

SI Appendix

Vitamin D deficiency and tuberculosis disease phenotype

Manish Pareek PhD MRCP DTM&H, *Clinical Lecturer*^{a,b,c,d}

John Innes FRCP *Honorary Consultant*^e

Saranya Sridhar DPhil *Research Associate*^b

Lisa Grass *Research Nurse*^b

David Connell MRCP *Clinical Research Fellow*^b

Gerrit Woltmann MD FRCP *Consultant Respiratory Physician*^f

Martin Wiselka PhD FRCP *Consultant*^d

Adrian R Martineau PhD MRCP *Clinical Professor*^g

Onn Min Kon MD FRCP *Consultant Respiratory Physician*^h

Martin Dedicoat PhD MRCP *Consultant*^e

Ajit Lalvani DM FRCP *Chair of Infectious Diseases and Director*^{b,h}

^aDepartment of Infection, Immunity and Inflammation, University of Leicester, UK

^bTuberculosis Research Centre, National Heart and Lung Institute, Imperial College London, UK

^cDepartment of Infectious Disease Epidemiology, Imperial College London, UK

^dDepartment of Infection and Tropical Medicine, University Hospitals Leicester NHS Trust, Leicester, UK

^eDepartment of Infection and Tropical Medicine, Heart of England NHS Foundation Trust, Bordesley Green East, Birmingham, UK

^fDepartment of Respiratory Medicine, Institute for Lung Health, University Hospitals Leicester NHS Trust, Leicester, UK

^gCentre for Primary Care and Public Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK

^hTuberculosis Service, Chest and Allergy Clinic, St. Mary's Hospital, Imperial College Healthcare NHS Trust, UK

Number of pages: 15 (including title page and references)

Number of supplementary figures: 3 **Number of supplementary tables:** 2

Text S1: Supplementary methods

Data collection

For both datasets information was retrieved on date of notification, age, sex, local/foreign-born, ethnicity, time since entry (only applies to foreign-born immigrants and calculated as time between arrival in the UK/US and diagnosis of active TB), site of disease, HIV status, culture result and drug sensitivity. Certain information was only available/retrievable from one dataset: specific organs involved (Birmingham only) and state of notification (US only).

Routine HIV testing of patients diagnosed with active TB is highly inconsistent in the UK and this variable was inconsistently recorded in the Birmingham dataset. Therefore, we considered patients either as HIV positive (if recorded) or as negative/unknown (for all other individuals).

In the UK datasets, ethnicity was defined in line with the Office for National Statistics classification¹ (five ethnic groups: White, Indian Subcontinent, Afro-Caribbean, Oriental/Other Asia, Other). Ethnicity in the US dataset was predefined into 5 categories (Hispanic or Latino, Non-Hispanic (NH) racial categories for White, American Indian or Alaska Native, Asian / Pacific Islander and Black/African American).

We accounted for the independent effects of ethnicity and geographical origin by stratifying individuals into the following groups: Local-born (UK or US) White, Local-born (US or UK) Non-White and Foreign-born Non-White (including individuals with and without pigmented skin of Indian Subcontinent and Oriental ethnicity).

To further compare/explore disease patterns in local-born (either UK or US) ethnic groups versus foreign-born ethnic groups we created a composite variable “country of birth/ethnicity” where we subcategorised individuals as being local/foreign-born and of different ethnic groups.

To investigate the effect of time since immigration on patterns of disease we also created a composite variable “country of birth/time since entry” as per previous authors² which subdivided

individuals into local and foreign-born, with the foreign-born further subdivided by time since entry.

Definitions of patterns of tuberculous disease

For all datasets in our analysis, all notified TB cases (confirmed, probable or possible) which resulted in treatment were included. Active TB cases were initially defined by the major sites of disease into three distinct clinical patterns: pulmonary only (only lung parenchyma involved), extrapulmonary only (only one/more extrapulmonary site(s) involved) and both pulmonary and extrapulmonary (pulmonary disease and one/more extrapulmonary sites involved concurrently). This broad classification was only used to describe overall disease distribution for the Birmingham and US CDC cohorts. However, for the more detailed comparative analyses (such as temporal trends, patterns between different demographic groups and regression modelling) we adopted the approach taken by Peto³ and others^{4 5} by excluding the concurrent pulmonary and extrapulmonary TB category which cannot be definitively assigned to a specific category (and therefore are likely to result in misclassification bias).

Extrapulmonary sites of disease (either with or without pulmonary TB) included: intrathoracic lymph nodes, extrathoracic lymph nodes, bone and/or joint, central nervous system, pleural, peritoneal, genitourinary, intestinal, miliary, pericardial, soft tissue and other. If an individual had more than one extrapulmonary site involved all sites were included and therefore numbers would sum to greater than the total.

Data analysis

Detailed below is a more comprehensive account of the data analysis undertaken in the study.

Continuous data were summarised with median and interquartile range (IQR), and compared using the non-parametric Mann-Whitney U-test. Categorical responses were expressed as a simple descriptive percentage and comparisons made using Pearson's chi-square test (or Fisher's exact test if appropriate).

Temporal trends for main sites of disease (pulmonary only and extrapulmonary only) were summarised by calculating the numbers and relative proportions of each disease type for distinct time periods for the Birmingham (six five-yearly time intervals - 1980-1984, 1985-1989, 1990-1994, 1995-1999, 2000-2004 and 2005-2009) and US CDC (four time intervals - 1993-1996, 1997-2000, 2001-2004 and 2005-2008) datasets; changes in proportions over time were calculated using the χ^2 test for trend.

Thereafter we calculated, and compared, the numbers (and proportion) of cases of each disease type stratified by local/Foreign-birth and White/Non-White ethnicity to create three main groups of interest: local-born White ethnicity, local-born Non-White ethnicity and foreign-born Non-White ethnicity. Two way comparisons were undertaken using Fisher's exact test. In addition for the foreign-born (non-White) population we also examined differences in the proportion of cases that were pulmonary and extrapulmonary stratified by time since entry to the receiving country using the χ^2 test for trend. Individuals who were foreign-born of White ethnicity were not specifically examined as they accounted for a relatively small proportion of cases in all the datasets.

To investigate factors associated with extrapulmonary involvement we compared individuals with exclusively extrapulmonary versus pulmonary patterns of disease. For the Birmingham dataset (but not US CDC), where individual level data was available, we undertook separate regression modelling for the whole cohort and for the non-White ethnic groups. Univariable association of factors individually associated with extrapulmonary tuberculosis only (compared against pulmonary cases only) was assessed using logistic regression and reported as crude odds ratios (OR) and 95% confidence intervals. To calculate adjusted odds ratios (and 95%

CI) we mutually adjusted for the following factors: year, and season, of notification, age, gender, ethnicity, country of birth, time since arrival and HIV status. Interaction terms (age and time since entry) were assessed and found to be non-significant and not included in the final multivariable models. For the US dataset, only aggregate data was available and therefore only univariable comparisons using logistic regression could be performed.

For the London and Leicester cohorts (analysed as one population), assessment of the impact of vitamin D status on disease patterns was undertaken through analysis of active TB patients who had serum 25-hydroxycholecalciferol [serum 25(OH)D] concentrations measured either before, or soon after, commencement of anti-tuberculous therapy using an immunoassay (Diasorin Liaison, Diasorin Ltd, Stillwater, Minnesota, USA and Centaur XP, Siemens Healthcare Diagnostics, New York, USA); both centres participate in the Vitamin D external equality assessment scheme DEQAS. The decision to measure serum 25(OH)D concentrations was determined by the treating physician. We used absolute serum 25-hydroxycholecalciferol concentrations [25(OH)D] except in individuals whose vitamin D measurement was below the lower limit of detection, where a value half of the lower limit was used.⁶ If necessary, serum 25(OH)D concentrations were converted from $\mu\text{g/L}$ to the standard nmol/L , by multiplying by 2.496. As 25(OH)D concentrations were found to be non-normally distributed, they were \log_{10} -transformed with geometric mean titres (and 95% confidence intervals) calculated by exponentiating the mean and of the lower and upper limits of the 95% confidence intervals of the \log_{10} -transformed titres. Comparisons between groups were made using a one-way analysis of variance (on the log transformed concentrations) with Bonferroni correction, if appropriate. The proportions of subjects who were severely vitamin D deficient [serum 25(OH)D <20nmol/L] were compared between groups using Fisher's exact test. We further assessed the factors associated with extrapulmonary TB in this dataset using logistic regression and reported results as odds ratios (with 95% confidence intervals); in the multivariable model we adjusted for the centre (ie. London or

Leicester), year, and season, of notification, age, gender, ethnic group, location of birth, time since arrival, HIV status and serum 25(OH)D concentration. In the multivariable model, serum 25(OH)D was used as a continuous variable transformed by \log_2 ; each unit change in \log_2 serum 25(OH)D corresponds to a doubling in serum 25(OH)D on the original scale which thereby allows a more clinically relevant, and intuitive, interpretation of the odds ratios. For the purposes of the analysis, subjects with missing data were excluded on a listwise basis.

Supplementary tables

Table S1. Demographic, clinical and temporal variables associated with different patterns of tuberculosis for the whole US cohort and the Non-White ethnic groups

Whole cohort (n=255489) ¹				Non-White ethnic groups (n=196177) ¹			
Variable	Extrapulmonary TB only (n=52040)	Unadjusted OR (95% CI)	p	Variable	Extrapulmonary TB only (n=42588)	Unadjusted OR (95% CI)	p
Period				Period			
1993-1996	15253/85806 (17.8)	1	<0.001	1993-1996	11642/62390 (18.7)	1	<0.001
1997-2000	13528/65930 (20.5)	1.19 (1.16-1.23)		1997-2000	10955/49774 (22)	1.23 (1.19-1.27)	
2001-2004	12250/54980 (22.3)	1.33 (1.29-1.36)		2001-2004	10425/43926 (23.7)	1.36 (1.32-1.40)	
2005-2008	11009/48773 (22.6)	1.35 (1.31-1.39)		2005-2008	9566/40087 (23.9)	1.37 (1.33-1.41)	
Age categories n (%)				Age categories n (%)			
Under 5	1949/9356 (20.8)	1	<0.001	Under 5	1680/8210 (20.5)	1	<0.001
5-14	1875/6756 (27.8)	1.46 (1.36-1.57)		5-14	1706/6092 (28.0)	1.51 (1.40-1.63)	
15-24	5034/23705 (21.2)	1.02 (0.97-1.09)		15-24	4678/21842 (21.4)	1.06 (1.00-1.13)	
25-44	19863/88342 (22.5)	1.10 (1.05-1.16)		25-44	17793/74678 (23.8)	1.22 (1.15-1.29)	
45-64	12739/70916 (18.0)	0.83 (0.79-0.88)		45-64	10398/52672 (19.7)	0.96 (0.90-1.01)	
>64	10580/56414 (18.8)	0.88 (0.83-0.93)		>64	6333/32683 (19.4)	0.93 (0.88-0.99)	
Gender				Gender			
Male	26995/159371 (16.9)	1	<0.001	Male	21607/120234 (18.0)	1	<0.001
Female	25045/96118 (26.1)	1.73 (1.69-1.76)		Female	20981/75943 (27.6)	1.74 (1.71-1.78)	
Country of birth/Ethnicity				Ethnicity			
US-born White	7672/52022 (14.7)	1	<0.001	American Indian or Alaska Native, NH	661/3156 (20.9)	1	<0.001
US-born American Indian	635/3079 (20.6)	1.50 (1.37-1.64)		Asian or Pacific Islander, NH	13742/53251 (25.8)	1.31 (1.20-1.43)	
US-born Asian	748/3010 (24.9)	1.91 (1.75-2.08)		Black or African American, NH	15884/78773 (20.2)	0.95 (0.87-1.04)	
US-born Black	11553/65093 (17.7)	1.25 (1.21-1.29)		Hispanic or Latino	12301/60997 (20.2)	0.95 (0.87-1.04)	
US-born Hispanic	3958/18380 (21.5)	1.59 (1.52-1.66)					
Foreign-born White	1780/7290 (24.4)	1.87 (1.76-1.98)		Country of birth/time since arrival			
Foreign-born American Indian	26/77 (33.8)	2.95 (1.84-4.73)		US-born	16894/89562 (18.9)	1	<0.001
Foreign-born Asian	12994/50241 (25.9)	2.02 (1.95-2.08)		Foreign-born/<1 year	2817/21219 (13.3)	0.66 (0.63-0.69)	
Foreign-born Black	4331/13680 (31.7)	2.68 (2.56-2.80)		Foreign-born/1-4 years	5947/21755 (27.3)	1.62 (1.56-1.67)	
Foreign-born Hispanic	8343/42617 (19.6)	1.41 (1.36-1.46)		Foreign-born/5-14 years	6590/23851 (27.6)	1.64 (1.59-1.70)	
				Foreign-born/≥15 years	5158/20390 (25.3)	1.46 (1.41-1.51)	
				Foreign-born/time not known	5182/19400 (26.7)	1.57 (1.51-1.63)	
Employment status				Employment status			
Unemployed	25750/132732 (19.4)	1	<0.001	Unemployed	20257/98781 (20.5)	1	<0.001
Employed	20445/92635 (22.1)	1.18 (1.15-1.20)					
Previous TB							
No	49779/239985 (20.7)	1	<0.001				

Yes	1803/13538 (13.3)	0.59 (0.56-0.62)	Employed	17570/73625 (23.9)	1.21 (1.19-1.24)	
HIV status			Previous TB			
Negative	18335/94526 (19.4)	1	No	40876/184563 (22.1)	1	<0.001
Positive	4848/23354 (20.8)	1.09 (1.05-1.13)	Yes	1328/10112 (13.1)	0.53 (0.50-0.56)	
Drug sensitivity			HIV status			
Not multi-drug resistant	34166/188723 (18.1)	1	Negative	38520/176150 (21.9)	1	0.001
Multi-drug resistant	319/2869 (11.1)	0.57 (0.50-0.64)	Positive	4068/20027 (20.3)	0.94 (0.90-0.97)	
Resident in correctional facility			Drug sensitivity			
No	50697/243677 (20.8)	1	Not multi-drug resistant	28112/144210 (19.5)	1	<0.001
Yes	979/9550 (10.3)	0.43 (0.41-0.46)	Multi-drug resistant	269/2442 (11.0)	0.51 (0.45-0.58)	
			Resident in correctional facility			
			No	41451/186698 (22.2)	1	<0.001
			Yes	863/7840 (11.0)	0.43 (0.40-0.47)	

Footnote

¹Comparison of extrapulmonary TB cases only compared to pulmonary TB only (concurrent pulmonary-extrapulmonary cases excluded)

Table S2. Demographic characteristics of the individuals with active TB included from the London and Leicester cohorts¹

Variable	Subjects included in study (n=462)	Subjects excluded from the study ² (n=263)	p
Age			
Median (interquartile range)	34 (26-47)	38 (28-53)	
Age categories			
16-25	101 (21.9%)	44 (16.7%)	0.03
26-35	144 (31.2%)	68 (25.9%)	
36-45	87 (18.8%)	47 (17.9%)	
46-55	63 (13.6%)	48 (18.3%)	
over 55	67 (14.5%)	56 (21.3%)	
Gender			
Male	263 (56.9%)	131 (49.8%)	0.06
Female	199 (43.1%)	132 (50.2%)	
Ethnicity			
White	52 (11.3%)	35 (13.3%)	0.21
Indian Subcontinent	276 (59.7%)	136 (51.7%)	
Afro-Caribbean	65 (14.1%)	52 (19.8%)	
Oriental/Other Asia	64 (13.9%)	37 (14.1%)	
Other	5 (1.1%)	3 (1.1%)	
Place of birth³			
United Kingdom	65 (14.5%)	42 (17.0%)	0.38
Foreign-born	383 (85.5%)	205 (83.0%)	
Time since entry to the UK (years)⁴			
Median (interquartile range)	7 (3-15)	8 (4-20)	
<1	14 (3.8%)	7 (3.6%)	0.80
1 - 5	141 (38.4%)	67 (34.7%)	
6-10	79 (21.5%)	47 (24.4%)	
>10	133 (36.2%)	72 (37.3%)	
HIV status⁵			
Negative	430 (98.2%)	200 (91.3%)	<0.01
Positive	8 (1.8%)	19 (8.7%)	
Type of disease			
Pulmonary only	161 (34.9%)	75 (28.5%)	0.06
Pulmonary and extrapulmonary	84 (18.2%)	40 (32.3%)	
Extrapulmonary only	217 (47.0%)	148 (56.3%)	
Serum 25(OH)D (nmol/L)			
Mean (95% confidence interval)	12.6 (11.8-13.5)		
Vitamin D category			
<20nmol/L	325 (70.4%)		
<50nmol/L	440 (95.2%)		
<75nmol/L	452 (97.8%)		

Footnote

¹All consecutively diagnosed cases of active TB recruited between November 1st 2007 and 30th June 2011 (for London) and January 1st 2010 and July 10th 2012 (for Leicester) (with complete demographic and clinical data) where a serum 25(OH)D sample had been drawn up to 90 days prior to diagnosis or within 30 days of treatment initiation.

²Subjects excluded due to not having had a vitamin D level checked (n=95) or the vitamin D level was done outside the pre-specified time window (n=168)

³Data missing for 14 subjects included in the study and 16 subjects excluded from the study

⁴Data missing for 16 subjects included in the study and 13 subjects excluded from the study

⁵Data missing for 24 subjects included in the study and 44 subjects excluded from the study

Supplementary figures

Figure S1A. Flow chart of cases included from Birmingham, UK dataset for data analysis

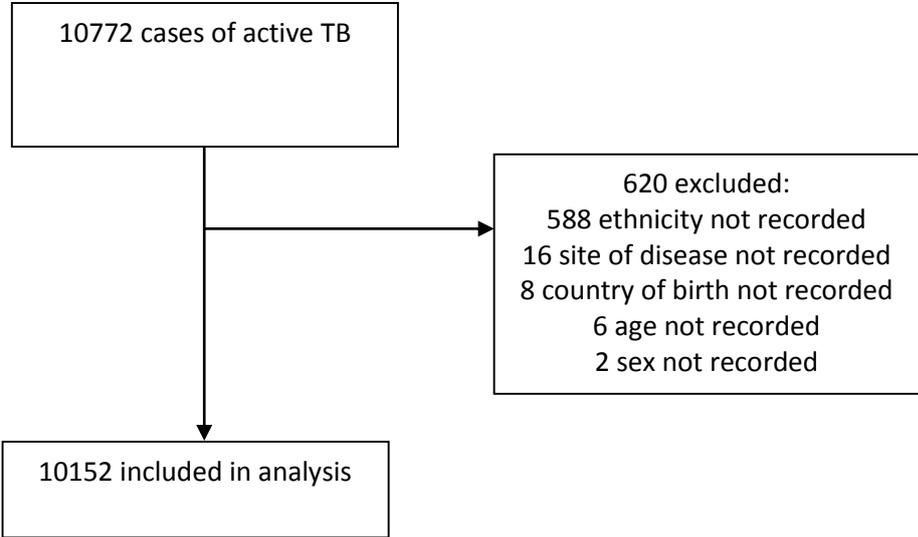
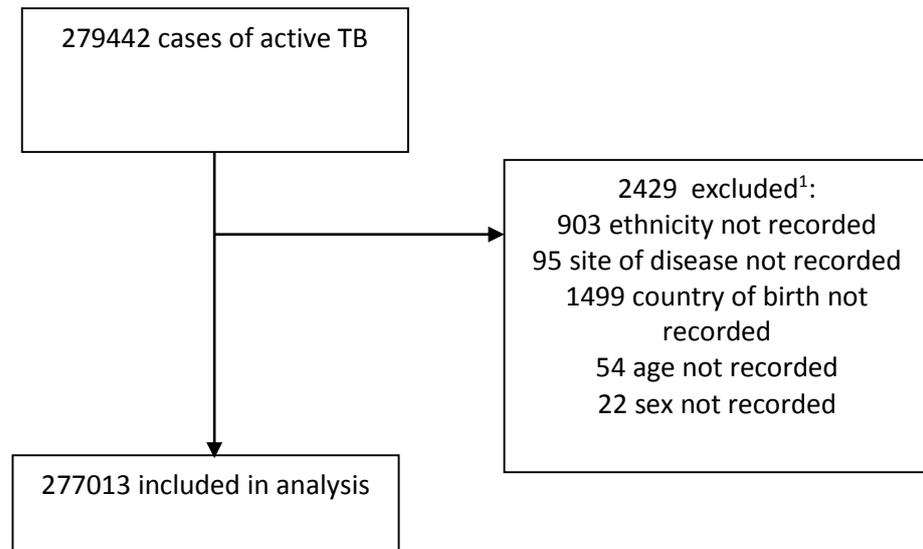


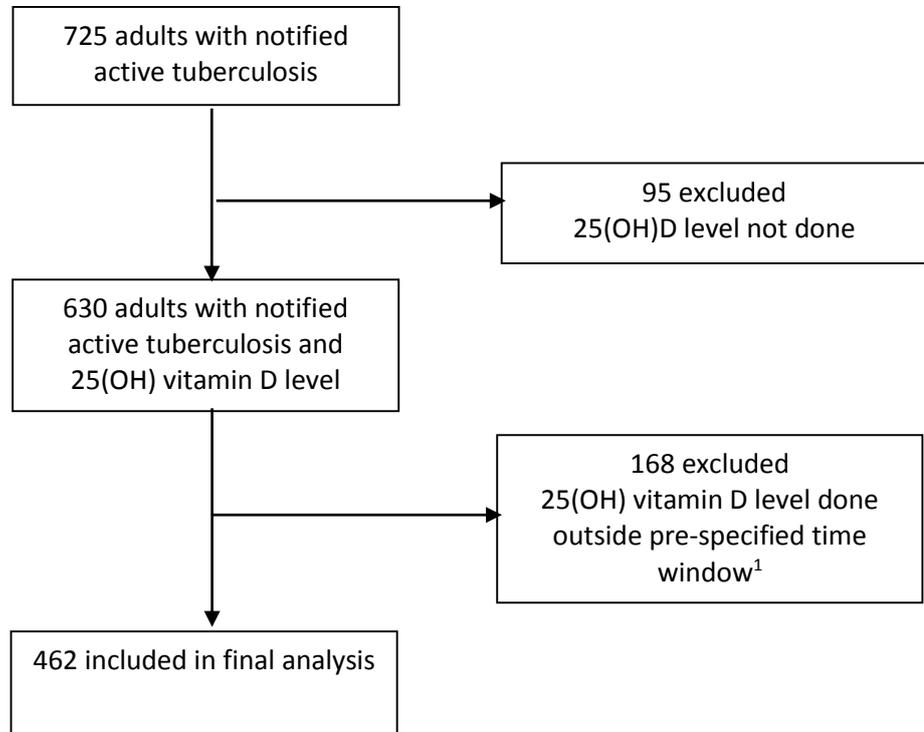
Figure S1B. Flow chart of cases included from US Centres for Disease Control dataset for data analysis



Footnote

¹Numbers sum to greater than 2429 (2573) as some individuals were excluded for more than one reason

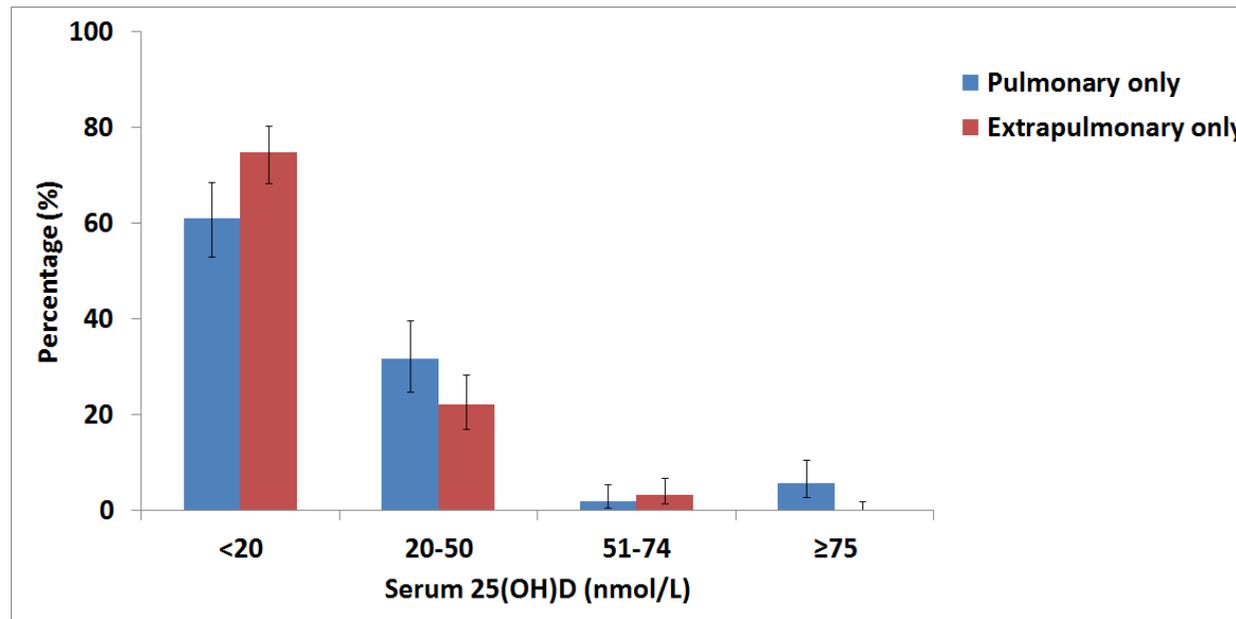
Figure S2. Flow chart of cases included from the London and Leicester cohorts for data analysis



Footnote

¹Individuals excluded if vitamin D sample taken more than 90 days before or 30 days after commencing anti-tuberculous therapy

Figure S3. Proportion of individuals with pulmonary and extrapulmonary tuberculosis with different vitamin D concentrations at diagnosis



Supplementary references

1. Office of National Statistics. *Ethnic Group Statistics: a guide for the collection and classification of ethnicity data*. London: ONS, 2009.
2. Kruijshaar ME, Abubakar I. Increase in extrapulmonary tuberculosis in England and Wales 1999–2006. *Thorax* 2009;64(12):1090-95.
3. Peto HM, Pratt RH, Harrington TA, LoBue PA, Armstrong LR. Epidemiology of extrapulmonary tuberculosis in the United States, 1993-2006. *Clin Infect Dis* 2009;49(9):1350-57.
4. te Beek LA, van der Werf MJ, Richter C, Borgdorff MW. Extrapulmonary tuberculosis by nationality, The Netherlands, 1993-2001. *Emerg Infect Dis* 2006;12(9):1375-82.
5. García-Rodríguez JF, Álvarez-Díaz H, Lorenzo-García MV, Mariño-Callejo A, Fernández-Rial Á, Sesma-Sánchez P. Extrapulmonary tuberculosis: epidemiology and risk factors. *Enferm Infecc Microbiol Clin* 2011;29(7):502-9.
6. Cole SR, Chu H, Nie L, Schisterman EF. Estimating the odds ratio when exposure has a limit of detection. *International Journal of Epidemiology* 2009;38(6):1674-80.