Mangtani, P., Nguipdop Djomo, P., Keogh, R., Sterne, J., Abubakar, I., Smith, P. G., Fine, P. E. M., Vynnycky, E., Watson, J. M., Elliman, D., Lipman, M., & Rodrigues, L. C. (2018). The duration of protection of school-age BCG vaccination in England: a population-based case-control study. *International Journal of Epidemiology*, *47*(1), 193-201. [dyx141]. https://doi.org/10.1093/ije/dyx141

Supplementary material

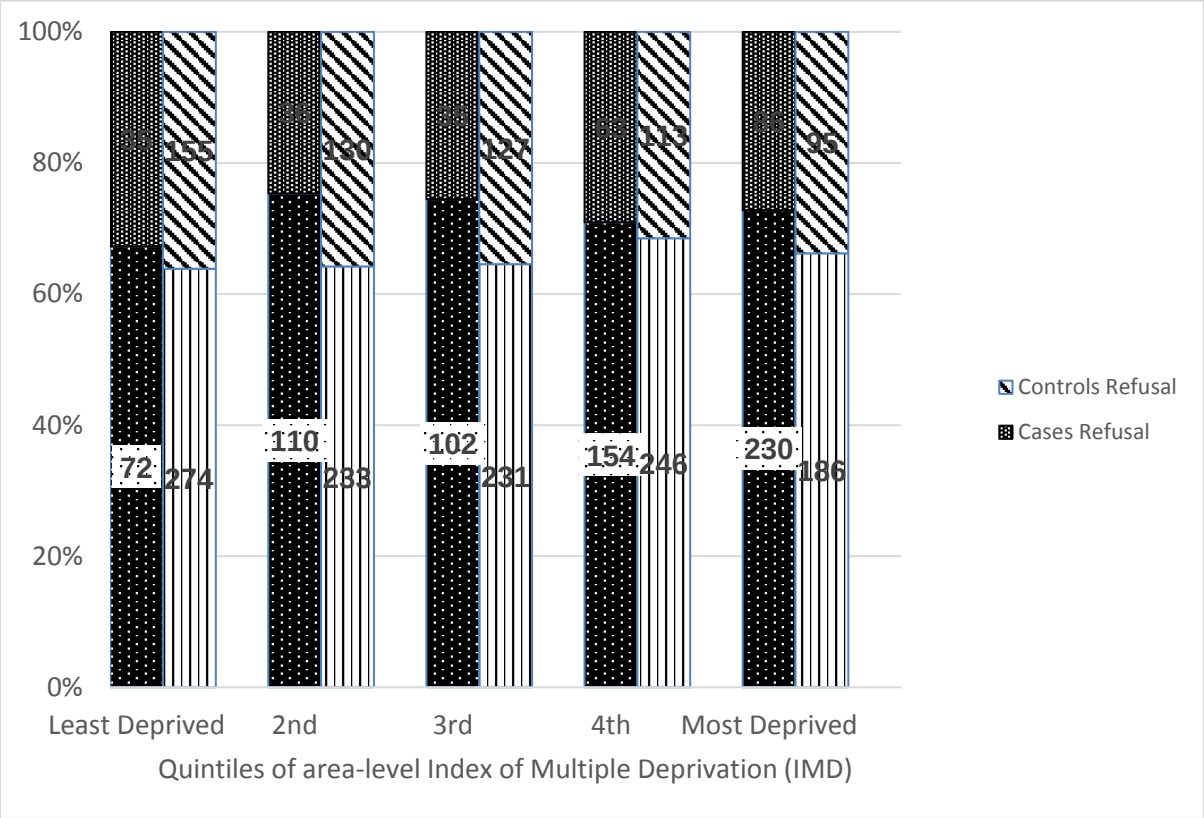
Mangtani P, Nguipdop Djomo P, Keogh R, et al. The duration of protection of school-age BCG vaccination in England: a population -based case-control study.

Supplementary Methods:

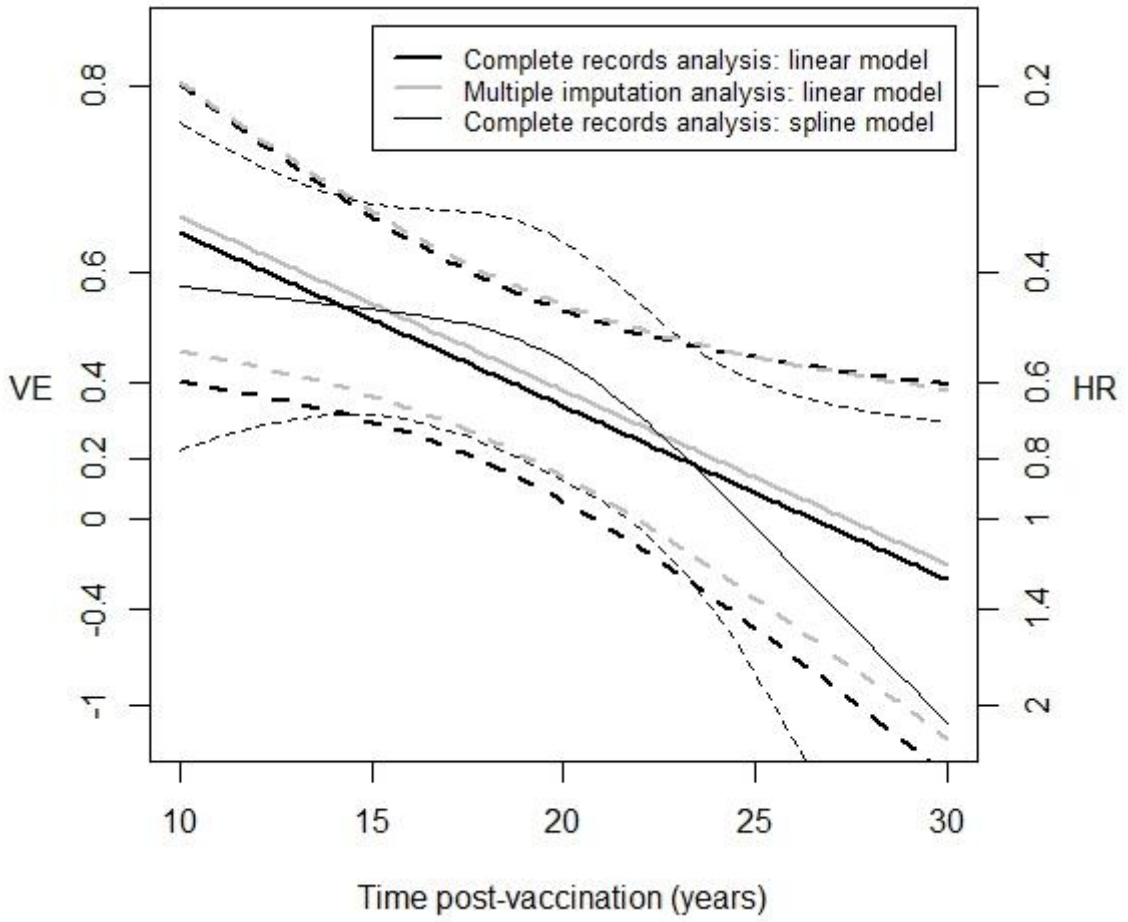
Details of multiple imputation conducted for missing data

As noted in the main text, 17% of individuals were missing one or more of the explanatory variables in the fully adjusted model. We used multiple imputation (MI) to impute the missing data. Missingness in several variables was handled by the chained equations approach[1] in which we specify an imputation model for each variable with missingness which includes as predictors all other explanatory variables in the main analysis model together with variables representing the outcome, about which we provide more detail below.
For imputation of missing data in the setting of Cox regression White and Royston[2] recommended including the event indicator (here, case-control status) and the Nelson-Aalen estimate of the cumulative hazard as predictors in the imputation model. Due to our study design it was not possible to obtain an estimate of the cumulative hazard and we therefore instead included the logarithm of age at TB diagnosis (for cases) or interview (for controls), together with the logarithm of age at entry to the analysis. Time since vaccination at the time of diagnosis (for cases) or interview (for controls) was also included as a predictor in all imputation models. For unvaccinated individuals and individuals with missing vaccination status the time-since-vaccination was taken to be the time since age 12. To accommodate the time-varying effect of the vaccine which is accommodated in the main analysis model we also included in the imputation models for vaccination status an interaction between case-control status and the time-since-vaccination; for the analysis using 4 time-since-vaccination periods we categorised time-since-vaccination into four periods and included an interaction between this and case-control status in the imputation model, and for the analysis in which the vaccine effectiveness was modelled as a continuous function of time-since vaccination using a linear model we used an interaction between time-since-vaccination and case-control status in the imputation model. After the imputation was performed, those imputed as having been vaccinated were assigned a vaccination age of 12, the median self-reported age.
In summary, the imputation model for each variable with missingness included as predictors (excluding the variable itself): sex; birth cohort (using 4 groups); deprivation group; education level; the life style variables variables (tobacco smoking, alcohol drinking and misuse/abuse of controlled drugs); history of homelessness; history of prison stays; regular travel abroad to high TB regions; case-control status; the logarithm of age at TB diagnosis (for cases) or interview (for controls); the logarithm of age at entry to the analysis; time-since vaccination at diagnosis (for cases) or interview (for controls), time since age 12 for the unvaccinated and those with missing vaccination status; and the interaction between case-control status and the function used to represent the time-varying association.
We used 20 imputed data sets[1]. The imputation was performed using Stata.

SFigure 1: Comparison of response rates by quintiles of deprivation in school-age BCG study cases and controls



SFigure 2: Results from modelling the time-varying effect of the vaccine as a linear function of time (on a log scale) and using a restricted cubic spline function of time (on a log scale) with 3 knots at 15, 20 and 25 years post vaccination.



STable 1: Comparison of recruitment by quintiles of deprivation in school-age BCG study cases and controls

	Cases available and contactable		Addresses contacted for potential controls	
IMD quintiles	**Total**	**Contacted**	**Total**	**contacted**
Least deprived	165	119 (72%)	1889	1634 (87%)
2	218	157 (72%)	1885	1599 (85%)
3	233	148 (64%)	1886	1631 (86%)
4	394	254 (64%)	1886	1659 (88%)
Most deprived	575	357 (62%)	1878	1653 (88%)

STable 2 Complete case analysis of the association between time since school aged BCG vaccination and risk of TB from analyses of all subjects where data were available on all the confounders used in each of the models

	Baseline[a] model (based on 669 cases and 1165 controls)			**Partially adjusted[b] model** (based on 638 cases and 1134 controls)		
	HR	**95% CI**	**p**	**HR**	**95% CI**	**p**
Unvaccinated	1 (ref)			1 (ref)		
Vaccinated: 10-15 yrs ago	0.40	(0.27,0.58)	<0.001	0.42	(0.27,0.64)	<0.001
Vaccinated: 15-20 yrs ago	0.35	(0.25,0.49)	<0.001	0.39	(0.26,0.58)	<0.001
Vaccinated: 20-25 yrs ago	0.55	(0.40,0.76)	<0.001	0.64	(0.45,0.92)	0.015
Vaccinated: 25-29 yrs ago	0.56	(0.34,0.93)	0.025	0.77	(0.45,1.32)	0.341

[a]Base model is stratified on birth cohort and adjusted for sex
[b]The partially adjusted model includes additional adjustment for confounding variables area-level deprivation and educational level.

SReferences

1. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Stat Med, 2011. **30**(4): p. 377-99.
2. White, I.R. and P. Royston, *Imputing missing covariate values for the Cox model.* Stat Med, 2009. **28**(15): p. 1982-98.