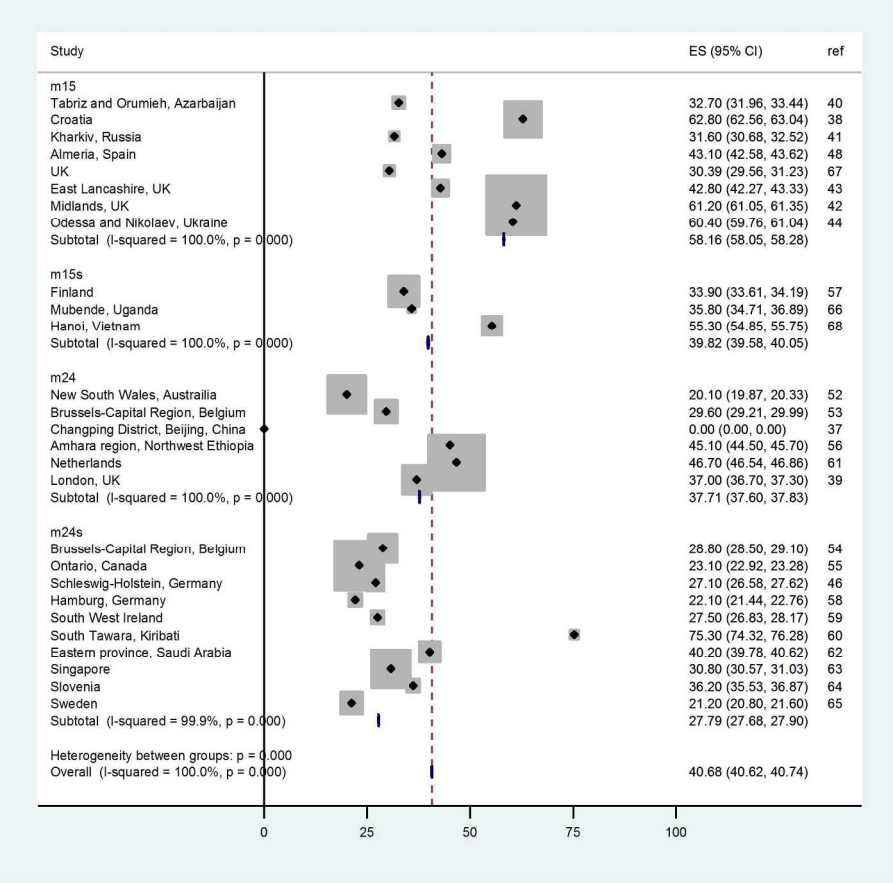
Figure 1: Results of systematic search, screening and data extraction.

Figure 2: Forest plot showing the proportion of clustering reported in each study by the number of MIRU-VNTR loci typed

The number of loci typed is categorised into 15 loci (m15), 15 loci with Spoligotyping (m15s), 24 loci (m24) and 24 loci with Spoligotyping (m24s). The study reference is shown in the right hand column.



190x254mm (300 x 300 DPI)



190x254mm (300 x 300 DPI)

Appendix 1: Medline/Embase search strategy

1. (tubercle adj3 (bacillus or bacilli)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

2. ((mycobacterium or mycobacteria) adj3 (bovis or africanum or microti or canetti)).mp.

3. exp tuberculosis/ or mycobacterium tuberculosis/ or tuberculosis.mp. or tb.mp. or Mtb.mp. or "M tuberculosis complex".mp.

4. or/1-3

5. Minisatellite Repeats/ or Genotype/ or Interspersed Repetitive Sequences/ or DNA Fingerprinting/ or Bacterial Typing Techniques/

6. "miru".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

7. "vntr".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

8. (miru adj3 vntr).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

9. (mycobacterial adj3 interspersed adj3 repetitive adj3 units).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

10. (dna adj3 fingerprinting).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

11. ((strain adj3 type) or (strain adj3 typing) or (strain adj3 types)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

12. ((molecular adj3 typing) or (molecular adj3 strain adj3 typ*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

13. (genotype or genotyping or genotypes).ti,ab.

14. (minisatellite adj3 repeat*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

15. molecular epidemiology/mt or (molecular adj3 epidemiology).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

16. or/5-15

17. exp disease outbreaks/ or (outbreak adj3 analysis).mp. or (outbreak adj3 investigation).mp. or (outbreak adj3 management).mp. or (tuberculosis adj3 outbreak).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

18. exp contact tracing/ or (contact adj3 tracing).mp. or (contact* adj3 traced).mp. or (contact adj3 screen*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

19. exp case management/ or (case adj3 management).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

20. exp Risk Factors/

21. (risk adj3 factor*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

22. exp Epidemiologic Factors/

23. infectious disease transmission.mp. or exp Disease Transmission, Infectious/

24. exp case management/ or (case adj3 management).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

25. program evaluation/ or evaluation studies as topic/ or (program adj3 evaluation).mp. or (programme adj3 evaluation).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

26. public health practice/ or (public adj3 health).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

27. ((tuberculosis adj3 control) or (tb adj3 control)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

28. (molecular adj3 surveillance).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

29. exp cluster analysis/ or (cluster* adj3 rate*).mp. or (cluster* adj3 growth).mp. or (cluster* adj3 analysis).mp. or (cluster adj3 investigation).mp. or (proportion adj3 cluster*).mp. or (molecular adj3 cluster*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

30. ((recent adj3 transmission) or (transmission adj3 event*) or (transmission adj3 rate*) or (chain adj3 transmission) or (transmission adj3 setting*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

31. or/17-30

32. 4 and 16

33. 32 and 31

34. limit 33 to yr="1998-Current"

35. limit 34 to english language

36. animals/

37. humans/

38. 36 not 37

39. 35 not 38

Appendix 2: STROME-ID scores for the included studies

Author	STROME-ID score[a]
Aleksic, E	24
Alliex-Beguec, C	32
Allix-Beguec, C	25
Alonso-Rodriguez, N	18
Asgharzadeh, M	28
Bidovec-Stojkovic, U	31
De Beer, JL	30
Dymova, MA	19
Evans, J	[b]
Grujav, U	32
Guang-ming, DAI	19
Hamblion, E	[b]
Hang, NTHL	31
Jonsson, J	22
Lim, LKY	30
Mandal, S	32
Muwonge, A	25
Nikolayevsky, V	23
Oelemann, M	34
Ojo, OO	36
Roetzer, A	16
Sails, A	23
Smit, PW	29
Tessema, B	26
Tuite, AR	31
Varghese, B	23
Zmak, L	19

[a]Individual studies score 1 for each element of checklist they had address

[b]Conference abstracts

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	5
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	5
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	15
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	15
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	18
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	15
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	18
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	7
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	7-8
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	8
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	8

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

BMJ Open

Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): A systematic review

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005636.R2
Article Type:	Research
Date Submitted by the Author:	23-Dec-2014
Complete List of Authors:	Mears, Jessica; University College London, Department of Infection and Population Health Abubakar, Ibrahim; University College London, Department of Infection and Population Health; Public Health England, Centre for Infectious Disease Surveillance and Control Cohen, Ted; Harvard School of Public Health, Harvard University, Division of Global Health Equity, Brigham and Women's Hospital and Department of Epidemiology McHugh, Timothy; Centre for Clinical Microbiology, Research Department of Infection, Royal Free Campus, University College London Sonnenberg, Pamela; University College London, Department of Infection and Population Health
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Infectious diseases, Public health
Keywords:	EPIDEMIOLOGY, Tuberculosis < INFECTIOUS DISEASES, MOLECULAR BIOLOGY

Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): A systematic review

Jessica Mears[1], Ibrahim Abubakar[1,2,3], Theodore Cohen[4], Timothy D McHugh[5] & Pam Sonnenberg[1,*]

[1] Department of Infection and Population Health, University College London

[2] Centre for Infectious Disease Surveillance and Control, Public Health England

[3] Clinical Trials Unit, Medical Research Council, London

[4]Division of Global Health Equity, Brigham and Women's Hospital and Department of Epidemiology, Harvard School of Public Health, Harvard University

[5] Centre for Clinical Microbiology, Department of Infection, University College London

*Corresponding author:

Dr Pam Sonnenberg

3rd Floor Mortimer Market Centre

University College London

London WC1E 6JB

p.sonnenberg@ucl.ac.uk

Word count: 2392

Abstract

Objectives: To systematically review the evidence for the impact of study design and setting on the interpretation of TB transmission using clustering derived from Mycobacterial Interspersed Repetitive Units – Variable Number Tandem Repeats (MIRU-VNTR) strain typing.

Data sources: Medline, Embase, CINHAL, Web of Science and Scopus were searched for articles published before 21st October 2014.

Review methods: Studies in humans that reported the proportion of clustering of TB isolates by MIRU-VNTR were included in the analysis. Univariable meta-regression analyses were conducted to assess the influence of study design and setting on the proportion of clustering.

Results: The search identified 27 eligible articles reporting clustering between 0% and 63%. The number of MIRU-VNTR loci typed, requiring consent to type patient isolates (as a proxy for sampling fraction), the TB incidence and the maximum cluster size explained 14%, 14%, 27% and 48% of between-study variation, respectively, and had a significant association with the proportion of clustering.

Conclusions: Although MIRU-VNTR typing is being adopted worldwide there is a paucity of data on how study design and setting may influence estimates of clustering. We have highlighted study design variables for consideration in the design and interpretation of future studies.

Strengths and Limitations of Study

- This is a timely evaluation of the impact of study design on estimates of TB clustering using MIRU-VNTR strain typing because it has been incorporated into national typing services globally.
- The strength of this meta-analysis was limited by the lack of detail reported by the included studies, highlighting the need for better quality reporting in primary studies.

Introduction

The introduction of molecular typing methods has improved our understanding of *Mycobacterium tuberculosis* (TB) transmission and has changed local and national control policies [1–5]. The proportion of cases that are clustered is often used to estimate the amount of ongoing transmission within the population, based on the assumption that cases with indistinguishable strain types are part of a chain of transmission. TB molecular typing methodology is changing rapidly and it is important that we better understand how to interpret the outputs and thus act.

TB molecular typing methods include Spoligotyping [6], insertion sequence *6110* (IS*6110*) restriction fragment length polymorphism (RFLP) analysis (the recent gold standard) [7], mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing [8], and whole genome sequencing [9–11]. Published reviews have identified factors that might influence or bias clustering by IS*6110* RFLP [12,13]. No study has repeated this analysis using more up-to-date typing methods, which is important for understanding of the epidemiology of TB and to shape the application of molecular typing to improve TB control.

Published meta-analyses and modelling studies using IS*6110* RFLP data show that the proportion of clustering observed can be affected by 1) study design (affecting the proportion of eligible cases that are included in the study); 2) features of the typing method (such as the ability to type isolates with low copy numbers); and 3) study setting (such as characteristics of the study population). For example, the proportion of clustering increases when the fraction of the total data sampled increases [13–15] and when study duration increases [16].

MIRU-VNTR is currently the preferred method of molecular typing [17–21], and can be used together with Spoligotyping [8]. Relative to IS*6110* RFLP, MIRU-VNTR does not have to exclude isolates with a low IS*6110* copy number, has a faster turnaround time, is high throughput and the numeric strain types are more easily compared. MIRU-VNTR strain typing is increasingly being adopted worldwide [1,22–27], yet unlike IS*6110* RFLP, the evidence for the interpretation of the findings such as the impact of study design and setting on clustering have not been reviewed. Although the two typing methods have been shown to have a similar discriminatory value, the markers evolve independently and at different rates, resulting in a difference in clustering between the two methods [28]. This suggests that there could be differences in the way study design, typing method and setting affects clustering by the two methods. We conducted a systematic review to assess the evidence for the impact of study design and setting on the interpretation of TB

transmission using clustering derived from MIRU-VNTR strain typing – as has been shown using IS*6110* RFLP typing.

Methods

Five electronic databases were searched (EMBASE, ISI Web of Science, CINHAL, Scopus and Medline (Ovid)) up to 20th October 2014. The search strategy combined the following terms with Boolean operators: Tuberculosis, strain typing, and transmission (Appendix 1). The search was limited to studies using the standard MIRU-VNTR method [8], in humans only, and in English.

All titles and abstracts from each of the searches were examined. The full text of each paper was obtained and reviewed if the study reported MIRU-VNTR strain typing of *M.tuberculosis* complex isolates with at least 15 of the standardised 24 loci (ETR A, B, C, D, E; MIRU 2, 10, 16, 20, 23, 24, 26, 27, 39, 40; VNTR 424, 1955, 2163b, 2347, 2401, 3171, 3690, 4052, 4156) [8,29,30].

Studies using fewer than 15 loci were not included because the level of discrimination is inadequate for epidemiological use (n=121) [8]. Studies that used loci different to the standardised 15 and 24 set were not included in the analysis in order to reduce the heterogeneity between studies (n=19). All publication types were included in this first screen to ensure that no relevant data were missed.

Reviews, letters, editorials, outbreaks or case reports (n=103) were excluded in the second screen. Studies that used incomplete sampling (e.g. random samples, studies using subsets of populations such as multidrug-resistant patients) (n=47) and studies that had a sample size of less than 50 (n=4) were also excluded.

A reviewer (JM) extracted the following data items from all included studies using a form developed in Excel (Microsoft 2010): publication details (year, authors, study country), study details (study duration, loci typed, secondary typing method, study population, whether participant consent was required (a characteristic of the study design that was used as proxy for sampling fraction, assuming that where consent was required the sampling fraction was low)), the number of clustered and unique isolates, and the covariates of interest: the maximum size of clusters; the proportion of clusters containing two cases; the proportion of the population that was culture positive; the proportion of culture positive isolates typed; risk factors for clustering; and the Hunter Gaston Discriminatory Index (HGDI) [31]). IA extracted data from 10% of the papers for external validity, disagreements were discussed and a consensus agreed upon.

The main outcome measure – the proportion of TB isolates clustered by MIRU-VNTR strain typing – was calculated as the number of clustered isolates/number of clustered+unique isolates. Where there were uncertainties JM consulted with IA.

Authors were contacted if TB incidence rate was not reported. Where no response was received WHO country estimates of TB incidence for the study year were used.[32] As so few studies reported the proportion coinfected with TB/HIV, these estimates for the study country were taken from an EU-wide survey and WHO country profiles.[33,34] Due to poor recording of the sampling fraction (the number of isolates typed/the total number of culture positive TB cases diagnosed during the study period (n=19)), whether the study required the consent of participants (yes/no) was included as a proxy for (low/high) sampling fraction. The risk of bias within each study was assessed using the STROME-ID checklist.[35]

Data were analysed in Stata 12. Where studies reported data from more than one set of loci, the method with the highest discriminatory value was included (i.e. MIRU-VNTR 24 would be chosen over MIRU-VNTR 15, and MIRU-VNTR 15 plus Spoligotyping would be chosen over MIRU-VNTR 15 alone) (n=8). This review was not concerned with summary measures of clustering, but factors that influenced clustering; therefore articles must have included at least one of the covariates. Continuous variables were transformed where the distribution was skewed. The proportion clustered was transformed using the Freeman Tukey transformation [36]. Study heterogeneity was assessed using a forest plot and the chi^2 test of heterogeneity. Univariable meta-regression analyses were carried out to determine the effect of the study design covariates on the proportion of clustered isolates. All covariates in the analysis were hypothesised to influence the proportion clustered *a priori*.

Sensitivity analyses were conducted to see the effect of removing studies reporting 0% clustering, with only extra-pulmonary TB cases, only *M.bovis* cases, studies using the 'old 12' MIRU loci as part of their 15 loci, and studies assessed as having a high likelihood of bias (STROME-ID score less than 20).

Results

The search identified 7274 references resulting in 27 studies (25 journal articles and 2 conference abstracts) included after deduplication and title/abstract/full text screening (Figure 1). The main characteristics of the included studies are shown in Table 1. Studies were published between 2007 and 2014 and the clustering reported varied from 0% [37] to 62.8% [38]. In all studies, clustered

isolates were defined as having identical strain types based on the MIRU-VNTR loci typed, with or without Spoligotyping. 17 studies included isolates from newly diagnosed TB cases, three studies reported including isolates from new and chronic cases of TB, and seven did not report this information. In addition, ten studies did not include repeat isolates from the same patient, one study included a repeat isolate from one patient, and the remaining 17 did not report whether repeat isolates were included or not. Furthermore, four studies included isolates with missing loci in the cluster analysis, whereas four excluded isolates with missing loci, and the remaining 20 did not report how they dealt with missing loci. The number of studies reporting each variable of interest is shown in Table 2. STROME-ID scores can be found in Appendix 2.

A forest plot shows the spread of clustering reported by number of loci and additional typing method (Figure 2). Significant heterogeneity was identified between the studies ($p<0.001$), suggesting that a meta-regression would be an appropriate analysis.

The univariable meta-regression shows evidence for the proportion of clustering to decrease as the number of MIRU-VNTR loci typed increased from 15 to 24 ($p=0.04$; Table 3), accounting for 14% of the between study variation, and to increase when the study participants consented to being included in the study ($p=0.03$), accounting for 14% of the between study variation. The proportion of clustering increased as the TB incidence in the population increased ($p=0.007$, Adj $R^2 = 26.7$). There was also evidence for the proportion of clustering to increase as the maximum cluster size increased ($p=0.001$), accounting for 48% of between study variation. There was no evidence of the other study design or study setting variables significantly influencing the proportion clustered. Though non-significant ($p>0.05$), the TB/HIV coinfection rate in the population explained 2% of the between study variation. Too few studies included information on the proportion of clusters containing two cases, proportion of the study sample with previous TB or with pulmonary TB, so these could not be included in the analysis (Table 2).

Sensitivity analyses to examine the effect of excluding studies reporting 0% clustering,[37] only *M.bovis* cases,[39] studies using the 'old 12' MIRU loci,[39–44] and studies assessed as having a high risk of bias,[37,45–48] did not generally change the results. The proportion of culture positive TB in the population remained insignificant but explained 2.6% of the between study variation when excluding 0% clustering ($p=0.278$ and Adj $R^2=2.62$). Similarly, the proportion of culture positive TB in the population remained insignificant but explained 2.6% of the between study variation when excluding studies with the highest risk of bias ($p=0.278$ and Adj $R^2=2.62$). The number of loci typed became non-significant, but explained 9.6% and 10.5% of the between study variation when

excluding studies using the 'old 12' loci and the highest risk of bias, respectively (p=0.106, Adj R^2=9.63; p=0.111, Adj R^2=10.51, respectively).

Discussion

This review identified 27 studies that met the inclusion criteria. We illustrate that the interpretation of studies using MIRU-VNTR to estimate clustering is subject to bias relating to study design and setting; however, there were insufficient data available to fully explore this impact.

As expected, we found that the proportion of clustering decreased with a greater number of MIRU-VNTR loci typed, with increasing TB incidence and with increasing maximum cluster size. We found that requiring consent to type patient isolates increased the proportion of clustering, which is not expected, given that the sampling fraction would be lower in these studies.

The other study design variables included in this analysis, such as study duration, did not significantly influence the proportion of isolates that were clustered, contrary to previous findings [12]. This is likely to be because of a lack of good quality evidence: of the 27 studies that met the inclusion criteria for the review, none reported all the variables of interest, reducing the power of the analysis and precluding multivariable meta-regression (Table 2). Importantly, key details of cluster analyses were not reported consistently across the studies, such as whether repeat isolates from the same patients were included, or typing profiles with missing loci were included, introducing new, unmeasured biases. In addition, the range of the variables may have been too limited to show any impact on clustering estimates. For example, the proportion of culture positive isolates typed ranged from 34.5% to 100%, with 17 of the 19 studies reporting this variable from 81.9% to 100%. Furthermore, most of the studies (17/27=63%) were from low TB burden settings and therefore may be reflecting the rate at which imported cases have matching strain types by chance, rather than rates of recent transmission.

The sensitivity analysis suggested that, when excluding the studies with the greatest risk of bias, the culture-positivity in the population might explain a small amount of the between study variation. This is consistent with estimates of the influence of sampling on the proportion of clustering using *IS*6110 RFLP typing [49]. In the sensitivity analysis excluding studies that used the 'old 12' loci, the effect of the number of loci typed becomes non-significant. This is likely because studies using the 'old 12' accounted for six out of ten studies reporting 15 loci, reducing the number of studies and the power of the model.

This study is a timely evaluation of the impact of study design on estimates of TB clustering using MIRU-VNTR strain typing because it has been incorporated into national typing services globally [23,50]. The findings are relevant where strain typing is used to evaluate TB control systems across different settings because the proportion of clustering is influenced by the number of loci typed, the TB incidence and the maximum cluster size. Given that strain typing methods are advancing beyond MIRU-VNTR typing and that the application of whole genome sequencing to TB control and public health strategies has been demonstrated [9–11,51], it is important that the biases in the analysis of such methods are explored and compared. Understanding how to design and compare research studies for public health will greatly improve the benefit gained from newer technologies.

The strength of this meta-analysis was limited by (a lack of) detail reported by the included studies. This review has highlighted the need for better quality reporting in primary studies to enable future reviews to be more robust. Recently published standards for reporting of molecular epidemiology for infectious diseases should improve the quality of reporting.[35] This review is further limited by our inability to access 58 of the title/abstract screened articles for full text screening.

The use of TB strain typing as a public health tool in TB control programmes is increasing globally. We have identified a lack of good quality studies that can contribute to our understanding in interpreting the molecular typing of TB. We have also shown that the proportion of clustering derived from MIRU-VTNR typing is influenced by the number of loci typed, whether consent is required to type isolates, TB incidence in the study setting, and the maximum cluster size, highlighting these as important considerations in the design and interpretation of future studies.

Conflict of interest

Nothing to declare.

Acknowledgements

We would like to acknowledge Ross Harris from the Statistics Unit at Public Health England for his advice on meta-regression.

Author contributions

All authors made substantial contributions to the conception and design of the review, and the analysis and interpretation of data. JM drafted the article and PS, IA, TM and TC revised it critically for important intellectual content. All authors approved the final version for publication.

Funding

JM is funded through a Public Health England and University College London Impact Studentship. IA is funded through a NIHR Senior Research Fellowship.

Ethics

Ethical approval was not required as this review analyses data that is in the public domain.

Data sharing

No additional data are available

References

1. Lambregts-van Weezenbeek CSB, Sebek MMGG, van Gerven PJHJ, de Vries G, Verver S, et al. (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. Int J Tuberc Lung Dis 7: S463–470.

2. Borgdorff MW, van den Hof S, Kremer K, Verhagen L, Kalisvaart N, et al. (2010) Progress towards tuberculosis elimination: secular trend, immigration and transmission. Eur Respir J 36: 339–347. doi:10.1183/09031936.00155409.

3. Kik SV, Verver S, Van Soolingen D, De Haas PEW, Cobelens FG, et al. (2008) Tuberculosis Outbreaks Predicted by Characteristics of First Patients in a DNA Fingerprint Cluster. Am J Respir Crit Care Med 178: 96–104. doi:10.1164/rccm.200708-1256OC.

4. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 31: 1677–1682.

5. De Vries G, van Hest RAH, Richardus JH (2007) Impact of mobile radiographic screening on tuberculosis among drug users and homeless persons. Am J Respir Crit Care Med 176: 201–207. doi:10.1164/rccm.200612-1877OC.

6. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35: 907–914.

7. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31: 406–409.

8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 44: 4498–4510. doi:10.1128/JCM.01392-06.

9. Schürch AC, van Soolingen D (2011) DNA fingerprinting of Mycobacterium tuberculosis: From phage typing to whole-genome sequencing. Infect Genet Evol. Available: http://www.ncbi.nlm.nih.gov/pubmed/22067515. Accessed 13 March 2012.

10. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364: 730–739. doi:10.1056/NEJMoa1003176.

11. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. (2013) Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13: 137–146. doi:10.1016/S1473-3099(12)70277-3.

12. Houben RMGJ, Glynn JR (2009) A systematic review and meta-analysis of molecular epidemiological studies of tuberculosis: development of a new tool to aid interpretation. Tropical Medicine & International Health 14: 892–909. doi:10.1111/j.1365-3156.2009.02316.x.

13. Fok A, Numata Y, Schulzer M, FitzGerald MJ (May) Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies [Review Article]. The International Journal of Tuberculosis and Lung Disease 12: 480–492.

14. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, Van Soolingen D (2011) Influence of Sampling on Clustering and Associations With Risk Factors in the Molecular Epidemiology of Tuberculosis. Am J Epidemiol. Available: http://aje.oxfordjournals.org/content/early/2011/05/23/aje.kwr061. Accessed 29 March 2012.

15. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, et al. (1999) Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis 3: 1055–1060.

16. Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD, et al. (2005) The Importance of Recent Infection with Mycobacterium tuberculosis in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study in Northern Malawi. J Infect Dis 192: 480–487. doi:10.1086/431517.

17. De Beer JL, Kremer K, Ködmön C, Supply P, van Soolingen D (2012) First Worldwide Proficiency Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex Strains. Journal of Clinical Microbiology 50: 662–669. doi:10.1128/JCM.00607-11.

18. Maes M, Kremer K, van Soolingen D, Takiff H, de Waard JH (2008) 24-locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis (Edinb) 88: 490–494. doi:10.1016/j.tube.2008.04.003.

19. Sougakoff W (2011) Molecular epidemiology of multidrug-resistant strains of Mycobacterium tuberculosis. Clinical Microbiology and Infection 17: 800–805. doi:10.1111/j.1469-0691.2011.03577.x.

20. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res 38: W326–331. doi:10.1093/nar/gkq351.

21. Supply P (2010) MIRU-VNTR typing: the new international standard for TB molecular epidemiology Symposium of the Institut Pasteur de Tunisia.

22. Van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, et al. (1999) Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J Infect Dis 180: 726–736. doi:10.1086/314930.

23. Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step approach for large-scale, prospective genotyping of Mycobacterium tuberculosis isolates in the United States. J Clin Microbiol 43: 688–695. doi:10.1128/JCM.43.2.688-695.2005.

24. Centers for Disease Control and Prevention (2005) New CDC Program for Rapid Genotyping of Mycobacterium Tuberculosis Isolates. JAMA 293: 2086–2086. doi:10.1001/jama.293.17.2086.

25. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, et al. (2000) A prospective evaluation of the clinical value of nation-wide DNA fingerprinting of tuberculosis isolates in Denmark. Int J Tuberc Lung Dis 4: 295–299.

26. Bauer J, Yang Z, Poulsen S, Andersen AB (1998) Results from 5 years of nationwide DNA fingerprinting of Mycobacterium tuberculosis complex isolates in a country with a low incidence of M. tuberculosis infection. J Clin Microbiol 36: 305–308.

27. Zolnir-Dovc M, Poljak M, Erzen D, Sorli J (2003) Molecular epidemiology of tuberculosis in Slovenia: results of a one-year (2001) nation-wide study. Scand J Infect Dis 35: 863–868.

28. Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Hoek KGP, et al. (2008) Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of Mycobacterium tuberculosis Beijing strains in a setting of high incidence of tuberculosis. J Clin Microbiol 46: 3338–3345. doi:10.1128/JCM.00770-08.

29. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, et al. (2001) Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. J Clin Microbiol 39: 3563–3571. doi:10.1128/JCM.39.10.3563-3571.2001.

30. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniewski FA (2006) Progression toward an improved DNA amplification-based typing technique in the study of Mycobacterium tuberculosis epidemiology. J Clin Microbiol 44: 2492–2498. doi:10.1128/JCM.01428-05.

31. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465–2466.

32. WHO | TB data (n.d.). WHO. Available: http://www.who.int/tb/country/en/index.html. Accessed 12 December 2012.

33. Kruijshaar ME, Pimpin L, Abubakar I, Rice B, Delpech V, et al. (2011) The burden of TB-HIV in the EU: how much do we know? A survey of surveillance practices and results. Eur Respir J 38: 1374–1381. doi:10.1183/09031936.00198310.

34. World Health Organization (n.d.) WHO Tuberculosis Country Profiles. Available: http://www.who.int/tb/country/data/profiles/en/.

35. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, et al. (2014) Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. The Lancet Infectious Diseases 14: 341–352. doi:10.1016/S1473-3099(13)70324-4.

36. Freeman MF, Tukey JW (1950) Transformations Related to the Angular and the Square Root. Ann Math Statist 21: 607–611. doi:10.1214/aoms/1177729756.

37. Guang-ming D, Zhi-guo Z, Peng-ju D, Qian Z, Li W, et al. (2013) Differences in the population of genetics of Mycobacterium tuberculosis between urban migrants and local residents in Beijing, China. Chinese Medical Journal 126: 4066–4071. doi:10.3760/cma.j.issn.0366-6999.20130216.

38. Zmak L, Obrovac M, Katalinic Jankovic V (2014) First insights into the molecular epidemiology of tuberculosis in Croatia during a three-year period, 2009 to 2011. Scandinavian Journal of Infectious Diseases 46: 123–129.

39. Mandal S, Bradshaw L, Anderson LF, Brown T, Evans JT, et al. (2011) Investigating transmission of Mycobacterium bovis in the United Kingdom in 2005 to 2008. J Clin Microbiol 49: 1943–1950. doi:10.1128/JCM.02299-10.

40. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR (2011) Tuberculosis transmission in Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. Infect Genet Evol 11: 124–131. doi:10.1016/j.meegid.2010.09.013.

41. Dymova MA, Liashenko OO, Poteiko PI, Krutko VS, Khrapov EA, et al. (2011) Genetic variation of Mycobacterium tuberculosis circulating in Kharkiv Oblast, Ukraine. BMC Infect Dis 11: 77. doi:10.1186/1471-2334-11-77.

42. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, et al. (2011) Molecular epidemiology of Mycobacterium tuberculosis in East Lancashire 2001-2009. Thorax 66: 709–713. doi:10.1136/thx.2011.158881.

43. Evans J (2010) Analysis of prevalent Mycobacterium tuberculosis strains in the United Kingdom: detection, distribution and expansion of MIRU-VNTR profiles containing high numbers of isolates. European Society of Clinical Microbiology and Infectious Diseases. Vienna, Austria.

44. Nikolayevskyy VV, Brown TJ, Bazhora YI, Asmolov AA, Balabanova YM, et al. (2007) Molecular epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in Mycobacterium tuberculosis strains from the southern Ukraine. Clin Microbiol Infect 13: 129–138. doi:10.1111/j.1469-0691.2006.01583.x.

45. L. Z, M. O, V. KJ (2014) First insights into the molecular epidemiology of tuberculosis in Croatia during a three-year period, 2009 to 2011. Scandinavian Journal of Infectious Diseases.

46. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, et al. (2011) Evaluation of Mycobacterium tuberculosis typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. J Clin Microbiol 49: 4173–4178. doi:10.1128/JCM.05293-11.

47. Dymova MA, Kinsht VN, Cherednichenko AG, Khrapov EA, Svistelnik AV, et al. (2011) Highest prevalence of the Mycobacterium tuberculosis Beijing genotype isolates in patients newly diagnosed with tuberculosis in the Novosibirsk oblast, Russian Federation. J Med Microbiol 60: 1003–1009. doi:10.1099/jmm.0.027995-0.

48. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, et al. (2009) Prospective universal application of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for fast identification of clustered and orphan cases. J Clin Microbiol 47: 2026–2032. doi:10.1128/JCM.02308-08.

49. Glynn JR, Vyonycky E, Fine PEM (1999) Influence of Sampling on Estimates of Clustering and Recent Transmission of Mycobacterium tuberculosis Derived from DNA Fingerprinting Techniques. American Journal of Epidemiology 149: 366 –371.

50. TB Strain Typing Project Board HPA (2011) TB Strain Typing Cluster Investigation Handbook for Health Protection Units 1st Edition. Available: https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealthProtectionProgrammes/Respiratory/TB/StrainTyping/. Accessed 30 November 2011.

51. Walker TM, Monk P, Grace Smith E, Peto TEA (2013) Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect. doi:10.1111/1469-0691.12183.

52. Gurjav U, Jelfs P, McCallum N, Marais BJ, Sintchenko V (2014) Temporal dynamics of Mycobacterium tuberculosis genotypes in New South Wales, Australia. BMC infectious diseases 14: 455–455. doi:10.1186/1471-2334-14-455.

53. Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 46: 1398–1406. doi:10.1128/JCM.02089-07.

54. Allix-Béguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008) Standardised PCR-based molecular epidemiology of tuberculosis. Eur Respir J 31: 1077–1084. doi:10.1183/09031936.00053307.

55. Tuite AR, Guthrie JL, Alexander DC, Whelan MS, Lee B, et al. (2013) Epidemiological evaluation of spatiotemporal and genotypic clustering of mycobacterium tuberculosis in Ontario, Canada. International Journal of Tuberculosis and Lung Disease 17: 1322–1327.

56. Tessema B, Beer J, Merker M, Emmrich F, Sack U, et al. (2013) Molecular epidemiology and transmission dynamics of Mycobacterium tuberculosis in Northwest Ethiopia: new phylogenetic lineages found in Northwest Ethiopia. Bmc Infectious Diseases 13: 131. doi:10.1186/1471-2334-13-131.

57. Smit PW, Haanpera M, Rantala P, Couvin D, Lyytikainen O, et al. (2013) Molecular Epidemiology of Tuberculosis in Finland, 2008-2011. Plos One 8: e85027. doi:10.1371/journal.pone.0085027.

58. Oelemann MC, Diel R, Vatin V, Haas W, Rüsch-Gerdes S, et al. (2007) Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol 45: 691–697. doi:10.1128/JCM.01393-06.

59. Ojo OO, Sheehan S, Corcoran DG, Nikolayevsky V, Brown T, et al. (2010) Molecular epidemiology of Mycobacterium tuberculosis clinical isolates in Southwest Ireland. Infect Genet Evol 10: 1110–1116. doi:10.1016/j.meegid.2010.07.008.

60. Aleksic E, Merker M, Cox H, Reiher B, Sekawi Z, et al. (2013) First Molecular Epidemiology Study of Mycobacterium tuberculosis in Kiribati. PLoS ONE 8. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8.

61. De Beer JL, van Ingen J, de Vries G, Erkens C, Sebek M, et al. (2013) Comparative study of IS6110 restriction fragment length polymorphism and variable-number tandem-repeat typing of Mycobacterium tuberculosis isolates in the Netherlands, based on a 5-year nationwide survey. J Clin Microbiol 51: 1193–1198. doi:10.1128/JCM.03061-12.

62. Varghese B, Supply P, Shoukri M, Allix-Beguec C, Memish Z, et al. (2013) Tuberculosis Transmission among Immigrants and Autochthonous Populations of the Eastern Province of Saudi Arabia. PLoS ONE 8. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2.

63. Lim LK-Y, Sng LH, Win W, Chee CB-E, Hsu LY, et al. (2013) Molecular Epidemiology of Mycobacterium tuberculosis Complex in Singapore, 2006-2012. Plos One 8: e84487. doi:10.1371/journal.pone.0084487.

64. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med 105 Suppl 1: S67–73. doi:10.1016/S0954-6111(11)70014-2.

65. Jonsson J., Hoffner S., Berggren I., Bruchfeld J., Ghebremichael S., et al. (2014) Comparison between RFLP and MIRU-VNTR genotyping of mycobacterium tuberculosis strains isolated in stockholm 2009 to 2011. PLoS ONE. Available: http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0095159&representation=PDF.

66. Muwonge A, Malama S, Johansen TB, Kankya C, Biffa D, et al. (2013) Molecular Epidemiology, Drug Susceptibility and Economic Aspects of Tuberculosis in Mubende District, Uganda. PLoS ONE 8. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697.

67. Hamblion EL, Wynne-Edwards E, Anderson C, Anderson SR (2011) A summary of strain typing and clustering of TB in London in 2010 and an analysis of the associated risk factors. Thorax 66: A88–A89. doi:10.1136/thoraxjnl-2011-201054c.50.

68. Hang NTL, Maeda S, Lien LT, Thuong PH, Hung NV, et al. (2013) Primary Drug-Resistant Tuberculosis in Hanoi, Viet Nam: Present Status and Risk Factors. Plos One 8: UNSP e71867. doi:10.1371/journal.pone.0071867.

Tables

Table 1: The study setting and design characteristics of the included articles

Ref	Study setting							Study design							Risk of bias [d]	Clustering (%)[e]
	Study area and country	TB incidence (per 100,000)	TB/HIV (per 100,000)[a]	Previous TB treatment (%)	Pulmonary TB (%)	Maximum cluster size	Clusters of size 2 (%)	Study duration (months)	Study size (clustered + unique isolates)	Culture positive in study population (%)	Culture positive isolates typed (%)	Typing method [b]	Loci typed [c]	Consent required		
[52]	New South Wales, Australia	6.7	0.2	0.0	63.7	.	.	36	1128	.	.	m24	N	no	low	20.1
[40]	Tabriz and Orumieh, Azarbaijan	26.0	.	5.2	87.0	5	81.8	12	156	.	94.5	m15	O	no	low	32.7
[53]	Brussels-Capital Region, Belgium	35.2	5.1	10.8	.	23	64.2	24	530	86.1	87.9	m24	N	no	low	29.6
[54]	Brussels-Capital Region, Belgium	35.2	5.1	.	100	.	.	39	802	81.8	84.7	m24s	N	no	low	28.8
[55]	Ontario, Canada	4.8	0.4	.	.	18	58.8	65	2016	.	.	m24s	N	no	low	23.1
[37]	Changping District, Beijing, China	.	0.3	.	100	0	.	30	318	31.5	94.6	m24	N	no	high	0.0
[38]	Croatia	19.0	0.1	.	.	45	48.3	36	1587	.	.	m15	N	no	high	62.8
[56]	Amhara region, Northwest Ethiopia	.	24.0	17.6	100	13	.	5	244	.	.	m24	N	yes	low	45.1
[57]	Finland	5.0	0.0	.	.	20	.	48	1048	75.4	99.4	m15s	.	no	low	33.9
[58]	Hamburg, Germany	12.7	.	.	.	.	45.5	12	154	78.2	91.1	m24s	N	no	low	22.1
[46]	Schleswig-Holstein, Germany	3.2	0.1	.	.	22	44.4	48	277	.	.	m24s	N	no	high	27.1
[59]	South West Ireland	15.3	3.3	.	82.7	12	.	36	171	79.5	96.1	m24s	N	no	low	27.5
[60]	South Tawara, Kiribati	370.0	.	4.1	100	25	55.6	24	73	45.4	98.6	m24s	N	yes	low	75.3
[61]	Netherlands	6.5	0.2	.	.	.	57.2	60	3978	.	100.1	m24	N	no	low	46.7
[41]	Kharkiv, Russia	94.0	3.8	63.3	100	10	50.0	3	98	.	100	m15	O	yes	high	31.6
[62]	Eastern province, Saudi Arabia	4.0	.	.	73.1	24	19.0	24	522	.	.	m24s	N	no	low	40.2

[63]	Singapore	40.5	1.2	.	.	21	48.0	24	1128	82.0	34.5	m24s	N	no	low	30.8
[64]	Slovenia	10.6	0.0	.	.	6	.	12	196	94.4	97.5	m24s	N	no	low	36.2
[48]	Almeria, Spain	26.0	6.0	.	.	8	.	27	281	.	81.9	m15	N	no	high	43.1
[65]	Sweden	4.8	0.1	.	.	10	.	36	406	.	.	m24s	N	no	low	21.2
[66]	Mubende, Uganda	.	86.0	31.1	87.8	11	70.0	6	67	21.5	90.5	m15s	N	yes	low	35.8
[42]	East Lancashire, UK	18.3	8.2	.	.	13	58.3	102	332	48.5	69.9	m15	O	no	low	42.8
[39]	UK	.	8.2	.	42.3	12	50.0	48	102	90.7	87.2	m15	O	no	low	30.4
[67]	London, UK	44.9	8.2	.	.	.	.	9	964	36.0	100	m24	N	no	.	37.0
[43]	Midlands, UK	15.0	8.2	.	.	.	.	48	4207	58.3	100	m15	O	no	.	61.2
[44]	Odessa and Nikolaev, Ukraine	80.4	3.9	34.2	100	.	.	4	225	.	.	m15	O	yes[f]	low	60.4
[68]	Hanoi, Vietnam	146.0	10.0	0.0	100	.	.	20	465	92.7	91.9	m15s	N	yes	low	55.3

[a] Estimates from of the prevalence of TB/HIV co-infection in the study country [33,34]

[b] 15=15 MIRU-VNTR loci (made up of the 'old 12' or 'new 12' defined in the footnote below), 24=24 MIRU-VNTR loci (ETR A, B, C, D, E; MIRU 2, 10, 16, 20, 23, 24, 26, 27, 39, 40; VNTR 424, 1955, 2163b, 2347, 2401, 3171, 3690, 4052, 4156), S=with Spoligotyping

[c] O= old 12 MIRU loci (MIRU 2, 4, 10, 16, 20, 23, 24, 26, 27,30, 31, 39, 40), N=new 12 MIRU loci (MIRU 10, 16, 26, 31, 40 + Mtub 04, 21, 39 + ETR A C + QUB 11b, 26)

[d] Risk of bias was assessed using the STROME-ID checklist. Studies scoring <20 were categorised as have a high risk of bias. See Appendix 2 for STROME-ID scores

[e] The proportion of clustering was calculated as the number of clustered isolates/number of clustered + unique isolates

[f] 11.3% did not consent to being part of the study. The other studies that required consent for isolates to be typed did not report the refusal rate

Table 2: The number of studies that reported the variables of interest

	Reported	Missing
Study setting		
TB incidence	8	15
TB/HIV co-infection	5	22
Previous TB treatment	9	18
Proportion pulmonary TB	14	13
Maximum cluster size	19	8
% clusters with 2 cases	14	13
Study design		
Study duration	27	0
Study size	27	0
% population that is culture positive	15	12
% culture positive typed	19	8
24 loci (compared to 15)	27	0
Repeat isolates	12	15
Missing loci	8	19
Double alleles	1	26
Consent required	6[a]	21
Epidemiological information	6	21

[a] Only one study reported the consent rate

Table 3: Univariable metaregression showing the coefficients for change in the proportion of clustering and the percentage of between-study variation explained by variables describing the study design and setting.

	n	Coefficient [a]	CI	p	Adj R^{2} [b]
Study setting					
TB incidence	23	0.14	0.04-0.24	0.007	26.74
TB/HIV co-infection	23	0.04	-0.03-0.11	0.246	2.00
Maximum cluster size	19	0.20	0.09-0.30	0.001	48.20
Study design					
Study duration	27	-0.02	-0.09-0.06	0.677	-3.37
% population that is culture positive	15	0.34	-1.23-1.96	0.661	-5.92
% culture positive typed	19	0.22	-1.08-1.52	0.725	-5.41
Study size	27	0.03	-0.11-0.16	0.702	-3.31
24 loci (compared to 15)	27	-0.30	-0.59--0.01	0.04	13.58
Consent required	27	0.38	0.04-0.72	0.029	14.41

[a] Coefficients for the change in the proportion of clustering for each covariate. E.g. for a one unit increase in maximum cluster size, the proportion of clustering increases by 0.2.

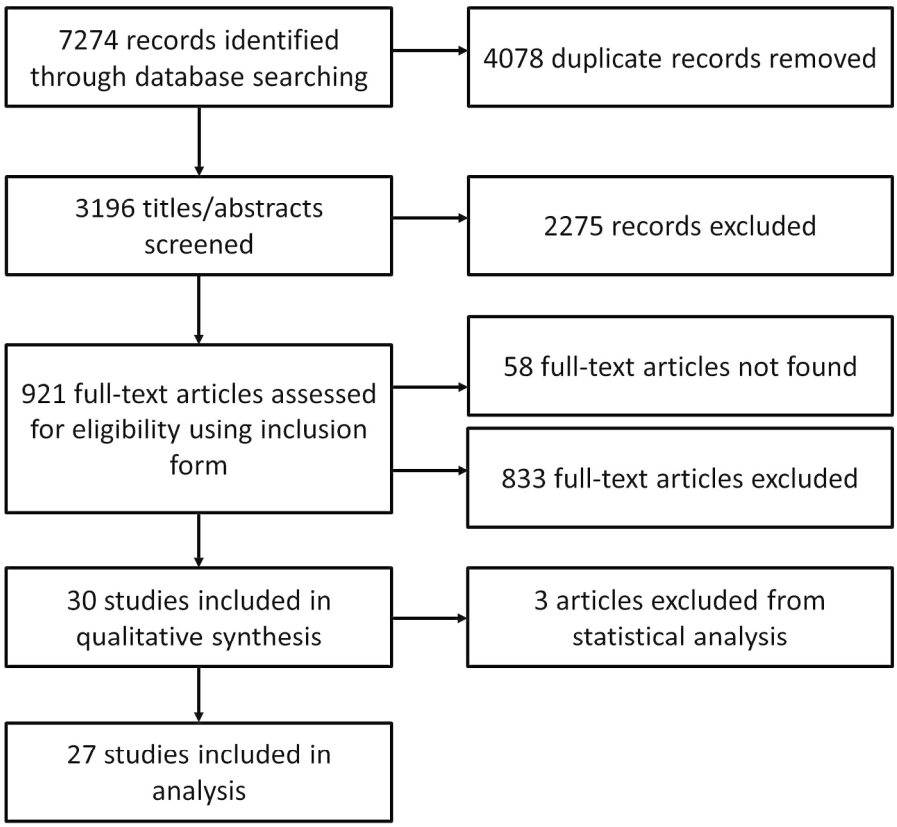
[b] The proportion of between-study variation explained by the univariate meta-regression.

Figure Caption

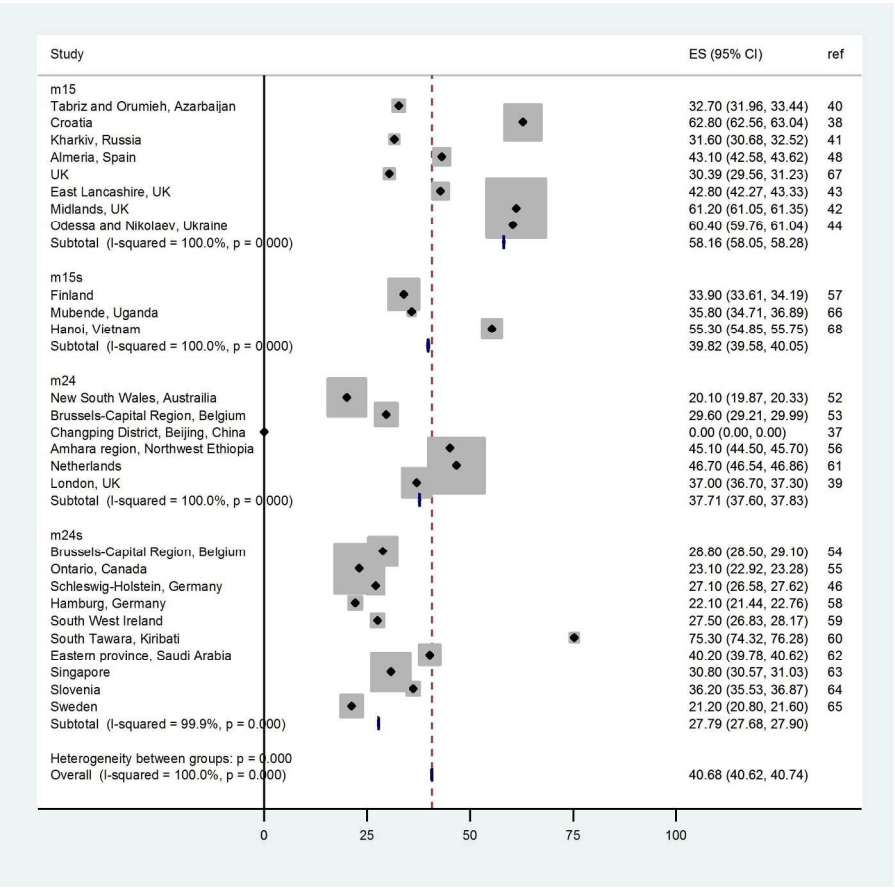
Figure 1: Results of systematic search, screening and data extraction.

Figure 2: Forest plot showing the proportion of clustering reported in each study by the number of MIRU-VNTR loci typed

The number of loci typed is categorised into 15 loci (m15), 15 loci with Spoligotyping (m15s), 24 loci (m24) and 24 loci with Spoligotyping (m24s). The study reference is shown in the right hand column.



190x254mm (300 x 300 DPI)



190x254mm (300 x 300 DPI)

Appendix 1: Medline/Embase search strategy

1. (tubercle adj3 (bacillus or bacilli)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

2. ((mycobacterium or mycobacteria) adj3 (bovis or africanum or microti or canetti)).mp.

3. exp tuberculosis/ or mycobacterium tuberculosis/ or tuberculosis.mp. or tb.mp. or Mtb.mp. or "M tuberculosis complex".mp.

4. or/1-3

5. Minisatellite Repeats/ or Genotype/ or Interspersed Repetitive Sequences/ or DNA Fingerprinting/ or Bacterial Typing Techniques/

6. "miru".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

7. "vntr".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

8. (miru adj3 vntr).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

9. (mycobacterial adj3 interspersed adj3 repetitive adj3 units).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

10. (dna adj3 fingerprinting).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

11. ((strain adj3 type) or (strain adj3 typing) or (strain adj3 types)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

12. ((molecular adj3 typing) or (molecular adj3 strain adj3 typ*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

13. (genotype or genotyping or genotypes).ti,ab.

14. (minisatellite adj3 repeat*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

15. molecular epidemiology/mt or (molecular adj3 epidemiology).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

16. or/5-15

17. exp disease outbreaks/ or (outbreak adj3 analysis).mp. or (outbreak adj3 investigation).mp. or (outbreak adj3 management).mp. or (tuberculosis adj3 outbreak).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

18. exp contact tracing/ or (contact adj3 tracing).mp. or (contact* adj3 traced).mp. or (contact adj3 screen*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

19. exp case management/ or (case adj3 management).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

20. exp Risk Factors/

21. (risk adj3 factor*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

22. exp Epidemiologic Factors/

23. infectious disease transmission.mp. or exp Disease Transmission, Infectious/

24. exp case management/ or (case adj3 management).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

25. program evaluation/ or evaluation studies as topic/ or (program adj3 evaluation).mp. or (programme adj3 evaluation).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

26. public health practice/ or (public adj3 health).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

27. ((tuberculosis adj3 control) or (tb adj3 control)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

28. (molecular adj3 surveillance).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

29. exp cluster analysis/ or (cluster* adj3 rate*).mp. or (cluster* adj3 growth).mp. or (cluster* adj3 analysis).mp. or (cluster adj3 investigation).mp. or (proportion adj3 cluster*).mp. or (molecular adj3 cluster*).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

30. ((recent adj3 transmission) or (transmission adj3 event*) or (transmission adj3 rate*) or (chain adj3 transmission) or (transmission adj3 setting*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]

31. or/17-30

32. 4 and 16

33. 32 and 31

34. limit 33 to yr="1998-Current"

35. limit 34 to english language

36. animals/

37. humans/

38. 36 not 37

39. 35 not 38

Appendix 2: STROME-ID scores for the included studies

Author	STROME-ID score[a]
Aleksic, E	24
Alliex-Beguec, C	32
Allix-Beguec, C	25
Alonso-Rodriguez, N	18
Asgharzadeh, M	28
Bidovec-Stojkovic, U	31
De Beer, JL	30
Dymova, MA	19
Evans, J	[b]
Grujav, U	32
Guang-ming, DAI	19
Hamblion, E	[b]
Hang, NTHL	31
Jonsson, J	22
Lim, LKY	30
Mandal, S	32
Muwonge, A	25
Nikolayevsky, V	23
Oelemann, M	34
Ojo, OO	36
Roetzer, A	16
Sails, A	23
Smit, PW	29
Tessema, B	26
Tuite, AR	31
Varghese, B	23
Zmak, L	19

[a]Individual studies score 1 for each element of checklist they had address

[b]Conference abstracts

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	5
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	5
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	15
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	15
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	18
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	15
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	18
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	7
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	7-8
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	8
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	8

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.