

BMJ Open

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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

4 Jessica Mears¹, Ibrahim Abubakar^{1,2}, Theodore Cohen³, Timothy D McHugh⁴ & Pam Sonnenberg^{1,*}

5 ¹ Department of Infection and Population Health, University College London

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

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

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

10 *Corresponding author:

11 Dr Pam Sonnenberg

12 3rd Floor Mortimer Market Centre

13 University College London

14 London WC1E 6JB

15 p.sonnenberg@ucl.ac.uk

16

17 **Word count: 1,733**

1
2
3 **18 Abstract**
4

5 19 Objectives: To systematically review the evidence for the impact of study design and setting on the
6 20 interpretation of TB transmission using clustering derived from Mycobacterial Interspersed
7 21 Repetitive Units – Variable Number Tandem Repeats (MIRU-VNTR) strain typing.
8
9

10 22 Data sources: Medline, Embase, CINHAL, Web of Science and Scopus were searched for articles
11 23 published before November 2012
12

13 24 Review methods: Studies in humans that reported the proportion of clustering of TB isolates by
14 25 MIRU-VNTR were included in the analysis. Univariable meta-regression analyses were conducted to
15 26 assess the influence of study design and setting on the proportion of clustering.
16
17

18 27 Results: The search identified 14 eligible articles reporting clustering between 22.1% and 61.2%. The
19 28 proportion of culture positive isolates and the number of MIRU-VNTR loci typed explained 49% and
20 29 34% of the between study variation, respectively, and had a significant association with the
21 30 proportion of clustering.
22

23 31 Conclusions: Although MIRU-VNTR typing is being adopted worldwide there is a paucity of data on
24 32 how study design and setting may influence estimates of clustering. We have highlighted study
25 33 design variables for consideration in the design and interpretation of future studies.
26
27

28 34

29 35 **Strengths and Limitations of Study**
30
31

- 32 36
- 33 37 • This is a timely evaluation of the impact of study design on estimates of TB clustering using
34 38 MIRU-VNTR strain typing because it has been incorporated into national typing services
35 39 globally.
36 40
 - 37 39 • There were insufficient data available to fully explore the impact of study design and setting
38 40 on estimates of clustering.
39 41
- 40 42

41 41

42 42

43 Introduction

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

50 TB molecular typing methods include Spoligotyping [6], insertion sequence 6110 (IS6110) restriction
51 fragment length polymorphism (RFLP) analysis (the recent gold standard) [7], mycobacterial
52 interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing [8], and whole
53 genome sequencing [9–11]. Published reviews have identified factors that might influence or bias
54 clustering by IS6110 RFLP [12,13]. No study has repeated this analysis using more up-to-date typing
55 methods, which is important for understanding of the epidemiology of TB and to shape the
56 application of molecular typing to improve TB control.

57 Published meta-analyses and modelling studies using IS6110 RFLP data show that the proportion of
58 clustering observed can be affected by 1) study design (affecting the proportion of eligible cases that
59 are included in the study); 2) features of the typing method (such as the ability to type isolates with
60 low copy numbers); and 3) study setting (such as characteristics of the study population). For
61 example, the proportion of clustering increases when the fraction of the total data sampled
62 increases [13–15] and when study duration increases [16].

63 MIRU-VNTR is currently the preferred method of molecular typing [17–21], and can be used
64 together with Spoligotyping [8]. Relative to IS6110 RFLP, MIRU-VNTR does not have to exclude
65 isolates with a low IS6110 copy number, has a faster turnaround time, is high throughput and the
66 numeric strain types are more easily compared. MIRU-VNTR strain typing is increasingly being
67 adopted worldwide [1,22–27], yet unlike IS6110 RFLP, the evidence for the interpretation of the
68 findings such as the impact of study design and setting on clustering have not been reviewed.
69 Although the two typing methods have been shown to have a similar discriminatory value, the
70 markers evolve independently and at different rates, resulting in a difference in clustering between
71 the two methods [28]. This suggests that there could be differences in the way study design, typing
72 method and setting affects clustering by the two methods. We conducted a systematic review to
73 assess the evidence for the impact of study design and setting on the interpretation of TB

1
2
3 74 transmission using clustering derived from MIRU-VNTR strain typing – as has been shown using
4 75 IS6110 RFLP typing.

6
7 **Methods**

8
9
10 77 Five electronic databases were searched (EMBASE, ISI Web of Science, CINHAI, Scopus and Medline
11 78 (Ovid)) up to 1 November 2012. The search strategy combined the following terms with Boolean
12 79 operators: Tuberculosis, strain typing, and transmission. The search was limited to studies using the
13 80 standard MIRU-VNTR method [8], in humans only, and in English.

14
15
16
17 81 All titles and abstracts from each of the searches were examined. The full text of each paper was
18 82 obtained and reviewed if the study reported MIRU-VNTR strain typing of *M.tuberculosis* complex
19 83 isolates with at least 15 of the standardised 24 loci [8,29,30].

20
21
22 84 Studies using fewer than 15 loci were not included because the level of discrimination is inadequate
23 85 for epidemiological use (n=97) [8]. Studies that used loci different to the standardised 15 and 24 set
24 86 were not included in the analysis in order to reduce the heterogeneity between studies (n=11). All
25 87 publication types were included in this first screen to ensure that no relevant data were missed.

26
27
28
29 88 Reviews, letters, editorials, outbreaks or case reports (n=99) were excluded in the second screen.
30 89 Studies that used incomplete sampling (e.g. random samples, studies using subsets of populations
31 90 such as MDR patients) (n=30) and studies that had a sample size of less than 50 (n=2) were also
32 91 excluded.

33
34
35
36
37 92 A reviewer extracted the following data items from all included studies using a form developed in
38 93 Excel (Microsoft 2010): publication details (year, authors, study country), study details (study
39 94 duration, loci typed, secondary typing method, study population), the proportion of total TB isolates
40 95 clustered by MIRU-VNTR strain typing, and the covariates of interest: the number of clustered and
41 96 unique isolates; the maximum size of clusters; the proportion of clusters containing two cases; the
42 97 prevalence of culture-positivity among TB patients included in the study; the proportion of culture
43 98 positive isolates typed; risk factors for clustering; and the Hunter Gaston Discriminatory Index (HGDI)
44 99 [31]).

45
46
47
48
49
50
51 100 Authors were contacted if TB incidence rate was not reported. Where no response was received
52 101 WHO country estimates of TB incidence for the study year were used [32].

53
54
55 102 Data were analysed in Stata 12. Where studies reported data from more than one set of loci, the
56 103 method with the highest discriminatory value was included (i.e. MIRU-VNTR 24 would be chosen

1
2
3 104 over MIRU-VNTR 15, and MIRU-VNTR 15 plus Spoligotyping would be chosen over MIRU-VNTR 15
4 105 alone) (n=5). This review was not concerned with summary measures of clustering, but factors that
5 106 influenced clustering; therefore articles must have included at least one of the covariates.
6 107 Continuous variables were transformed where the distribution was skewed. The proportion
7 108 clustered was transformed using the Freeman Tukey transformation [33]. Univariable meta-
8 109 regression analyses were carried out to determine the effect of the study design covariates on the
9 110 proportion of clustered isolates. All covariates in the analysis were hypothesised to influence the
10 111 proportion clustered *a priori*.

112 Results

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

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

125 Discussion

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

130 As expected, we found that the proportion of clustering decreased with a greater number of MIRU-
131 VNTR loci typed. Our finding that the prevalence of culture-positivity among TB patients included in
132 the study influences the estimates of transmission within a population is counterintuitive and not
133 consistent with estimates of the influence of sampling on the proportion of clustering using *IS6110*
134 RFLP typing [36]. This may reflect the relationship between TB burden and resource poor/rich

1
2
3 135 settings and the consequent availability of culture diagnostic laboratory services; i.e. in resource
4 136 poor settings where there is a high burden of TB (and, therefore, high rates of clustering) the
5 137 prevalence of culture positive TB cases is low. The finding may also be due to chance, with only 8
6 138 studies included in the analysis of this variable.

9
10 139 The other study design variables included in this analysis, such as study duration, did not significantly
11 140 influence the proportion of isolates that were clustered, contrary to previous findings [12]. This is
12 141 likely to be because of a lack of good quality evidence: only 14 studies met the inclusion criteria for
13 142 the review and of those only three reported all the variables of interest, reducing the power of the
14 143 analysis and precluding multivariable meta-regression. In addition, the range of the variables may
15 144 have been too limited to show any impact on clustering estimates. For example, the proportion of
16 145 culture positive isolates typed had a narrow range from 81.9% to 100%. Furthermore, most of the
17 146 studies were from low TB burden settings and therefore may be reflecting the rate at which
18 147 imported cases have matching strain types by chance, rather than rates of recent transmission.

19
20 148 This study is a timely evaluation of the impact of study design on estimates of TB clustering using
21 149 MIRU-VNTR strain typing because it has been incorporated into national typing services globally
22 150 [23,37]. The findings are relevant where strain typing is used to evaluate TB control systems across
23 151 different settings because the proportion of clustering is influenced by the prevalence of culture
24 152 positive TB cases in the study setting. Given that strain typing methods are advancing beyond MIRU-
25 153 VNTR typing and that the application of whole genome sequencing to TB control and public health
26 154 strategies has been demonstrated [9–11,38], it is important that the biases in the analysis of such
27 155 methods are explored and compared. Understanding how to design and compare research studies
28 156 for public health will greatly improve the benefit gained from newer technologies.

29
30 157 This review has highlighted the need for better quality reporting in primary studies to enable future
31 158 reviews to be more robust. A lack of standards for the molecular epidemiology of infectious diseases
32 159 may explain the poor quality of reporting; this field would benefit from the introduction of such
33 160 standards (STROBE-ID, submitted).

34
35 161 The use of TB strain typing as a public health tool in TB control programmes is increasing globally.
36 162 We have identified a lack of good quality studies that can contribute to our understanding in
37 163 interpreting the molecular typing of TB. We have also shown that the proportion of clustering
38 164 derived from MIRU-VNTR typing is influenced by the number of loci typed and the prevalence of
39 165 culture-positivity among TB patients included in the study, highlighting these as important
40 166 considerations in the design and interpretation of future studies.

1
2
3 167 **Conflict of interest**
4

5 168 Nothing to declare.
6
7

8 169 **Acknowledgements**
9

10 170 We would like to acknowledge Ross Harris from the Statistics Unit at Public Health England for his
11
12 171 advice on meta-regression.
13

14 172 **Author contributions**
15

16
17 173 All authors made substantial contributions to the conception and design of the review, and the
18
19 174 analysis and interpretation of data. JM drafted the article and PS, IA, TM and TC revised it critically
20
21 175 for important intellectual content. All authors approved the final version for publication.
22

23 176 **Funding**
24

25 177 JM is funded through a Public Health England and University College London Impact Studentship. IA
26
27 178 is funded through a NIHR Senior Research Fellowship.
28

29 179 **Ethics**
30

31 180 Ethical approval was not required as this review analyses data that is in the public domain.
32

33 181 **Data sharing**
34

35 182 No additional data are available
36
37

38 183 **References**
39

- 40 184 1. Lambregts-van Weezenbeek CSB, Sebek MMGG, Van Gerven PJHJ, De Vries G, Verver S, et al.
41 185 (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands:
42 186 6 years' experience with nation-wide cluster feedback and cluster monitoring. *Int J Tuberc Lung*
43 187 *Dis* 7: S463–470.
44
45 188 2. Borgdorff MW, Van den Hof S, Kremer K, Verhagen L, Kalisvaart N, et al. (2010) Progress
46 189 towards tuberculosis elimination: secular trend, immigration and transmission. *Eur Respir J* 36:
47 190 339–347. doi:10.1183/09031936.00155409.
48
49 191 3. Kik SV, Verver S, Van Soolingen D, De Haas PEW, Cobelens FG, et al. (2008) Tuberculosis
50 192 Outbreaks Predicted by Characteristics of First Patients in a DNA Fingerprint Cluster. *Am J*
51 193 *Respir Crit Care Med* 178: 96–104. doi:10.1164/rccm.200708-1256OC.
52
53 194 4. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain
54 195 typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology

- 1
2
3 196 laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J*
4 197 *Clin Microbiol* 31: 1677–1682.
- 5
6 198 5. De Vries G, Van Hest RAH, Richardus JH (2007) Impact of mobile radiographic screening on
7 199 tuberculosis among drug users and homeless persons. *Am J Respir Crit Care Med* 176: 201–
8 200 207. doi:10.1164/rccm.200612-1877OC.
- 9
10 201 6. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, et al. (1997) Simultaneous
11 202 detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and
12 203 epidemiology. *J Clin Microbiol* 35: 907–914.
- 13
14 204 7. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification
15 205 of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized
16 206 methodology. *J Clin Microbiol* 31: 406–409.
- 17
18 207 8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsç-Gerdes S, et al. (2006) Proposal for
19 208 standardization of optimized mycobacterial interspersed repetitive unit-variable-number
20 209 tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 44: 4498–4510.
21 210 doi:10.1128/JCM.01392-06.
- 22
23 211 9. Schürch AC, Van Soolingen D (2011) DNA fingerprinting of *Mycobacterium tuberculosis*: From
24 212 phage typing to whole-genome sequencing. *Infect Genet Evol*. Available:
25 213 <http://www.ncbi.nlm.nih.gov/pubmed/22067515>. Accessed 13 March 2012.
- 26
27 214 10. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and
28 215 social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730–739.
29 216 doi:10.1056/NEJMoa1003176.
- 30
31 217 11. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. (2013) Whole-genome sequencing to
32 218 delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet*
33 219 *Infect Dis* 13: 137–146. doi:10.1016/S1473-3099(12)70277-3.
- 34
35 220 12. Houben RMGJ, Glynn JR (2009) A systematic review and meta-analysis of molecular
36 221 epidemiological studies of tuberculosis: development of a new tool to aid interpretation.
37 222 *Tropical Medicine & International Health* 14: 892–909. doi:10.1111/j.1365-3156.2009.02316.x.
- 38
39 223 13. Fok A, Numata Y, Schulzer M, FitzGerald MJ (May) Risk factors for clustering of tuberculosis
40 224 cases: a systematic review of population-based molecular epidemiology studies [Review
41 225 Article]. *The International Journal of Tuberculosis and Lung Disease* 12: 480–492.
- 42
43 226 14. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, Van Soolingen D (2011) Influence of
44 227 Sampling on Clustering and Associations With Risk Factors in the Molecular Epidemiology of
45 228 Tuberculosis. *Am J Epidemiol*. Available:
46 229 <http://aje.oxfordjournals.org/content/early/2011/05/23/aje.kwr061>. Accessed 29 March 2012.
- 47
48 230 15. Glynn JR, Bauer J, De Boer AS, Borgdorff MW, Fine PE, et al. (1999) Interpreting DNA fingerprint
49 231 clusters of *Mycobacterium tuberculosis*. *European Concerted Action on Molecular*
50 232 *Epidemiology and Control of Tuberculosis*. *Int J Tuberc Lung Dis* 3: 1055–1060.
- 51
52 233 16. Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD, et al. (2005) The Importance of
53 234 Recent Infection with *Mycobacterium tuberculosis* in an Area with High HIV Prevalence: A

- 1
2
3 235 Long-Term Molecular Epidemiological Study in Northern Malawi. *J Infect Dis* 192: 480–487.
4 236 doi:10.1086/431517.
- 5
6 237 17. De Beer JL, Kremer K, Ködmön C, Supply P, Van Soolingen D (2012) First Worldwide Proficiency
7 238 Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex
8 239 Strains. *Journal of Clinical Microbiology* 50: 662–669. doi:10.1128/JCM.00607-11.
- 9
10 240 18. Maes M, Kremer K, Van Soolingen D, Takiff H, De Waard JH (2008) 24-locus MIRU-VNTR
11 241 genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao
12 242 Amerindians in Venezuela. *Tuberculosis (Edinb)* 88: 490–494. doi:10.1016/j.tube.2008.04.003.
- 13
14 243 19. Sougakoff W (2011) Molecular epidemiology of multidrug-resistant strains of Mycobacterium
15 244 tuberculosis. *Clinical Microbiology and Infection* 17: 800–805. doi:10.1111/j.1469-
16 245 0691.2011.03577.x.
- 17
18 246 20. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for
19 247 polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. *Nucleic Acids Res* 38:
20 248 W326–331. doi:10.1093/nar/gkq351.
- 21
22 249 21. Supply P (2010) MIRU-VNTR typing: the new international standard for TB molecular
23 250 epidemiology Symposium of the Institut Pasteur de Tunisia.
- 24
25 251 22. Van Soolingen D, Borgdorff MW, De Haas PE, Sebek MM, Veen J, et al. (1999) Molecular
26 252 epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997.
27 253 *J Infect Dis* 180: 726–736. doi:10.1086/314930.
- 28
29 254 23. Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step
30 255 approach for large-scale, prospective genotyping of Mycobacterium tuberculosis isolates in the
31 256 United States. *J Clin Microbiol* 43: 688–695. doi:10.1128/JCM.43.2.688-695.2005.
- 32
33 257 24. New CDC Program for Rapid Genotyping of Mycobacterium Tuberculosis Isolates (2005). *JAMA*
34 258 293: 2086–2086. doi:10.1001/jama.293.17.2086.
- 35
36 259 25. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, et al. (2000) A prospective evaluation
37 260 of the clinical value of nation-wide DNA fingerprinting of tuberculosis isolates in Denmark. *Int J*
38 261 *Tuberc Lung Dis* 4: 295–299.
- 39
40 262 26. Bauer J, Yang Z, Poulsen S, Andersen AB (1998) Results from 5 years of nationwide DNA
41 263 fingerprinting of Mycobacterium tuberculosis complex isolates in a country with a low
42 264 incidence of M. tuberculosis infection. *J Clin Microbiol* 36: 305–308.
- 43
44 265 27. Zolnir-Dovc M, Poljak M, Erzen D, Sorli J (2003) Molecular epidemiology of tuberculosis in
45 266 Slovenia: results of a one-year (2001) nation-wide study. *Scand J Infect Dis* 35: 863–868.
- 46
47 267 28. Hanekom M, Van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Hoek KGP, et al. (2008)
48 268 Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-
49 269 repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of
50 270 Mycobacterium tuberculosis Beijing strains in a setting of high incidence of tuberculosis. *J Clin*
51 271 *Microbiol* 46: 3338–3345. doi:10.1128/JCM.00770-08.
- 52
53 272 29. Supply P, Lesjean S, Savine E, Kremer K, Van Soolingen D, et al. (2001) Automated high-
54 273 throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based

- 1
2
3 274 on mycobacterial interspersed repetitive units. *J Clin Microbiol* 39: 3563–3571.
4 275 doi:10.1128/JCM.39.10.3563-3571.2001.
- 5
6 276 30. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniewski FA (2006) Progression toward an
7 277 improved DNA amplification-based typing technique in the study of *Mycobacterium*
8 278 tuberculosis epidemiology. *J Clin Microbiol* 44: 2492–2498. doi:10.1128/JCM.01428-05.
- 9
10 279 31. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems:
11 280 an application of Simpson's index of diversity. *J Clin Microbiol* 26: 2465–2466.
- 12
13 281 32. WHO | TB data (n.d.). WHO. Available: <http://www.who.int/tb/country/en/index.html>.
14 282 Accessed 12 December 2012.
- 15
16 283 33. Freeman MF, Tukey JW (1950) Transformations Related to the Angular and the Square Root.
17 284 *Ann Math Statist* 21: 607–611. doi:10.1214/aoms/1177729756.
- 18
19 285 34. Oelemann MC, Diel R, Vatin V, Haas W, Rüscher-Gerdes S, et al. (2007) Assessment of an
20 286 optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing
21 287 system combined with spoligotyping for population-based molecular epidemiology studies of
22 288 tuberculosis. *J Clin Microbiol* 45: 691–697. doi:10.1128/JCM.01393-06.
- 23
24 289 35. Evans J (2010) Analysis of prevalent *Mycobacterium tuberculosis* strains in the United
25 290 Kingdom: detection, distribution and expansion of MIRU-VNTR profiles containing high
26 291 numbers of isolates. European Society of Clinical Microbiology and Infectious Diseases. Vienna,
27 292 Austria.
- 28
29 293 36. Glynn JR, Vyonycky E, Fine PEM (1999) Influence of Sampling on Estimates of Clustering and
30 294 Recent Transmission of *Mycobacterium tuberculosis* Derived from DNA Fingerprinting
31 295 Techniques. *American Journal of Epidemiology* 149: 366–371.
- 32
33 296 37. TB Strain Typing Project Board HPA (2011) TB Strain Typing Cluster Investigation Handbook for
34 297 Health Protection Units. Available:
35 298 <https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealth>
36 299 [ProtectionProgrammes/Respiratory/TB/StrainTyping/](https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealth). Accessed 30 November 2011.
- 37
38 300 38. Walker TM, Monk P, Grace Smith E, Peto TEA (2013) Contact investigations for outbreaks of
39 301 *Mycobacterium tuberculosis*: advances through whole genome sequencing. *Clin Microbiol*
40 302 *Infect*. doi:10.1111/1469-0691.12183.
- 41
42 303 39. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR (2011) Tuberculosis transmission in
43 304 Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. *Infect Genet Evol*
44 305 11: 124–131. doi:10.1016/j.meegid.2010.09.013.
- 45
46 306 40. Allix-Béguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008) Standardised PCR-based
47 307 molecular epidemiology of tuberculosis. *Eur Respir J* 31: 1077–1084.
48 308 doi:10.1183/09031936.00053307.
- 49
50 309 41. Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of
51 310 standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat
52 311 typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 46: 1398–1406.
53 312 doi:10.1128/JCM.02089-07.

- 1
2
3 313 42. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, et al. (2011) Evaluation of Mycobacterium
4 314 tuberculosis typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. *J Clin*
5 315 *Microbiol* 49: 4173–4178. doi:10.1128/JCM.05293-11.
- 7 316 43. Ojo OO, Sheehan S, Corcoran DG, Nikolayevsky V, Brown T, et al. (2010) Molecular
8 317 epidemiology of Mycobacterium tuberculosis clinical isolates in Southwest Ireland. *Infect*
9 318 *Genet Evol* 10: 1110–1116. doi:10.1016/j.meegid.2010.07.008.
- 11 319 44. Dymova MA, Liashenko OO, Poteiko PI, Krutko VS, Khrapov EA, et al. (2011) Genetic variation
12 320 of Mycobacterium tuberculosis circulating in Kharkiv Oblast, Ukraine. *BMC Infect Dis* 11: 77.
13 321 doi:10.1186/1471-2334-11-77.
- 15 322 45. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-
16 323 locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. *Respir Med*
17 324 105 Suppl 1: S67–73. doi:10.1016/S0954-6111(11)70014-2.
- 19 325 46. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, et al. (2009)
20 326 Prospective universal application of mycobacterial interspersed repetitive-unit-variable-
21 327 number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for
22 328 fast identification of clustered and orphan cases. *J Clin Microbiol* 47: 2026–2032.
23 329 doi:10.1128/JCM.02308-08.
- 26 330 47. Hamblion EL, Wynne-Edwards E, Anderson C, Anderson SR (2011) A summary of strain typing
27 331 and clustering of TB in London in 2010 and an analysis of the associated risk factors. *Thorax* 66:
28 332 A88–A89. doi:10.1136/thoraxjnl-2011-201054c.50.
- 30 333 48. Mandal S, Bradshaw L, Anderson LF, Brown T, Evans JT, et al. (2011) Investigating Transmission
31 334 of Mycobacterium Bovis in the UK, 2005-2008. *J Clin Microbiol*. Available:
32 335 <http://jcm.asm.org/content/early/2011/03/23/JCM.02299-10>. Accessed 24 April 2012.
- 34 336 49. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, et al. (2011) Molecular epidemiology of
35 337 Mycobacterium tuberculosis in East Lancashire 2001-2009. *Thorax* 66: 709–713.
36 338 doi:10.1136/thx.2011.158881.
- 38 339 50. Nikolayevskyy VV, Brown TJ, Bazhora YI, Asmolov AA, Balabanova YM, et al. (2007) Molecular
39 340 epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in
40 341 Mycobacterium tuberculosis strains from the southern Ukraine. *Clin Microbiol Infect* 13: 129–
41 342 138. doi:10.1111/j.1469-0691.2006.01583.x.
- 44 343

Tables

Table 1: Studies included in the analysis

Reference	Author	Country	Study site ^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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

^co= 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)

^destimates from the literature of the prevalence of TB/HIV co-infection reported in the study area

For peer review only

Table 2: Univariable meta-regression 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 [‡]
<u>Study design</u>					
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
<u>Study setting</u>					
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

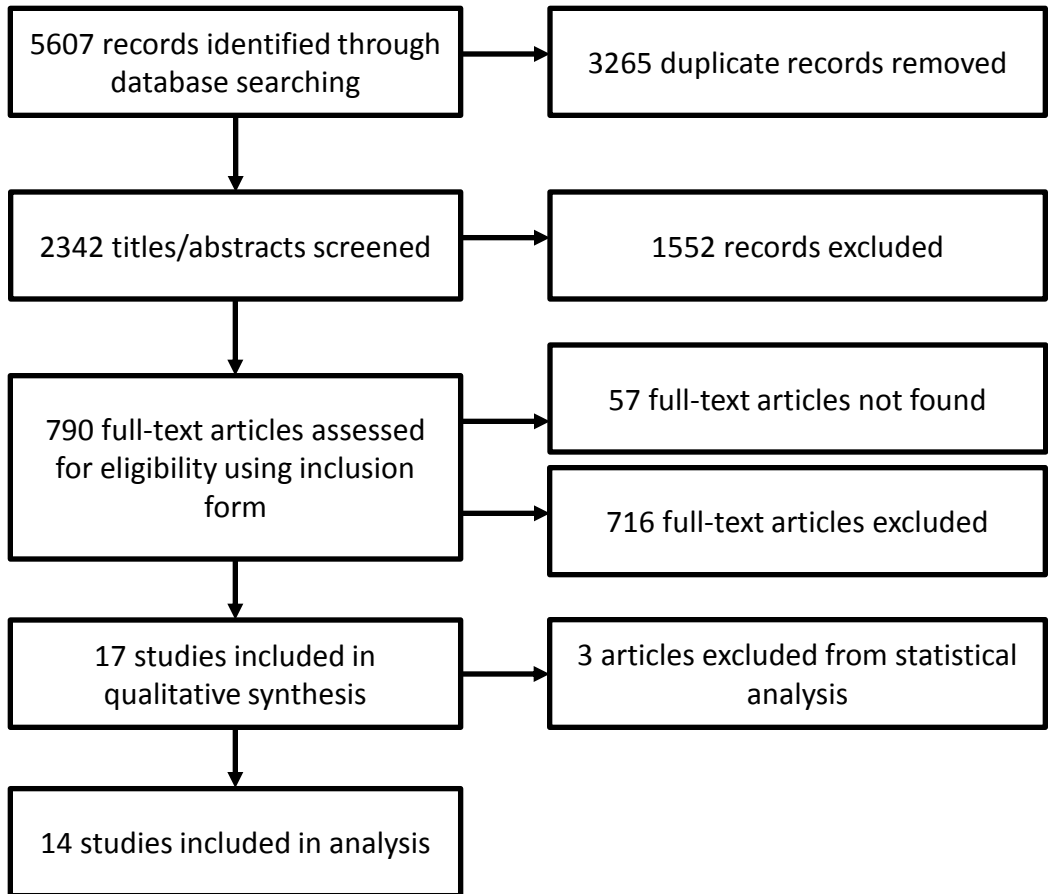
^π Coefficients for the change in the proportion of clustering for each covariate. E.g. for a one-month increase in study duration, the proportion of clustering increases by 0.003.

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

Figure Caption

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58



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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 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

For peer review only



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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

4 Jessica Mears¹, Ibrahim Abubakar^{1,2,3}, Theodore Cohen⁴, Timothy D McHugh⁵ & Pam Sonnenberg^{1,*}

5 ¹ Department of Infection and Population Health, University College London

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

7 ³ Clinical Trials Unit, Medical Research Council, London

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

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

11 *Corresponding author:

12 Dr Pam Sonnenberg

13 3rd Floor Mortimer Market Centre

14 University College London

15 London WC1E 6JB

16 p.sonnenberg@ucl.ac.uk

17

18 **Word count: 2439**

1
2
3 19 **Abstract**
4

5 20 Objectives: To systematically review the evidence for the impact of study design and setting on the
6
7 21 interpretation of TB transmission using clustering derived from Mycobacterial Interspersed
8
9 22 Repetitive Units – Variable Number Tandem Repeats (MIRU-VNTR) strain typing.

10
11 23 Data sources: Medline, Embase, CINHAI, Web of Science and Scopus were searched for articles
12
13 24 published before 21st October 2014.

14
15 25 Review methods: Studies in humans that reported the proportion of clustering of TB isolates by
16
17 26 MIRU-VNTR were included in the analysis. Univariable meta-regression analyses were conducted to
18
19 27 assess the influence of study design and setting on the proportion of clustering.

20
21 28 Results: The search identified 27 eligible articles reporting clustering between 0% and 63%. The
22
23 29 number of MIRU-VNTR loci typed, requiring consent to type patient isolates (as a proxy for sampling
24
25 30 fraction), the TB incidence and the maximum cluster size explained 14%, 14%, 27% and 48%,
26
27 31 respectively, and had a significant association with the proportion of clustering .

28
29 32 Conclusions: Although MIRU-VNTR typing is being adopted worldwide there is a paucity of data on
30
31 33 how study design and setting may influence estimates of clustering. We have highlighted study
32
33 34 design variables for consideration in the design and interpretation of future studies.
34
35

36 **Strengths and Limitations of Study**

- 37
38 37 • This is a timely evaluation of the impact of study design on estimates of TB clustering using
39
40 38 MIRU-VNTR strain typing because it has been incorporated into national typing services
41
42 39 globally.
43
44 40 • The strength of this meta-analysis was limited by the lack of detail reported by the included
45
46 41 studies, highlighting the need for better quality reporting in primary studies.
47
48 42
49
50
51
52
53
54
55
56
57
58
59
60

44 Introduction

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

51 TB molecular typing methods include Spoligotyping [6], insertion sequence *6110* (*IS6110*) restriction
52 fragment length polymorphism (RFLP) analysis (the recent gold standard) [7], mycobacterial
53 interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing [8], and whole
54 genome sequencing [9–11]. Published reviews have identified factors that might influence or bias
55 clustering by *IS6110* RFLP [12,13]. No study has repeated this analysis using more up-to-date typing
56 methods, which is important for understanding of the epidemiology of TB and to shape the
57 application of molecular typing to improve TB control.

58 Published meta-analyses and modelling studies using *IS6110* RFLP data show that the proportion of
59 clustering observed can be affected by 1) study design (affecting the proportion of eligible cases that
60 are included in the study); 2) features of the typing method (such as the ability to type isolates with
61 low copy numbers); and 3) study setting (such as characteristics of the study population). For
62 example, the proportion of clustering increases when the fraction of the total data sampled
63 increases [13–15] and when study duration increases [16].

64 MIRU-VNTR is currently the preferred method of molecular typing [17–21], and can be used
65 together with Spoligotyping [8]. Relative to *IS6110* RFLP, MIRU-VNTR does not have to exclude
66 isolates with a low *IS6110* copy number, has a faster turnaround time, is high throughput and the
67 numeric strain types are more easily compared. MIRU-VNTR strain typing is increasingly being
68 adopted worldwide [1,22–27], yet unlike *IS6110* RFLP, the evidence for the interpretation of the
69 findings such as the impact of study design and setting on clustering have not been reviewed.
70 Although the two typing methods have been shown to have a similar discriminatory value, the
71 markers evolve independently and at different rates, resulting in a difference in clustering between
72 the two methods [28]. This suggests that there could be differences in the way study design, typing
73 method and setting affects clustering by the two methods. We conducted a systematic review to
74 assess the evidence for the impact of study design and setting on the interpretation of TB

1
2
3 75 transmission using clustering derived from MIRU-VNTR strain typing – as has been shown using
4 76 IS6110 RFLP typing.

77 **Methods**

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

82 All titles and abstracts from each of the searches were examined. The full text of each paper was
83 obtained and reviewed if the study reported MIRU-VNTR strain typing of *M.tuberculosis* complex
84 isolates with at least 15 of the standardised 24 loci (ETR A, B, C, D, E; MIRU 2, 10, 16, 20, 23, 24, 26,
85 27, 39, 40; VNTR 424, 1955, 2163b, 2347, 2401, 3171, 3690, 4052, 4156) [8,29,30].

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

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

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

1
2
3 104 The main outcome measure – the proportion of TB isolates clustered by MIRU-VNTR strain typing –
4 105 was calculated as the number of clustered isolates/number of clustered+unique isolates. Where
5
6 106 there were uncertainties JM consulted with IA
7

8
9 107 Authors were contacted if TB incidence rate was not reported. Where no response was received
10 108 WHO country estimates of TB incidence for the study year were used [32]. As so few studies
11 109 reported the proportion coinfecting with TB/HIV, these estimates for the study country were taken
12 110 from an EU-wide survey and WHO country profiles.[33,34] Due to poor recording of the sampling
13 111 fraction (the number of isolates typed/ the total number of culture positive TB cases diagnosed
14 112 during the study period (n=19)), whether the study required the consent of participants (yes/no) was
15 113 included as a proxy for (high/low) sampling fraction. The risk of bias within each study was assessed
16 114 using the STROME-ID checklist. [35]
17
18
19
20
21

22 115 Data were analysed in Stata 12. Where studies reported data from more than one set of loci, the
23 116 method with the highest discriminatory value was included (i.e. MIRU-VNTR 24 would be chosen
24 117 over MIRU-VNTR 15, and MIRU-VNTR 15 plus Spoligotyping would be chosen over MIRU-VNTR 15
25 118 alone) (n=8). This review was not concerned with summary measures of clustering, but factors that
26 119 influenced clustering; therefore articles must have included at least one of the covariates.
27 120 Continuous variables were transformed where the distribution was skewed. The proportion
28 121 clustered was transformed using the Freeman Tukey transformation [36]. Study heterogeneity was
29 122 assessed using a forest plot and the χ^2 test of heterogeneity. Univariable meta-regression analyses
30 123 were carried out to determine the effect of the study design covariates on the proportion of
31 124 clustered isolates. All covariates in the analysis were hypothesised to influence the proportion
32 125 clustered *a priori*.
33
34
35
36
37
38
39

40
41 126 Sensitivity analyses were conducted to see the effect of removing studies reporting 0% clustering,
42 127 with only extra-pulmonary TB cases, only *M.bovis* cases, studies using the 'old 12' MIRU loci as part
43 128 of their 15 loci, and studies assessed as having a high likelihood of bias (STROME-ID score less than
44 129 20).
45
46
47

48 130 **Results**

49
50
51 131 The search identified 7274 references resulting in 27 studies (25 journal articles and 2 conference
52 132 abstracts) included after deduplication and title/abstract/full text screening (Figure 1). The main
53 133 characteristics of the included studies are shown in Table 1. Studies were published between 2007
54 134 and 2014 and the clustering reported varied from 0% [37] to 62.8% [38]. In all studies, clustered
55
56
57
58
59
60

1
2
3 135 isolates were defined as having identical strain types based on the MIRU-VNTR loci typed, with or
4 136 without Spoligotyping. 17 studies included isolates from newly diagnosed TB cases, three studies
5 137 reported including isolates from new and chronic cases of TB, and seven did not report this
6 138 information. In addition, ten studies did not include repeat isolates from the same patient, one study
7 139 included a repeat isolate from one patient, and the remaining 17 did not report whether repeat
8 140 isolates were included or not. Furthermore, four studies included isolates with missing loci in the
9 141 cluster analysis, whereas four excluded isolates with missing loci, and the remaining 20 did not
10 142 report how they dealt with missing loci. The number of studies reporting each variable of interest is
11 143 shown in Table 2.

12
13
14
15
16
17
18 144 A forest plot shows the spread of clustering reported by number of loci and additional typing
19 145 method (Figure 2). Significant heterogeneity was identified between the studies ($p < 0.001$),
20 146 suggesting that a meta-regression would be an appropriate analysis.

21
22
23
24 147 The univariable meta-regression shows evidence for the proportion of clustering to decrease as the
25 148 number of MIRU-VNTR loci typed increased from 15 to 24 ($p = 0.04$; Table 3), accounting for 14% of
26 149 the between study variation, and to increase when the study participants consented to being
27 150 included in the study ($p = 0.03$), accounting for 14% of the between study variation. The proportion of
28 151 clustering increased as the TB incidence in the population increased ($p = 0.007$, $\text{Adj } R^2 = 26.7$). There
29 152 was also evidence for the proportion of clustering to increase as the maximum cluster size increased
30 153 ($p = 0.001$), accounting for 48% of between study variation. There was no evidence of the other study
31 154 design or study setting variables significantly influencing the proportion clustered. Though non-
32 155 significant ($p > 0.05$), the TB/HIV coinfection rate in the population explained 2% of the between
33 156 study variation. Too few studies included information on the proportion of clusters containing two
34 157 cases, proportion of the study sample with previous TB or with pulmonary TB, and the proportion of
35 158 the population with culture positive TB, so these could not be included in the analysis (Table 2).

36
37
38
39
40
41
42
43
44 159 Sensitivity analyses to examine the effect of excluding studies reporting 0% clustering,[37] only
45 160 *M.bovis* cases,[39] studies using the 'old 12' MIRU loci,[39–44] and studies assessed as having a high
46 161 risk of bias,[37,45–48] did not generally change the results. The proportion of culture positive TB in
47 162 the population remained insignificant but explained 2.6% of the between study variation when
48 163 excluding 0% clustering ($p = 0.278$ and $\text{Adj } R^2 = 2.62$). Similarly, the proportion of culture positive TB in
49 164 the population remained insignificant but explained 2.6% of the between study variation when
50 165 excluding studies with the highest risk of bias ($p = 0.278$ and $\text{Adj } R^2 = 2.62$). The number of loci typed
51 166 became non-significant, but explained 9.6% and 10.5% of the between study variation when
52
53
54
55
56
57
58
59
60

1
2
3 167 excluding studies using the 'old 12' loci and the highest risk of bias, respectively ($p=0.106$, Adj
4 168 $R^2=9.63$; $p=0.111$, Adj $R^2=10.51$, respectively).

6
7 169 **Discussion**

8
9 170 This review identified 27 studies that met the inclusion criteria. We illustrate that the interpretation
10 171 of studies using MIRU-VNTR to estimate clustering is subject to bias relating to study design and
11 172 setting; however, there were insufficient data available to fully explore this impact.

12
13 173 As expected, we found that the proportion of clustering decreased with a greater number of MIRU-
14 174 VNTR loci typed, with increasing TB incidence and with increasing maximum cluster size. We found
15 175 that requiring consent to type patient isolates reduced the proportion of clustering, which is
16 176 expected, given that the sampling fraction would be lower in these studies.

17 177 The other study design variables included in this analysis, such as study duration, did not significantly
18 178 influence the proportion of isolates that were clustered, contrary to previous findings [12]. This is
19 179 likely to be because of a lack of good quality evidence: of the 27 studies that met the inclusion
20 180 criteria for the review, none reported all the variables of interest, reducing the power of the analysis
21 181 and precluding multivariable meta-regression (Table 2). Importantly, key details of cluster analyses
22 182 were not reported consistently across the studies, such as whether repeat isolates from the same
23 183 patients were included, or typing profiles with missing loci were included, introducing new,
24 184 unmeasured biases. In addition, the range of the variables may have been too limited to show any
25 185 impact on clustering estimates. For example, the proportion of culture positive isolates typed ranged
26 186 from 34.5% to 100%, with 17 of the 19 studies reporting this variable from 81.9% to 100%.
27 187 Furthermore, most of the studies (17/27=63%) were from low TB burden settings and therefore may
28 188 be reflecting the rate at which imported cases have matching strain types by chance, rather than
29 189 rates of recent transmission.

30 190 The sensitivity analysis suggested that, when excluding the studies with the greatest risk of bias, the
31 191 culture-positivity in the population might explain a small amount of the between study variation.
32 192 This is counterintuitive and not consistent with estimates of the influence of sampling on the
33 193 proportion of clustering using *IS6110* RFLP typing [49]. This may reflect the relationship between TB
34 194 burden and resource poor/rich settings and the consequent availability of culture diagnostic
35 195 laboratory services; i.e. in resource poor settings where there is a high burden of TB (and, therefore,
36 196 high rates of clustering) the prevalence of culture positive TB cases is low. The finding may also be
37 197 due to chance, with only 14 studies included in the analysis of this variable. In the sensitivity analysis

1
2
3 198 excluding studies that used the 'old 12' loci, the effect of the number of loci typed becomes non-
4 199 significant. This is likely because studies using the 'old 12' accounted for six out of ten studies
5
6 200 reporting 15 loci, reducing the number of studies and the power of the model.
7

8
9 201 This study is a timely evaluation of the impact of study design on estimates of TB clustering using
10 202 MIRU-VNTR strain typing because it has been incorporated into national typing services globally
11 203 [23,50]. The findings are relevant where strain typing is used to evaluate TB control systems across
12 204 different settings because the proportion of clustering is influenced by the number of loci typed, the
13 205 TB incidence and the maximum cluster size. Given that strain typing methods are advancing beyond
14 206 MIRU-VNTR typing and that the application of whole genome sequencing to TB control and public
15 207 health strategies has been demonstrated [9–11,51], it is important that the biases in the analysis of
16 208 such methods are explored and compared. Understanding how to design and compare research
17 209 studies for public health will greatly improve the benefit gained from newer technologies.
18
19

20 210 The strength of this meta-analysis was limited by (a lack of) detail reported by the included studies.
21 211 This review has highlighted the need for better quality reporting in primary studies to enable future
22 212 reviews to be more robust. Recently published standards for reporting of molecular epidemiology
23 213 for infectious diseases should improve the quality of reporting.[35] This review is further limited by
24 214 our inability to access 58 of the title/abstract screened articles for full text screening.
25
26

27 215 The use of TB strain typing as a public health tool in TB control programmes is increasing globally.
28 216 We have identified a lack of good quality studies that can contribute to our understanding in
29 217 interpreting the molecular typing of TB. We have also shown that the proportion of clustering
30 218 derived from MIRU-VNTR typing is influenced by the number of loci typed, whether consent is
31 219 required to type isolates, TB incidence in the study setting, and the maximum cluster size,
32 220 highlighting these as important considerations in the design and interpretation of future studies.
33
34

35 221 **Conflict of interest**

36 222 Nothing to declare.
37
38

39 223 **Acknowledgements**

40 224 We would like to acknowledge Ross Harris from the Statistics Unit at Public Health England for his
41 225 advice on meta-regression.
42
43

44 226 **Author contributions**

45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 227 All authors made substantial contributions to the conception and design of the review, and the
4 228 analysis and interpretation of data. JM drafted the article and PS, IA, TM and TC revised it critically
5 229 for important intellectual content. All authors approved the final version for publication.
6
7

8 230 **Funding**

9
10
11 231 JM is funded through a Public Health England and University College London Impact Studentship. IA
12 232 is funded through a NIHR Senior Research Fellowship.
13

14 233 **Ethics**

15
16
17 234 Ethical approval was not required as this review analyses data that is in the public domain.
18
19

20 235 **Data sharing**

21
22 236 No additional data are available
23
24

25 237 **References**

- 26
27 238 1. Lambregts-van Weezenbeek CSB, Sebek MMGG, van Gerven PJHJ, de Vries G, Verver S, et al.
28 239 (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands:
29 240 6 years' experience with nation-wide cluster feedback and cluster monitoring. *Int J Tuberc Lung*
30 241 *Dis* 7: S463–470.
31
32 242 2. Borgdorff MW, van den Hof S, Kremer K, Verhagen L, Kalisvaart N, et al. (2010) Progress
33 243 towards tuberculosis elimination: secular trend, immigration and transmission. *Eur Respir J* 36:
34 244 339–347. doi:10.1183/09031936.00155409.
35
36 245 3. Kik SV, Verver S, Van Soolingen D, De Haas PEW, Cobelens FG, et al. (2008) Tuberculosis
37 246 Outbreaks Predicted by Characteristics of First Patients in a DNA Fingerprint Cluster. *Am J*
38 247 *Respir Crit Care Med* 178: 96–104. doi:10.1164/rccm.200708-1256OC.
39
40 248 4. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain
41 249 typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology
42 250 laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J*
43 251 *Clin Microbiol* 31: 1677–1682.
44
45 252 5. De Vries G, van Hest RAH, Richardus JH (2007) Impact of mobile radiographic screening on
46 253 tuberculosis among drug users and homeless persons. *Am J Respir Crit Care Med* 176: 201–
47 254 207. doi:10.1164/rccm.200612-1877OC.
48
49 255 6. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous
50 256 detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and
51 257 epidemiology. *J Clin Microbiol* 35: 907–914.
52
53 258 7. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification
54 259 of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized
55 260 methodology. *J Clin Microbiol* 31: 406–409.
56
57
58
59
60

- 1
2
3 261 8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsç-Gerdes S, et al. (2006) Proposal for
4 262 standardization of optimized mycobacterial interspersed repetitive unit-variable-number
5 263 tandem repeat typing of Mycobacterium tuberculosis. *J Clin Microbiol* 44: 4498–4510.
6 264 doi:10.1128/JCM.01392-06.
- 8 265 9. Schürch AC, van Soolingen D (2011) DNA fingerprinting of Mycobacterium tuberculosis: From
9 266 phage typing to whole-genome sequencing. *Infect Genet Evol*. Available:
10 267 <http://www.ncbi.nlm.nih.gov/pubmed/22067515>. Accessed 13 March 2012.
- 12 268 10. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and
13 269 social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730–739.
14 270 doi:10.1056/NEJMoa1003176.
- 16 271 11. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. (2013) Whole-genome sequencing to
17 272 delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet*
18 273 *Infect Dis* 13: 137–146. doi:10.1016/S1473-3099(12)70277-3.
- 20 274 12. Houben RMGJ, Glynn JR (2009) A systematic review and meta-analysis of molecular
21 275 epidemiological studies of tuberculosis: development of a new tool to aid interpretation.
22 276 *Tropical Medicine & International Health* 14: 892–909. doi:10.1111/j.1365-3156.2009.02316.x.
- 24 277 13. Fok A, Numata Y, Schulzer M, FitzGerald MJ (May) Risk factors for clustering of tuberculosis
25 278 cases: a systematic review of population-based molecular epidemiology studies [Review
26 279 Article]. *The International Journal of Tuberculosis and Lung Disease* 12: 480–492.
- 28 280 14. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, Van Soolingen D (2011) Influence of
29 281 Sampling on Clustering and Associations With Risk Factors in the Molecular Epidemiology of
30 282 Tuberculosis. *Am J Epidemiol*. Available:
31 283 <http://aje.oxfordjournals.org/content/early/2011/05/23/aje.kwr061>. Accessed 29 March 2012.
- 33 284 15. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, et al. (1999) Interpreting DNA fingerprint
34 285 clusters of Mycobacterium tuberculosis. *European Concerted Action on Molecular*
35 286 *Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis* 3: 1055–1060.
- 37 287 16. Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD, et al. (2005) The Importance of
38 288 Recent Infection with Mycobacterium tuberculosis in an Area with High HIV Prevalence: A
39 289 Long-Term Molecular Epidemiological Study in Northern Malawi. *J Infect Dis* 192: 480–487.
40 290 doi:10.1086/431517.
- 42 291 17. De Beer JL, Kremer K, Ködmön C, Supply P, van Soolingen D (2012) First Worldwide Proficiency
43 292 Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex
44 293 Strains. *Journal of Clinical Microbiology* 50: 662–669. doi:10.1128/JCM.00607-11.
- 46 294 18. Maes M, Kremer K, van Soolingen D, Takiff H, de Waard JH (2008) 24-locus MIRU-VNTR
47 295 genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao
48 296 Amerindians in Venezuela. *Tuberculosis (Edinb)* 88: 490–494. doi:10.1016/j.tube.2008.04.003.
- 50 297 19. Sougakoff W (2011) Molecular epidemiology of multidrug-resistant strains of Mycobacterium
51 298 tuberculosis. *Clinical Microbiology and Infection* 17: 800–805. doi:10.1111/j.1469-
52 299 0691.2011.03577.x.

- 1
2
3 300 20. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for
4 301 polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. *Nucleic Acids Res* 38:
5 302 W326–331. doi:10.1093/nar/gkq351.
- 6
7 303 21. Supply P (2010) MIRU-VNTR typing: the new international standard for TB molecular
8 304 epidemiology Symposium of the Institut Pasteur de Tunisia.
- 9
10 305 22. Van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, et al. (1999) Molecular
11 306 epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997.
12 307 *J Infect Dis* 180: 726–736. doi:10.1086/314930.
- 13
14 308 23. Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step
15 309 approach for large-scale, prospective genotyping of Mycobacterium tuberculosis isolates in the
16 310 United States. *J Clin Microbiol* 43: 688–695. doi:10.1128/JCM.43.2.688-695.2005.
- 17
18 311 24. Centers for Disease Control and Prevention (2005) New CDC Program for Rapid Genotyping of
19 312 Mycobacterium Tuberculosis Isolates. *JAMA* 293: 2086–2086. doi:10.1001/jama.293.17.2086.
- 20
21 313 25. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, et al. (2000) A prospective evaluation
22 314 of the clinical value of nation-wide DNA fingerprinting of tuberculosis isolates in Denmark. *Int J*
23 315 *Tuberc Lung Dis* 4: 295–299.
- 24
25 316 26. Bauer J, Yang Z, Poulsen S, Andersen AB (1998) Results from 5 years of nationwide DNA
26 317 fingerprinting of Mycobacterium tuberculosis complex isolates in a country with a low
27 318 incidence of M. tuberculosis infection. *J Clin Microbiol* 36: 305–308.
- 28
29 319 27. Zolnir-Dovc M, Poljak M, Erzen D, Sorli J (2003) Molecular epidemiology of tuberculosis in
30 320 Slovenia: results of a one-year (2001) nation-wide study. *Scand J Infect Dis* 35: 863–868.
- 31
32 321 28. Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Hoek KGP, et al. (2008)
33 322 Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-
34 323 repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of
35 324 Mycobacterium tuberculosis Beijing strains in a setting of high incidence of tuberculosis. *J Clin*
36 325 *Microbiol* 46: 3338–3345. doi:10.1128/JCM.00770-08.
- 37
38 326 29. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, et al. (2001) Automated high-
39 327 throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based
40 328 on mycobacterial interspersed repetitive units. *J Clin Microbiol* 39: 3563–3571.
41 329 doi:10.1128/JCM.39.10.3563-3571.2001.
- 42
43 330 30. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniowski FA (2006) Progression toward an
44 331 improved DNA amplification-based typing technique in the study of Mycobacterium
45 332 tuberculosis epidemiology. *J Clin Microbiol* 44: 2492–2498. doi:10.1128/JCM.01428-05.
- 46
47 333 31. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems:
48 334 an application of Simpson's index of diversity. *J Clin Microbiol* 26: 2465–2466.
- 49
50 335 32. WHO | TB data (n.d.). WHO. Available: <http://www.who.int/tb/country/en/index.html>.
51 336 Accessed 12 December 2012.
- 52
53
54
55
56
57
58
59
60

- 1
2
3 337 33. Kruijshaar ME, Pimpin L, Abubakar I, Rice B, Delpech V, et al. (2011) The burden of TB-HIV in
4 338 the EU: how much do we know? A survey of surveillance practices and results. *Eur Respir J* 38:
5 339 1374–1381. doi:10.1183/09031936.00198310.
- 6
7 340 34. World Health Organization (n.d.) WHO Tuberculosis Country Profiles. Available:
8 341 <http://www.who.int/tb/country/data/profiles/en/>.
- 9
10 342 35. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, et al. (2014) Strengthening the Reporting of
11 343 Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE
12 344 statement. *The Lancet Infectious Diseases* 14: 341–352. doi:10.1016/S1473-3099(13)70324-4.
- 13
14 345 36. Freeman MF, Tukey JW (1950) Transformations Related to the Angular and the Square Root.
15 346 *Ann Math Statist* 21: 607–611. doi:10.1214/aoms/1177729756.
- 16
17 347 37. Guang-ming D, Zhi-guo Z, Peng-ju D, Qian Z, Li W, et al. (2013) Differences in the population of
18 348 genetics of *Mycobacterium tuberculosis* between urban migrants and local residents in Beijing,
19 349 China. *Chinese Medical Journal* 126: 4066–4071. doi:10.3760/cma.j.issn.0366-6999.20130216.
- 20
21 350 38. Zmak L, Obrovac M, Katalinic Jankovic V (2014) First insights into the molecular epidemiology
22 351 of tuberculosis in Croatia during a three-year period, 2009 to 2011. *Scandinavian Journal of*
23 352 *Infectious Diseases* 46: 123–129.
- 24
25 353 39. Mandal S, Bradshaw L, Anderson LF, Brown T, Evans JT, et al. (2011) Investigating transmission
26 354 of *Mycobacterium bovis* in the United Kingdom in 2005 to 2008. *J Clin Microbiol* 49: 1943–
27 355 1950. doi:10.1128/JCM.02299-10.
- 28
29 356 40. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR (2011) Tuberculosis transmission in
30 357 Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. *Infect Genet Evol*
31 358 11: 124–131. doi:10.1016/j.meegid.2010.09.013.
- 32
33 359 41. Dymova MA, Liashenko OO, Poteiko PI, Krutko VS, Khrapov EA, et al. (2011) Genetic variation
34 360 of *Mycobacterium tuberculosis* circulating in Kharkiv Oblast, Ukraine. *BMC Infect Dis* 11: 77.
35 361 doi:10.1186/1471-2334-11-77.
- 36
37 362 42. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, et al. (2011) Molecular epidemiology of
38 363 *Mycobacterium tuberculosis* in East Lancashire 2001-2009. *Thorax* 66: 709–713.
39 364 doi:10.1136/thx.2011.158881.
- 40
41 365 43. Evans J (2010) Analysis of prevalent *Mycobacterium tuberculosis* strains in the United
42 366 Kingdom: detection, distribution and expansion of MIRU-VNTR profiles containing high
43 367 numbers of isolates. *European Society of Clinical Microbiology and Infectious Diseases*. Vienna,
44 368 Austria.
- 45
46 369 44. Nikolayevskyy VV, Brown TJ, Bazhora YI, Asmolov AA, Balabanova YM, et al. (2007) Molecular
47 370 epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in
48 371 *Mycobacterium tuberculosis* strains from the southern Ukraine. *Clin Microbiol Infect* 13: 129–
49 372 138. doi:10.1111/j.1469-0691.2006.01583.x.
- 50
51 373 45. L. Z, M. O, V. KJ (2014) First insights into the molecular epidemiology of tuberculosis in Croatia
52 374 during a three-year period, 2009 to 2011. *Scandinavian Journal of Infectious Diseases*.
- 53
54
55
56
57
58
59
60

- 1
2
3 375 46. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, et al. (2011) Evaluation of Mycobacterium
4 376 tuberculosis typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. *J Clin*
5 377 *Microbiol* 49: 4173–4178. doi:10.1128/JCM.05293-11.
- 6
7 378 47. Dymova MA, Kinsht VN, Cherednichenko AG, Khrapov EA, Svistelnik AV, et al. (2011) Highest
8 379 prevalence of the Mycobacterium tuberculosis Beijing genotype isolates in patients newly
9 380 diagnosed with tuberculosis in the Novosibirsk oblast, Russian Federation. *J Med Microbiol* 60:
10 381 1003–1009. doi:10.1099/jmm.0.027995-0.
- 11
12 382 48. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, et al. (2009)
13 383 Prospective universal application of mycobacterial interspersed repetitive-unit-variable-
14 384 number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for
15 385 fast identification of clustered and orphan cases. *J Clin Microbiol* 47: 2026–2032.
16 386 doi:10.1128/JCM.02308-08.
- 17
18 387 49. Glynn JR, Vyonycky E, Fine PEM (1999) Influence of Sampling on Estimates of Clustering and
19 388 Recent Transmission of Mycobacterium tuberculosis Derived from DNA Fingerprinting
20 389 Techniques. *American Journal of Epidemiology* 149: 366–371.
- 21
22 390 50. TB Strain Typing Project Board HPA (2011) TB Strain Typing Cluster Investigation Handbook for
23 391 Health Protection Units 1st Edition. Available:
24 392 <https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealth>
25 393 [ProtectionProgrammes/Respiratory/TB/StrainTyping/](https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealth). Accessed 30 November 2011.
- 26
27 394 51. Walker TM, Monk P, Grace Smith E, Peto TEA (2013) Contact investigations for outbreaks of
28 395 Mycobacterium tuberculosis: advances through whole genome sequencing. *Clin Microbiol*
29 396 *Infect*. doi:10.1111/1469-0691.12183.
- 30
31 397 52. Gurjav U, Jelfs P, McCallum N, Marais BJ, Sintchenko V (2014) Temporal dynamics of
32 398 Mycobacterium tuberculosis genotypes in New South Wales, Australia. *BMC infectious*
33 399 *diseases* 14: 455–455. doi:10.1186/1471-2334-14-455.
- 34
35 400 53. Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of
36 401 standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat
37 402 typing of Mycobacterium tuberculosis. *J Clin Microbiol* 46: 1398–1406.
38 403 doi:10.1128/JCM.02089-07.
- 39
40 404 54. Allix-Béguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008) Standardised PCR-based
41 405 molecular epidemiology of tuberculosis. *Eur Respir J* 31: 1077–1084.
42 406 doi:10.1183/09031936.00053307.
- 43
44 407 55. Tuite AR, Guthrie JL, Alexander DC, Whelan MS, Lee B, et al. (2013) Epidemiological evaluation
45 408 of spatiotemporal and genotypic clustering of mycobacterium tuberculosis in Ontario, Canada.
46 409 *International Journal of Tuberculosis and Lung Disease* 17: 1322–1327.
- 47
48 410 56. Tessema B, Beer J, Merker M, Emmrich F, Sack U, et al. (2013) Molecular epidemiology and
49 411 transmission dynamics of Mycobacterium tuberculosis in Northwest Ethiopia: new
50 412 phylogenetic lineages found in Northwest Ethiopia. *Bmc Infectious Diseases* 13: 131.
51 413 doi:10.1186/1471-2334-13-131.

- 1
2
3 414 57. Smit PW, Haanpera M, Rantala P, Couvin D, Lyytikainen O, et al. (2013) Molecular
4 415 Epidemiology of Tuberculosis in Finland, 2008-2011. Plos One 8: e85027.
5 416 doi:10.1371/journal.pone.0085027.
- 6
7 417 58. Oelemann MC, Diel R, Vatin V, Haas W, Rüscher-Gerdes S, et al. (2007) Assessment of an
8 418 optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing
9 419 system combined with spoligotyping for population-based molecular epidemiology studies of
10 420 tuberculosis. J Clin Microbiol 45: 691–697. doi:10.1128/JCM.01393-06.
- 11
12 421 59. Ojo OO, Sheehan S, Corcoran DG, Nikolayevsky V, Brown T, et al. (2010) Molecular
13 422 epidemiology of Mycobacterium tuberculosis clinical isolates in Southwest Ireland. Infect
14 423 Genet Evol 10: 1110–1116. doi:10.1016/j.meegid.2010.07.008.
- 15
16 424 60. Aleksic E, Merker M, Cox H, Reiher B, Sekawi Z, et al. (2013) First Molecular Epidemiology Study
17 425 of Mycobacterium tuberculosis in Kiribati. PLoS ONE 8. Available:
18 426 [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8)
19 427 [84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8](http://www.scopus.com/inward/record.url?eid=2-s2.0-84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8).
- 20
21 428 61. De Beer JL, van Ingen J, de Vries G, Erkens C, Sebek M, et al. (2013) Comparative study of
22 429 IS6110 restriction fragment length polymorphism and variable-number tandem-repeat typing
23 430 of Mycobacterium tuberculosis isolates in the Netherlands, based on a 5-year nationwide
24 431 survey. J Clin Microbiol 51: 1193–1198. doi:10.1128/JCM.03061-12.
- 25
26 432 62. Varghese B, Supply P, Shoukri M, Allix-Beguec C, Memish Z, et al. (2013) Tuberculosis
27 433 Transmission among Immigrants and Autochthonous Populations of the Eastern Province of
28 434 Saudi Arabia. PLoS ONE 8. Available: [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2)
29 435 [84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2](http://www.scopus.com/inward/record.url?eid=2-s2.0-84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2).
- 30
31 436 63. Lim LK-Y, Sng LH, Win W, Chee CB-E, Hsu LY, et al. (2013) Molecular Epidemiology of
32 437 Mycobacterium tuberculosis Complex in Singapore, 2006-2012. Plos One 8: e84487.
33 438 doi:10.1371/journal.pone.0084487.
- 34
35 439 64. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-
36 440 locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med
37 441 105 Suppl 1: S67–73. doi:10.1016/S0954-6111(11)70014-2.
- 38
39 442 65. Jonsson J., Hoffner S., Berggren I., Bruchfeld J., Ghebremichael S., et al. (2014) Comparison
40 443 between RFLP and MIRU-VNTR genotyping of mycobacterium tuberculosis strains isolated in
41 444 stockholm 2009 to 2011. PLoS ONE. Available:
42 445 <http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0095159&representation=PDF>.
43 446
- 44
45 447 66. Muwonge A, Malama S, Johansen TB, Kankya C, Biffa D, et al. (2013) Molecular Epidemiology,
46 448 Drug Susceptibility and Economic Aspects of Tuberculosis in Mubende District, Uganda. PLoS
47 449 ONE 8. Available: [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697)
48 450 [84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697](http://www.scopus.com/inward/record.url?eid=2-s2.0-84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697).
- 49
50 451 67. Hamblion EL, Wynne-Edwards E, Anderson C, Anderson SR (2011) A summary of strain typing
51 452 and clustering of TB in London in 2010 and an analysis of the associated risk factors. Thorax 66:
52 453 A88–A89. doi:10.1136/thoraxjnl-2011-201054c.50.
- 53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

454 68. Hang NTL, Maeda S, Lien LT, Thuong PH, Hung NV, et al. (2013) Primary Drug-Resistant
455 Tuberculosis in Hanoi, Viet Nam: Present Status and Risk Factors. Plos One 8: UNSP e71867.
456 doi:10.1371/journal.pone.0071867.
457

For peer review only

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

[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
<u>Study setting</u>		
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
<u>Study design</u>		
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 meta-regression 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.

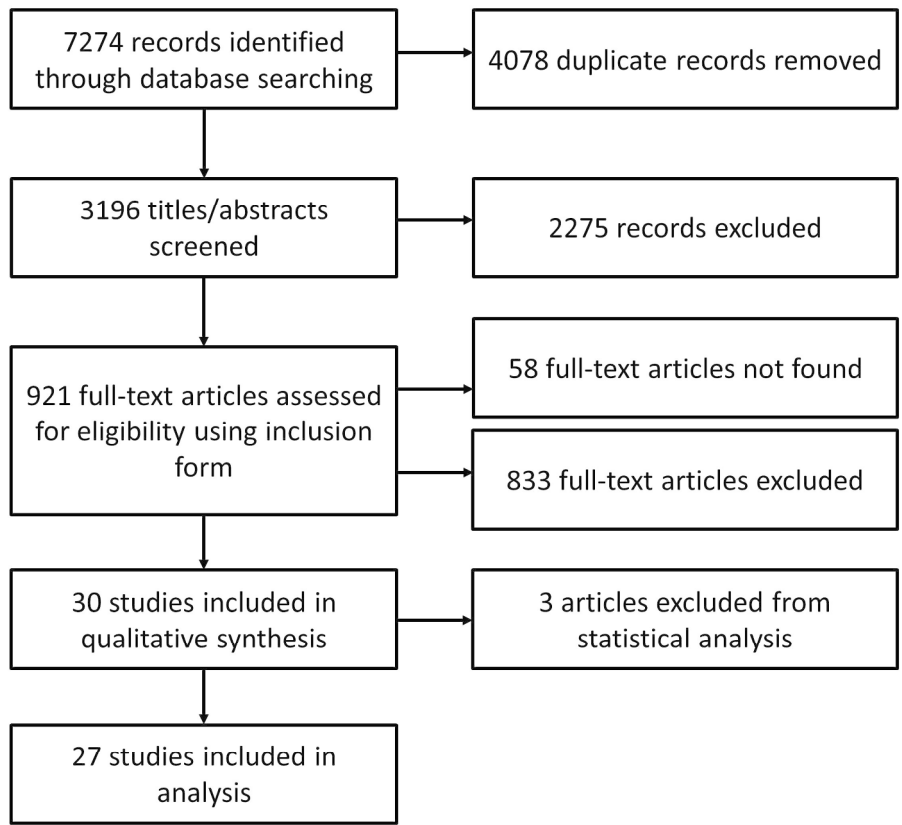
1
2
3
4
5 **Figure Caption**
6

7 **Figure 1: Results of systematic search, screening and data extraction.**
8

9 **Figure 2: Forest plot showing the proportion of clustering reported in each study by the number of MIRU-**
10 **VNTR loci typed**
11

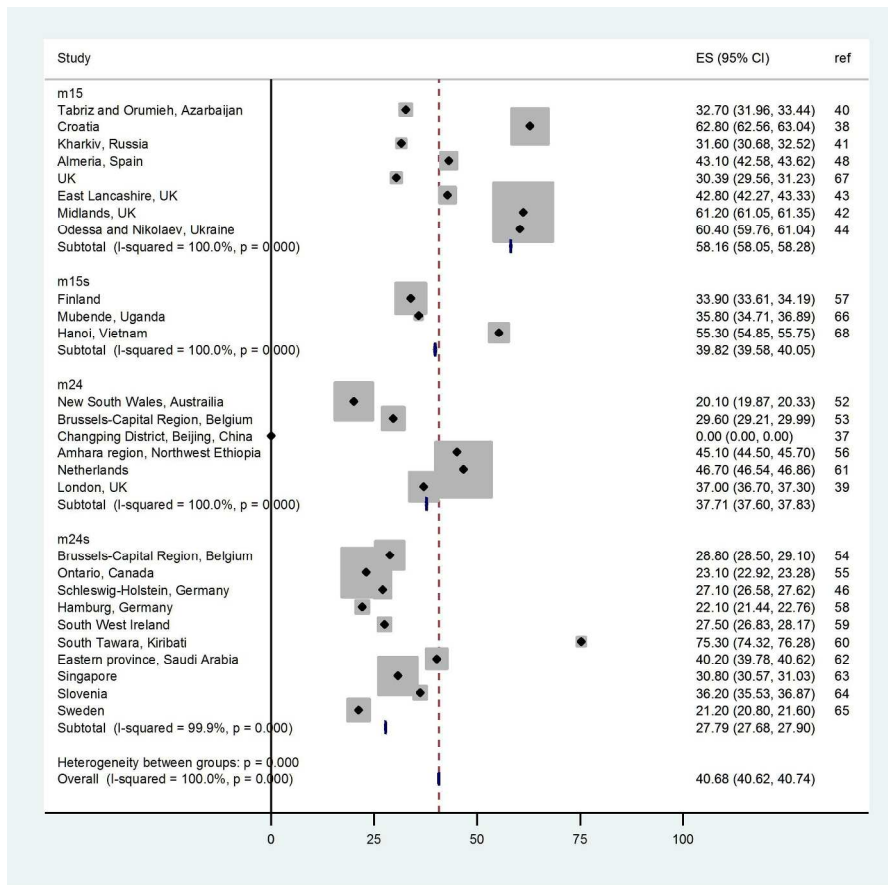
12 The number of loci typed is categorised into 15 loci (m15), 15 loci with Spoligotyping (m15s), 24 loci (m24) and
13 24 loci with Spoligotyping (m24s). The study reference is shown in the right hand column.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



190x254mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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

- 1
2
3 32. 4 and 16
4 33. 32 and 31
5
6 34. limit 33 to yr="1998-Current"
7
8 35. limit 34 to english language
9
10 36. animals/
11 37. humans/
12 38. 36 not 37
13 39. 35 not 38
14
15
16

17 Appendix 2: STROME-ID scores for the included studies

19 Author	STROME-ID score ^a
20 Aleksic, E	24
21 Allix-Beguec, C	32
22 Allix-Beguec, C	25
23 Alonso-Rodriguez, N	18
24 Asgharzadeh, M	28
25 Bidovec-Stojkovic, U	31
26 De Beer, JL	30
27 Dymova, MA	19
28 Evans, J	^b
29 Grujav, U	32
30 Guang-ming, DAI	19
31 Hamblion, E	^b
32 Hang, NTHL	31
33 Jonsson, J	22
34 Lim, LKY	30
35 Mandal, S	32
36 Muwonge, A	25
37 Nikolayevsky, V	23
38 Oelemann, M	34
39 Ojo, OO	36
40 Roetzer, A	16
41 Sails, A	23
42 Smit, PW	29
43 Tessema, B	26
44 Tuite, AR	31
45 Varghese, B	23
46 Zmak, L	19

57 ^aIndividual studies score 1 for each element of checklist they had address

58 ^bConference abstracts



PRISMA 2009 Checklist

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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

Page 1 of 2

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.

Page 2 of 2

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

4 Jessica Mears¹, Ibrahim Abubakar^{1,2,3}, Theodore Cohen⁴, Timothy D McHugh⁵ & Pam Sonnenberg^{1,*}

5 ¹ Department of Infection and Population Health, University College London

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

7 ³ Clinical Trials Unit, Medical Research Council, London

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

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

11 *Corresponding author:

12 Dr Pam Sonnenberg

13 3rd Floor Mortimer Market Centre

14 University College London

15 London WC1E 6JB

16 p.sonnenberg@ucl.ac.uk

17

18 **Word count: 2392**

1
2
3 **19 Abstract**
4

5 20 Objectives: To systematically review the evidence for the impact of study design and setting on the
6 21 interpretation of TB transmission using clustering derived from Mycobacterial Interspersed
7 22 Repetitive Units – Variable Number Tandem Repeats (MIRU-VNTR) strain typing.
8
9

10 23 Data sources: Medline, Embase, CINHAI, Web of Science and Scopus were searched for articles
11 24 published before 21st October 2014.
12

13 25 Review methods: Studies in humans that reported the proportion of clustering of TB isolates by
14 26 MIRU-VNTR were included in the analysis. Univariable meta-regression analyses were conducted to
15 27 assess the influence of study design and setting on the proportion of clustering.
16
17
18

19 28 Results: The search identified 27 eligible articles reporting clustering between 0% and 63%. The
20 29 number of MIRU-VNTR loci typed, requiring consent to type patient isolates (as a proxy for sampling
21 30 fraction), the TB incidence and the maximum cluster size explained 14%, 14%, 27% and 48% of
22 31 between-study variation, respectively, and had a significant association with the proportion of
23 32 clustering.
24
25
26
27
28

29 33 Conclusions: Although MIRU-VNTR typing is being adopted worldwide there is a paucity of data on
30 34 how study design and setting may influence estimates of clustering. We have highlighted study
31 35 design variables for consideration in the design and interpretation of future studies.
32
33
34

35 36

36
37 **37 Strengths and Limitations of Study**
38

- 39 38 • This is a timely evaluation of the impact of study design on estimates of TB clustering using
40 39 MIRU-VNTR strain typing because it has been incorporated into national typing services
41 40 globally.
42
43 41 • The strength of this meta-analysis was limited by the lack of detail reported by the included
44 42 studies, highlighting the need for better quality reporting in primary studies.
45
46
47

48 43

49 44
50
51
52
53
54
55
56
57
58
59
60

45 Introduction

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

52 TB molecular typing methods include Spoligotyping [6], insertion sequence *6110* (*IS6110*) restriction
53 fragment length polymorphism (RFLP) analysis (the recent gold standard) [7], mycobacterial
54 interspersed repetitive units-variable number tandem repeats (MIRU-VNTR) typing [8], and whole
55 genome sequencing [9–11]. Published reviews have identified factors that might influence or bias
56 clustering by *IS6110* RFLP [12,13]. No study has repeated this analysis using more up-to-date typing
57 methods, which is important for understanding of the epidemiology of TB and to shape the
58 application of molecular typing to improve TB control.

59 Published meta-analyses and modelling studies using *IS6110* RFLP data show that the proportion of
60 clustering observed can be affected by 1) study design (affecting the proportion of eligible cases that
61 are included in the study); 2) features of the typing method (such as the ability to type isolates with
62 low copy numbers); and 3) study setting (such as characteristics of the study population). For
63 example, the proportion of clustering increases when the fraction of the total data sampled
64 increases [13–15] and when study duration increases [16].

65 MIRU-VNTR is currently the preferred method of molecular typing [17–21], and can be used
66 together with Spoligotyping [8]. Relative to *IS6110* RFLP, MIRU-VNTR does not have to exclude
67 isolates with a low *IS6110* copy number, has a faster turnaround time, is high throughput and the
68 numeric strain types are more easily compared. MIRU-VNTR strain typing is increasingly being
69 adopted worldwide [1,22–27], yet unlike *IS6110* RFLP, the evidence for the interpretation of the
70 findings such as the impact of study design and setting on clustering have not been reviewed.
71 Although the two typing methods have been shown to have a similar discriminatory value, the
72 markers evolve independently and at different rates, resulting in a difference in clustering between
73 the two methods [28]. This suggests that there could be differences in the way study design, typing
74 method and setting affects clustering by the two methods. We conducted a systematic review to
75 assess the evidence for the impact of study design and setting on the interpretation of TB

1
2
3 76 transmission using clustering derived from MIRU-VNTR strain typing – as has been shown using
4 77 IS6110 RFLP typing.

78 **Methods**

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

83 All titles and abstracts from each of the searches were examined. The full text of each paper was
84 obtained and reviewed if the study reported MIRU-VNTR strain typing of *M.tuberculosis* complex
85 isolates with at least 15 of the standardised 24 loci (ETR A, B, C, D, E; MIRU 2, 10, 16, 20, 23, 24, 26,
86 27, 39, 40; VNTR 424, 1955, 2163b, 2347, 2401, 3171, 3690, 4052, 4156) [8,29,30].

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

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

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

1
2
3 105 The main outcome measure – the proportion of TB isolates clustered by MIRU-VNTR strain typing –
4 106 was calculated as the number of clustered isolates/number of clustered+unique isolates. Where
5
6 107 there were uncertainties JM consulted with IA.
7

8
9 108 Authors were contacted if TB incidence rate was not reported. Where no response was received
10 109 WHO country estimates of TB incidence for the study year were used.[32] As so few studies reported
11 110 the proportion coinfecting with TB/HIV, these estimates for the study country were taken from an
12 111 EU-wide survey and WHO country profiles.[33,34] Due to poor recording of the sampling fraction
13 112 (the number of isolates typed/the total number of culture positive TB cases diagnosed during the
14 113 study period (n=19)), whether the study required the consent of participants (yes/no) was included
15 114 as a proxy for (low/high) sampling fraction. The risk of bias within each study was assessed using the
16 115 STROME-ID checklist.[35]
17

18
19
20
21
22 116 Data were analysed in Stata 12. Where studies reported data from more than one set of loci, the
23 117 method with the highest discriminatory value was included (i.e. MIRU-VNTR 24 would be chosen
24 118 over MIRU-VNTR 15, and MIRU-VNTR 15 plus Spoligotyping would be chosen over MIRU-VNTR 15
25 119 alone) (n=8). This review was not concerned with summary measures of clustering, but factors that
26 120 influenced clustering; therefore articles must have included at least one of the covariates.
27 121 Continuous variables were transformed where the distribution was skewed. The proportion
28 122 clustered was transformed using the Freeman Tukey transformation [36]. Study heterogeneity was
29 123 assessed using a forest plot and the χ^2 test of heterogeneity. Univariable meta-regression analyses
30 124 were carried out to determine the effect of the study design covariates on the proportion of
31 125 clustered isolates. All covariates in the analysis were hypothesised to influence the proportion
32 126 clustered *a priori*.
33
34
35
36
37
38
39

40
41 127 Sensitivity analyses were conducted to see the effect of removing studies reporting 0% clustering,
42 128 with only extra-pulmonary TB cases, only *M.bovis* cases, studies using the 'old 12' MIRU loci as part
43 129 of their 15 loci, and studies assessed as having a high likelihood of bias (STROME-ID score less than
44 130 20).
45
46
47

48 131 **Results**

49
50
51 132 The search identified 7274 references resulting in 27 studies (25 journal articles and 2 conference
52 133 abstracts) included after deduplication and title/abstract/full text screening (Figure 1). The main
53 134 characteristics of the included studies are shown in Table 1. Studies were published between 2007
54 135 and 2014 and the clustering reported varied from 0% [37] to 62.8% [38]. In all studies, clustered
55
56
57
58
59
60

1
2
3 136 isolates were defined as having identical strain types based on the MIRU-VNTR loci typed, with or
4 137 without Spoligotyping. 17 studies included isolates from newly diagnosed TB cases, three studies
5 138 reported including isolates from new and chronic cases of TB, and seven did not report this
6 139 information. In addition, ten studies did not include repeat isolates from the same patient, one study
7 140 included a repeat isolate from one patient, and the remaining 17 did not report whether repeat
8 141 isolates were included or not. Furthermore, four studies included isolates with missing loci in the
9 142 cluster analysis, whereas four excluded isolates with missing loci, and the remaining 20 did not
10 143 report how they dealt with missing loci. The number of studies reporting each variable of interest is
11 144 shown in Table 2. STROME-ID scores can be found in Appendix 2.

12
13
14
15
16
17
18 145 A forest plot shows the spread of clustering reported by number of loci and additional typing
19 146 method (Figure 2). Significant heterogeneity was identified between the studies ($p < 0.001$),
20 147 suggesting that a meta-regression would be an appropriate analysis.

21
22
23
24 148 The univariable meta-regression shows evidence for the proportion of clustering to decrease as the
25 149 number of MIRU-VNTR loci typed increased from 15 to 24 ($p = 0.04$; Table 3), accounting for 14% of
26 150 the between study variation, and to increase when the study participants consented to being
27 151 included in the study ($p = 0.03$), accounting for 14% of the between study variation. The proportion of
28 152 clustering increased as the TB incidence in the population increased ($p = 0.007$, $\text{Adj } R^2 = 26.7$). There
29 153 was also evidence for the proportion of clustering to increase as the maximum cluster size increased
30 154 ($p = 0.001$), accounting for 48% of between study variation. There was no evidence of the other study
31 155 design or study setting variables significantly influencing the proportion clustered. Though non-
32 156 significant ($p > 0.05$), the TB/HIV coinfection rate in the population explained 2% of the between
33 157 study variation. Too few studies included information on the proportion of clusters containing two
34 158 cases, proportion of the study sample with previous TB or with pulmonary TB, so these could not be
35 159 included in the analysis (Table 2).

36
37
38
39
40
41
42
43
44 160 Sensitivity analyses to examine the effect of excluding studies reporting 0% clustering,[37] only
45 161 *M.bovis* cases,[39] studies using the 'old 12' MIRU loci,[39–44] and studies assessed as having a high
46 162 risk of bias,[37,45–48] did not generally change the results. The proportion of culture positive TB in
47 163 the population remained insignificant but explained 2.6% of the between study variation when
48 164 excluding 0% clustering ($p = 0.278$ and $\text{Adj } R^2 = 2.62$). Similarly, the proportion of culture positive TB in
49 165 the population remained insignificant but explained 2.6% of the between study variation when
50 166 excluding studies with the highest risk of bias ($p = 0.278$ and $\text{Adj } R^2 = 2.62$). The number of loci typed
51 167 became non-significant, but explained 9.6% and 10.5% of the between study variation when
52
53
54
55
56
57
58
59
60

1
2
3 168 excluding studies using the 'old 12' loci and the highest risk of bias, respectively ($p=0.106$, Adj
4 169 $R^2=9.63$; $p=0.111$, Adj $R^2=10.51$, respectively).

5
6
7 170 **Discussion**

8
9 171 This review identified 27 studies that met the inclusion criteria. We illustrate that the interpretation
10 172 of studies using MIRU-VNTR to estimate clustering is subject to bias relating to study design and
11 173 setting; however, there were insufficient data available to fully explore this impact.

12
13
14 174 As expected, we found that the proportion of clustering decreased with a greater number of MIRU-
15 175 VNTR loci typed, with increasing TB incidence and with increasing maximum cluster size. We found
16 176 that requiring consent to type patient isolates increased the proportion of clustering, which is not
17 177 expected, given that the sampling fraction would be lower in these studies.

18
19
20 178 The other study design variables included in this analysis, such as study duration, did not significantly
21 179 influence the proportion of isolates that were clustered, contrary to previous findings [12]. This is
22 180 likely to be because of a lack of good quality evidence: of the 27 studies that met the inclusion
23 181 criteria for the review, none reported all the variables of interest, reducing the power of the analysis
24 182 and precluding multivariable meta-regression (Table 2). Importantly, key details of cluster analyses
25 183 were not reported consistently across the studies, such as whether repeat isolates from the same
26 184 patients were included, or typing profiles with missing loci were included, introducing new,
27 185 unmeasured biases. In addition, the range of the variables may have been too limited to show any
28 186 impact on clustering estimates. For example, the proportion of culture positive isolates typed ranged
29 187 from 34.5% to 100%, with 17 of the 19 studies reporting this variable from 81.9% to 100%.
30 188 Furthermore, most of the studies (17/27=63%) were from low TB burden settings and therefore may
31 189 be reflecting the rate at which imported cases have matching strain types by chance, rather than
32 190 rates of recent transmission.

33
34
35 191 The sensitivity analysis suggested that, when excluding the studies with the greatest risk of bias, the
36 192 culture-positivity in the population might explain a small amount of the between study variation.
37 193 This is consistent with estimates of the influence of sampling on the proportion of clustering using
38 194 /S6110 RFLP typing [49]. In the sensitivity analysis excluding studies that used the 'old 12' loci, the
39 195 effect of the number of loci typed becomes non-significant. This is likely because studies using the
40 196 'old 12' accounted for six out of ten studies reporting 15 loci, reducing the number of studies and
41 197 the power of the model.

1
2
3 198 This study is a timely evaluation of the impact of study design on estimates of TB clustering using
4 199 MIRU-VNTR strain typing because it has been incorporated into national typing services globally
5 200 [23,50]. The findings are relevant where strain typing is used to evaluate TB control systems across
6 201 different settings because the proportion of clustering is influenced by the number of loci typed, the
7 202 TB incidence and the maximum cluster size. Given that strain typing methods are advancing beyond
8 203 MIRU-VNTR typing and that the application of whole genome sequencing to TB control and public
9 204 health strategies has been demonstrated [9–11,51], it is important that the biases in the analysis of
10 205 such methods are explored and compared. Understanding how to design and compare research
11 206 studies for public health will greatly improve the benefit gained from newer technologies.

12
13
14
15
16
17
18 207 The strength of this meta-analysis was limited by (a lack of) detail reported by the included studies.
19 208 This review has highlighted the need for better quality reporting in primary studies to enable future
20 209 reviews to be more robust. Recently published standards for reporting of molecular epidemiology
21 210 for infectious diseases should improve the quality of reporting.[35] This review is further limited by
22 211 our inability to access 58 of the title/abstract screened articles for full text screening.

23
24
25
26
27 212 The use of TB strain typing as a public health tool in TB control programmes is increasing globally.
28 213 We have identified a lack of good quality studies that can contribute to our understanding in
29 214 interpreting the molecular typing of TB. We have also shown that the proportion of clustering
30 215 derived from MIRU-VNTR typing is influenced by the number of loci typed, whether consent is
31 216 required to type isolates, TB incidence in the study setting, and the maximum cluster size,
32 217 highlighting these as important considerations in the design and interpretation of future studies.

33 34 35 36 37 38 218 **Conflict of interest**

39
40 219 Nothing to declare.

41 42 43 220 **Acknowledgements**

44
45 221 We would like to acknowledge Ross Harris from the Statistics Unit at Public Health England for his
46 222 advice on meta-regression.

47 48 49 223 **Author contributions**

50
51 224 All authors made substantial contributions to the conception and design of the review, and the
52 225 analysis and interpretation of data. JM drafted the article and PS, IA, TM and TC revised it critically
53 226 for important intellectual content. All authors approved the final version for publication.

1
2
3 227 **Funding**
4

5 228 JM is funded through a Public Health England and University College London Impact Studentship. IA
6
7 229 is funded through a NIHR Senior Research Fellowship.
8

9
10 230 **Ethics**
11

12 231 Ethical approval was not required as this review analyses data that is in the public domain.
13

14 232 **Data sharing**
15

16 233 No additional data are available
17

18
19 234 **References**
20

- 21 235 1. Lambregts-van Weezenbeek CSB, Sebek MMGG, van Gerven PJHJ, de Vries G, Verver S, et al.
22 236 (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands:
23 237 6 years' experience with nation-wide cluster feedback and cluster monitoring. *Int J Tuberc Lung*
24 238 *Dis* 7: S463–470.
25
26 239 2. Borgdorff MW, van den Hof S, Kremer K, Verhagen L, Kalisvaart N, et al. (2010) Progress
27 240 towards tuberculosis elimination: secular trend, immigration and transmission. *Eur Respir J* 36:
28 241 339–347. doi:10.1183/09031936.00155409.
29
30 242 3. Kik SV, Verver S, Van Soolingen D, De Haas PEW, Cobelens FG, et al. (2008) Tuberculosis
31 243 Outbreaks Predicted by Characteristics of First Patients in a DNA Fingerprint Cluster. *Am J*
32 244 *Respir Crit Care Med* 178: 96–104. doi:10.1164/rccm.200708-1256OC.
33
34 245 4. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain
35 246 typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology
36 247 laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J*
37 248 *Clin Microbiol* 31: 1677–1682.
38
39 249 5. De Vries G, van Hest RAH, Richardus JH (2007) Impact of mobile radiographic screening on
40 250 tuberculosis among drug users and homeless persons. *Am J Respir Crit Care Med* 176: 201–
41 251 207. doi:10.1164/rccm.200612-1877OC.
42
43 252 6. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous
44 253 detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and
45 254 epidemiology. *J Clin Microbiol* 35: 907–914.
46
47 255 7. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification
48 256 of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized
49 257 methodology. *J Clin Microbiol* 31: 406–409.
50
51 258 8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, et al. (2006) Proposal for
52 259 standardization of optimized mycobacterial interspersed repetitive unit-variable-number
53 260 tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 44: 4498–4510.
54 261 doi:10.1128/JCM.01392-06.
55
56
57
58
59
60

- 1
2
3 262 9. Schürch AC, van Soolingen D (2011) DNA fingerprinting of Mycobacterium tuberculosis: From
4 263 phage typing to whole-genome sequencing. *Infect Genet Evol*. Available:
5 264 <http://www.ncbi.nlm.nih.gov/pubmed/22067515>. Accessed 13 March 2012.
- 7 265 10. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and
8 266 social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730–739.
9 267 doi:10.1056/NEJMoa1003176.
- 11 268 11. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. (2013) Whole-genome sequencing to
12 269 delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet*
13 270 *Infect Dis* 13: 137–146. doi:10.1016/S1473-3099(12)70277-3.
- 15 271 12. Houben RMGJ, Glynn JR (2009) A systematic review and meta-analysis of molecular
16 272 epidemiological studies of tuberculosis: development of a new tool to aid interpretation.
17 273 *Tropical Medicine & International Health* 14: 892–909. doi:10.1111/j.1365-3156.2009.02316.x.
- 19 274 13. Fok A, Numata Y, Schulzer M, FitzGerald MJ (May) Risk factors for clustering of tuberculosis
20 275 cases: a systematic review of population-based molecular epidemiology studies [Review
21 276 Article]. *The International Journal of Tuberculosis and Lung Disease* 12: 480–492.
- 23 277 14. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, Van Soolingen D (2011) Influence of
24 278 Sampling on Clustering and Associations With Risk Factors in the Molecular Epidemiology of
25 279 Tuberculosis. *Am J Epidemiol*. Available:
26 280 <http://aje.oxfordjournals.org/content/early/2011/05/23/aje.kwr061>. Accessed 29 March 2012.
- 29 281 15. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, et al. (1999) Interpreting DNA fingerprint
30 282 clusters of Mycobacterium tuberculosis. *European Concerted Action on Molecular*
31 283 *Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis* 3: 1055–1060.
- 33 284 16. Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD, et al. (2005) The Importance of
34 285 Recent Infection with Mycobacterium tuberculosis in an Area with High HIV Prevalence: A
35 286 Long-Term Molecular Epidemiological Study in Northern Malawi. *J Infect Dis* 192: 480–487.
36 287 doi:10.1086/431517.
- 38 288 17. De Beer JL, Kremer K, Ködmön C, Supply P, van Soolingen D (2012) First Worldwide Proficiency
39 289 Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex
40 290 Strains. *Journal of Clinical Microbiology* 50: 662–669. doi:10.1128/JCM.00607-11.
- 42 291 18. Maes M, Kremer K, van Soolingen D, Takiff H, de Waard JH (2008) 24-locus MIRU-VNTR
43 292 genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao
44 293 Amerindians in Venezuela. *Tuberculosis (Edinb)* 88: 490–494. doi:10.1016/j.tube.2008.04.003.
- 46 294 19. Sougakoff W (2011) Molecular epidemiology of multidrug-resistant strains of Mycobacterium
47 295 tuberculosis. *Clinical Microbiology and Infection* 17: 800–805. doi:10.1111/j.1469-
48 296 0691.2011.03577.x.
- 50 297 20. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for
51 298 polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. *Nucleic Acids Res* 38:
52 299 W326–331. doi:10.1093/nar/gkq351.
- 54 300 21. Supply P (2010) MIRU-VNTR typing: the new international standard for TB molecular
55 301 epidemiology Symposium of the Institut Pasteur de Tunisia.

- 1
2
3 302 22. Van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, et al. (1999) Molecular
4 303 epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997.
5 304 *J Infect Dis* 180: 726–736. doi:10.1086/314930.
- 7 305 23. Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step
8 306 approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the
9 307 United States. *J Clin Microbiol* 43: 688–695. doi:10.1128/JCM.43.2.688-695.2005.
- 11 308 24. Centers for Disease Control and Prevention (2005) New CDC Program for Rapid Genotyping of
12 309 *Mycobacterium Tuberculosis* Isolates. *JAMA* 293: 2086–2086. doi:10.1001/jama.293.17.2086.
- 14 310 25. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, et al. (2000) A prospective evaluation
15 311 of the clinical value of nation-wide DNA fingerprinting of tuberculosis isolates in Denmark. *Int J*
16 312 *Tuberc Lung Dis* 4: 295–299.
- 18 313 26. Bauer J, Yang Z, Poulsen S, Andersen AB (1998) Results from 5 years of nationwide DNA
19 314 fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low
20 315 incidence of *M. tuberculosis* infection. *J Clin Microbiol* 36: 305–308.
- 22 316 27. Zolnir-Dovc M, Poljak M, Erzen D, Sorli J (2003) Molecular epidemiology of tuberculosis in
23 317 Slovenia: results of a one-year (2001) nation-wide study. *Scand J Infect Dis* 35: 863–868.
- 25 318 28. Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Hoek KGP, et al. (2008)
26 319 Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-
27 320 repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of
28 321 *Mycobacterium tuberculosis* Beijing strains in a setting of high incidence of tuberculosis. *J Clin*
29 322 *Microbiol* 46: 3338–3345. doi:10.1128/JCM.00770-08.
- 31 323 29. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, et al. (2001) Automated high-
32 324 throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based
33 325 on mycobacterial interspersed repetitive units. *J Clin Microbiol* 39: 3563–3571.
34 326 doi:10.1128/JCM.39.10.3563-3571.2001.
- 36 327 30. Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniowski FA (2006) Progression toward an
37 328 improved DNA amplification-based typing technique in the study of *Mycobacterium*
38 329 *tuberculosis* epidemiology. *J Clin Microbiol* 44: 2492–2498. doi:10.1128/JCM.01428-05.
- 40 330 31. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems:
41 331 an application of Simpson's index of diversity. *J Clin Microbiol* 26: 2465–2466.
- 43 332 32. WHO | TB data (n.d.). WHO. Available: <http://www.who.int/tb/country/en/index.html>.
44 333 Accessed 12 December 2012.
- 46 334 33. Kruijshaar ME, Pimpin L, Abubakar I, Rice B, Delpech V, et al. (2011) The burden of TB-HIV in
47 335 the EU: how much do we know? A survey of surveillance practices and results. *Eur Respir J* 38:
48 336 1374–1381. doi:10.1183/09031936.00198310.
- 50 337 34. World Health Organization (n.d.) WHO Tuberculosis Country Profiles. Available:
51 338 <http://www.who.int/tb/country/data/profiles/en/>.

- 1
2
3 339 35. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, et al. (2014) Strengthening the Reporting of
4 340 Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE
5 341 statement. *The Lancet Infectious Diseases* 14: 341–352. doi:10.1016/S1473-3099(13)70324-4.
- 6
7 342 36. Freeman MF, Tukey JW (1950) Transformations Related to the Angular and the Square Root.
8 343 *Ann Math Statist* 21: 607–611. doi:10.1214/aoms/1177729756.
- 9
10 344 37. Guang-ming D, Zhi-guo Z, Peng-ju D, Qian Z, Li W, et al. (2013) Differences in the population of
11 345 genetics of *Mycobacterium tuberculosis* between urban migrants and local residents in Beijing,
12 346 China. *Chinese Medical Journal* 126: 4066–4071. doi:10.3760/cma.j.issn.0366-6999.20130216.
- 13
14 347 38. Zmak L, Obrovac M, Katalinic Jankovic V (2014) First insights into the molecular epidemiology
15 348 of tuberculosis in Croatia during a three-year period, 2009 to 2011. *Scandinavian Journal of*
16 349 *Infectious Diseases* 46: 123–129.
- 17
18
19 350 39. Mandal S, Bradshaw L, Anderson LF, Brown T, Evans JT, et al. (2011) Investigating transmission
20 351 of *Mycobacterium bovis* in the United Kingdom in 2005 to 2008. *J Clin Microbiol* 49: 1943–
21 352 1950. doi:10.1128/JCM.02299-10.
- 22
23 353 40. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR (2011) Tuberculosis transmission in
24 354 Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. *Infect Genet Evol*
25 355 11: 124–131. doi:10.1016/j.meegid.2010.09.013.
- 26
27 356 41. Dymova MA, Liashenko OO, Poteiko PI, Krutko VS, Khrapov EA, et al. (2011) Genetic variation
28 357 of *Mycobacterium tuberculosis* circulating in Kharkiv Oblast, Ukraine. *BMC Infect Dis* 11: 77.
29 358 doi:10.1186/1471-2334-11-77.
- 30
31 359 42. Sails AD, Barrett A, Sarginson S, Magee JG, Maynard P, et al. (2011) Molecular epidemiology of
32 360 *Mycobacterium tuberculosis* in East Lancashire 2001-2009. *Thorax* 66: 709–713.
33 361 doi:10.1136/thx.2011.158881.
- 34
35 362 43. Evans J (2010) Analysis of prevalent *Mycobacterium tuberculosis* strains in the United
36 363 Kingdom: detection, distribution and expansion of MIRU-VNTR profiles containing high
37 364 numbers of isolates. *European Society of Clinical Microbiology and Infectious Diseases*. Vienna,
38 365 Austria.
- 39
40
41 366 44. Nikolayevskyy VV, Brown TJ, Bazhora YI, Asmolov AA, Balabanova YM, et al. (2007) Molecular
42 367 epidemiology and prevalence of mutations conferring rifampicin and isoniazid resistance in
43 368 *Mycobacterium tuberculosis* strains from the southern Ukraine. *Clin Microbiol Infect* 13: 129–
44 369 138. doi:10.1111/j.1469-0691.2006.01583.x.
- 45
46 370 45. L. Z, M. O, V. KJ (2014) First insights into the molecular epidemiology of tuberculosis in Croatia
47 371 during a three-year period, 2009 to 2011. *Scandinavian Journal of Infectious Diseases*.
- 48
49 372 46. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, et al. (2011) Evaluation of *Mycobacterium*
50 373 *tuberculosis* typing methods in a 4-year study in Schleswig-Holstein, Northern Germany. *J Clin*
51 374 *Microbiol* 49: 4173–4178. doi:10.1128/JCM.05293-11.
- 52
53 375 47. Dymova MA, Kinsht VN, Cherednichenko AG, Khrapov EA, Svistelnik AV, et al. (2011) Highest
54 376 prevalence of the *Mycobacterium tuberculosis* Beijing genotype isolates in patients newly
55 377 diagnosed with tuberculosis in the Novosibirsk oblast, Russian Federation. *J Med Microbiol* 60:
56 378 1003–1009. doi:10.1099/jmm.0.027995-0.

- 1
2
3 379 48. Alonso-Rodriguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, et al. (2009)
4 380 Prospective universal application of mycobacterial interspersed repetitive-unit-variable-
5 381 number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for
6 382 fast identification of clustered and orphan cases. *J Clin Microbiol* 47: 2026–2032.
7 383 doi:10.1128/JCM.02308-08.
- 8
9 384 49. Glynn JR, Vyonycky E, Fine PEM (1999) Influence of Sampling on Estimates of Clustering and
10 385 Recent Transmission of Mycobacterium tuberculosis Derived from DNA Fingerprinting
11 386 Techniques. *American Journal of Epidemiology* 149: 366–371.
- 12
13 387 50. TB Strain Typing Project Board HPA (2011) TB Strain Typing Cluster Investigation Handbook for
14 388 Health Protection Units 1st Edition. Available:
15 389 <https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealth>
16 390 [ProtectionProgrammes/Respiratory/TB/StrainTyping/](https://hpaintranet.hpa.org.uk/Content/ProgrammesProjects/HPAProgrammes/HPAKeyHealth). Accessed 30 November 2011.
- 17
18 391 51. Walker TM, Monk P, Grace Smith E, Peto TEA (2013) Contact investigations for outbreaks of
19 392 Mycobacterium tuberculosis: advances through whole genome sequencing. *Clin Microbiol*
20 393 *Infect.* doi:10.1111/1469-0691.12183.
- 21
22 394 52. Gurjav U, Jelfs P, McCallum N, Marais BJ, Sintchenko V (2014) Temporal dynamics of
23 395 Mycobacterium tuberculosis genotypes in New South Wales, Australia. *BMC infectious*
24 396 *diseases* 14: 455–455. doi:10.1186/1471-2334-14-455.
- 25
26 397 53. Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of
27 398 standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat
28 399 typing of Mycobacterium tuberculosis. *J Clin Microbiol* 46: 1398–1406.
29 400 doi:10.1128/JCM.02089-07.
- 30
31 401 54. Allix-Béguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008) Standardised PCR-based
32 402 molecular epidemiology of tuberculosis. *Eur Respir J* 31: 1077–1084.
33 403 doi:10.1183/09031936.00053307.
- 34
35 404 55. Tuite AR, Guthrie JL, Alexander DC, Whelan MS, Lee B, et al. (2013) Epidemiological evaluation
36 405 of spatiotemporal and genotypic clustering of mycobacterium tuberculosis in Ontario, Canada.
37 406 *International Journal of Tuberculosis and Lung Disease* 17: 1322–1327.
- 38
39 407 56. Tessema B, Beer J, Merker M, Emmrich F, Sack U, et al. (2013) Molecular epidemiology and
40 408 transmission dynamics of Mycobacterium tuberculosis in Northwest Ethiopia: new
41 409 phylogenetic lineages found in Northwest Ethiopia. *Bmc Infectious Diseases* 13: 131.
42 410 doi:10.1186/1471-2334-13-131.
- 43
44 411 57. Smit PW, Haanpera M, Rantala P, Couvin D, Lyytikäinen O, et al. (2013) Molecular
45 412 Epidemiology of Tuberculosis in Finland, 2008-2011. *Plos One* 8: e85027.
46 413 doi:10.1371/journal.pone.0085027.
- 47
48 414 58. Oelemann MC, Diel R, Vatin V, Haas W, Rüsck-Gerdes S, et al. (2007) Assessment of an
49 415 optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing
50 416 system combined with spoligotyping for population-based molecular epidemiology studies of
51 417 tuberculosis. *J Clin Microbiol* 45: 691–697. doi:10.1128/JCM.01393-06.

- 1
2
3 418 59. Ojo OO, Sheehan S, Corcoran DG, Nikolayevsky V, Brown T, et al. (2010) Molecular
4 419 epidemiology of Mycobacterium tuberculosis clinical isolates in Southwest Ireland. Infect
5 420 Genet Evol 10: 1110–1116. doi:10.1016/j.meegid.2010.07.008.
- 6
7 421 60. Aleksic E, Merker M, Cox H, Reiher B, Sekawi Z, et al. (2013) First Molecular Epidemiology Study
8 422 of Mycobacterium tuberculosis in Kiribati. PLoS ONE 8. Available:
9 423 [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8)
10 424 [84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8](http://www.scopus.com/inward/record.url?eid=2-s2.0-84873163328&partnerID=40&md5=3994b8e5638129b621abc4d7d6d5e3b8).
- 11
12 425 61. De Beer JL, van Ingen J, de Vries G, Erkens C, Sebek M, et al. (2013) Comparative study of
13 426 IS6110 restriction fragment length polymorphism and variable-number tandem-repeat typing
14 427 of Mycobacterium tuberculosis isolates in the Netherlands, based on a 5-year nationwide
15 428 survey. J Clin Microbiol 51: 1193–1198. doi:10.1128/JCM.03061-12.
- 16
17 429 62. Varghese B, Supply P, Shoukri M, Allix-Beguec C, Memish Z, et al. (2013) Tuberculosis
18 430 Transmission among Immigrants and Autochthonous Populations of the Eastern Province of
19 431 Saudi Arabia. PLoS ONE 8. Available: [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2)
20 432 [84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2](http://www.scopus.com/inward/record.url?eid=2-s2.0-84885784886&partnerID=40&md5=4fdbf4015a999a9fcd1a1c31207a75a2).
- 21
22 433 63. Lim LK-Y, Sng LH, Win W, Chee CB-E, Hsu LY, et al. (2013) Molecular Epidemiology of
23 434 Mycobacterium tuberculosis Complex in Singapore, 2006-2012. Plos One 8: e844487.
24 435 doi:10.1371/journal.pone.0084487.
- 25
26 436 64. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-
27 437 locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med
28 438 105 Suppl 1: S67–73. doi:10.1016/S0954-6111(11)70014-2.
- 29
30 439 65. Jonsson J., Hoffner S., Berggren I., Bruchfeld J., Ghebremichael S., et al. (2014) Comparison
31 440 between RFLP and MIRU-VNTR genotyping of mycobacterium tuberculosis strains isolated in
32 441 stockholm 2009 to 2011. PLoS ONE. Available:
33 442 <http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0095159&representation=PDF>.
34 443
- 35
36 444 66. Muwonge A, Malama S, Johansen TB, Kankya C, Biffa D, et al. (2013) Molecular Epidemiology,
37 445 Drug Susceptibility and Economic Aspects of Tuberculosis in Mubende District, Uganda. PLoS
38 446 ONE 8. Available: [http://www.scopus.com/inward/record.url?eid=2-s2.0-](http://www.scopus.com/inward/record.url?eid=2-s2.0-84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697)
39 447 [84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697](http://www.scopus.com/inward/record.url?eid=2-s2.0-84878608813&partnerID=40&md5=babbd6d006ca64e327fb19e01b6bc697).
- 40
41 448 67. Hamblion EL, Wynne-Edwards E, Anderson C, Anderson SR (2011) A summary of strain typing
42 449 and clustering of TB in London in 2010 and an analysis of the associated risk factors. Thorax 66:
43 450 A88–A89. doi:10.1136/thoraxjnl-2011-201054c.50.
- 44
45 451 68. Hang NTL, Maeda S, Lien LT, Thuong PH, Hung NV, et al. (2013) Primary Drug-Resistant
46 452 Tuberculosis in Hanoi, Viet Nam: Present Status and Risk Factors. Plos One 8: UNSP e71867.
47 453 doi:10.1371/journal.pone.0071867.
- 48
49 454
- 50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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
<u>Study setting</u>		
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
<u>Study design</u>		
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 meta-regression 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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

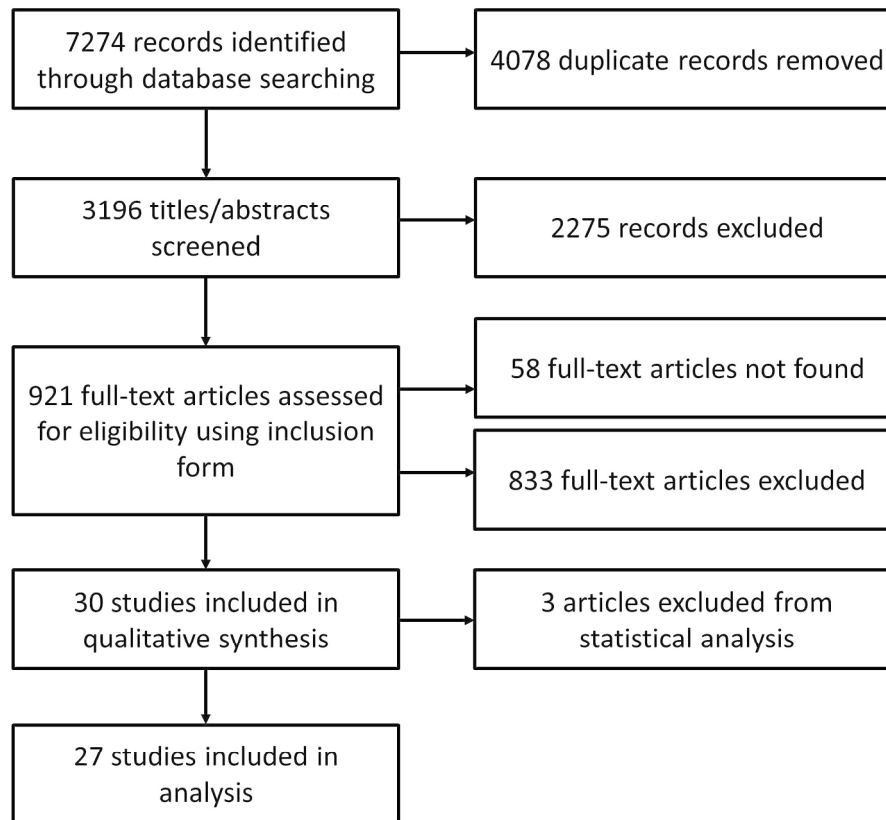
Figure Caption

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

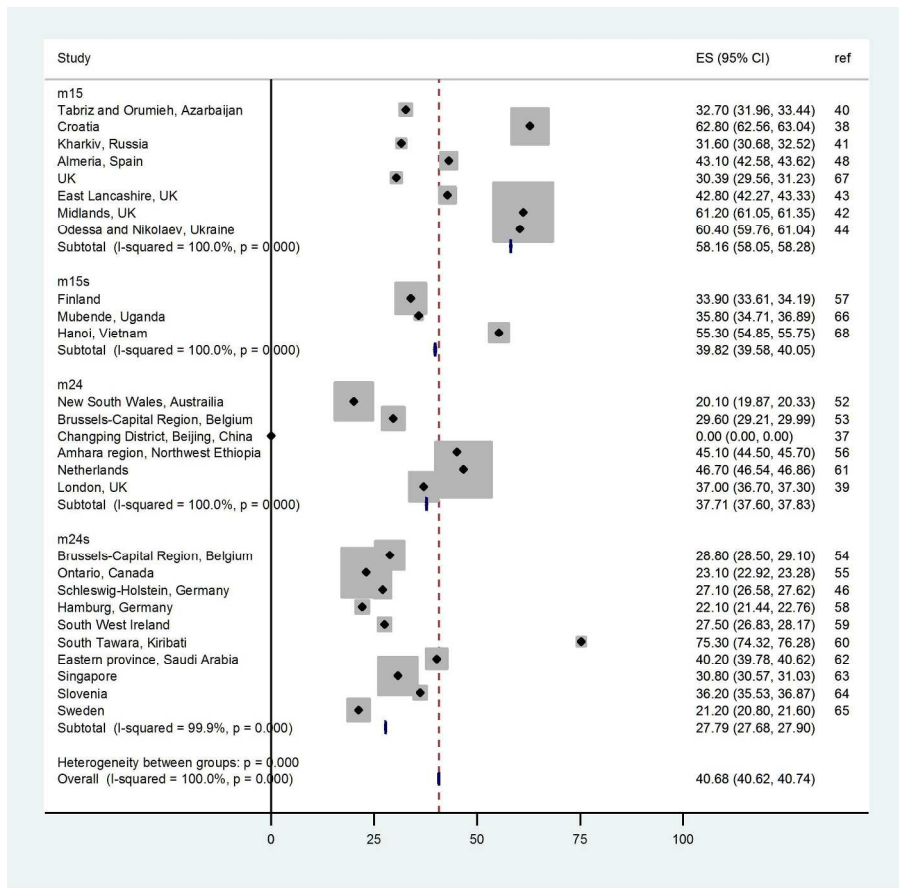
The number of loci typed is categorised into 15 loci (m15), 15 loci with Spoligotyping (m15s), 24 loci (m24) and 24 loci with Spoligotyping (m24s). The study reference is shown in the right hand column.

For peer review only



190x254mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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

- 1
2
3 32. 4 and 16
4 33. 32 and 31
5
6 34. limit 33 to yr="1998-Current"
7
8 35. limit 34 to english language
9
10 36. animals/
11 37. humans/
12 38. 36 not 37
13 39. 35 not 38
14
15
16

17 Appendix 2: STROME-ID scores for the included studies

19 Author	STROME-ID score ^a
20 Aleksic, E	24
21 Allix-Beguec, C	32
22 Allix-Beguec, C	25
23 Alonso-Rodriguez, N	18
24 Asgharzadeh, M	28
25 Bidovec-Stojkovic, U	31
26 De Beer, JL	30
27 Dymova, MA	19
28 Evans, J	^b
29 Grujav, U	32
30 Guang-ming, DAI	19
31 Hamblion, E	^b
32 Hang, NTHL	31
33 Jonsson, J	22
34 Lim, LKY	30
35 Mandal, S	32
36 Muwonge, A	25
37 Nikolayevsky, V	23
38 Oelemann, M	34
39 Ojo, OO	36
40 Roetzer, A	16
41 Sails, A	23
42 Smit, PW	29
43 Tessema, B	26
44 Tuite, AR	31
45 Varghese, B	23
46 Zmak, L	19

47
48
49
50
51
52
53
54
55
56
57 ^aIndividual studies score 1 for each element of checklist they had address

58
59 ^bConference abstracts
60



PRISMA 2009 Checklist

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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

Page 1 of 2

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.

Page 2 of 2

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>